Optimaliteit, onzekerheid, en dynamisch programmeren met onderprevisies

Optimality, Uncertainty, and Dynamic Programming with Lower Previsions

Matthias C. M. Troffaes

Promotor: prof. dr. ir. G. de Cooman
Proefschrift ingediend tot het behalen van de graad van
Doctor in de Toegepaste Wetenschappen

Vakgroep Elektrische Energie, Systemen, en Automatisering
Voorzitter: Prof. dr. ir. J. Melkebeek
Faculteit Toegepaste Wetenschappen
Academiejaar 2004–2005

To my princess
Sabine

Optimality, Uncertainty, and Dynamic Programming with Lower Previsions
Matthias C. M. Troffaes

Supervisor:
prof. dr. ir. Gert de Cooman

Address:
Onderzoeksgroep SYSTeMS
Vakgroep Elektrische Energie, Systemen, en Automatisering (TW08)
Universiteit Gent
Technologiepark Zwijnaarde 914
B-9052 Zwijnaarde
België

This book was typeset using LaTeX.

ISBN 90-8578-007-1
NUR 916, 992
Wettelijk depot: D/2005/10.500/7

Acknowledgements

I especially thank Gert and Sabine for their wonderful support, both of a technical and emotional nature, during the four and a half years I have been working on my Ph.D. Others I wish to thank are, in random order: Carien, Lev, Enrique, George, Teddy, Stefanie, Jonathan, Godfried, Marco, Christiane, Elena, Pedro, Griet, Vivian, Fabio, Fabio, Guillaume, Erik, Werner, Sofie, Hektor, Katrijn, Katrijn, Sven, Rene, Igor, Viviana, Scott, Sabien, Madhu, Christine, Serafín, Jean-François, Nele, Nele, Elmar, Karolyne, Frank, Krista, Jean-Marc, Kristof, Dirk, Michael, Lies, Paolo, Thomas, Thomas, Vicky, Cedric, Marc, Valery, and many others, whose names I unfortunately cannot recall at the moment (sorry girls). Still in random order, I also wish to thank Peter Walley for "Statistical Reasoning with Imprecise Probabilities", George Romero for "Night of the Living Dead" and "Dawn of the Dead", Rockstar for GTA III and GTA Vice City, Id Software for Doom III, CryTek for Far Cry, and RedHat and Warren Togami for Fedora. And finally, for their music that kept me going and going and going and going while writing this piece, I thank Moby, Schubert, Passarani, Richard Strauss, Converter, Winterkälte, J. S. Bach and C. Ph. E. Bach, the Ultra Milkmaids, Telepherique, Morgenstern, van Beethoven, and Cubanate; and I also thank whoever invented the harpsichord.

This dissertation presents research results of project G.0139.01 of the Fund for Scientific Research, Flanders (Belgium), and of the Belgian Programme on Interuniversity Attraction Poles, initiated by the Belgian Federal Science Policy Office. The scientific responsibility rests with its author.

Matthias C. M. Troffaes

March 25, 2005
Gent

Contents

List of Symbols

$\mathbb{N}$	$\{0, 1, 2, 3, \ldots\}$ (the set of natural numbers)
$\mathbb{N} \setminus \{0\}$	$\{1, 2, 3, \ldots\}$ (the set of natural numbers without zero)
$\mathbb{N}^*$	$\{0, 1, 2, 3, \ldots, +\infty\} = \mathbb{N} \cup \{+\infty\}$
$\mathbb{R}$	the set of real numbers
$\mathbb{R}^*$	$\mathbb{R} \cup \{-\infty, +\infty\}$ (the set of extended real numbers)
$\complement A$	the complement of a set A
$\wp(A)$	the power set of a set A
$(A, \leq)$	a directed set
x_n	a sequence (map from $\mathbb{N}$ to a set)
x_α	a net (map from a directed set $(A, \leq)$ to a set)

$\mathcal{F}$	a field
$\mathbb{A}(\mathcal{F})$	the set of atoms of a field $\mathcal{F}$
μ	a set function
$\underline{\mathbf{P}}_\mu$	the lower probability (lower prevision) induced by a set function
$\overline{\mathbf{P}}_\mu$	the upper probability (upper prevision) induced by a set function
$\mathbf{P}_\mu$	the probability (prevision) induced by a set function
μ^*	the outer set function induced by a set function μ
μ_*	the inner set function induced by a set function μ
$\mathcal{C}_\mu$	Carathéodory field of μ
$\mu_\mathcal{C}$	Carathéodory extension of μ
$\mathcal{J}_\mu$	Jordan field of μ
$\mu_\mathcal{J}$	Jordan extension of μ

ν	a set function to be interpreted as a lower probability (such as a 2-monotone set function), or a necessity distribution
π	a set function to be interpreted as an upper probability (such as a 2-alternating set function), or a possibility distribution
Π	possibility measure
N	necessity measure
$G_{*\nu,f}$, $F^*_{\nu,f}$	the lower decreasing distribution function and upper distribution function of f with respect to ν
$\mathcal{P}(\mathcal{F})$	the set of all probability charges on a field $\mathcal{F}$
F	a cumulative distribution function
μ_F	the probability charge induced by F
(F_*, F^*)	a p-box
$\underline{\mathbf{P}}_{(F_*,F^*)}$	the lower prevision induced by a p-box (F_*, F^*)
$\mathrm{S}\underline{\int} f\,\mathrm{d}\mu$	lower S-integral of f with respect to μ
$\mathrm{D}\overline{\int} f\,\mathrm{d}\mu$	Dunford integral of f with respect to μ
$\mathrm{C}\int f\,\mathrm{d}\nu$	Choquet integral of f with respect to ν
$\mathrm{R}\underline{\int}_a^b f(x)\,\mathrm{d}x$	lower Riemann integral of f over $[a,b]$
$\text{R-S}\underline{\int}_a^b f(x)\,\mathrm{d}F(x)$	lower Riemann-Stieltjes integral of f with respect to F over $[a,b]$
ρ_F	Riemann-Stieltjes charge
λ_F	Lebesgue-Stieltjes measure
X, Y, Z	a random variable
$\mathcal{X}, \mathcal{Y}, \mathcal{Z}$	the set of possible values of a random variable
f, g, h	a gamble or a random quantity
$f(X), g(X), h(X)$	a gamble or a random quantity on X
$\mathcal{L}(X)$	the set of all gambles on X
$\mathrm{span}(\mathcal{F})$	the set of all $\mathcal{F}$-simple gambles on X
$\mathcal{L}_{\mathcal{F}}(X)$	the set of all $\mathcal{F}$-measurable gambles on X
$\mathcal{L}_{\mu}(X)$	the set of all μ-integrable gambles on X
$\mathcal{R}(X)$	the set of all random quantities on X
P, Q, R	an (extended) prevision

$\underline{P}$, $\underline{Q}$, $\underline{R}$	an (extended) lower prevision
$\overline{P}$, $\overline{Q}$, $\overline{R}$	an (extended) upper prevision
$\mathcal{M}$	a set of (extended) linear previsions
$\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{K}}$	the natural extension of $\underline{P}$ to $\mathcal{K}$
$\underline{\mathbf{E}}_{\underline{P}}$	the natural extension of $\underline{P}$ to the set of all gambles
$\mathbf{E}_{\underline{P}}$	the linear extension of $\underline{P}$
$\mathcal{P}(X)$	the set of all linear previsions on $\mathcal{L}(X)$
$\mathcal{P}^{\mathcal{K}}(X)$	the set of all linear (extended) previsions on some subset $\mathcal{K}$ of $\mathcal{L}(X)$ ($\mathcal{R}(X)$)
$\boldsymbol{\mu}_P^{\mathcal{F}}$	probability charge μ on $\mathcal{F}$ obtained by restriction of P
$\mathbf{M}_{\underline{P}}^{\mathcal{K}}$	the set of all linear (extended) previsions on $\mathcal{K}$ that dominate $\underline{P}$
$\underline{\mathbf{E}}_{\mathcal{M}}^{\mathcal{K}}$	the lower envelope of a set $\mathcal{M}$ of linear previsions on domain $\mathcal{K}$
$\|\bullet\|_{\underline{P}}$	$\underline{P}$-norm
$\mathcal{K}_{\underline{P}}$	$\underline{P}$-space
$\mathcal{N}_{\underline{P}}$	set of all $\underline{P}$-null sets
$\mathcal{R}_{\underline{P}}^{0}$	set of all $\underline{P}$-null random variables
$f = g$ a.e. $\underline{P}$	f is equal to g almost everywhere with respect to $\underline{P}$
$f \leq g$ a.e. $\underline{P}$	f is less or equal to g almost everywhere with respect to $\underline{P}$
$\mathcal{K}_{\underline{P}}^{\sharp}(X)$	set of $\underline{P}$-essentially bounded random variables on X
$\underline{P}^{\sharp}$	extension of $\underline{P}$ to $\mathcal{K}_{\underline{P}}^{\sharp}(X)$
$f_\alpha \xrightarrow{\underline{P}} f$	the net f_α converges $\underline{P}$-hazily to f
$\mathcal{K}_{\underline{P}}^{\mathbf{x}}(X)$	set of $\underline{P}$-previsible random variables on X
$\underline{P}^{\mathbf{x}}$	extension of $\underline{P}$ to $\mathcal{K}_{\underline{P}}^{\mathbf{x}}(X)$
$\left\|f\right\|_{\underline{P}}^{\mathbf{x}}$	the $\underline{P}$-previsible norm of f
$f_{a,b}(x)$	the (a, b)-cut of f
R	a choice relation
opt	an optimality operator, or a social choice function

$f >_{\underline{P}} g$	f is strictly preferred to g with respect to $\underline{P}$
$\mathrm{opt}_{\not<_{\underline{P}}}(A)$, $\max_{\geqslant_{\underline{P}}}(A)$	set of $\underline{P}$-maximal actions in A
$\mathrm{opt}_{\mathcal{M}}(A)$	set of $\mathcal{M}$-maximal actions in A
$\max_{\sqsupseteq_{\underline{P}}}(A)$	set of $\underline{P}$-maximin actions in A
$\max_{\sqsupseteq_{\overline{P}}}(A)$	set of $\underline{P}$-maximax actions in A
$\max_{\supseteq_{\underline{P}}}(A)$	set of weakly $\underline{P}$-maximal actions in A
$(x, k, u_\bullet)$	a path from initial state x at time k through application of the control sequence u_k, $u_{k+1}, \ldots$
$\mathcal{U}(x, k)$	the collection of all admissible paths starting in x at k
$(x, k, u_\bullet)_\ell \oplus (x_\ell, \ell, u_\bullet)_\ell$	concatenation of two paths
$(x, k, u_\bullet)_\ell \oplus \mathcal{V}$	concatenation of a path with every element of a set of paths
$J(x, k, u_\bullet)$	the gain random quantity induced by a path $(x, k, u_\bullet)$
$\mathcal{J}(x, k)$	the collection of all gain random quantities for admissible paths from initial state x at time k
$\underline{\mathbb{E}}(\bullet \vert x_1 \ldots x_k)$	marginal extension
π_k	a control law
$\pi_k(x_k \ldots x_{\ell-1})$	control law after observation of $x_k \ldots x_{\ell-1}$
$J_{\pi_k(x_k \ldots x_{\ell-1})}$	the gain gamble after observation of $x_k \ldots x_{\ell-1}$
$\Pi_k(x_k \ldots x_\ell, u_k \ldots u_{\ell-1})$	the set of control laws after observation of $x_k \ldots x_\ell$ and application of $u_k \ldots u_{\ell-1}$

Hoofdstuk 1

Een Nederlandstalige samenvatting van dit proefschrift (A Dutch Summary of this Dissertation)

If you can read English, please skip this chapter and proceed to p. 35.

Kun je geen Engels lezen, dan heb je het geluk om in dit hoofdstuk—in navolging van Artikel 102 van het FTW-reglement voor het doctoraatsexamen in de toegepaste wetenschappen—een korte Nederlandstalige samenvatting te vinden, die je in staat stelt om inzicht te krijgen in het volledige proefschrift en die voldoende gedetailleerd is om de essentie, de originaliteit, en de wetenschappelijke waarde van dit proefschrift te kunnen begrijpen.

1.1 Inleiding

Het doel van dit proefschrift is tweevoudig. Ten eerste, willen we onderzoeken in hoeverre onderprevisies andere bekende onzekerheidsmodellen verenigen en uitbreiden, en hoe ze tot een aanvaardbare notie van optimali-

teit leiden. Ten tweede, willen we onderzoeken hoe een dynamisch systeem, waarvan de onzekerheid beschreven wordt door zo'n geünificeerd model, optimaal geregeld kan worden, en we willen eveneens onderzoeken of het mogelijk is deze optimale regeling te vinden met behulp van een efficiënte methode, namelijk, dynamisch programmeren.

We verwijzen naar Tabel 1.1 voor een vertaling van de technische termen.

1.2 Onderprevisies en optimaliteit

In deze paragraaf introduceren we de basisconcepten van de theorie van de onderprevisies, die vooral ontwikkeld werd door Walley [86]. We demonstreren eveneens het unificerend karakter van deze theorie, en we geven aan hoe onderprevisies aanleiding geven tot een aantal welbekende noties van optimaliteit.

1.2.1 Een gedragsgericht onzekerheidsmodel

Toevallige veranderlijken, gokken, en prijzen

Een *toevallige veranderlijke* is een, mogelijk onzekere, maar waarneembare eigenschap van een systeem. Je kunt bijvoorbeeld denken aan de nog niet gekende uitkomst van een experiment. De verzameling van waarden die een toevallige veranderlijke X kan aannemen, noteren we met een kalligrafische letter $\mathcal{X}$, en een bepaalde waarde van X noemen we een *realisatie* van X en wordt genoteerd door een kleine letter x. Observeren we dat de realisatie van X gegeven is door x, dan schrijven we $X = x$.

We willen optimale regeling onder onzekerheid bestuderen. Specifieker, we willen onze kennis over de onzekere realisatie van de winst van een systeem modelleren, en beslissingen nemen—optimale regeling reduceert uiteindelijk tot het nemen van een beslissing—gebaseerd op deze kennis. In wat volgt zullen we een gedragsgericht onzekerheidsmodel suggereren. Deze modelleerkeuze is uitermate geschikt met het oog op optimale regeling, gezien optimale regeling op zichzelf natuurlijk een vorm van gedrag is. Bijgevolg leidt elk voldoend gesofistikeerd gedragsgericht kennismodel op een natuurlijke wijze tot een notie van optimaliteit. Dit zullen we demonstreren in paragraaf 1.2.4. Deze manier om naar kennis, en in het bijzonder, onzekerheid, te kijken, is natuurlijk niet nieuw: zie bijvoorbeeld Ramsey [65],

Tabel 1.1: Vertaling van de technische termen

Engels	Nederlands
behavioural	gedragsgericht
random variable	toevallige veranderlijke
realisation	realisatie
gamble	gok
random quantity	reëelwaardige functie
event	gebeurtenis
lower prevision	onderprevisie
upper prevision	bovenprevisie
prevision	previsie
conjugate upper prevision	toegevoegde bovenprevisie
self-conjugate	zelftoegevoegd
avoiding sure loss	zeker verlies vermijden
incuring sure loss	zeker verlies oplopen
coherent	coherent
vacuous	niets-zeggend
probability charge	waarschijnlijkheidslading
measure	maat
set function	verzamelingenfunctie
inner set function	binnenverzamelingenfunctie
possibility measure	possibiliteitsmaat
p-box	p-doos
natural extension	natuurlijke uitbreiding
linear extension	lineaire uitbreiding
lower envelope	onderomhullende
extended lower prevision	veralgemeende onderprevisie
act-state (in)dependence	actie-kennis(on)afhankelijkheid
Markov decision process	markovbeslisproces
Markov chain	markovketen
update	herzien

De Finetti [26], von Neumann en Morgenstern [83], Savage [69], Anscombe en Aumann [2], Khaneman en Tversky [47], Walley [85], von Winterfeld en Edwards [84], Seidenfeld, Schervish, en Kadane [72], en vele anderen.

We beginnen met een definitie: een *gok* f op een toevallige veranderlijke X is een reëelwaardige winst, uitgedrukt in een eenheid van utiliteit die vast verondersteld wordt, die begrensd is als functie van X. Wiskundig is een gok f op X een begrensde $\mathcal{X}$–$\mathbb{R}$-afbeelding, en een gok wordt dus geïnterpreteerd als een begrensde onzekere winst: als $X = x$, ontvangen we $f(x)$. Wanneer meerdere toevallige veranderlijken in het spel zijn, dan schrijven we soms ook $f(X)$ in plaats van f om te benadrukken dat f een gok is op de toevallige veranderlijke X. De verzameling van alle gokken op X noteren we als $\mathcal{L}(X)$. De beperking om alleen naar begrensde onzekere winsten te kijken is tot op zekere hoogte een wiskundig gemak. We zullen verderop de theorie grotendeels veralgemenen om ook over onbegrensde onzekere winsten iets te kunnen zeggen. Een reëel getal $a \in \mathbb{R}$ kan bijvoorbeeld geïdentificeerd worden met een constante gok $a(x) := a$ voor alle $x \in \mathcal{X}$. Een andere speciale klasse van gokken zijn deze die corresponderen tot zogenaamde *gebeurtenissen*. Een gebeurtenis op X is een deelverzameling van $\mathcal{X}$. Met een gebeurtenis A op X kunnen we een $\{0,1\}$-waardige gok I_A associëren, die ons één eenheid utiliteit geeft als de realisatie x van X tot A behoort, en anders niets. Deze gok I_A noemen we de *indicator* van A.

De *onderprevisie* $\underline{P}(f)$ van een gok f is gedefinieerd als de supremum koopprijs voor f: $\underline{P}(f)$ is de hoogste prijs s zodat voor elke $t < s$, we bereid zijn t te betalen alvorens X te observeren, als we garantie krijgen $f(x)$ te ontvangen na observatie van $X = x$. Wiskundig is een onderprevisie op X een reëelwaardige afbeelding gedefinieerd op een deelverzameling $\operatorname{dom}\underline{P}$, het domein van $\underline{P}$, van $\mathcal{L}(X)$. Een onderprevisie hoeft niet op alle gokken gedefinieerd te zijn. Verderop zullen we beschrijven hoe een onderprevisie kan uitgebreid worden tot de verzameling van alle gokken.

We kunnen een gok f natuurlijk ook interpreteren als een onzeker begrensd verlies. De *bovenprevisie* $\overline{P}(f)$ van f is dan de infimum verkoopprijs voor f: het is de laagste prijs s zodat voor elke $t > s$, we bereid zijn t te ontvangen alvorens X te observeren, als we garantie krijgen $f(x)$ te verliezen na observatie van $X = x$. Gezien een winst r equivalent is met een verlies $-r$, moet er gelden dat $\overline{P}(f) = -\underline{P}(-f)$: voor elke onderprevisie $\underline{P}$ bestaat er een zogenaamde *toegevoegde bovenprevisie* $\overline{P}$ op $\operatorname{dom}\overline{P} = -\operatorname{dom}\underline{P}$ die hetzelfde

gedrag modelleert. We kunnen ons daarom beperken tot de studie van onderprevisies, en we spreken af dat als $\underline{P}$ een onderprevisie voorstelt, $\overline{P}$ altijd zijn toegevoegde voorstelt.

Het kan gebeuren dat $\underline{P}$ *zelftoegevoegd* is: dit gebeurt precies wanneer $\operatorname{dom}\underline{P} = \operatorname{dom}\overline{P}$ en $\underline{P}(f) = \overline{P}(f)$ voor alle gokken $f \in \operatorname{dom}\underline{P}$. In dat geval schrijven we P in plaats van $\underline{P}$ of $\overline{P}$, als het duidelijk is uit de context of we het over aankoopprijzen of over verkoopprijzen hebben. Een zelftoegevoegde onderprevisie noemen we ook gewoonweg een *previsie*, in navolging van De Finetti [26], die previsies gebruikte om een gedragsgerichte interpretatie te geven voor de klassieke waarschijnlijkheidsleer, gebaseerd op de notie van verwachtingswaarde.

Het vermijden van zeker verlies, en coherentie

Onderprevisies impliceren dus een bereidheid bepaalde gokken te kopen voor een bepaalde prijs. Indien het mogelijk is om een combinatie van gokken in $\operatorname{dom}\underline{P}$ te vinden, en aanvaardbare aankoopprijzen voor deze gokken (strikt lager dan hun supremum aankoopprijs), zodanig dat we met zekerheid een strikt positieve hoeveelheid utiliteit verliezen als we die gokken kopen voor die aanvaardbare aankoopprijzen, dan zeggen we dat $\underline{P}$ *zeker verlies oploopt*. Zo'n combinatie van gokken en aankoopprijzen wordt in het Engels ook wel een 'Dutch book' genoemd. In het andere geval zeggen we dat $\underline{P}$ *zeker verlies vermijdt*. Uiteraard willen we dat onderprevisies zeker verlies vermijden; dit idee werd voor het eerst geopperd door Ramsey [65, p. 182].

Het kan ook gebeuren dat we bereid zijn een gok f voor een hogere prijs te kopen dan $\underline{P}(f)$, na overweging van andere combinaties van gokken en aanvaardbare aankoopprijzen. Is dit nimmer het geval, dan zeggen we dat $\underline{P}$ *coherent* is; dit idee is afkomstig van Williams [92]. Coherente onderprevisies voldoen aan een hele resem mooie eigenschappen; je kunt die vinden op blz. 55 (Theorem 3.5). Indien het domein van $\underline{P}$ een lineaire ruimte is, dan is $\underline{P}$ coherent als en slechts als ze de afbeelding inf (die een gok f afbeeldt op zijn infimum) puntsgewijs domineert, positief homogeen is, en superadditief is. Coherentie impliceert eveneens het vermijden van zeker verlies.

Voorbeelden van coherente onderprevisies

Als belangrijkste voorbeelden van coherente onderprevisies, vermelden we de volgende:

- waarschijnlijkheidsladingen—dit zijn eindig additieve waarschijnlijkheidsmaten, en waarschijnlijkheidsmaten zelf zijn hier dus een speciaal geval van: zij μ een waarschijnlijkheidslading, en definieer $\mathbf{P}_\mu(I_A) := -\mathbf{P}_\mu(-I_A) := \mu(A)$ voor alle A in het domein van μ, dan is $\mathbf{P}_\mu$ een coherente previsie.

- 2-monotone verzamelingenfuncties—hierbij is een verzamelingenfunctie een $[0,1]$-waardige monotone afbeelding ν, gedefinieerd op een veld van gebeurtenissen op $\mathcal{X}$, waarbij $\nu(\emptyset) = 0$ en $\nu(\mathcal{X}) = 1$, en deze wordt 2-monotoon genoemd als ze voldoet aan de ongelijkheid $\nu(A \cup B) + \nu(A \cap B) \geq \nu(A) + \nu(B)$ voor alle A en B in haar domein. Zij ν een 2-monotone verzamelingenfunctie en definiëren we $\underline{\mathbf{P}}_\nu(I_A) := \nu(A)$ voor alle A in het domein van ν, dan is $\underline{\mathbf{P}}_\nu$ een coherente onderprevisie.

- possibiliteitsmaten—dit zijn gewoonweg supremumbehoudende verzamelingenfuncties. Zij π een possibiliteitsmaat, en definiëren we $\overline{\mathbf{P}}_\pi(I_A) := \pi(A)$ voor alle A in het domein van π, dan is $\overline{\mathbf{P}}_\nu$ een coherente bovenprevisie.

Andere voorbeelden van coherente onderprevisies zijn:

- niets-zeggende onderprevisies ten opzichte van een niet-lege deelverzameling A van $\mathcal{X}$—zij zijn gedefinieerd door $\underline{P}_A(f) := \inf_{x \in A} f(x)$ voor alle gokken f op X.

- geneste verzamelingenfuncties—dit zijn verzamelingenfuncties gedefinieerd op een ketting van gebeurtenissen. Deze induceren coherente previsies.

- minitieve en maxitieve verzamelingenfuncties—dit zijn minimum- en maximumbehoudende verzamelingenfuncties. Minitieve induceren coherente onderprevisies, en maxitieve induceren coherente bovenprevisies.

- cumulatieve distributiefuncties op een gesloten interval $[a,b]$—zij F een reëelwaardige functie op $[a,b]$, en definiëren we $\mathbf{P}_F(I_{[a,x]}) := F(x)$

voor alle x in $[a,b]$, dan is $\mathbf{P}_F$ een coherente previsie als en slechts als $0 \leq F(x) \leq F(b) = 1$ voor alle x in $[a,b]$, en F monotoon stijgend is.

- p-dozen—zij F_* en F^* twee reëelwaardige functies op een gesloten interval $[a,b]$, en definiëer $\underline{\mathbf{P}}_{(F_*,F^*)}(I_{[a,x]}) := F_*(x)$ en $-\underline{\mathbf{P}}_{(F_*,F^*)}(-I_{[a,x]}) := F^*(x)$ voor alle x in $[a,b]$. Dan is $\underline{\mathbf{P}}_{(F_*,F^*)}$ een coherente onderprevisie als en slechts als $0 \leq F_*(x) \leq F^*(x) \leq F_*(b) = F^*(b) = 1$ voor alle x in $[a,b]$, en zowel F_* als F^* monotoon stijgend zijn.

1.2.2 Natuurlijke uitbreiding

Definitie en voorbeelden

De *natuurlijke uitbreiding*, ingevoerd door Walley [86], zegt hoe een onderprevisie uitbreidt naar een groter domein, *i.e.*, naar een grotere verzameling gokken, en is essentieel om besluiten te trekken omtrent systemen waarvan de onzekerheid beschreven wordt door onderprevisies, zoals optimaliteitsbesluiten. Indien een onderprevisie zeker verlies vermijdt, dan is haar natuurlijke uitbreiding naar een verzameling gokken gedefinieerd als haar meest conservatieve coherente uitbreiding naar die verzameling; in geval ze zeker verlies oploopt, heeft ze geen enkele coherente uitbreiding (zie ook Walley [86]). Als het domein van een onderprevisie eindig is, dan kan haar natuurlijke uitbreiding berekend worden aan de hand van een lineair programma (zie vergelijking (4.1) op blz. 96 voor details). Dit programma komt in essentie neer op het vinden van de hoogst mogelijke aanvaardbare aankoopprijs voor een bepaalde gok, op basis van aanvaardbare aankoopprijzen voor gokken in het domein van de onderprevisie.

Walley [86] beschouwt enkel natuurlijke uitbreiding tot de verzameling van alle gokken. We kunnen echter gemakkelijk aantonen dat natuurlijke uitbreiding tot een arbitraire verzameling van gokken (die het domein van de onderprevisie bevat) alle oorspronkelijke eigenschappen van natuurlijke uitbreiding behoudt. Dit is het gevolg van de *transitiviteit* van natuurlijke uitbreiding: een natuurlijke uitbreiding van een natuurlijke uitbreiding van een onderprevisie is opnieuw een natuurlijke uitbreiding van een onderprevisie (zie Corollary 4.9 op blz. 98 voor details).

Vele uitbreidingsmethoden uit de literatuur zijn speciale gevallen van natuurlijke uitbreiding. Definiëren we de *lineaire uitbreiding* van een onderpre-

visie als haar natuurlijke uitbreiding tot het grootst mogelijke domein waar deze uitbreiding zelftoegevoegd is, dan heeft deze lineaire uitbreiding alle eigenschappen van een (additieve) integraal. Dit nodigt uit tot de volgende definitie: noem een gok $\underline{P}$-*integreerbaar* indien hij behoort tot het domein van de lineaire uitbreiding van de onderprevisie $\underline{P}$. Dan kunnen we de volgende nieuwe resultaten bewijzen:

- Zij μ een waarschijnlijkheidslading. Een gok is $\mathbf{P}_\mu$-integreerbaar als en slechts als ze Dunford-integreerbaar is ten opzichte van μ, in welk geval de natuurlijke uitbreiding van $\mathbf{P}_\mu$ voor deze gok samenvalt met zijn Dunford-integraal (de Dunford-integraal werd ingevoerd door Dunford [31, p. 443, Sect. 3] en Dunford en Schwartz [30, Part I, Chapter III, Definition 2.17, p. 112]). Merk echter op dat de Dunford-integraal niet enkel gedefinieerd is voor gokken, maar ook voor mogelijk onbegrensde reëelwaardige functies.

- Zij μ een waarschijnlijkheidslading. De onder-S-integraal ten opzichte van μ valt samen met de natuurlijke uitbreiding van $\mathbf{P}_\mu$ (de S-integraal werd ingevoerd, onder variërende voorwaarden, door Moore en Smith [57, Section 5, p. 114, ll. 10–13], Kolmogoroff [50, Zweites Kapitel, §2, p. 663, Nr. 12], Hildebrandt [42, Sect. 1(f), p. 869], Gould [37, Definition 4.3, p. 201, en Definition 6.1& Theorem 6.2, p. 213], en Bhaskara Rao en Bhaskara Rao [9, Section 4.5]; wij gebruiken enkel de boven- en onder-S-integraal zoals gedefinieerd door Bhaskara Rao en Bhaskara Rao [9, Section 4.5]). Bijgevolg is een gok S-integreerbaar als en slechts als hij Dunford-integreerbaar is, en vallen de twee integralen samen op gokken (dit laatste feit was reeds bewezen op een compleet andere manier door Bhaskara Rao en Bhaskara Rao [9, Section 4.5]).

- Riemann- en Riemann-Stieltjes-integratie zijn eveneens een bijzonder geval van natuurlijke uitbreiding: de onder-Riemann-Stieltjes-integraal ten opzichte van een coherente cumulatieve distributiefunctie F, gedefinieerd analoog aan de manier waarop Darboux [14, Section II, pp. 64–71] de onder-Riemann-integraal definieert (zie Hildebrandt [43, Chapter II, pp. 27–32, in het bijzonder Definition 2.1, Definition 2.2, Theorem 3.2 en Theorem 3.10] voor een discussie), valt samen met de natuurlijke uitbreiding van de waarschijnlijkheidslading μ_F—die gedefinieerd is op het veld (*niet* het σ-veld) gegenereerd door alle in-

tervallen waarvan de indicator Riemann-Stieltjes-integreerbaar is, en die uniek bepaald is door de identiteit $\mu_F([x,y]) = F(y) - F(x)$ voor alle $a \leq x \leq y \leq b$ zodat $I_{[x,y]}$ Riemann-Stieltjes-integreerbaar is—als en slechts als F in elk punt ofwel linkscontinu ofwel rechtscontinu is. Noteer dat μ_F de restrictie is van de zogenaamde Lebesgue-Stieltjes-maat (zie Halmos [40, Section 15.9]) tot het veld gegenereerd door alle 'Riemann-Stieltjes-integreerbare intervallen'. Het is mogelijk om *alle* waarschijnlijkheidsladingen te karakteriseren waarvan de natuurlijke uitbreiding samenvalt met een gegeven onder-Riemann-Stieltjes-integraal: zie Theorem 4.52 op blz. 145 voor details. Als een bijzonder geval hiervan vermelden we dat de onder-Riemann-integraal over het eenheidsinterval $[0,1]$ samenvalt met de natuurlijke uitbreiding van de restrictie van de Lebesgue-maat tot het veld gegenereerd door alle intervallen in $[0,1]$.

- Als een gevolg van het voorgaande puntje, kunnen we eveneens gemakkelijk bewijzen dat de natuurlijke uitbreiding van een coherente cumulatieve distributiefunctie F op $[a,b]$ samenvalt met de onder-Riemann-Stieltjes-integraal als en slechts als $F(a) = 0$ en F rechtscontinu is in elk punt van het interval $[a,b)$.

- Eveneens als gevolg van de karakterisatie van de Riemann-integraal als speciaal geval van natuurlijke uitbreiding, vinden we twee nieuwe uitdrukkingen voor de Choquet-integraal (een niet-additieve integraal voor 2-monotone verzamelingenfuncties, zie Choquet [11, Section 48.1, p. 265] voor een algemene definitie), in termen van een Dunford-integraal, en in termen van een S-integraal.

We vermelden eveneens een aantal belangrijke reeds gekende resultaten uit de literatuur, betreffende 2-monotone verzamelingenfuncties en de Choquet-integraal:

- De Choquet-integraal van een 2-monotone verzamelingenfunctie ν valt samen met de natuurlijke uitbreiding van $\underline{\mathbf{P}}_\nu$ (Walley [85]).

- De binnenverzamelingenfunctie van een 2-monotone verzamelingenfunctie ν is 2-monotoon, en valt samen met de natuurlijke uitbreiding van $\underline{\mathbf{P}}_\nu$ tot de klasse van alle gebeurtenissen (Walley [85] en Walley [86, Corollary 3.1.9, p. 127]).

Al deze resultaten geven enerzijds aan hoe natuurlijke uitbreiding onder bepaalde voorwaarden kan berekend worden met behulp van traditionele integratietechnieken, en anderzijds hoe de theorie van de onderprevisies, en natuurlijke uitbreiding in het bijzonder, vele gekende onzekerheidstheorieen (en aanverwante wiskundige modellen, zoals integralen) omvat. Dit ondersteunt de visie op de theorie van de onderprevisies als unificerend model. Uiteraard zijn er ook onzekerheidsmodellen die niet omvat worden door onderprevisies; we vermelden hierbij als belangrijk voorbeeld het model gesuggereerd door Seidenfeld, Schervish, en Kadane [72].

Dualiteit

De natuurlijke uitbreiding van een onderprevisie valt samen met de onderomhullende van alle coherente previsies die haar domineren. Dit resultaat werd bewezen door Walley [86, Sections 3.3.3&3.4.1, pp. 134–136] in geval van natuurlijke uitbreiding tot alle gokken. We kunnen dit resultaat veralgemenen voor natuurlijke uitbreiding tot een kleinere verzameling gokken. Een bijzonder interessant gevolg van deze veralgemening, is dat indien alle gokken in het domein van een onderprevisie meetbaar zijn ten opzichte van een veld, we de natuurlijke uitbreiding tot alle gokken kunnen schrijven als een onderomhullende van onder-S-integralen ten opzichte van alle waarschijnlijkheidsladingen op dit veld waarvan de S-integraal de onderprevisie domineert; zie Corollary 4.88 op blz. 195 voor details. Dit resulteert in een eenvoudigere duale uitdrukking voor natuurlijke uitbreiding.

1.2.3 Cauchy-uitbreiding

Er is één welbekende interessante uitbreidingstechniek, ontwikkeld door Dunford [31] in integraaltheorie, die niet door natuurlijke uitbreiding gedekt wordt. Deze uitbreidingstechniek verdient onze aandacht omdat hij uitbreidt tot reëelwaardige functies die niet noodzakelijk begrensd zijn. Vele problemen in regeltheorie vereisen de behandeling van onbegrensde reële veranderlijken: bijvoorbeeld, een kwadratische kostenfunctie is onbegrensd als functie van de toestand en de regeling van het systeem. Daar wij optimale regeling op het oog hebben, dringt een studie van onbegrensde veranderlijken zich op. De uitbreiding van Dunford's idee (oorspronkelijk ontworpen ter uitbreiding van integralen tot onbegrensde functies, en gebaseerd op

Cauchy-rijen) tot onderprevisies laten we achterwege in deze samenvatting, en we beperken ons tot een losse formulering van de belangrijkste resultaten van deze studie:

- De theorie van de onderprevisies kan worden uitgebreid tot supremum aankoopprijzen voor reëelwaardige functies die niet noodzakelijk begrensd zijn, als we ook $-\infty$ en $+\infty$ als mogelijke supremum aankoopprijs beschouwen ($-\infty$ betekent nooit aankopen ongeacht de prijs, en $+\infty$ betekent altijd aankopen ongeacht de prijs). Een afbeelding van een verzameling reëelwaardige functies tot supremum aankoopprijzen noemen we een veralgemeende onderprevisie. De begrippen 'zeker verlies vermijden', 'coherentie', en 'natuurlijke uitbreiding' veralgemenen op een bijna voor de hand liggende wijze. Het dualiteitsresultaat, *i.e.*, natuurlijke uitbreiding als onderomhullende van veralgemeende previsies, veralgemeent slechts in beperkte mate: technische beperkingen op het domein van de veralgemeende onderprevisies zijn vereist.

- De natuurlijke uitbreiding van een onderprevisie tot een veralgemeende onderprevisie neemt de waarde $-\infty$ aan op elke reëelwaardige functie die onbegrensd is naar onder. Dat betekent dat de natuurlijke uitbreiding heel erg conservatief is: op basis van natuurlijke uitbreiding zijn we nooit bereid een reëelwaardige functie aan te kopen die onbegrensd is naar onder, ongeacht de prijs.

- De uitbreiding van een onderprevisie tot een veralgemeende onderprevisie met behulp van Cauchy-rijen (en die we daarom de *Cauchy-uitbreiding* noemen) is een reëelwaardige coherente veralgemeende onderprevisie, en is niet even conservatief als de natuurlijke uitbreiding.

- De Cauchy-uitbreiding valt samen met de Choquet-integraal voor onbegrensde veranderlijken, in het geval we vertrekken van een 2-monotone verzamelingenfunctie.

- In het algemeen geval, kan de Cauchy-uitbreiding geschreven worden als een onderomhullende van Dunford-integralen (naar analogie met de onderomhullende van onder-S-integralen in geval van gewone natuurlijke uitbreiding).

1.2.4 Optimaliteit

Beschouw een statisch systeem, waarop we een regeling a kunnen toepassen, vrij te kiezen uit een verzameling A van beschikbare regelingen. Elke regeling a in A induceert een reëelwaardige winst J_a. Vaak worden de winsten J_a beinvloed door veranderlijken die niet goed gekend zijn. In de onderstelling dat we deze veranderlijken kunnen modelleren door een toevallige veranderlijke X, beschouwen we de winsten J_a dus als een reëelwaardige functie van X.

Een traditionele manier om optimaliteit te definiëren, in geval we onzeker zijn over de winst van een systeem, bestaat erin alle onzekerheden te modelleren aan de hand van een waarschijnlijkheidsverdeling op X, en de *verwachte* winst te maximaliseren. Deze aanpak leidt onder andere tot de volgende problemen:

- Ongeacht de waarschijnlijkheidsverdeling, vinden we bijna altijd een unieke optimale regeling, wat verwonderlijk is in geval we weinig informatie hebben: maximaliseren van verwachte winst kan geen onbeslistheid modelleren.

- Indien het niet duidelijk is hoe we een waarschijnlijkheidsverdeling kunnen opstellen op basis van de gegevens, dan kan onze uiteindelijke (en noodzakelijkerwijs arbitraire) keuze van deze verdeling een onidentificeerbaar effect hebben op wat we optimaal noemen: het maximaliseren van verwachte winst omvat geen robuustheid.

- Het maximaliseren van verwachte winst reflecteert niet noodzakelijkerwijs de symmetrie van het oorspronkelijke probleem, tenzij we ons beperken tot een waarschijnlijkheidsverdeling die deze symmetrie heeft. Helaas is er vaak geen waarschijnlijkheidsverdeling die zowel de symmetrie van het probleem als de gegeven informatie reflecteert. Beschouw bijvoorbeeld een muntstuk, waarvan we enkel weten dat het kop of munt kan vallen. De gegeven informatie is symmetrisch wat betreft kop en munt, en de enige verdeling die deze symmetrie reflecteert is $p(\text{kop}) = p(\text{munt}) = \frac{1}{2}$. Deze verdeling weerspiegelt niet de gegeven informatie, die niets zegt over de waarschijnlijkheid van kop en munt.

In deze paragraaf geven we een overzicht van de belangrijkste noties van optimaliteit die we kunnen associëren met onderprevisies, om alternatieven

te vinden die aan bovenstaande argumenten weerstaan. Om deze verschillende noties met elkaar te kunnen vergelijken, veronderstellen we dat $\underline{P}$ een reëelwaardige coherente veralgemeende onderprevisie is, gedefinieerd op een lineaire tralie die op zijn minst de verzameling van alle constante gokken en alle winsten J_a, $a \in A$, bevat. Het domein van $\underline{P}$ voorzien we van de topologie bepaald door het volgende convergentiecriterium: we zeggen dat een rij f_n in het domein van $\underline{P}$ convergeert naar een element f in het domein van $\underline{P}$, indien $f_n(x)$ naar $f(x)$ convergeert voor alle mogelijke waarden x van X, en $\overline{P}(|f_n - f|)$ naar nul convergeert. Noteer dat deze topologie een combinatie is van de zwakke topologie en de topologie geïnduceerd door de semi-norm $\overline{P}(|\bullet|)$. De verzameling van alle veralgemeende previsies op het domein van $\underline{P}$ die $\underline{P}$ puntsgewijs domineren noteren we door $\mathcal{M}$: uit voorgaande resultaten volgt dat $\mathcal{M}$ niet-leeg, convex, en compact is ten opzichte van de topologie van puntsgewijze convergentie op leden van het domein van $\underline{P}$, en bovendien is $\underline{P}$ precies de onderomhullende van $\mathcal{M}$: $\underline{P}(f) = \min_{Q \in \mathcal{M}} Q(f)$ (het minimum wordt bereikt precies wegens de compactheid van $\mathcal{M}$ ten opzichte van de topologie van puntsgewijze convergentie op leden van het domein van $\underline{P}$).

$\underline{P}$-maximaliteit

Dit criterium, ingevoerd door Walley [86, Section 3.9.2, p. 161], is gebaseerd op paarsgewijze keuze. Voor twee regelingen a en b in A zeggen we dat we *a strikt verkiezen boven b*, en we schrijven $a >_{\underline{P}} b$, indien $\underline{P}(J_a - J_b) > 0$, of indien $J_a \geq J_b$ en $J_a \neq J_b$. De eerste voorwaarde betekent dat we bereid zijn een strikt positieve prijs te betalen om de winst J_a te krijgen en J_b te verliezen, wat duidelijk wijst op een strikte voorkeur voor a ten opzichte van b. Anderzijds, als $J_a \geq J_b$ en $J_a \neq J_b$, dan resulteert a in een hogere winst dan b, en daarom zullen we ook in dat geval a strikt verkiezen boven b.

De relatie $>_{\underline{P}}$ is een strikte partiële ordening, en er is veelal geen grootste element in A ten opzichte van $>_{\underline{P}}$. In plaats van op zoek te gaan naar een grootste element, gaan we dus beter op zoek naar ongedomineerde elementen: we noemen een regeling *a $\underline{P}$-maximaal* in A indien er geen regeling in A is die we strikt verkiezen boven a: $\max_{>_{\underline{P}}}(A) = \{a \in A : (\forall b \in A)(a \not<_{\underline{P}} b)\}$. Het bestaan van $\underline{P}$-maximale elementen is gegarandeerd indien $\{J_a : a \in A\}$ compact is.

$\mathcal{M}$-maximaliteit

$\mathcal{M}$-maximaliteit is een speciaal geval van wat Levi E-admissibiliteit noemt [54, Section 4.8], en kan gezien worden als een robuuste versie van het maximaliseren van verwachte winst: we noemen een regeling $\mathcal{M}$-*maximaal* indien ze Q-maximaal is voor minstens één Q in $\mathcal{M}$.

Als we $\mathcal{M}$ interpreteren als een verzameling van mogelijke verwachtingsoperatoren—maar we weten niet precies welk model in $\mathcal{M}$ het correcte is—dan selecteert $\mathcal{M}$-maximaliteit precies die regelingen die maximaal zijn onder één van de mogelijke modellen Q in $\mathcal{M}$.

Deze notie van optimaliteit valt samen met $\underline{P}$-maximaliteit indien A niet meer dan twee elementen bevat, of indien de verzameling van alle winsten $\{J_a : a \in A\}$ convex is. In het algemeen is de verzameling $\mathcal{M}$-maximale regelingen een deel van de verzameling $\underline{P}$-maximale regelingen.

$\underline{P}$-maximin en $\underline{P}$-maximax

Nog een populaire veralgemening van het maximaliseren van verwachte winst bestaat uit het ordenen van regelingen volgens de onderprevisie (of bovenprevisie) van hun winsten: noem een regeling $\underline{P}$-*maximin* in A als ze maximaal is ten opzichte van de ordening $\sqsupset_{\underline{P}}$ gedefinieerd door $a \sqsupset_{\underline{P}} b$ indien $\underline{P}(J_a) > \underline{P}(J_b)$, of indien $J_a \geq J_b$ en $J_a \neq J_b$. We noemen een regeling $\underline{P}$-*maximax* in A als ze maximaal is ten opzichte van de ordening $\sqsupset_{\overline{P}}$ gedefinieerd door $a \sqsupset_{\overline{P}} b$ indien $\overline{P}(J_a) > \overline{P}(J_b)$, of indien $J_a \geq J_b$ en $J_a \neq J_b$.

Een axiomatische studie van $\underline{P}$-maximin werd gegeven door Gilboa en Schmeidler [34].

Het is welbekend dat $\underline{P}$-maximin en $\underline{P}$-maximax regelingen eveneens $\underline{P}$-maximaal zijn. $\underline{P}$-maximin regelingen zijn niet noodzakelijk $\mathcal{M}$-maximaal, maar $\underline{P}$-maximax regelingen wel.

Intervaldominantie en zwakke $\underline{P}$-maximaliteit

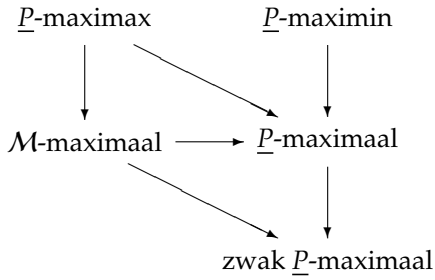
Een laatste veralgemening die we vermelden bestaat uit het ordenen van regelingen op de volgende manier: noem een regeling *zwak* $\underline{P}$-*maximaal* in A als ze maximaal is ten opzichte van de ordening $\supset_{\underline{P}}$ gedefinieerd door $a \supset_{\underline{P}} b$ indien $\underline{P}(J_a) > \overline{P}(J_b)$, of indien $J_a \geq J_b$ en $J_a \neq J_b$.

De ongelijkheid $\underline{P}(J_a) > \overline{P}(J_b)$ wordt ook wel *intervaldominantie* genoemd: ze zegt dat het interval $[\underline{P}(J_a), \overline{P}(J_a)]$ volledig rechts van $[\underline{P}(J_b), \overline{P}(J_b)]$ ligt. Zoals

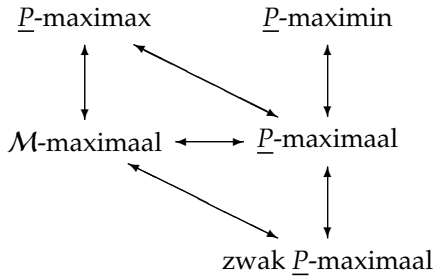
de naam reeds doet vermoeden is zwakke $\underline{P}$-maximaliteit zwakker dan $\underline{P}$-maximaliteit: elke $\underline{P}$-maximale regeling is eveneens zwak $\underline{P}$-maximaal.

Verbanden en belangrijkste eigenschappen

- In het algemeen gelden de volgende implicaties:

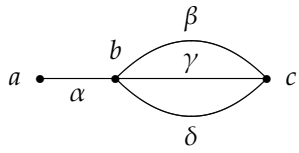


- Als $\underline{P}$ zelftoegevoegd is, dan is $\mathcal{M} = \{\underline{P}\}$, en alle optimaliteitscriteria vallen samen met het maximaliseren van de verwachte winst (als we $\underline{P}$ als verwachtingsoperator beschouwen):



- Alle criteria afgeleid uit paarsgewijze voorkeur (*i.e.*, $\underline{P}$-maximaliteit, zwakke $\underline{P}$-maximaliteit, $\underline{P}$-maximin, en $\underline{P}$-maximax) voldoen aan de volgende eigenschap: als $\{J_a : a \in A\}$ compact is, dan is er voor elke niet-optimale regeling a een optimale regeling b zodat b strikt verkozen wordt boven a. We zullen later zien dat deze eigenschap cruciaal is ter veralgemening van dynamisch programmeren.

- Voor al deze criteria, behalve voor $\underline{P}$-maximin en $\underline{P}$-maximax, geldt: als $\underline{P}$ puntsgewijs gedomineerd wordt door $\underline{Q}$, dan impliceert optimaliteit ten opzichte van $\underline{Q}$ eveneens optimaliteit ten opzichte van $\underline{P}$. Dus,



Figuur 1.1: Het optimaliteitsprincipe

hoe sterker onze disposities, *i.e.*, hoe hoger onze aankoopprijzen, hoe kleiner onze verzameling optimale regelingen, en dus, hoe éénduidiger onze optimale beslissingen bepaald zijn. $\underline{P}$-maximin en $\underline{P}$-maximax voldoen hier niet aan: ze impliceren een duidelijke, vaak zelfs éénduidige beslissing, zelfs als onze aankoopprijzen zeer laag zijn. Net zoals bij het maximaliseren van verwachte winst, falen zij in het modelleren van onbeslistheid.

1.3 Dynamisch programmeren

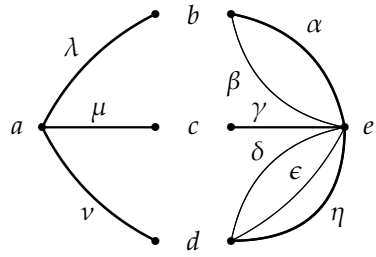
In deze paragraaf, beantwoorden we de tweede grote vraag van dit proefschrift: we onderzoeken in welke mate dynamisch programmeren kan toegepast worden op dynamische discretetijdssystemen waarvan de onzekerheid omtrent winst en dynamica beschreven wordt door onderprevisies.

1.3.1 Het deterministische geval met onzekere winst

In geval enkel de winst onzeker is, kunnen we, in het algemeen, dynamisch programmeren enkel gebruiken om $\underline{P}$-maximale en $\mathcal{M}$-maximale paden te vinden (zie De Cooman en Troffaes [24, 23]). Eerst beschrijven we wat dynamisch programmeren precies inhoudt, en dan onderzoeken we onder welke voorwaarden dit algoritme kan toegepast worden, en voor welke noties van optimaliteit deze voorwaarden voldaan zijn.

Inleiding

Dynamisch programmeren is een efficiënte recursieve methode om optimale paden van een systeem te bepalen, en werd ontworpen door Bellman [4]. De werking van het algoritme is gebaseerd op het optimaliteitsprincipe. Laat ons dit principe uitleggen aan de hand van Figuren 1.1 en 1.2. Figuur 1.1 schetst



Figuur 1.2: Dynamisch programmeren

een systeem dat op drie manieren van toestand a naar toestand c, via toestand b, kan evolueren: via de paden $\alpha\beta$, $\alpha\gamma$, of $\alpha\delta$. De winsten geassocieerd met elk van deze paden noteren we door $J_{\alpha\beta}$, $J_{\alpha\gamma}$, en $J_{\alpha\delta}$ respectievelijk. Onderstel dat $\alpha\gamma$ optimaal is: dus $J_{\alpha\gamma} > J_{\alpha\beta}$ en $J_{\alpha\gamma} > J_{\alpha\delta}$. Dan volgt dat pad γ optimaal is om van b naar c te gaan. Inderdaad, daar $J_{\alpha\nu} = J_\alpha + J_\nu$ voor $\nu \in \{\beta, \gamma, \delta\}$ (we onderstellen dat winsten additief zijn langs paden) kunnen we uit bovenstaande ongelijkheden afleiden dat $J_\gamma > J_\beta$ en $J_\gamma > J_\delta$. Deze eenvoudige observatie, die Bellman het *optimaliteitsprincipe* noemde, vormt de basis voor de recursieve techniek van dynamisch programmeren.

Om te zien hoe dit in zijn werk gaat, beschouw Figuur 1.2. Onderstel dat we een optimale weg om van a naar e te gaan willen vinden. Na één tijdstap, kunnen we de toestanden b, c, en d bereiken vanuit a, en de optimale paden vanuit deze toestanden naar de finale toestand e zijn gekend: α, γ en η, respectievelijk. Om nu de optimale paden van a naar e te vinden, moeten we enkel de winsten $J_\lambda + J_\alpha$, $J_\mu + J_\gamma$, en $J_\nu + J_\eta$ van de optimale paden $\lambda\alpha$, $\mu\gamma$, en $\nu\eta$ vergelijken, gezien het optimaliteitsprincipe zegt dat de paden $\lambda\beta$, $\nu\delta$, en $\nu\epsilon$ niet optimaal kunnen zijn: waren ze dat toch, dan zouden ook β, δ, en ϵ optimaal moeten zijn. Deze observatie, is wat in essentie gekend is als de *Bellman-vergelijking*, en stelt ons in staat een regelprobleem behoorlijk efficient op te lossen aan de hand van een recursieve procedure, door achterwaarts de optimale paden te berekenen vanuit de finale toestand.

Uiteraard gebeurt het vaak dat de winst onzeker is. Het is gebruikelijk deze onzekerheid te modelleren aan de hand van een waarschijnlijkheidsverdeling, en de verwachte winst te maximaliseren. Dynamisch programmeren blijft dan nog steeds mogelijk, ten gevolge van de lineariteit van de verwachtingsoperator. We hebben echter reeds betoogd dat deze weg in sommige

situaties niet tot de gewenste resultaten leidt. De vraag rijst dus of dynamisch programmeren ook kan toegepast worden als we de onzekerheid, en dus ook optimaliteit, beschrijven aan de hand van onderprevisies, die, zoals we eveneens reeds betoogd hebben, de klassieke aanpak veralgemenen tot situaties waar we te weinig informatie hebben om een verdeling te identificeren.

Op dit punt merken we op dat ook andere auteurs, zoals Satia en Lave [68], White en Eldeib [90], Givan, Leach, en Dean [36], en Harmanec [41], dynamisch programmeren veralgemeend hebben tot systemen met onzekere winst en/of onzekere dynamica, waar onzekerheid gemodelleerd wordt door onderprevisies (of, wat min of meer equivalent is, verzamelingen van waarschijnlijkheidsverdelingen). Maar geen van deze auteurs vraagt zich af in welke zin hun veralgemeende methoden tot optimale paden leidt. Wij benaderen het probleem vanuit een ander perspectief: we definiëren eerst een notie van optimaliteit, en pas dan onderzoeken we of dynamisch programmeren mogelijk is, in plaats van blindweg Bellman's algoritme te veralgemenen zonder aan te tonen in welke zin de gevonden paden optimaal zijn.

Discretetijdssystemen

Dynamica Voor a en b in $\mathbb{N}$, noteren we de verzameling van alle natuurlijke getallen c die voldoen aan $a \leq c \leq b$ als $[a, b]$. De vergelijking $x_{k+1} = f(x_k, u_k, k)$ beschrijft een discretetijdssysteem met $k \in \mathbb{N}$, $x_k \in \mathcal{X}$, en $u_k \in \mathcal{U}$. De verzameling $\mathcal{X}$ is de toestandsruimte, en $\mathcal{U}$ is de regelruimte De afbeelding $f : \mathcal{X} \times \mathcal{U} \times \mathbb{N} \to \mathcal{X}$ beschrijft de evolutie van de toestand in de tijd: gegeven toestand $x_k \in \mathcal{X}$ en regeling $u_k \in \mathcal{U}$ op tijdstip $k \in \mathbb{N}$, geeft $f(x_k, u_k, k)$ de volgende toestand x_{k+1} van het systeem. We leggen een eindtijd N op: na dit tijdstip zijn we niet meer geïnteresseerd in de dynamica van het systeem. Het mag eveneens voorkomen dat niet alle toestanden en regelingen toegestaan zijn op alle tijdstippen: we eisen dat x_k behoort tot de verzameling van toegestane toestanden $\mathcal{X}_k$ voor elk tijdstip $k \in [0, N]$, en dat u_k behoort tot de verzameling van toegestane regelingen $\mathcal{U}_k$ voor elk tijdstip $k \in [0, N-1]$, waar $\mathcal{X}_k \subseteq \mathcal{X}$ en $\mathcal{U}_k \subseteq \mathcal{U}$ uiteraard gegeven zijn.

Paden Een pad is een triplet $(x, k, u_\bullet)$, waar $x \in \mathcal{X}$ een toestand is, $k \in [0, N]$ een tijdstip, en $u_\bullet : [k, N-1] \to \mathcal{U}$ een rij van regelingen. Deze gegevens bepalen een uniek toestandstraject $x_\bullet : [k, N] \to \mathcal{X}$, recursief gedefinieerd

door $x_k = x$ en $x_{\ell+1} = f(x_\ell, u_\ell, \ell)$ voor elke $\ell \in [k, N-1]$. Een pad is *toelaatbaar* als $x_\ell \in \mathcal{X}_\ell$ voor elke $\ell \in [k, N]$, en $u_\ell \in \mathcal{U}_\ell$ voor elke $\ell \in [k, N-1]$. De unieke afbeelding van de lege verzameling naar $\mathcal{U}$ noteren we als $u_\emptyset$. Als $k = N$, dan doet $u_\bullet$ niets, en is gelijk aan $u_\emptyset$. Het unieke pad dat vertrekt en eindigt in toestand x op tijdstip $k = N$ kunnen we dan noteren als $(x, N, u_\emptyset)$.

De verzameling toelaatbare paden vanuit toestand $x \in \mathcal{X}$ op tijdstip $k \in [0, N]$ noteren we als $\mathcal{U}(x, k)$. Bijvoorbeeld, $\mathcal{U}(x, N) = \{(x, N, u_\emptyset)\}$ als $x \in \mathcal{X}_N$, en $\mathcal{U}(x, N) = \emptyset$ in het andere geval.

We kunnen ook paden beschouwen met eindtijd M verschillend van N, en we noteren die als $(x, k, u_\bullet)_M$ (in de onderstelling dat $k \leq M \leq N$). We kunnen $(x, k, u_\bullet)_k$ identificeren met $(x, k, u_\emptyset)_k$: het unieke pad, lengte nul, dat op tijdstip k start en eindigt in x. Zij $0 \leq k \leq \ell \leq m$. We kunnen twee paden $(x, k, u_\bullet)_\ell$ en $(y, \ell, v_\bullet)_m$ aaneenschakelen als $y = x_\ell$, en deze aaneenschakeling noteren we als $(x, k, u_\bullet, \ell, v_\bullet)_m$ of als $(x, k, u_\bullet)_\ell \oplus (y, \ell, v_\bullet)_m$.

De verzameling van alle toelaatbare paden die vanuit x vertrekken op tijdstip k en eindigen op tijdstip $\ell \in [k, N]$, noteren we als $\mathcal{U}(x, k)_\ell$. Voor een pad $(x, k, u_\bullet)_\ell \in \mathcal{U}(x, k)_\ell$ en een verzameling paden $\mathcal{V} \subseteq \mathcal{U}(x_\ell, \ell)$, noteren we ook $(x, k, u_\bullet)_\ell \oplus \mathcal{V} = \{(x, k, u_\bullet)_\ell \oplus (x_\ell, \ell, v_\bullet) \colon (x_\ell, \ell, v_\bullet) \in \mathcal{V}\}$.

Winstfuncties Het toepassen van regeling $u \in \mathcal{U}$ op het systeem in toestand $x \in \mathcal{X}$ op tijdstip $k \in [0, N-1]$, resulteert in een reëelwaardige winst $g(x, u, k, \omega)$. Bereiken we de toestand $x \in \mathcal{X}$ op het eindtijdstip N, dan krijgen we eveneens een winst, gegeven door $h(x, \omega)$. De parameter ω verzamelt alle veranderlijken die de winst beïnvloeden. Kenden we de exacte waarde van ω, dan zouden we eveneens de exacte waarde van de winsten kennen. De waarde van ω is echter onzeker, en dus beschouwen we ω als de uitkomst van een toevallige veranderlijke Ω die waarden aanneemt in een verzameling $\mathbf{\Omega}$. De winsten zijn dus eveneens onzeker, en we beschouwen ze als reëelwaardige functies van Ω. Het is belangrijk op te merken dat Ω enkel de winsten beïnvloedt, en geen effect heeft op de systeemdynamica. We onderstellen ook dat onze kennis over Ω niet beïnvloed wordt door de begintoestand van het systeem, of de regelingen die we erop uitoefenen—dit noemt men *actie-kennisonafhankelijkheid*.

We zullen eveneens enkel het belangrijke geval beschouwen waar de winstfuncties additief zijn langs paden van het systeem: aan het pad $(x, k, u_\bullet)$ koppelen we een winstfunctie $J(x, k, u_\bullet, \omega) = \sum_{i=k}^{N-1} g(x_i, u_i, i, \omega) + h(x_N, \omega)$,

voor elke $\omega \in \mathbf{\Omega}$. Als $M < N$, dan definiëren we eveneens $J(x, k, u_\bullet, \omega)_M = \sum_{i=k}^{M-1} g(x_i, u_i, i, \omega)$. Aan de lege regeling koppelen we winst nul: we definiëren $J(x, k, u_\bullet, \omega)_k = 0$, voor éénder welk tijdstip k.

Gegeven een systeem in initiële toestand $x \in \mathcal{X}$ op tijdstip $k \in [0, N]$, kunnen we een regeling $u_\bullet \colon [k, N-1] \to \mathcal{U}$ vinden, die resulteert in een toelaatbaar pad $(x, k, u_\bullet)$, en zodat de winstfunctie $J(x, k, u_\bullet, \omega)$ optimaal is?

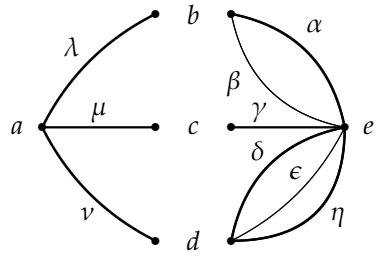
Is ω gekend, dan reduceert dit probleem zich tot een klassiek regelprobleem, en kan opgelost worden met behulp van dynamisch programmeren. We onderstellen hier dat de beschikbare kennis over Ω gemodelleerd wordt aan de hand van een coherente veralgemeende onderprevisie $\underline{P}$ gedefiniëerd op een voldoend grote verzameling $\operatorname{dom} \underline{P}$ van reëelwaardige functies op Ω.

Noteer dat voor een gegeven pad $(x, k, u_\bullet)$, de winstfunctie $J(x, k, u_\bullet, \omega)$ gezien kan worden als een reëelwaardige afbeelding op $\mathbf{\Omega}$, die we noteren als $J(x, k, u_\bullet)$. Er geldt dat $J(x, k, u_\bullet, \ell, v_\bullet)_m = J(x, k, u_\bullet)_\ell + J(x_\ell, \ell, v_\bullet)_m$ voor $k \leq \ell \leq m \leq N$, en $J(x, k, u_\bullet)_k = 0$. We definiëren eveneens de verzameling $\mathcal{J}(x, k)$ van alle winstfuncties gekoppeld aan toelaatbare paden vanuit $x \in \mathcal{X}_k$ op tijdstip $k \in [0, N]$: $\mathcal{J}(x, k) = \{J(x, k, u_\bullet) \colon (x, k, u_\bullet) \in \mathcal{U}(x, k)\}$. Dus onderstellen we dat $\underline{P}$ een reëelwaardige coherente veralgemeende onderprevisie is, en dat $\operatorname{dom} \underline{P}$ een lineaire tralie is die op zijn minst alle constante gokken bevat, en alle winstfuncties $g(x_k, u_k, k)$ en $h(x_N)$, voor alle $k \in [0, N-1]$, alle x_k in $\mathcal{X}_k$, alle u_k in $\mathcal{U}_k$, en alle x_N in $\mathcal{X}_N$. Het domein $\operatorname{dom} \underline{P}$ van $\underline{P}$ kunnen we dan voorzien van de topologie beschreven in paragraaf 1.2.4.

Voorwaarden voor dynamisch programmeren onder een algemene notie van optimaliteit

In paragraaf 1.2.4, hebben we vijf verschillende manieren beschreven waarop we met een onderprevisie een optimaal pad—een pad is een actie in de context van dynamische systemen—kunnen associëren. We zoeken nu uit, voor welke van deze verschillende types optimaliteit, we dynamisch programmeren mogen toepassen ter oplossing van het corresponderende regelprobleem.

Laat ons daartoe Bellman's analyse iets nauwkeuriger onderzoeken, en uitvissen aan welke eigenschappen een generieke notie van optimaliteit moet voldoen, om dynamisch programmeren mogelijk te maken. Beschouw een eigenschap, genaamd $*$-optimaliteit, die een pad in een gegeven verzameling $\mathcal{P}$ van paden ofwel heeft, ofwel niet heeft. Indien een pad in $\mathcal{P}$ deze



Figuur 1.3: Een algemenere vorm van dynamisch programmeren

eigenschap heeft, dan zeggen we dat het $*$-optimaal is in $\mathcal{P}$. We noteren de verzameling van $*$-optimale elementen van $\mathcal{P}$ door $\text{opt}_*(\mathcal{P})$. Per definitie geldt dat $\text{opt}_*(\mathcal{P}) \subseteq \mathcal{P}$. Verderop zullen we onze bevindingen toepassen op de reeds beschreven concrete optimaleitsnoties.

Het optimaliteitsprincipe Beschouw Figuur 1.3. We willen de $*$-optimale paden van a naar e bepalen. Na één tijdstap kunnen we de toestanden b, c, en d bereiken vanuit a. De $*$-optimale paden vanuit deze toestanden zijn α, γ, en δ en η, respectievelijk. Opdat Bellman's algoritme zou werken, moeten de $*$-optimale paden vanuit a naar e, *a priori* gegeven door $\text{opt}_*(\{\lambda\alpha, \lambda\beta, \mu\gamma, \nu\delta, \nu\epsilon, \nu\eta\})$, ook gegeven zijn door $\text{opt}_*(\{\lambda\alpha, \mu\gamma, \nu\delta, \nu\eta\})$, *i.e.*, de $*$-optimale paden in de verzameling van aaneenschakelingen van λ, μ, en ν met de respectieve $*$-optimale paden α, γ, en δ en η. We moeten dus uitsluiten dat de aaneenschakelingen $\lambda\beta$ en $\nu\epsilon$ met de niet-$*$-optimale paden β en ϵ $*$-optimaal kunnen zijn. Dit komt neer op de vereiste dat de operator opt_* aan een veralgemening van Bellman's optimaliteitsprincipe moet voldoen, en ons zodoende doet besluiten dat als β en ϵ niet $*$-optimaal zijn, dan $\lambda\beta$ en $\nu\epsilon$ eveneens niet $*$-optimaal. Concreet zeggen we dat $*$-optimaliteit *voldoet aan het optimaliteitsprincipe* indien voor alle $k \in [0, N]$, $x \in \mathcal{X}_k$, $\ell \in [k, N]$, en $(x, k, u_\bullet) \in \mathcal{U}(x, k)$ de volgende implicatie geldt: als $(x, k, u_\bullet)$ $*$-optimaal is in $\mathcal{U}(x, k)$, dan is $(x_\ell, \ell, u_\bullet)$ $*$-optimaal in $\mathcal{U}(x_\ell, \ell)$. Dit kan ook geschreven worden als:

$$\text{opt}_*(\mathcal{U}(x,k)) \subseteq \bigcup_{(x,k,u_\bullet)_\ell \in \mathcal{U}(x,k)_\ell} (x,k,u_\bullet)_\ell \oplus \text{opt}_*(\mathcal{U}(x_\ell,\ell)) \,.$$

De Bellman-vergelijking zegt dat we gelijkheid krijgen als we de $*$-optimaliteitsoperator laten inwerken op het rechterlid. Dit wordt normaal geformuleerd met $\ell = k + 1$. Echter, en dit misschien tot de grote verwondering voor lezers die vertrouwd zijn met de traditionele vorm van dynamisch programmeren, hiertoe moet opt_* aan nog een eigenschap voldoen.

Ongevoeligheid voor het weglaten van niet-optimale paden Inderdaad, het weglaten van de niet-$*$-optimale paden $\lambda\beta$ en $\nu\epsilon$ uit de verzameling van kandidaat $*$-optimale paden mag geen effect hebben op de uiteindelijke verzameling $*$-optimale paden: er moet gelden dat

$$\text{opt}_*\left(\{\lambda\alpha, \lambda\beta, \mu\gamma, \nu\delta, \nu\epsilon, \nu\eta\}\right) = \text{opt}_*\left(\{\lambda\alpha, \mu\gamma, \nu\delta, \nu\eta\}\right).$$

Dit is uiteraard voldaan voor de traditionele vorm van optimaliteit—maximaliseren van verwachte winst—maar dit hoeft niet voldaan te zijn voor de meer abstracte types optimaliteit die we zullen beschouwen. De gelijkheid is gegarandeerd als opt_* ongevoelig is voor het weglaten van niet-$*$-optimale elementen uit $\{\lambda\alpha, \lambda\beta, \mu\gamma, \nu\delta, \nu\epsilon, \nu\eta\}$, in de volgende zin: beschouw een niet-lege verzameling S en een optimaliteitsoperator opt_* gedefinieerd op de verzameling $\wp(S)$ van deelverzamelingen van S zodat $\text{opt}_*(T) \subseteq T$ voor alle $T \subseteq S$. Dan noemen we opt_* *ongevoelig voor het weglaten van niet-$*$-optimale elementen van S* indien $\text{opt}_*(S) = \text{opt}_*(T)$ voor alle T zodat $\text{opt}_*(S) \subseteq T \subseteq S$.

In geval $*$-optimaliteit geassocieerd is met een (familie van) strikt partiële ordening(en), dan is aan deze voorwaarde voldaan indien elk niet-$*$-optimaal pad gedomineerd wordt door een optimaal pad. Concreet, zij S een niet-lege verzameling voorzien van een familie strikte partiële ordeningen $>_j$, $j \in J$. Definieer voor elke $T \subseteq S$ de verzameling $\text{opt}_{>_j}(T) := \left\{a \in T \colon (\forall b \in T)(b \not>_j a)\right\}$ als de verzameling van maximale elementen van T ten opzichte van $>_j$, en zij $\text{opt}_J(T) := \bigcup_{j\in J} \text{opt}_{>_j}(T)$, en noem elementen van deze verzameling J-optimaal in T. Dan zijn $\text{opt}_{>_j}$, $j \in J$ en opt_J optimaliteitsoperatoren. Bovendien, als voor $j \in J$

$$(\forall a \in S \setminus \text{opt}_{>_j}(S))(\exists b \in \text{opt}_{>_j}(S))(b >_j a), \tag{1.1}$$

dan is $\text{opt}_{>_j}$ ongevoelig voor het weglaten van niet-$>_j$-maximale elementen van S. Als bovenstaande eigenschap geldt voor alle $j \in J$, dan is opt_J ongevoelig voor het weglaten van niet-J-optimale elementen van S.

De veralgemeende Bellman-vergelijking Zij $k \in [0, N]$ en $x \in \mathcal{X}_k$. Indien $*$-optimaliteit voldoet aan het optimaliteitsprincipe, en indien opt_* ongevoelig is voor het weglaten van niet-$*$-optimale elementen van $\mathcal{U}(x,k)$, dan kunnen we aantonen dat voor alle $\ell \in [k, N]$ geldt dat

$$\mathrm{opt}_*\left(\mathcal{U}(x,k)\right) = \mathrm{opt}_*\left(\bigcup_{(x,k,u)_\ell \in \mathcal{U}(x,k)_\ell} (x,k,u)_\ell \oplus \mathrm{opt}_*\left(\mathcal{U}(x_\ell, \ell)\right)\right),$$

dus, in dat geval is een pad $*$-optimaal als en slechts als ze een $*$-optimale aaneenschakeling is van een toelaatbaar pad $(x,k,u_\bullet)_\ell$ en een $*$-optimaal pad in $\mathcal{U}(x_\ell, \ell)$.

Laat ons nu deze algemene resultaten toepassen op de concrete types optimaliteit die we reeds behandeld hebben. Voor alle vijf optimaliteitsoperatoren $\mathrm{opt}_{>_{\underline{P}}}$, $\mathrm{opt}_{\mathcal{M}}$, $\mathrm{opt}_{\sqsupset_{\underline{P}}}$, $\mathrm{opt}_{\sqsupset_{\overline{P}}}$, en $\mathrm{opt}_{\supset_{\underline{P}}}$, gaan we na of we de Bellman-vergelijking kunnen gebruiken ter oplossing van het corresponderende regelprobleem.

Onderzoek van de voorwaarden voor dynamisch programmeren bij een aantal concrete vormen van optimaliteit

$\underline{P}$-maximaliteit We beschouwen eerst de optimaliteitsoperator $\mathrm{opt}_{>_{\underline{P}}}$ die uit een verzameling paden, die paden selecteert, die maximaal zijn ten opzichte van de strikt partiële ordening $>_{\underline{P}}$. Merk eerst en vooral op dat $>_{\underline{P}}$ een vectorordening is op $\mathrm{dom}\,\underline{P}$: als $f >_{\underline{P}} g$, dan $f + h >_{\underline{P}} g + h$, voor eender welke reëelwaardige functies f, g, en h in $\mathrm{dom}\,\underline{P}$. Uitgaande van deze observatie, is het is eenvoudig aan te tonen dat $\underline{P}$-maximaliteit voldoet aan het optimaliteitsprincipe.

In paragraaf 1.2.4 hebben we reeds vermeld dat als $\mathcal{J}(x,k)$ compact is, elk niet-$\underline{P}$-maximaal element in $\mathcal{J}(x,k)$ gedomineerd wordt door een $\underline{P}$-maximaal element in $\mathcal{J}(x,k)$, of wat equivalent is, elk niet-$\underline{P}$-maximaal pad in $\mathcal{U}(x,k)$ gedomineerd wordt door een $\underline{P}$-maximaal pad in $\mathcal{U}(x,k)$. Bijgevolg, wegens een eigenschap die we eveneens reeds vermeld hebben bij de behandeling van ongevoeligheid voor het weglaten van niet-$*$-optimale elementen, volgt dat als $\mathcal{J}(x,k)$ compact is, dan is $\mathrm{opt}_{>_{\underline{P}}}$ ongevoelig voor het weglaten van niet-$\underline{P}$-maximale paden van $\mathcal{U}(x,k)$.

We besluiten: zij $k \in [0, N]$ en $x \in \mathcal{X}_k$, en onderstel dat $\mathcal{J}(x,k)$ compact is,

dan geldt voor alle $\ell \in [k, N]$ dat

$$\mathrm{opt}_{>_{\underline{P}}}(\mathcal{U}(x,k)) = \mathrm{opt}_{>_{\underline{P}}}\left(\bigcup_{(x,k,u)_\ell \in \mathcal{U}(x,k)_\ell} (x,k,u)_\ell \oplus \mathrm{opt}_{>_{\underline{P}}}(\mathcal{U}(x_\ell,\ell))\right).$$

Deze Bellman-vergelijking voor $\underline{P}$-maximaliteit resulteert in een recursieve methode om alle $\underline{P}$-maximale paden te bepalen. Inderdaad, neem $\mathrm{opt}_{>_{\underline{P}}}(\mathcal{U}(x,N)) = \{u_\emptyset\}$ voor elke $x \in \mathcal{X}_N$, en $\mathrm{opt}_{>_{\underline{P}}}(\mathcal{U}(x,k))$ kan recursief bepaald worden uit de Bellman-vergelijking. Noteer dat deze vergelijking ook een methode geeft ter constructie van een $\underline{P}$-maximale terugkoppelwet: voor elke $x \in \mathcal{X}_k$, kies éénder welke $(x,k,u^*_\bullet(x,k)) \in \mathrm{opt}_{>_{\underline{P}}}(\mathcal{U}(x,k))$. Dan realiseert $\phi(x,k) = u^*_k(x,k)$ een $\underline{P}$-maximale terugkoppelwet.

$\mathcal{M}$-maximaliteit Gelijkaardige overwegingen leiden tot het volgende besluit: de optimaliteitsoperator $\mathrm{opt}_{\mathcal{M}}$ voldoet aan het optimaliteitsprincipe, en als $\mathcal{J}(x,k)$ compact is, dan is $\mathrm{opt}_{\mathcal{M}}$ ongevoelig voor het weglaten van niet-$\mathcal{M}$-maximale paden van $\mathcal{U}(x,k)$. Bijgevolg kunnen we ook voor $\mathcal{M}$-maximaliteit een Bellman-vergelijking neerschrijven.

$\underline{P}$-maximin, $\underline{P}$-maximax, en zwakke $\underline{P}$-maximaliteit Als $\mathcal{J}(x,k)$ compact is, dan is de optimaliteitsoperator $\mathrm{opt}_{\sqsupset_{\underline{P}}}$ ongevoelig voor het weglaten van niet-$\underline{P}$-maximin paden van $\mathcal{U}(x,k)$, en gelijkaardige observaties gelden voor $\mathrm{opt}_{\sqsupset_{\overline{P}}}$ en $\mathrm{opt}_{\supset_{\underline{P}}}$. Maar helaas, zoals met tegenvoorbeelden kan aangetoond worden, geldt het optimaliteitsprincipe noch voor $\underline{P}$-maximin, noch voor $\underline{P}$-maximax, noch voor zwakke $\underline{P}$-maximaliteit. Bijgevolg vinden we geen Bellman-vergelijking. In essentie, komt dit omdat de strikt partiële ordeningen $\sqsupset_{\underline{P}}$, $\sqsupset_{\overline{P}}$, en $\supset_{\underline{P}}$ geen vectorordeningen zijn—ze zijn niet verenigbaar met de additiviteit van winstfuncties langs paden.

Dit toont ook aan dat de Bellman-vergelijking voorgesteld door Harmanec [41], op basis van zwakke $\underline{P}$-maximaliteit, niet noodzakelijk tot zwak-$\underline{P}$-maximale paden leidt.

Toepassing: het aligneren van genetische sequenties onder onzekere evolutionaire afstand

Het aligneren van genetische sequenties is een veel gebruikte techniek (zie bijvoorbeeld Mount [58]). Om maar een paar voorbeelden te geven, aan de

hand van het aligneren van sequenties kunnen we de evolutionaire relatie tussen soorten bepalen, en evolutionaire bomen reconstrueren. Een alignering kan ook functionele gebieden in genetische sequenties onthullen, wat op zijn beurt kan leiden tot nieuwe geneesmiddelen, of kan helpen te beslissen welke behandeling het meest geschikt is voor een bepaalde patiënt met een wel bepaald genotype. Het aligneren van genetische sequenties wordt ook gebruikt ter voorspelling van structurele en biochemische eigenschappen van de sequenties zelf.

Het aligneringsprobleem wordt over het algemeen geformuleerd als een optimalisatieprobleem: positieve scores worden toegekend aan goede overeenkomsten, en negatieve scores aan slechte overeenkomsten en gaten. Deze scores worden samengevat in wat een *scorematrix* genoemd wordt. We willen dan die alignering vinden met de grootste totale score. Deze aanpak heeft twee voordelen: (i) het stelt ons in staat op een objectieve manier de beste alignering te vinden uit alle mogelijke aligneringen, en (ii) de hoogste score, die correspondeert met de beste alignering, geeft ons een objectieve kwaliteitsmaat voor deze alignering. Needleman en Wunsch [60] ontwikkelden een efficiënt dynamisch programmeeralgorithme ter alignering van paren van genetische sequenties.

De alignering hangt natuurlijk sterk af van de keuze van de scorematrix: hoe moeten we goede overeenkomsten belonen, en slechte overeenkomsten en gaten straffen? In de praktijk zijn er een hele resem scorematrices in omloop, en de precieze keuze van de scorematrix hangt af van extra onderstellingen omtrent de bestudeerde genetische sequenties. Bijvoorbeeld, als we PAM scorematrices gebruiken, die geïntroduceerd werden door Dayhoff, Schwartz, en Orcutt [15], en waartoe wij ons zullen beperken, dan maakt men de volgende—op het eerste gezicht ongetwijfeld absurde—onderstellingen:

- de evolutionaire afstand van de sequenties tot hun dichtste gemeenschappelijke voorouder is gekend,
- evolutie is in evenwicht,
- in dit evenwichtspunt is er evolutionaire omkeerbaarheid—eender welke puntmutatie is even waarschijnlijk als haar inverse mutatie,
- puntmutaties op verschillende locaties in een sequentie zijn identiek en onafhankelijk verdeeld, en

- puntmutaties op verschillende tijdstippen zijn identiek en onafhankelijk verdeeld.

Verschillende evolutionaire afstanden T induceren dan verschillende scorematrices PAM(T). Uiteraard zijn bovenstaande onderstellingen niet courant: veel onderzoek is gewijd aan het veralgemenen van deze onderstellingen. Wij zullen ons concentreren op het veralgemenen van de eerste onderstelling.

Inderdaad, het schatten van de evolutionaire afstand is een groot probleem, vooral wanneer men korte sequenties vergelijkt: 'estimation bias usually occurs when the sequence length is short so that stochastic effects are strong'[1] (Gu en Li [39, p. 5899, rechterkolom, ll. 25–27]). En vaak kunnen we enkel steunen op de sequenties zelf ter schatting van de evolutionaire afstand.

Één aanpak is de afstand te schatten op basis van de gelijkaardigheid van de twee sequenties. Typisch kiest men PAM250 als de sequenties voor 20% gelijk zijn, PAM120 als ze 40% gelijk zijn, PAM60 als ze 60% gelijk zijn, enzovoort. Het is echter niet helemaal duidelijk hoe het gelijkheidspercentage kan afgeleid worden alvorens te aligneren, en zelfs dan nog riskeren we voor korte sequenties een serieuze schattingsfout.

Een andere aanpak bestaat eruit het optimalisatieprobleem voor een verzameling PAM-matrices op te lossen, of zelfs met andere methoden, en dan die methode uit te kiezen die de hoogste score geeft. De prestatie van verschillende aligneringsmethodes is reeds goed bestudeerd, en één van de interessante resultaten uit deze studie is dat 'for different pairs many different methods create the best alignments', en daarom, 'if a method that could select the best alignment method for each pair existed, a significant improvement of the alignment quality could be gained'[2] (zie Elofsson [32]). Helaas is het in de praktijk computationeel onmogelijk per sequentiepaar een groot aantal methodes uit te proberen en alle parameters af te stemmen voor elk van deze methodes.

In plaats daarvan, onderzoeken wij, met de dynamische programmeermethode ontwikkeld in dit proefschrift, of een fout in de schatting van evo-

[1] [. . .] 'een schattingsfout komt gebruikelijk voor wanneer de sequenties kort zijn, waardoor stochastische effecten sterk zijn'

[2] [. . .] 'voor verschillende paren, leiden verschillende methodes tot een beste alignering', en daarom, 'ware er een methode ter selectie van de beste methode, zouden we een serieuze vooruitgang boeken in aligneringskwaliteit'.

lutionaire afstand ook leidt tot een fout in de optimale alignering. In het bijzonder, veralgemenen we het welbekende Needleman-Wunsch algoritme (zie Needleman en Wunsch [60]) om te bepalen of een alignering, of delen ervan, ongevoelig zijn voor variaties in evolutionaire afstand in een opgegeven interval.

Het belangrijkste resultaat van deze veralgemening, na toepassing op een aantal testgevallen, is dat de keuze van de PAM-matrix doorgaans irrelevant is voor de alignering van de grootste delen van genetische sequenties. Met andere woorden, vaak maakt het niet uit welke PAM-matrix je precies selecteert, als je enkel in de alignering geïnteresseerd bent. Merk misschien nog op dat, hoewel de alignering zelf doorgaans niet echt veel varieert, de optimale score doorgaans wel sterk varieert.

1.3.2 Het niet-deterministische geval: optimaal sturen en tegelijk leren over de dynamica

We beschrijven nu hoe een systeem, met onzekere dynamica, optimaal gestuurd kan worden terwijl we tegelijk leren over de dynamica, aan de hand van conditionele onderprevisies, en een nieuwe notie van optimaliteit die gedeeltelijke actie-kennisafhankelijkheid toelaat. De voorwaarden voor dynamisch programmeren zijn voldaan als we het leren voorstellen door een imprecies Dirichlet-model.

Inleiding

Één van de belangrijkste modellen voor niet-deterministische systemen zijn markovbeslisprocessen; dit zijn geregelde markovketens (zie Markov [55]). Markovbeslisprocessen modelleren onzekerheid omtrent dynamica via zogenaamde *transitiewaarschijnlijkheden*. Onderstellen we de winst voor elke transitie onder elke regelactie gekend, dan bekomen we een optimale regeling door de verwachte winst te maximaliseren. Dit maximalisatieprobleem (en vele varianten ervan) kan efficiënt opgelost worden met dynamisch programmeren; zie bijvoorbeeld Bertsekas [8] voor een uitstekend overzicht.

Reeds vroeg in de ontwikkeling van markovbeslisprocessen, realiseerde men zich dat de transitiewaarschijnlijkheden zelf vaak onzeker zijn, omdat ze in de praktijk vaak moeilijk te meten zijn. Om aan dit probleem hoofd te bieden, worden in de literatuur twee aanpakken gesuggereerd en bestudeerd:

(i) leren—we herzien onze kennis over de transitiewaarschijnlijkheden naarmate we transities observeren; zie bijvoorbeeld Bellman [3], Martin [56], en Satia en Lave [68].

(ii) verzamelingen—we onderstellen enkel dat de transitiewaarschijnlijkheden tot een bepaalde convexe verzameling behoren; zie bijvoorbeeld Wolfe en Dantzig [93], Satia en Lave [68], White en Eldeib [90], Givan, Leach, en Dean [36], Harmanec [41], en Kozine en Utkin [52].

Beide aanpakken hebben hun nadelen. De leeraanpak vergt a priori kennis over de transitiewaarschijnlijkheden. Is deze kennis foutief, dan is de optimale regelwet eveneens foutief in de initiële fase van het regelproces. De verzamelingaanpak heeft als nadeel dat er niets geleerd wordt, en men dus mogelijk interessante informatie over de dynamica negeert, die in vele gevallen wel beschikbaar is. Bovendien heeft deze aanpak een ietwat problematische relatie met optimaliteit: met een verzameling transitiewaarschijnlijkheden kunnen we enkel een interval voor de verwachte winst berekenen, voor elke regelwet. Bijna alle auteurs beschouwen daarom enkel maximin of maximax oplossingen. Ze ontwerpen algoritmes, gebaseerd op dynamisch programmeren, die ofwel de minimaal verwachte winst maximaliseren (pessimistisch, maximin), ofwel de maximaal verwachte winst maximaliseren (optimistisch, maximax), en dus alle tussenoplossingen negeren.

Een opmerkelijke uitzondering is Harmanec [41], die een dynamisch programmeeralgoritme voorstelt ter berekening van alle maximale regelwetten ten opzichte van een partiële ordening, die wij hier intervaldominantie hebben genoemd. Op die manier beperkt hij zich niet tot enkel extreme oplossingen. Helaas stelde Harmanec [41] zich niet de vraag in welke zin zijn dynamisch programmeermethode tot optimale regelwetten leidde—zoals reeds vermeld, leidt zij zelfs niet eens tot maximale regelwetten ten opzichte van de gegeven ordening. Dus, pakken we het probleem opnieuw aan vanuit een ander perspectief: we definiëren *eerst* een notie van optimaliteit, en we onderzoeken dan of er aan de voorwaarden voor dynamisch programmeren voldaan is, zijnde (i) het optimaliteitsprincipe, en (ii) ongevoeligheid voor het weglaten van niet-optimale elementen.

Voorheen hebben we reeds een partiële ordening gesuggereerd die aan het optimaliteitsprincipe en aan de ongevoeligheidseigenschap voldeed (onder een compactheidsvoorwaarde). Zoals eveneens opgemerkt werd door

Harmanec [41], veralgemeent deze partiële ordening niet zomaar tot niet-deterministische dynamische systemen. Zoals we straks zullen aangeven, is de reden dat er zogenaamde actie-kennisafhankelijkheid is in markovbeslisprocessen.

Ons doel is het combineren van de leeraanpak en de verzamelingaanpak om de problemen te vermijden waaraan beide aanpakken elk lijden. We willen de verzameling transitiewaarschijnlijkheden herzien op basis van observaties van voorgaande transities. Een markovbeslisproces kan beschouwd worden als een verzameling onafhankelijke multinomiale samplemodellen (zie Martin [56]), en we hebben een welbekende wiskundige methode om zulke modellen te herzien: het imprecies Dirichlet-model, ontwikkeld door Walley [87]. Hiertoe moeten we echter eerst de partiële preferentierelatie veralgemenen tot het geval van actie-kennisafhankelijkheid.

We kunnen dan aantonen dat er behoorlijk algemene voorwaarden zijn waaronder aan het optimaliteitsprincipe en aan de ongevoeligheidseigenschap voldaan is; echter, zoals in het klassieke geval, met gekende transitiewaarschijnlijkheden, eveneens het geval is, moet de regelwet van de volledige toestandsgeschiedenis afhangen (zie Bertsekas [8]), en daarom blijft een directe implementatie van de gesuggereerde techniek beperkt tot relatief kleine systemen.

Een bijzonder geval: leren met het imprecies Dirichlet-model

In deze samenvatting beschouwen we enkel het geval waarin het leren gemodelleerd wordt aan de hand van het imprecies Dirichlet-model; alle basisideeën zijn hierin reeds aanwezig, en het laat ons eveneens toe de meeste technische details—in het bijzonder, marginale uitbreiding en optimalisatie onder gedeeltelijke actie-kennisonafhankelijkheid—grofweg over te slaan. Het enige wat we nog aanmerken wat betreft het algemene geval, is dat het optimalisatieprincipe niet altijd geldt: het leermodel moet aan bepaalde structurele eigenschappen voldoen. Als bij wonder zijn deze voldaan in geval we het imprecies Dirichlet-model gebruiken.

Regelwetten en winstgokken Zij $\mathcal{X}$ de eindige toestandsruimte, en $\mathcal{U}$ de eindige regelruimte. De toevallige veranderlijke die de toestand voorstelt op tijdstip k noteren we als X_k. Na tijdstip N zijn we niet meer geïnteresseerd

in de dynamica van het systeem. Beschouw het systeem op tijdstip k. We kunnen

- $X_k = x_k$ observeren,
- de regeling $\mu_k(x_k) \in \mathcal{U}$ toepassen en $X_{k+1} = x_{k+1}$ observeren,
- de regeling $\mu_{k+1}(x_k x_{k+1}) \in \mathcal{U}$ toepassen en $X_{k+2} = x_{k+2}$ observeren,
- *etc.*,
- de regeling $\mu_{N-1}(x_k x_{k+1} \dots x_{N-1}) \in \mathcal{U}$ toepassen en $X_N = x_N$ observeren.

Deze operatie wordt gekarakteriseerd door een eindige rij van functies $\pi_k = (\mu_k, \mu_{k+1}, \dots, \mu_{N-1})$, waar $\mu_\ell \colon \mathcal{X}^{\ell-k+1} \to \mathcal{U}$. We noemen π_k een *regelwet* startend op k. De verzameling van alle regelwetten startend op k noteren we als Π_k.

Met elke regelwet $\pi_k \in \Pi_k$ associëren we een *winstgok vanaf tijdstip ℓ na observatie van $x_k \dots x_{\ell-1}$* (met $\ell \geq k$; als $\ell = k$ dan is $x_k \dots x_{\ell-1}$ een lege rij, *i.e.*, er is geen observatie, en we schrijven ook $J_{\pi_k}(x_k, \dots, x_N)$ in dat geval),

$$J_{\pi_k(x_k \dots x_{\ell-1})}(x_\ell, \dots, x_N) = \sum_{q=\ell}^{N-1} g_q(x_q, \mu_q(x_k \dots x_q), x_{q+1}) + g_N(x_N)$$

Dit is een gok op $(X_\ell, \dots, X_N)$. Elke transitie leidt tot een winst: startend op tijdstip q in toestand x_q, en het toepassen van regeling $u_q \in \mathcal{U}$ waarbij we in toestand $x_{q+1} \in \mathcal{X}$ terechtkomen, krijgen we een winst $g_q(x_q, u_q, x_{q+1})$. Komen we toe in de eindtoestand x_N op tijdstip N, dan krijgen we een extra winst $g_N(x_N)$. Noteer dat $J_{\pi_k(x_k \dots x_{\ell-1})}$ enkel afhangt van π_k via $\mu_\ell(x_k \dots x_{\ell-1} X_\ell)$, $\dots$, $\mu_{N-1}(x_k \dots x_{\ell-1} X_\ell \dots X_{N-1})$. Deze rij, die overeenkomt met de regelwet π_k na observatie van $x_k \dots x_{\ell-1}$, noteren we als $\pi_k(x_k \dots x_{\ell-1})$.

We willen die regelwet vinden die de winstgok maximaliseert.

Leren van de dynamica via het imprecies Dirichlet-model Een eenvoudige manier om de dynamica te leren, is het Dirichlet-model (zie Martin [56]). In dit model is de verwachtingswaarde van een gok f op $X_{\ell+1}$, na observatie van $x_k \dots x_\ell$ en onder regelwet π_k, gegeven door

$$E_{\pi_k}(f | x_k \dots x_\ell s\theta) = \sum_{x_{\ell+1} \in \mathcal{X}} f(x_{\ell+1}) \frac{s\theta^{\mu_\ell(x_k \dots x_\ell)}_{x_\ell x_{\ell+1}} + n^{\mu_\ell(x_k \dots x_\ell)}_{x_\ell x_{\ell+1}}(x_k \dots x_\ell, \pi_k)}{s + N^{\mu_\ell(x_k \dots x_\ell)}_{x_\ell}(x_k \dots x_\ell, \pi_k)}$$

Het symbool $n^u_{xy}(x_k \dots x_\ell, \pi_k)$ stelt het aantal transities voor vanuit toestand x naar y via regeling u, in de rij $x_k \dots x_\ell$ onder regelwet π_k, en het symbool $N^u_x(x_k \dots x_\ell, \pi_k)$ stelt het aantal transities voor vanuit toestand x via regeling u (ongeacht de eindtoestand):

$$\begin{aligned} n^u_{xy}(x_k, \pi_k) &= 0, \\ n^u_{xy}(x_k \dots x_{\ell+1}, \pi_k) &= n^u_{xy}(x_k \dots x_\ell, \pi_k) \\ &\quad + \begin{cases} 1, & \text{if } x_\ell = x, x_{\ell+1} = y \text{ en } \mu_\ell(x_k \dots x_\ell) = u \\ 0, & \text{in het andere geval} \end{cases} \end{aligned}$$

en,

$$N^u_x(x_k \dots x_\ell, \pi_k) = \sum_{y \in \mathcal{X}} n^u_{xy}(x_k \dots x_\ell, \pi_k).$$

De hyperparameters van dit model zijn s en θ^u_{xy}, voor elke x en $y \in \mathcal{X}$ en elke $u \in \mathcal{U}$. De hyperparameter $s > 0$ bepaalt de leersnelheid (lagere s betekent sneller leren), en de hyperparameters θ^u_{xy}, $\sum_{y \in \mathcal{X}} \theta^u_{xy} = 1$, modelleren de initiële kennis over de transitiewaarschijnlijkheden; θ^u_{xy} is de verwachte transitiewaarschijnlijkheid om van toestand x naar y te gaan via regeling u.

Na het toepassen van de regel van Bayes, vinden we

$$\begin{aligned} \mathbb{E}_{\pi_k}(f | x_k \dots x_\ell s\theta) = E_{\pi_k}(\bullet | x_k \dots x_\ell s\theta) \circ E_{\pi_k}(\bullet | x_k \dots x_\ell X_{\ell+1} s\theta) \circ \cdots \\ \cdots \circ E_{\pi_k}(\bullet | x_k \dots x_\ell X_{\ell+1} \dots X_{N-1} s\theta)(f) \end{aligned}$$

als de verwachtingswaarde van een willekeurige gok f op $(X_{\ell+1}, \dots, X_N)$ (zie ook Bertsekas [8]).

Het *imprecies Dirichlet-model* (Walley [87]) bestaat er nu uit de verzameling van alle Dirichlet-modellen te beschouwen voor alle mogelijke waarden van de hyperparameters θ^u_{xy} (de leerparameter s wordt constant verondersteld). Om technische redenen zullen wij enkel waarden $\theta^u_{xy} \geq \epsilon$ beschouwen, voor een willekeurig kleine $\epsilon > 0$: dit betekent dat de initiële verwachte transitiewaarschijnlijkheden minstens ϵ zijn.

Optimaliteit Noteer dat de hyperparameters s en θ^u_{xy}, die dus initiële informatie over de dynamica voorstellen, niet beïnvloed worden door de regelwet π_k: zij zijn *actie-kennisonafhankelijk*. Deze observatie is feitelijk *de* reden waarom dynamisch programmeren mogelijk blijkt onder een robuuste versie van

dit klassieke model. Inderdaad, op basis van bovenstaande uitdrukking voor de verwachtingswaarde, kunnen we regelwetten ordenen—-uiteraard heeft het enkel zin regelwetten te ordenen met dezelfde regelgeschiedenis. Noteer dus $\Pi_k(x_k \dots x_\ell, u_k \dots u_{\ell-1})$ als de verzameling van regelwetten π_k in Π_k die voldoen aan

$$\pi_k(x_k) = u_k, \quad \pi_k(x_k x_{k+1}) = u_{k+1}, \quad \dots, \quad \pi_{\ell-1}(x_k \dots x_{\ell-1}) = u_{\ell-1},$$

en laat ons $\Pi_k(x_k)$ met Π_k identificeren. Concreet, onderstel dat π_k en $\rho_k \in \Pi_k(x_k \dots x_\ell, u_k \dots u_{\ell-1})$. Dan verkiezen we π_k boven ρ_k na observatie van $x_k \dots x_\ell$ en toepassing van $u_k \dots u_{\ell-1}$, en we schrijven $\pi_k >_{x_k \dots x_\ell, u_k \dots u_{\ell-1}} \rho_k$, indien

$$\inf_{\theta^u_{xy} \geq \epsilon} (\mathbb{E}_{\pi_k(x_k \dots x_\ell)}(J_{\pi_k(x_k \dots x_\ell)} | x_k \dots x_\ell s\theta) - \mathbb{E}_{\rho_k(x_k \dots x_\ell)}(J_{\rho_k(x_k \dots x_\ell)} | x_k \dots x_\ell s\theta)) > 0. \tag{1.2}$$

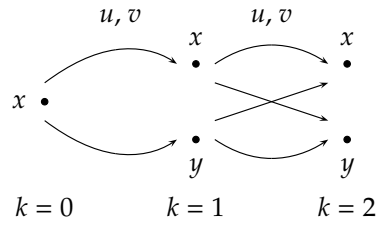
Eens $x_k \dots x_\ell$ en $u_k \dots u_{\ell-1}$ vastliggen, hangt de ordening enkel af van π_k en ρ_k via $\pi_k(x_k \dots x_\ell)$ en $\rho_k(x_k \dots x_\ell)$.

We bekomen een optimaliteitsoperator door die regelwetten te selecteren die optimaal zijn ten opzichte van de bovenstaande strikte partiële ordening. Concreet, noemen we een regelwet π_k *optimaal* indien hij maximaal is in $\Pi_k(x_k)$ ten opzichte van $>_{x_k}$ voor elke $x_k \in \mathcal{X}$. We noemen π_k *optimaal vanaf tijdstip* ℓ indien ze een maximaal element is van $\Pi_k(x_k \dots x_\ell, \mu_k(x_k) \dots \mu_{\ell-1}(x_k \dots x_{\ell-1}))$ ten opzichte van de partiële ordening $>_{x_k \dots x_\ell, \mu_k(x_k) \dots \mu_{\ell-1}(x_k \dots x_{\ell-1})}$ voor elke $x_k \dots x_\ell$.

Deze definitie is zinvol: Π_k is eindig, en het bestaan van maximale regelwetten ten opzichte van $>_{x_k \dots x_\ell, \mu_k(x_k) \dots \mu_{\ell-1}(x_k \dots x_{\ell-1})}$ is dus gegarandeerd. Men kan eveneens bewijzen dat er steeds regelwetten bestaan die simultaan maximaal zijn ten opzichte van $>_{x_k \dots x_\ell, \mu_k(x_k) \dots \mu_{\ell-1}(x_k \dots x_{\ell-1})}$ voor alle $x_k \dots x_\ell$.

Het optimaliteitsprincipe We kunnen nu het volgende bewijzen. Zij $k < N$ en $\pi_k \in \Pi_k$. Voor elke $k \leq \ell < N$, geldt de volgende implicatie: als π_k optimaal is vanaf tijdstip ℓ dan is π_k eveneens optimaal vanaf tijdstip $\ell + 1$.

Dynamisch programmeren We beschrijven kort hoe dit optimaliteitsprincipe leidt tot een dynamisch programmeeralgoritme. Beschouw hiertoe het beslisproces geschetst in Figuur 1.4. Op tijdstip k kunnen we kiezen tussen twee regelacties: u en v. Beschouw de regelwet π_0 die v toepast op tijdstip 0,



Figuur 1.4: Een eenvoudig beslisproces

en u als $x_1 = x$ en v als $x_1 = y$ op tijdstip 1:

$$\begin{aligned}\mu_0(x) &= v,\\ \mu_1(xx) &= u,\\ \mu_1(xy) &= v.\end{aligned}$$

Het optimaliteitsprincipe dicteert: als π_0 optimaal is vanaf tijdstip 0, dan is $\pi_0(x)$, die u toepast als $x_1 = x$ en v als $x_1 = y$ op tijdstip 1, optimaal vanaf tijdstip 1.

Dit leidt tot een vereenvoudiging bij het berekenen van de optimale regelwetten vanaf tijdstip 0. Inderdaad, onderstel bijvoorbeeld dat ρ_1, gegeven door $\nu_1(x) = v$ en $\nu_1(y) = u$, niet optimaal is vanaf tijdstip 1. Wegens het optimaliteitsprincipe zullen σ_0 en σ_0', gegeven door

$$\begin{aligned}\kappa_0(x) &= u, & \kappa_0'(x) &= v\\ \kappa_1(xx) &= \nu_1(x) = v, & \kappa_1'(xx) &= \nu_1(x) = v,\\ \kappa_1(xy) &= \nu_1(y) = u, & \kappa_1'(xy) &= \nu_1(y) = u,\end{aligned}$$

ook niet optimaal zijn vanaf tijdstip 0, gezien ρ_1 anders optimaal zou moeten zijn vanaf tijdstip 0. Dus, als we de regelwetten kennen die optimaal zijn vanaf tijdstip $\ell + 1$, dan kunnen we deze informatie gebruiken om de verzameling van regelwetten die mogelijk optimaal zijn vanaf tijdstip ℓ te reduceren.

Natuurlijk kunnen we dit enkel doen als een reductie in de zoekruimte de uiteindelijke verzameling van optimale elementen niet wijzigt: onze notie van optimaliteit moet ongevoelig zijn voor het weglaten van niet-optimale elementen. De verzameling van winstgokken is eindig, en dus compact. Bijgevolg voldoet onze notie van optimaliteit, die gebaseerd is op een ordening,

aan de ongevoeligheidseigenschap.

Het aantal regelwetten groeit exponentieel met de lengte van de paden die we beschouwen, maar wegens het optimaliteitsprincipe en de ongevoeligheidseigenschap moeten we de meeste van deze paden niet beschouwen: dynamisch programmeren, zoals hier ruwweg beschreven werd, leidt tot een exponentiële versnelling in het zoeken naar de verzameling van alle optimale regelwetten.

Chapter 2

Introduction

The aim of this dissertation is two-fold. First, we wish to investigate to what extent lower previsions unify and extend various other well-known uncertainty models, and how they lead to an acceptable notion of optimality. Secondly, we wish to investigate how a dynamical system, whose uncertainties are described by such a unified uncertainty model, can be controlled in an optimal way, and we want to find out whether it is possible to find such a control policy through an efficient method, namely, dynamic programming.

Uncertainty and Optimality

In Part I, we introduce the basic concepts of the theory of lower previsions (which was developed mainly by Walley [86]), we demonstrate the unifying rôle of theory of lower previsions in uncertainty modelling, and we explain how lower previsions naturally lead to various well-known notions of optimality.

The first four sections of Chapter 3, Sections 3.1–3.4, are concerned with the precise definition of random variables, gambles, lower previsions, avoiding sure loss, and coherence, which are essential for a good understanding of the remaining chapters. The remainder of this chapter, Section 3.5, presents an overview of the most important uncertainty models that fall within the theory of lower previsions. From a practical point of view, the most important ones are probability charges, 2-monotone set functions, and possibility measures.

Chapter 4 is concerned solely with the technique of *natural extension* of lower previsions: natural extension dictates whether and how a lower prevision must extend to a larger domain, and is essential in making inferences from lower previsions, such as optimality judgements. In Section 4.1 we present a precise definition of natural extension, that generalises the notion of natural extension given by Walley [86]. This section is essential for a good understanding of the remaining chapters. In Section 4.2, we introduce a new notion of integration based on natural extension, called *linear extension*.

Section 4.3 presents various examples of natural extension, and relates them to other well-known extension procedures known from the literature. Some of these results are new: in particular, we prove that the S-integral, the Riemann integral, and the Riemann-Stieltjes integral (Darboux's type), are instances of linear extension, and that the corresponding lower and upper integrals are instances of natural extension (Sections 4.3.5&4.3.6); we show that the natural extension of a cumulative distribution function is, under fairly general assumptions, given by the Riemann-Stieltjes integral (Section 4.3.7); and finally, we recover two new expressions of the Choquet integral in terms of the S-integral, and in terms of the Dunford integral (Section 4.3.10). The most important known results collected in Section 4.3 are that the linear extension of a probability charge corresponds to the Dunford integral restricted to gambles, or equivalently, the S-integral (Section 4.3.2), and that the natural extension of a 2-monotone set function is given by the Choquet integral restricted to gambles (Section 4.3.10).

In Section 4.4 we generalise a number of well-known results about sets of probability measures and lower previsions: in particular, we show that if all gambles in the domain of a lower prevision $\underline{P}$ are measurable with respect to a field $\mathcal{F}$, then that lower prevision $\underline{P}$ avoids sure loss if and only if there is a probability charge on $\mathcal{F}$ whose S-integral dominates $\underline{P}$, and that in such a case, the natural extension of $\underline{P}$ to the set of all gambles is given by the lower envelope of the lower S-integrals with respect to all probability charges on $\mathcal{F}$ that dominate $\underline{P}$; this new result yields a simpler representation of coherent lower previsions, and a simpler expression for natural extension, in case all gambles in the domain of $\underline{P}$ are measurable with respect to a non-trivial field, *i.e.*, a field smaller than the power set (Section 4.4.2).

There is one well-known extension procedure, due to Dunford [31], which is not covered by natural extension, but which still deserves our attention,

as it allows for an extension to random quantities that needn't be bounded. Many problems in optimal control concern unbounded random quantities: for instance, a cost function is often chosen to be a quadratic (and hence, unbounded) function of the state and of the control. Since we aim to study optimal control in what follows, we must first extend the theory of lower previsions, including natural extension, to random quantities that are not necessarily bounded (Sections 5.1–5.2). The technical matter of porting Dunford's idea (originally devised to extend integrals to unbounded functions) to lower previsions is dealt with in Section 5.4, but first we discuss a simpler and less general method in Section 5.3, which at the same time introduces part of the mathematical machinery required in Section 5.4. We provide a more convenient characterisation of our extension in Section 5.5, we show that our extension coincides with the Choquet integral for unbounded random quantities, in case we start out with a 2-monotone set function (Section 5.5.4), and we establish that our extension can always be written as a lower envelope of Dunford integrals (Section 5.6). The main results of Chapter 5 have been published by Troffaes and De Cooman [78, 79, 80].

As the results of Chapters 3–5 clearly identify the unifying character of lower previsions and natural extension, we shall only consider uncertainty modelled by lower previsions in the remaining chapters.

Chapter 6 presents an overview of various notions of optimality that can be derived from a lower prevision and studies their properties (Sections 6.3–6.6). An incentive for studying these generalisations of the classical maximising expected utility approach (which is briefly presented in Section 6.1) is given in Section 6.2.1. The most important new result in this chapter is a technical lemma about preorders in Section 6.2.3 (a slightly less general version has already been published by De Cooman and Troffaes [24]), which allows us to prove a property, called insensitivity to omission of non-optimal elements, that turns out to be essential for the dynamic programming approach to work.

Dynamic Programming

In Part II, we set out to answer the second question: we investigate the applicability of dynamic programming to dynamical systems whose uncertainties, in gain or dynamics, are described by lower previsions.

Chapter 7 studies deterministic systems whose uncertain gain is described by a lower prevision. In Section 7.2 we identify two conditions for the dynamic programming approach to work: (i) the principle of optimality, which is due to Bellman [4], and (ii) insensitivity to omission of non-optimal elements, which is a new condition (in case the gain is not uncertain, this condition is satisfied whenever an optimal solution exists). All notions of optimality introduced in Chapter 6 satisfy the insensitivity property, but, as we show in Section 7.2, only $\underline{P}$-maximality and $\mathcal{M}$-maximality satisfy the principle of optimality. We present a simple example in Section 7.3, and a more realistic example, robust sequence alignment, in Section 7.4, where we find that the exact choice of the PAM matrix in (amino-acid) sequence alignment usually does not matter when aligning short sequences, contrary to what is often claimed (as for instance by Mount [58]). The basic ideas behind the dynamic programming method are due to De Cooman and Troffaes [24, 23]. The robust sequence alignment example has also been published earlier by Troffaes [76].

In Chapter 8, we study systems whose dynamics is uncertain, and we describe a way to accomplish simultaneous optimal control and learning about the dynamics, by means of conditional lower previsions (Section 8.2), and a new notion of optimality, which allows for partial act-state dependence (Section 8.3). In Section 8.4 we identify a fairly general class of dynamical systems, whose uncertain dynamics is described by conditional lower previsions; this allows us to model learning about the dynamics at the same time. In Section 8.5 we state conditions for the principle of optimality to hold, and we note that the insensitivity property is almost trivially satisfied for the systems under study. As a result, we can construct a dynamic programming algorithm: this is demonstrated in Section 8.7 through a numerical example, where the learning is modelled through an imprecise Dirichlet model. The results of this chapter, without proofs, were published earlier by Troffaes [77].

The appendix contains results about the extended real number system, that are probably not new, but certainly very hard to find. We need these results only in Chapter 5.

Part I

Uncertainty and Optimality

Chapter 3

Lower and Upper Previsions

PROBABILITY DOES NOT EXIST.

Bruno de Finetti [27], Volume I, p. x

In this and the next chapter we introduce the main ideas behind the theory of coherent lower and upper previsions. We refer to Walley [86] for a more in depth discussion.

3.1 Random Variables

We call a *random variable* any, possibly uncertain, but observable property of a system. One could think of a random variable as an experiment's outcome that is not necessarily known to the modeller. For example, the amount of rainfall $R(d, w)$ during a particular day d, measured at a particular weather station w, is a random variable; it takes values in the set $\{r\colon r \in \mathbb{R}, r \geq 0\}$. But also the statement "tomorrow, it will rain in Ghent" is a random variable, with possible values "true" and "false". The set of possible values of a random variable X is denoted by a calligraphic letter $\mathcal{X}$. A particular value of X is called a *realisation* of X and it is denoted by a lower-case letter x. It is convenient to denote the event of observing the realisation of X to be x as $X = x$.

Let us now describe a few things one can do with random variables. At the same time we fix some notation that will be used extensively further on.

Any map f defined on the set of outcomes $\mathcal{X}$ of a random variable X we can again consider as a random variable. We call such a mapping a *function of* X, and it represents nothing but a relabelling (not necessarily one-to-one) of the outcomes of X. If those outcomes are observable, so must be, in principle, any relabelling of them, and hence, the function f again constitutes a random variable. For instance, "tomorrow, it will rain in Ghent" could be defined as Boolean function of the amount of rainfall $R(d, w)$ for d tomorrow and measured at Ghent's weather station w. If we consider more than one random variable we may write $f(X)$ in order to emphasise that f is a function of X only. The realisation of f is uniquely determined by the realisation x of X and is therefore denoted by $f(x)$.

Obviously, the composition $g \circ f$ of mappings f and g is a random variable if f is defined on the set of outcomes of a random variable. As $(g \circ f)(x)$ is defined as $g(f(x))$, we shall also write $(g \circ f)(X)$ as $g(f(X))$.

Considering the combined outcomes of a collection of random variables X_1, X_2, $\ldots$, and X_n we again obtain a random variable which we denote by $(X_1, X_2, \ldots, X_n)$ and whose possible outcomes are (not necessarily all) elements of $\mathcal{X}_1 \times \mathcal{X}_2 \times \cdots \times \mathcal{X}_n$ and are generically denoted by $(x_1, \ldots, x_n)$. From a function $f(X_1, X_2, \ldots, X_n)$ we may obtain a new function of for instance $(X_1, \ldots, X_j)$ (with $j < n$) by fixing the values of $X_{j+1}, \ldots X_n$ at their respective realisations $x_{j+1}, \ldots, x_n$. This function is denoted by $f(X_1, \ldots, X_j, x_{j+1}, \ldots, x_n)$.

3.2 Belief, Behaviour and Optimality

Our aim in this work is to study optimal control under uncertainty. More particularly, we wish to model our beliefs about the uncertain realisation of the gain of a system and make decisions—optimal control eventually reduces to decision making—based on these beliefs. In the following we shall propose a behavioural belief model. Before going into a detailed description of this model, we first discuss very briefly a few important how's, why's, pro's and con's of behavioural belief models: why are behavioural belief models well suited for studying optimal control under uncertainty, and what are the limits of behavioural belief models?

Beliefs about X can be modelled through behaviour. For instance, if we strongly believe that the outcome of the random variable "tomorrow, it will rain in Ghent" will be "true" then probably we will take an umbrella with us

for tomorrow's city-trip to Ghent (even though we might not be 100% certain of rain). Thus, clearly our behaviour is a reflection of our beliefs (possibly involving uncertainty). This has lead some people, such as Ramsey [65], De Finetti [26], von Neumann and Morgenstern [83], Savage [69], Anscombe and Aumann [2], Khaneman and Tversky [47], Walley [85], von Winterfeld and Edwards [84], Seidenfeld, Schervish and Kadane [72], and many others, to the idea of taking behaviour—in whatever form—as the primitive notion when modelling belief, *de facto* taking some form of behaviour as a definition of belief. It is also the course we shall pursue.

Optimal control in itself is actually a form of behaviour, and hence, any sufficiently sophisticated behavioural belief model naturally leads to a notion of optimality induced by belief. In optimal control we are presented with a set of controls, for instance

$$\mathcal{U} = \{\text{take umbrella}, \text{don't take umbrella}\},$$

and we are faced with the problem of identifying the subset of best controls $\text{opt}(\mathcal{U}) \subseteq \mathcal{U}$. Observe that this identification clearly reflects our beliefs: if we strongly believe that the value of the random variable "tomorrow, it will rain in Ghent" will be "true" then clearly

$$\text{opt}(\mathcal{U}) = \{\text{take umbrella}\},$$

provided we prefer not to get wet. Thus, we might in fact take optimality, or more precisely, the *optimality operator*, also called *choice function*, $\text{opt}: \wp(\mathcal{U}) \to \wp(\mathcal{U})$ that selects the best options from a set, as the primitive notion when modelling belief. We shall however consider a simpler way of modelling behaviour and derive optimality from it. It turns out that the simple behavioural belief model we shall use still encompasses quite a large number of uncertainty models and optimality operators that exist in the literature.

A simple but fairly general way to model our beliefs about the random variable X consists of considering our dispositions towards transactions whose value depends on the outcome of X; such transactions occur typically in optimal control problems: often, we want to optimise a gain whose value is a function of a random variable X. Again, to see how such dispositions relate to belief, assume for instance that we have strong belief that we shall

observe $X = x$. We should then be inclined to accept, prior to observation of X, any transaction that incurs a positive gain if $X = x$. On the other hand, if we strongly believe that $X = x$ will not be observed, we should not care too much, prior to observation of X, about the gains or losses connected with $X = x$. Hence, beliefs about X naturally translate into behavioural dispositions towards transactions whose value depends on the outcome of X. In the theory of lower previsions we shall consider a very particular type of transactions: buying gains that are a bounded function of X.

Any behavioural belief model can be made operational since measuring belief is simply obtained by measuring behaviour. We refer to von Winterfeld and Edwards [84] for an extensive discussion.

On the other hand, not all beliefs are reflected by behaviour, and certainly not by the restricted types of behaviour we shall study. Despite its unifying character, this is a hard limit on the applicability of the theory used in this work. This limit is at least three-fold. Firstly, we shall only express belief regarding random variables. Secondly, it turns out that only using transactions in describing optimality operators forces us to restrict to a very particular type of optimality operators. Finally, we shall be concerned only with transactions expressed in terms of a linear utility scale.

Indeed, not all our beliefs concern random variables. For instance, I believe that most of my beliefs are not concerned at all with dispositions towards transactions whose value depends on the outcome of a random variable (although I usually have a lot of fun in trying to express my beliefs by such transactions). This belief I cannot express as dispositions towards transactions whose value depends on the outcome of a random variable. More seriously, any belief concerning an entity whose value is not observable falls beyond the scope of the behavioural belief models that are based on random variables. For instance, what about a transaction whose outcome depends on the fact whether or not the electron is a fundamental particle? Or, what about Pascal's [63] wager, who considers a transaction whose outcome depends on the fact whether God exists or not? As Pascal puts it:

> [...] — Examinons donc ce point, et disons: « Dieu est, ou il n'est pas. » Mais de quel côté pencherons-nous? La raison n'y peut rien déterminer: il y a un chaos infini qui nous sépare. Il se joue un jeu, à l'extrémité de cette distance infinie, où il arrivera croix ou pile. Que gagerez-vous? Par raison, vous ne pouvez défendre nul des deux. Ne blâmez donc pas de fausseté ceux qui ont pris un choix; car vous n'en

> savez rien. — « Non; mais je les blâmerai d'avoir fait, non ce choix, mais un choix; car, encore que celui qui prend croix et l'autre soient en paraille faute, ils sont tous deux en faute: *le juste est de ne point parier* [our emphasis]. » — Oui; mais il faut parier. [...]
>
> Blaise Pascal [63, Vol. 1, Sect. 233-418, pp. 101–102]

Unless we have an experiment that determines an objective (or at least intersubjective) answer to such questions, transactions whose outcome depends on these answers cannot be executed. Therefore, they fall beyond the scope of behavioural belief models.

A second reason why not all beliefs are reflected by the restricted types of behaviour we shall study, is that the optimality operator induced by transactions is uniquely determined by its restriction to pairs of controls only. That is, our best options within pairs of controls uniquely determine our best options within any set of controls. There are reasonable optimality operators that do not satisfy this property; see for instance Schervish, Seidenfeld, Kadane and Levi [71], and Chapter 6 of this work.

Finally, we shall only be concerned with transactions expressed in a fixed and unique linear utility scale: this means roughly that twice executing a transaction doubles its value. This is taken for granted if transactions only involve a sufficiently small amount of money (see De Finetti [27, Volume I, Section 3.2]) or an exchange of lottery tickets (see Walley [86, Section 2.2]). However, they often fail when the results of transactions involve drinking beer, eating gnocci, sunbathing, or making love. So, there are behavioural dispositions, reflecting beliefs, which cannot be completely expressed in terms of a linear utility scale. They fall beyond the scope of the behavioural belief models we shall study.

3.3 Gambles and Prices

3.3.1 Definitions

A *gamble* f on a random variable X is a real-valued gain (expressed in units of a linear utility scale which is assumed to be fixed) that is a bounded function of X. In particular, it is a real-valued random variable. Mathematically, it is a bounded $\mathcal{X}$–$\mathbb{R}$-mapping, and it is thus interpreted as a bounded uncertain gain: if x turns out to be the realisation of X, then we receive an amount

$f(x)$ of utility. As before, when we consider gambles on different random variables, we may write $f(X)$ in order to emphasise that f is a gamble on X.

Restricting to bounded uncertain gains is to some extent a mathematical convenience. We shall propose a way to generalise large parts of the theory to the unbounded case further on.

The set of all gambles on X is denoted by $\mathcal{L}(X)$. It is a linear lattice—an ordered linear space such that every two vectors have a supremum and an infimum—with respect to the point-wise addition, the point-wise scalar multiplication and the point-wise ordering, which are defined as

$$\begin{aligned}(f+g)(x) &:= f(x)+g(x), \text{ and}\\ (\lambda f)(x) &:= \lambda f(x)\end{aligned}$$

for any pair of gambles f and g on X, any real λ and all $x \in \mathcal{X}$, and

$$f \leq g \text{ if for all } x \in \mathcal{X}\colon\ f(x) \leq g(x).$$

The supremum $f \vee g$ of two gambles f and g on X is given by their point-wise maximum, and their infimum $f \wedge g$ is given by their point-wise minimum: for all $x \in \mathcal{X}$,

$$\begin{aligned}(f \vee g)(x) &= \max\{f(x), g(x)\}, \text{ and}\\ (f \wedge g)(x) &= \min\{f(x), g(x)\}.\end{aligned}$$

These operators should not be confused with the supremum $\sup f$ and infimum $\inf f$ of a gamble f on X, which are defined as

$$\begin{aligned}\sup f &:= \min\{a \in \mathbb{R}\colon a \geq f\}, \text{ and}\\ \inf f &:= \max\{a \in \mathbb{R}\colon a \leq f\}.\end{aligned}$$

We also define the absolute value $\left|f\right|$ of a gamble f as $\left|f\right|(x) := \left|f(x)\right|$ for all $x \in \mathcal{X}$, and the negation $-f$ of f as $(-f)(x) := -(f(x))$ for all $x \in \mathcal{X}$. As usual, $f + (-g)$ is abbreviated to $f - g$.

It is convenient to identify a real number $a \in \mathbb{R}$ with the constant gamble $a(x) := a$ for all $x \in \mathcal{X}$. For instance, the expression $a \geq f$, where f is a gamble on $\mathcal{X}$ and a is a real number, means $a \geq f(x)$ for all $x \in \mathcal{X}$; we already used this in our definition of supremum. The set of constant gambles on X will be

denoted by $\mathbb{R}(X)$.

Another special class of gambles are those that correspond to so-called events. An *event* on X simply is a subset of $\mathcal{X}$. With an event A on X we can associate a $\{0,1\}$-valued gamble I_A that gives us one unit of utility if the realisation x of X belongs to A, and nothing otherwise:

$$I_A(x) = \begin{cases} 1, & \text{if } x \in A, \\ 0, & \text{otherwise.} \end{cases}$$

This gamble I_A is called the *indicator* of A. For a collection $\mathcal{A}$ of events on X, we define

$$I_{\mathcal{A}} := \{I_A : A \in \mathcal{A}\}.$$

The *lower prevision* $\underline{P}(f)$ of a gamble f is defined as the supremum buying price for f; $\underline{P}(f)$ is the highest price s such that for any $t < s$, we are willing to pay t prior to observation of X, if we are guaranteed to receive $f(x)$ when observing $X = x$. Mathematically, we define a lower prevision on X as a real-valued mapping defined on some subset $\operatorname{dom}\underline{P}$, the domain of $\underline{P}$, of $\mathcal{L}(X)$. Indeed, we do not require a lower prevision to be defined on the set of all gambles. Further on, we shall describe methods for extending any lower prevision to the set of all gambles.

We can also interpret f as an uncertain bounded loss: if x turns out to be the true value of X, we lose an amount $f(x)$ of linear utility. The *upper prevision* $\overline{P}(f)$ of the gamble f is then the infimum selling price for f; it is the lowest price s such that for any $t > s$, we are willing to receive t prior to observation of X, if we are guaranteed to lose $f(x)$ when observing $X = x$. Since a gain r is equivalent to a loss $-r$ it should hold that $\overline{P}(f) = -\underline{P}(-f)$: from any lower prevision $\underline{P}$ we can infer a so-called *conjugate upper prevision* $\overline{P}$ on $\operatorname{dom}\overline{P} = -\operatorname{dom}\underline{P}$ which represents the same behavioural dispositions. We can therefore restrict our attention to the study of lower previsions only, without loss of generality. Also, if we use the notation $\underline{P}$ for a lower prevision, $\overline{P}$ always denotes its conjugate.

It may happen that $\underline{P}$ is *self-conjugate*, that is, $\operatorname{dom}\underline{P} = \operatorname{dom}\overline{P}$ and $\underline{P}(f) = \overline{P}(f)$ for all gambles $f \in \operatorname{dom}\underline{P}$. In such a case, we may simply write P instead of $\underline{P}$ or $\overline{P}$ whenever it is clear from the context whether we are considering either buying or selling prices (or both). We call a self-conjugate lower

prevision P simply a *prevision*, and $P(f)$ represents a so-called *fair price* for the gamble f: we are willing to buy f for any price $s < P(f)$, and we are willing to sell f for any price $s > P(f)$. Previsions, interpreted as fair prices, were introduced by De Finetti [26], who used them to provide a behavioural foundation for probability theory based on expectation.

It is convenient to identify indicator gambles with their corresponding events. In particular, we shall denote the lower prevision of an indicator gamble $\underline{P}(I_A)$ simply as $\underline{P}(A)$. We call $\underline{P}(A)$ the lower probability of the event A, and if the domain of $\underline{P}$ only contains indicator gambles, we simply call $\underline{P}$ a *lower probability*. Similarly, we denote $\overline{P}(I_A)$ as $\overline{P}(A)$ and call it the upper probability of A. If the domain of $\overline{P}$ only contains indicator gambles it is called an *upper probability*. If P is a (self-conjugate lower) prevision then $P(A)$ is called the probability of A. If the domain of a prevision P only contains indicator gambles along with their negations then P is called a *probability* .

3.3.2 Lower and Upper Previsions by Chance

As a simple example of lower and upper previsions, we illustrate how chance (aleatory probability) can be related to lower and upper previsions of $\{0,1\}$-valued gambles, or, lower and upper probabilities.

Let's for instance assume that X corresponds to the colour of a marble randomly drawn from a bag of N marbles. For simplicity, assume that each marble can be either red or blue, so $\mathcal{X} = \{\text{red}, \text{blue}\}$. Let A be the event of drawing a red marble, and suppose we learn somehow that at least n_* of the marbles are red. Now consider the gamble I_A. If we assume that each marble of the bag is equally likely to be drawn, then we gain one unit of utility in at least n_* of the N equiprobable cases; we could say that the statistical chance of drawing red is at least $\frac{n_*}{N}$. In an infinite sequence of independent trials of the marble experiment, the relative frequency by which we draw red will converge almost surely to the chance of drawing red by the strong law of large numbers. Therefore, repeatedly buying I_A for any price strictly less than $\frac{n_*}{N}$ will result almost surely in a net gain in such a sequence. This suggests $\underline{P}(A) = \frac{n_*}{N}$ as a supremum buying price. In this example, the lower probability of A is nothing but a lower bound for the chance of A.

Alternatively, suppose we learn that at most n^* of the marbles are red and consider the gamble $-I_A$. We expect to lose one unit of utility in at most n^* of

the N equiprobable cases. In an infinite sequence of independent trials of the marble experiment, again we should be disposed to buying $-I_A$ for any price strictly less than $-\frac{n^*}{N}$. This suggests $\underline{P}(-I_A) = -\frac{n^*}{N}$, or equivalently, $\overline{P}(A) = \frac{n^*}{N}$. The upper probability of A is nothing but an upper bound for the chance of A.

We've shown how our belief about the chance of particular events translates naturally into lower and upper previsions. This is sometimes called the *principle of direct inference*; see Walley [86, Section 7.2.4] for a more detailed discussion and a different approach not relying on sequences of independent trials. If the chance of an event A is exactly known, the fair price for I_A is exactly equal to this chance. Of course, lower and upper previsions can be used as models for beliefs in far more general situations.

3.3.3 Inference Rules

Some behavioural dispositions are evident: for instance, an indicator gamble does not take a negative value, and therefore, we should be disposed to buy it (at least) for any strictly negative price. Also, some behavioural dispositions imply other behavioural dispositions. To give a simple example, if we are disposed to buy a gamble for a price, then we should also be disposed to buy it for any lower price (we already used this implicitly in accepting the *supremum* buying price as a summary of all our buying prices). We shall accept the following axioms of rationality governing dispositions towards buying and selling gambles.

Axiom 3.1 (Axioms of Rationality)**.** For arbitrary gambles f and g on X and arbitrary real numbers s and t the following should hold.

(i) We are disposed to buy f for any price strictly less than $\inf f$ (accepting a sure gain).

(ii) We are disposed not to buy f for any price strictly larger than $\sup f$ (avoiding a sure loss).

(iii) If we are disposed to buy f for s then we should be disposed to buy λf for λs, for any strictly positive $\lambda \in \mathbb{R}$ (scale independence).

(iv) If we are disposed to buy f for s and g for t then we should be disposed to buy $f + g$ for $s + t$ (accepting combined transactions).

(v) If we are disposed to buy f for s and $g \geq f$ then we should be disposed to buy g for s (monotonicity).

The above axioms can easily be motivated through the linearity of the utility scale in which gambles and prices are expressed; see for instance Walley [86, Section 2.4.4] for a justification using probability currency as a linear utility scale. Also see De Finetti [27, Volume I, Section 3.2.5] for a simple solution in case gambles and prices are expressed in a precise but non-linear utility scale.

3.4 Criteria of Rationality

3.4.1 Avoiding Sure Loss

Buying $f(X)$ for a price strictly higher than $\sup f$ makes no sense: with certainty, such transaction incurs a strictly positive loss. For example, suppose that for some non-empty subset A of $\mathcal{X}$ our lower prevision for I_A is equal to 1.2; $\underline{P}(I_A) = 1.2$. This means that we are disposed to buy the uncertain gain I_A for a price 1.1 since 1.1 is strictly less than 1.2. But, the reward of I_A is 1 at most, so this behavioural disposition incurs a sure loss of at least $1.1 - 1 = 0.1$. This is irrational behaviour, so our lower prevision $\underline{P}(I_A)$ should not be equal to 1.2; in fact, it should not be strictly higher than 1.

More generally, a combined buy of a finite collection of gambles is not acceptable if this transaction leads to a sure loss. For example, suppose we have that $\underline{P}(I_A) = 0.7$ and $\underline{P}(I_B) = 0.7$, with A and B non-empty and disjoint subsets of $\mathcal{X}$. This means that we are disposed to buy both I_A and I_B for 0.6. Therefore, we are disposed to buy $I_A + I_B = I_{A \cup B}$ for $0.6 + 0.6 = 1.2$. But again, the reward for $I_{A \cup B}$ is 1 at most, so these behavioural dispositions lead to a loss of at least $1.2 - 1 = 0.2$. Again, this behaviour is irrational: $\underline{P}(I_A) + \underline{P}(I_B)$ should not be equal to 1.2; it should not be higher than 1 in fact. These examples call for the following definition, which apparently goes back to Ramsey [65, p. 182].

Definition 3.2. Let $\underline{P}$ be a lower prevision on X. Then the following conditions are equivalent; if any (hence all) of them are satisfied, then $\underline{P}$ is said to *avoid sure loss*.

(A) For any $n \in \mathbb{N} \setminus \{0\}$, and $f_1, \ldots, f_n$ in $\operatorname{dom} \underline{P}$, it holds that

$$\sum_{i=1}^{n} \underline{P}(f_i) \leq \sup\left[\sum_{i=1}^{n} f_i\right]. \tag{3.1}$$

(B) For any $n \in \mathbb{N} \setminus \{0\}$, non-negative $\lambda_1, \ldots, \lambda_n$ in $\mathbb{R}$, and $f_1, \ldots, f_n$ in $\operatorname{dom} \underline{P}$, it holds that

$$\sum_{i=1}^{n} \lambda_i \underline{P}(f_i) \leq \sup\left[\lambda_i \sum_{i=1}^{n} f_i\right]. \tag{3.2}$$

Explanation. The equivalence of (A) and (B) is proved in Walley's book [86, Lemma 2.4.4]. Let's explain (A).

Suppose that Eq. (3.1) does not hold for some $n \in \mathbb{N} \setminus \{0\}$ and $f_1, \ldots, f_n$ in $\operatorname{dom} \underline{P}$. Define the strictly positive number

$$\epsilon := \sum_{i=1}^{n} \underline{P}(f_i) - \sup\left[\sum_{i=1}^{n} f_i\right] > 0.$$

Let $\delta := \frac{\epsilon}{n+1}$. Since $\delta > 0$, we are disposed to buy each f_i for $\underline{P}(f_i) - \delta$. Hence, by the axioms of rationality, Axiom 3.1(iv) in particular, we are disposed to buy $\sum_{i=1}^{n} f_i$ for $\sum_{i=1}^{n} \underline{P}(f_i) - n\delta$. But, this transaction always leads to a strictly positive loss since

$$\sum_{i=1}^{n} \underline{P}(f_i) - n\delta > \sum_{i=1}^{n} \underline{P}(f_i) - \epsilon = \sup\left[\sum_{i=1}^{n} f_i\right]$$

by the definition of ϵ. This violates Axiom 3.1(ii). □

3.4.2 Coherence

It might occur that we are disposed to buy a gamble f for a higher price than the one implied by $\underline{P}(f)$ after consideration of buying prices of other gambles. For example, consider the lower prevision $\underline{P}$ defined by $\underline{P}(I_A) = 0.3$, $\underline{P}(I_B) = 0.4$ and $\underline{P}(I_C) = 0.5$, where A, B and C are non-empty subsets of $\mathcal{X}$ with $A \cap B = \emptyset$ and $A \cup B \subseteq C$. This means for instance that we are disposed to buy $I_A + I_B = I_{A \cup B}$ for $0.25 + 0.35 = 0.6$. But observe that $I_{A \cup B} \leq I_C$ since $A \cup B \subseteq C$: I_C gives us a reward at least as high as $I_{A \cup B}$. Therefore, we should also be disposed to buy I_C for 0.6; the supremum buying price $\underline{P}(I_C) = 0.5$ is

too low. These observations are formalised and generalised in the following definition, which apparently goes back to Williams [92].

Definition 3.3. Let $\underline{P}$ be a lower prevision on X. Then the following conditions are equivalent; if any (hence all) of them are satisfied, then $\underline{P}$ is called *coherent*:

(A) For any $m \in \mathbb{N}$, $n \in \mathbb{N}$, and gambles $f_0, f_1, \ldots, f_m$ in $\operatorname{dom} \underline{P}$, it holds that

$$\sum_{i=1}^{n} \underline{P}(f_i) - m\underline{P}(f_0) \leq \sup\left[\sum_{i=1}^{n} f_i - mf_0\right]. \tag{3.3}$$

(B) For any $n \in \mathbb{N}$, non-negative $\lambda_0, \lambda_1, \ldots, \lambda_n$ in $\mathbb{R}$, and gambles $f_0, f_1, \ldots, f_n$ in $\operatorname{dom} \underline{P}$, it holds that

$$\sum_{i=1}^{n} \lambda_i \underline{P}(f_i) - \lambda_0 \underline{P}(f_0) \leq \sup\left[\sum_{i=1}^{n} \lambda_i f_i - \lambda_0 f_0\right]. \tag{3.4}$$

Explanation. The equivalence of (A) and (B) is mentioned, but not proved, in Walley [86, Lemma 2.5.4]. Let's first explain (A), and then prove that (A) $\Longrightarrow$ (B); the converse implication is immediate.

Suppose (3.3) is violated for some $n \in \mathbb{N}$, $m \in \mathbb{N}$, and gambles $f_0, f_1, \ldots, f_n$ in $\operatorname{dom} \underline{P}$. The case $m = 0$ means that $\underline{P}$ does not avoid sure loss. We already demonstrated that this is irrational (see Section 3.4.1). So suppose $m \neq 0$. Let

$$\sigma := \sup\left[\sum_{i=1}^{n} f_i - mf_0\right], \qquad \epsilon := \sum_{i=1}^{n} \underline{P}(f_i) - m\underline{P}(f_0) - \sigma > 0,$$

and $\delta := \frac{\epsilon}{n+1}$. Since $\delta > 0$, we are disposed to buy each f_i for $\underline{P}(f_i) - \delta$. Hence, by Axiom 3.1(iv), we are disposed to buy $\sum_{i=1}^{n} f_i$ for $\sum_{i=1}^{n} \underline{P}(f_i) - n\delta$, or equivalently, to buy $\sum_{i=1}^{n} f_i - \sigma$ for $\sum_{i=1}^{n} \underline{P}(f_i) - n\delta - \sigma$. Now, observe that

$$\sum_{i=1}^{n} f_i - \sigma = \sum_{i=1}^{n} f_i - \sup\left[\sum_{i=1}^{n} f_i - mf_0\right] \leq mf_0,$$

and hence, by Axiom 3.1(v), we should also be disposed to buy the larger reward mf_0 for the same price $\sum_{i=1}^{n} \underline{P}(f_i) - n\delta - \sigma$, and therefore, by Axiom 3.1(iii),

also to buy f_0 for $\frac{1}{m}\left(\sum_{i=1}^{n}\underline{P}(f_i) - n\delta - \sigma\right)$. But

$$\frac{1}{m}\left(\sum_{i=1}^{n}\underline{P}(f_i) - n\delta - \sigma\right) > \frac{1}{m}\left(\sum_{i=1}^{n}\underline{P}(f_i) - \epsilon - \sigma\right) = \underline{P}(f_0),$$

by the choice of δ, which means that $\underline{P}(f_0)$ is too low.

(A) $\Longrightarrow$ (B) Assume that (A) holds:

$$\sum_{i=1}^{n}\underline{P}(f_i) - m\underline{P}(f_0) \leq \sup\left[\sum_{i=1}^{n} f_i - mf_0\right],$$

for any $n \in \mathbb{N}$, $m \in \mathbb{N}$, and gambles $f_0, f_1, \ldots, f_n$ in $\operatorname{dom}\underline{P}$; and suppose *ex absurdo* that

$$\sum_{i=1}^{n}\lambda_i\underline{P}(f_i) - \lambda_0\underline{P}(f_0) > \sup\left[\sum_{i=1}^{n}\lambda_i f_i - \lambda_0 f_0\right],$$

for a particular choice of $n \in \mathbb{N}$, non-negative $\lambda_0, \lambda_1, \ldots, \lambda_n$ in $\mathbb{R}$, and gambles $f_0, f_1, \ldots, f_n$ in $\operatorname{dom}\underline{P}$. Define

$$\delta := \sum_{i=1}^{n}\lambda_i\underline{P}(f_i) - \lambda_0\underline{P}(f_0) - \sup\left[\sum_{i=1}^{n}\lambda_i f_i - \lambda_0 f_0\right] > 0$$

$\alpha := \sup\left[\sum_{i=0}^{n}|f_j|\right] + \sum_{i=0}^{n}|\underline{P}(f_j)| \geq 0$ and, $\epsilon := \frac{\delta}{2\alpha+1} > 0$. Since $\mathbb{Q}$ is dense in $\mathbb{R}$, there are non-negative rational numbers $\rho_i \in \mathbb{Q}$ such that $\lambda_i \leq \rho_i \leq \lambda_i + \epsilon$ for every $i \in \{0, \ldots, n\}$. By Lemma 3.4 we find that

$$\begin{aligned}
\rho_i f_i &\leq \lambda_i f_i + \epsilon|f_i|,\\
-\rho_0 f_0 &\leq -\lambda_0 f_0 + \epsilon|f_0|,\\
-\rho_i\underline{P}(f_i) &\leq -\lambda_i\underline{P}(f_i) + \epsilon\left|\underline{P}(f_i)\right|\\
\rho_0\underline{P}(f_0) &\leq \lambda_0\underline{P}(f_0) + \epsilon\left|\underline{P}(f_0)\right|
\end{aligned}$$

for every $j \in \{1, \ldots, n\}$. Let $k \in \mathbb{N}$ be a common denominator of $\rho_0, \ldots, \rho_n$ and let $m_i = k\rho_i \in \mathbb{N}$ for every $j \in \{0, \ldots, n\}$. We find that

$$\sup\left[\sum_{i=1}^{n}\rho_i f_i - \rho_0 f_0\right] - \left(\sum_{i=1}^{n}\rho_i\underline{P}(f_i) - \rho_0\underline{P}(f_0)\right)$$

$$
\begin{aligned}
&\leq \sup\left[\sum_{i=1}^{n} \lambda_i f_i - \lambda_0 f_0 + \sum_{i=0}^{n} \epsilon |f_i|\right] \\
&\quad - \left(\sum_{i=1}^{n} \lambda_i \underline{P}(f_i) - \lambda_0 \underline{P}(f_0)\right) + \sum_{i=0}^{n} \epsilon \left|\underline{P}(f_i)\right| \\
&\leq \sup\left[\sum_{i=1}^{n} \lambda_i f_i - \lambda_0 f_0\right] - \sum_{i=1}^{n} \lambda_i \underline{P}(f_i) + \lambda_0 \underline{P}(f_0) \\
&\quad + \epsilon \left(\sup\left[\sum_{i=0}^{n} |f_i|\right] + \sum_{i=0}^{n} \left|\underline{P}(f_i)\right|\right) \\
&= -\delta + \epsilon\alpha < -\delta/2.
\end{aligned}
$$

We conclude that

$$
\sup\left[\sum_{i=1}^{n} m_i f_i - m_0 f_0\right] - \left(\sum_{i=1}^{n} m_i \underline{P}(f_i) - m_0 \underline{P}(f_0)\right) < -k\delta/2 < 0.
$$

We have reached a contradiction. □

Lemma 3.4. *Let a, b and ϵ be real numbers and assume that $\epsilon \geq 0$. If $a \leq b \leq a + \epsilon$ then for every real number c we have that $ac - \epsilon|c| \leq bc \leq ac + \epsilon|c|$.*

Proof. If $c \geq 0$ then we have that $ac \leq bc \leq (a + \epsilon)c$ which implies that

$$
ac - \epsilon|c| \leq ac \leq bc \leq (a + \epsilon)c \leq ac + \epsilon|c|.
$$

On the other hand, if $c < 0$ then we have that $ac \geq bc \geq (a + \epsilon)c$ which implies that

$$
ac + \epsilon|c| \geq ac \geq bc \geq (a + \epsilon)c \geq ac - \epsilon|c|.
$$

In both cases we find that $ac - \epsilon|c| \leq bc \leq ac + \epsilon|c|$. □

Coherence is the minimal requirement we shall impose on lower previsions. Clearly, avoiding sure loss is weaker than coherence. So, why did we introduce avoiding sure loss, separately from coherence? Because it is sufficient for a lower prevision to avoid sure loss: in Theorem 4.3 on p. 96, we shall see that, any lower prevision that avoids sure loss, can be corrected to a behaviourally equivalent coherent lower prevision. This statement will be made precise in Section 4.1.

For previsions—self-conjugate lower previsions—avoiding sure loss and coherence are equivalent. De Finetti actually defines coherence for previsions as avoiding sure loss; see De Finetti [27, Vol. I, p. 87, Sect. 3.3.5]. Regarding terminology, Walley [86, Section 2.8.1, p. 86] calls a coherent prevision a linear prevision—but his definition is slightly more general: it extends to lower previsions that are not self-conjugate. We shall say that a lower prevision is *linear* if it is coherent and self-conjugate. Denote the set of all linear previsions with domain $\mathcal{K} \subseteq \mathcal{L}(X)$ by $\mathcal{P}^{\mathcal{K}}(X)$. The set $\mathcal{P}^{\mathcal{L}(X)}(X)$ is simply denoted by $\mathcal{P}(X)$.

Coherence has the following implications; we refer to Walley [86, Theorem 2.6.1] for a proof.

Theorem 3.5. *Let $\underline{P}$ be a coherent lower prevision. Let f and g be gambles on X. Let a be a constant gamble. Let λ and κ be reals with $\lambda \geq 0$ and $0 \leq \kappa \leq 1$. Let f_α be a net of gambles. Then the following statements hold whenever every term is well defined.*

(i) $\inf f \leq \underline{P}(f) \leq \overline{P}(f) \leq \sup f$

(ii) $\underline{P}(a) = \overline{P}(a) = a$

(iii) $\underline{P}(f + a) = \underline{P}(f) + a, \qquad \overline{P}(f + a) = \overline{P}(f) + a$

(iv) $f \leq g + a \implies \underline{P}(f) \leq \underline{P}(g) + a$ *and* $\overline{P}(f) \leq \overline{P}(g) + a$

(v) $\underline{P}(f) + \underline{P}(g) \leq \underline{P}(f + g) \leq \underline{P}(f) + \overline{P}(g) \leq \overline{P}(f + g) \leq \overline{P}(f) + \overline{P}(g)$

(vi) $\underline{P}(\lambda f) = \lambda \underline{P}(f), \qquad \overline{P}(\lambda f) = \lambda \overline{P}(f)$

(vii) $\kappa \underline{P}(f) + (1 - \kappa)\underline{P}(g) \leq \underline{P}(\kappa f + (1 - \kappa)g) \leq \kappa \underline{P}(f) + (1 - \kappa)\overline{P}(g) \leq \overline{P}(\kappa f + (1 - \kappa)g) \leq \kappa \overline{P}(f) + (1 - \kappa)\overline{P}(g)$

(viii) $\underline{P}(|f|) \geq \underline{P}(f), \qquad \overline{P}(|f|) \geq \overline{P}(f)$

(ix) $\left|\underline{P}(f) - \underline{P}(g)\right| \leq \overline{P}(|f - g|), \qquad \left|\overline{P}(f) - \overline{P}(g)\right| \leq \overline{P}(|f - g|)$

(x) $\underline{P}(|f + g|) \leq \underline{P}(|f|) + \overline{P}(|g|), \qquad \overline{P}(|f + g|) \leq \overline{P}(|f|) + \overline{P}(|g|)$

(xi) $\underline{P}(f \vee g) + \underline{P}(f \wedge g) \leq \underline{P}(f) + \overline{P}(g) \leq \overline{P}(f \vee g) + \overline{P}(f \wedge g)$,
$\underline{P}(f) + \underline{P}(g) \leq \underline{P}(f \vee g) + \overline{P}(f \wedge g) \leq \overline{P}(f) + \overline{P}(g)$, *and*
$\underline{P}(f) + \underline{P}(g) \leq \overline{P}(f \vee g) + \underline{P}(f \wedge g) \leq \overline{P}(f) + \overline{P}(g)$.

(xii) $\lim_\alpha \overline{P}(|f_\alpha - f|) = 0 \implies \lim_\alpha \underline{P}(f_\alpha) = \underline{P}(f)$ *and* $\lim_\alpha \overline{P}(f_\alpha) = \overline{P}(f)$.

(xiii) $\underline{P}$ *and* $\overline{P}$ *are uniformly continuous with respect to the topology of uniform convergence on their respective domains: for any* $\epsilon > 0$ *and any* f *and* g *in* $\operatorname{dom}\underline{P}$*, if* $\sup|f - g| < \epsilon$*, then* $|\underline{P}(f) - \underline{P}(g)| < \epsilon$*.*

If the domain of a lower prevision is a linear space, then we can characterise coherence in a simpler way. Again, we refer to Walley [86, Theorem 2.5.5, p. 75] for a proof.

Theorem 3.6. *Let* $\underline{P}$ *be a lower prevision on* X*, and assume that* $\operatorname{dom}\underline{P}$ *is a linear subspace of* $\mathcal{L}(X)$*. Then* $\underline{P}$ *is coherent if and only if for any gambles* f *and* g *in* $\operatorname{dom}\underline{P}$ *and any strictly positive real number* λ *the following conditions are met.*

(i) $\underline{P}(f) \geq \inf f$

(ii) $\underline{P}(\lambda f) = \lambda \underline{P}(f)$

(iii) $\underline{P}(f + g) \geq \underline{P}(f) + \underline{P}(g)$

For previsions, the characterisation is even simpler; see De Finetti [27, Sect. 3.5.1, p. 74] or Walley [86, Theorem 2.8.4, p. 88] for a proof.

Theorem 3.7. *Let* P *be a prevision on* X*, and assume that* $\operatorname{dom} P$ *is a linear subspace of* $\mathcal{L}(X)$*. Then* P *is a linear prevision (i.e.,* P *is coherent) if and only if for any gambles* f *and* $g \in \operatorname{dom} P$ *the following conditions are met.*

(i) $P(f) \geq \inf f$

(ii) $P(f + g) = P(f) + P(g)$

Because the conditions on linear subspaces are much easier to check, a very common method for proving coherence of lower previsions (or linearity of previsions) defined on arbitrary subsets of $\mathcal{L}(X)$ consists in proving that it is the restriction of a coherent lower prevision (or linear prevision) on a linear subspace of $\mathcal{L}(X)$. This simple observation is sufficiently important to call it a lemma.

Definition 3.8. Let $\underline{P}$ and $\underline{Q}$ be lower previsions on X. Then $\underline{P}$ is said to be a *restriction* of $\underline{Q}$, and $\underline{Q}$ is said to be an *extension* of $\underline{P}$, whenever $\operatorname{dom}\underline{P} \subseteq \operatorname{dom}\underline{Q}$ and $\underline{P}(f) = \underline{Q}(f)$ for all $f \in \operatorname{dom}\underline{P}$.

Lemma 3.9. *The following statements hold.*

(i) The restriction of a lower prevision avoiding sure loss also avoids sure loss.

(ii) The restriction of a coherent lower prevision is also coherent.

(iii) The restriction of a linear prevision to a prevision is also linear.

Proof. Immediately from Definition 3.2 and Definition 3.3. □

Lemma 3.9 will be useful in particular for proving the coherence of (lower and upper) probabilities, that is, lower previsions defined on indicator gambles. Typically, a lower probability is extended to a lower prevision on a linear space by means of an integral.

Another way to prove coherence of a lower prevision is to express it as a convex combination, a lower envelope, or a point-wise limit of lower previsions that are already known to be coherent.

Lemma 3.10. *Suppose* $\Gamma = \{\underline{P}_1, \underline{P}_2, \ldots, \underline{P}_n\}$ *is a finite collection of lower previsions defined on a common domain* $\mathcal{K}$*. Let* $\underline{Q}$ *be a convex combination of* Γ*:*

$$\underline{Q}(f) := \sum_{i=1}^{n} \lambda_i \underline{P}_i(f), \text{ for all } f \in \mathcal{K},$$

for some $\lambda_1, \ldots, \lambda_n \geq 0$ *and* $\sum_{i=1}^{n} \lambda_i = 1$*. Then the following statements hold.*

(i) If all lower previsions in Γ *avoid sure loss, then* $\underline{Q}$ *avoids sure loss.*

(ii) If all lower previsions in Γ *are coherent, then* $\underline{Q}$ *is coherent.*

(iii) If all lower previsions in Γ *are linear previsions, then* $\underline{Q}$ *is a linear prevision.*

Proof. See Walley [86, Theorem 2.6.4, p. 79] for a proof of (i) and (ii). To prove (iii) it suffices to check self-conjugacy of $\underline{Q}$:

$$\underline{Q}(-f) = \sum_{i=1}^{n} \lambda_i P_i(-f) = -\sum_{i=1}^{n} \lambda_i P_i(f) = -\underline{Q}(f).$$

□

Lemma 3.11. *Suppose* Γ *is a collection of lower previsions defined on a common domain* $\mathcal{K}$*. Let* $\underline{Q}$ *be the lower envelope of* Γ*:*

$$\underline{Q}(f) := \inf_{\underline{P} \in \Gamma} \underline{P}(f), \text{ for any } f \in \mathcal{K}.$$

Then the following statements hold.

(i) *If all lower previsions in* Γ *avoid sure loss, then* $\underline{Q}$ *avoids sure loss.*

(ii) *If all lower previsions in* Γ *are coherent, then* $\underline{Q}$ *is coherent.*

(iii) *If all lower previsions in* Γ *are linear, then* $\underline{Q}$ *is coherent but not linear, unless* Γ *is a singleton.*

Proof. See Walley [86, Theorem 2.6.3] for a proof of (i) and (ii). To prove (iii) it suffices to check that $\underline{Q}$ is not self-conjugate whenever Γ contains more than one element. Choose two distinct linear previsions P_1 and P_2 in Γ. Since they are distinct, there is a gamble $f \in \mathcal{K}$ such that for instance $P_1(f) < P_2(f)$. This means that

$$\underline{Q}(f) \leq P_1(f) < P_2(f) \leq \overline{Q}(f),$$

therefore $\underline{Q}$ is not self-conjugate, which establishes the proof. □

Lemma 3.12. *Suppose* $\underline{P}_\alpha$ *is a net of lower previsions defined on a common domain* $\mathcal{K}$*. Suppose that* $\underline{P}_\alpha$ *converges point-wise to a lower prevision* $\underline{Q}$ *defined on* $\mathcal{K}$*. Then the following statements hold.*

(i) *If all* $\underline{P}_\alpha$ *avoid sure loss then* $\underline{Q}$ *avoids sure loss.*

(ii) *If all* $\underline{P}_\alpha$ *are coherent then* $\underline{Q}$ *is coherent.*

(iii) *If all* $\underline{P}_\alpha$ *are linear then* $\underline{Q}$ *is linear.*

Proof. See Walley [86, Theorem 2.6.5, p. 79] for a proof of (i) and (ii). To prove (iii) it suffices to check self-conjugacy of $\underline{Q}$:

$$\underline{Q}(f) = \lim_\alpha \underline{P}_\alpha(f) = \lim_\alpha -\underline{P}_\alpha(-f) = -\lim_\alpha \underline{P}_\alpha(-f) = -\underline{Q}(-f) = \overline{Q}(f).$$

This establishes the lemma. □

Finally, continuity may also help in proving that a lower prevision is coherent. This establishes a partial converse of Lemma 3.9.

Lemma 3.13. *Suppose* $\underline{P}$ *is continuous on its domain with respect to the topology of uniform convergence, and coherent on a dense subset* $\mathcal{K}$ *of* $\operatorname{dom} \underline{P}$*. Then* $\underline{P}$ *is coherent.*

Proof. Let $\epsilon > 0$. Since $\mathcal{K}$ is dense in $\operatorname{dom} \underline{P}$, we have that for any gamble $f \in \operatorname{dom} \underline{P}$ there is an $f_\epsilon \in \mathcal{K}$ such that $\sup \left|f - f^\epsilon\right| < \epsilon$. Therefore, for any $m \in \mathbb{N}$, $n \in \mathbb{N} \setminus \{0\}$ and $f_0, f_1, \ldots, f_n$ in $\operatorname{dom} \underline{P}$ it holds that

$$\begin{aligned}\sum_{i=1}^{n} \underline{P}(f_i) - m\underline{P}(f_0) &\leq (n+m)\epsilon + \sum_{i=1}^{n} \underline{P}(f_i^\epsilon) - m\underline{P}(f_0^\epsilon) \\ &\leq (n+m)\epsilon + \sup\left[\sum_{i=1}^{n} f_i^\epsilon - m f_0^\epsilon\right] \\ &\leq 2(n+m)\epsilon + \sup\left[\sum_{i=1}^{n} f_i - m f_0\right].\end{aligned}$$

Since this holds for any $\epsilon > 0$, it must also hold for $\epsilon = 0$. Coherence follows. □

The following is a variation on the same theme.

Lemma 3.14. *Let $\underline{P}$ be a coherent lower prevision. Then $\underline{P}$ has a unique extension to a coherent lower prevision defined on the uniform closure of* $\operatorname{dom} \underline{P}$. *Moreover, if $\underline{P}$ is self-conjugate, then also this unique extension is self-conjugate.*

Proof. In Chapter 4, Section 4.1, Theorem 4.3, on p. 96, it will be proved that any coherent lower prevision $\underline{P}$ has a coherent extension to an arbitrary (but larger) domain. Since the proof is so much shorter using this result, we shall cheat, and use it; of course, Theorem 4.3 does not rely on this lemma. So, there are coherent extensions $\underline{Q}$ of $\underline{P}$ to the uniform closure of $\operatorname{dom} \underline{P}$.

Let $\underline{Q}_1$ and $\underline{Q}_2$ be two coherent lower previsions defined on the uniform closure of $\operatorname{dom} \underline{P}$, and suppose that $\underline{Q}_1(f) = \underline{Q}_2(f) = \underline{P}(f)$ for all f in $\operatorname{dom} \underline{P}$. We must show that $\underline{Q}_1(g) = \underline{Q}_2(g)$ for all g in $\operatorname{dom} \underline{Q}_1 = \operatorname{dom} \underline{Q}_2$. For every such g, there is a sequence f_n in $\operatorname{dom} \underline{P}$ that converges uniformly to g. Since $\underline{Q}_1$ and $\underline{Q}_2$ are coherent, it follows from Theorem 3.5(xii) that

$$\underline{Q}_1(g) = \lim_{n\in\mathbb{N}} \underline{Q}_1(f_n) = \lim_{n\in\mathbb{N}} \underline{P}(f_n) = \lim_{n\in\mathbb{N}} \underline{Q}_2(f_n) = \underline{Q}_2(g),$$

and hence, $\underline{Q}_1(g) = \underline{Q}_2(g)$ for any gamble g in their domain (incidentally, note that the limit is independent of the choice of the sequence f_n converging uniformly to g). This proves uniqueness.

If $\underline{P}$ is self-conjugate, then, again take for every g in $\operatorname{dom} \underline{Q}$ a sequence f_n in $\operatorname{dom} \underline{P}$ that converges uniformly to g. Since $\underline{Q}$ is coherent, it follows from

Theorem 3.5(xii) that

$$\underline{Q}(g) = \lim_{n\in\mathbb{N}} \underline{Q}(f_n) = \lim_{n\in\mathbb{N}} \underline{P}(f_n) = \lim_{n\in\mathbb{N}} \overline{P}(f_n) = \lim_{n\in\mathbb{N}} \overline{Q}(f_n) = \overline{Q}(g).$$

□

3.5 Examples of Coherent Lower Previsions

3.5.1 Vacuous Lower Previsions

We may know nothing at all about the realisation of X. We express this by stating that we are only disposed to buy a gamble f on X for a price that is a lower bound of it:

$$\underline{P}_{\mathcal{X}}(f) = \inf f.$$

This lower prevision, defined on the set of all gambles $\mathcal{L}(X)$, is called the *vacuous lower prevision* on X. It is coherent (check the conditions of Theorem 3.6).

More generally, we might only be sure that the realisation of X belongs to a non-empty subset A of $\mathcal{X}$. We then should only be willing to buy a gamble f on X for a price than is a lower bound of $f(X)$ restricted to A:

$$\underline{P}_A(f) = \inf_{x\in A} f(x) = \max\{a \in \mathbb{R} \colon \forall x \in A,\, a \leq f(x)\}. \tag{3.5}$$

Again this lower prevision is defined on the set of all gambles $\mathcal{L}(X)$ and is coherent. We call it the *vacuous lower prevision on X relative to A*. Its conjugate is given by

$$\overline{P}_A(f) = \sup_{x\in A} f(x) = \min\{a \in \mathbb{R} \colon \forall x \in A,\, a \geq f(x)\}.$$

Within the class of vacuous lower previsions, only the ones relative to singletons are linear. They model exact knowledge of the value of X. Lower envelopes of vacuous lower previsions are again vacuous, but convex combinations usually are not. In Section 4.3.12 we shall show that convex combinations of vacuous lower previsions correspond to the natural extension of so-called belief functions.

3.5.2 Probability Charges

Probability charges, also called finitely additive probability measures, are a very common way of representing uncertainty about a random variable. They are slightly more general than so-called probability measures, used in the classical theory of probability. In Section 4.4, we shall establish that avoiding sure loss, coherence, and natural extension (see Section 4.1 further on), can be expressed using probability charges only, whence their importance in the theory of lower previsions.

First, we identify a sufficiently large collection of events of interest. Conveniently, we assume that this collection has the structure of a field.

Definition 3.15. A *field $\mathcal{F}$ on $\mathcal{X}$* is a collection of subsets of $\mathcal{X}$ that contains the empty set and is closed under finite unions and complementation. A *σ-field $\mathcal{F}$ on $\mathcal{X}$* is a field on $\mathcal{X}$ that is also closed under countable unions. An *ample field on $\mathcal{X}$* is a field on $\mathcal{X}$ that is also closed under arbitrary unions.

For example, if $\mathcal{X}$ is a topological space, the smallest σ-field that contains all open sets is called the *Borel σ-field* on $\mathcal{X}$ and is denoted by $\mathcal{B}(\mathcal{X})$. The members of this field are called *Borel sets*. For instance, $\mathcal{B}(\mathbb{R})$ corresponds to the smallest σ-field that contains all open intervals (a, b) (where $a, b \in \mathbb{R}$ and $a < b$). In measure theory, this is the standard way to equip $\mathbb{R}$ with a σ-field.

A simpler example is the smallest field that contains all closed intervals of a compact interval $[a, b]$. We denote this field by $\mathcal{F}_{[]}([a, b])$. It contains all finite unions of intervals in $[a, b]$, and turns out to be a handy tool in the study of Riemann-Stieltjes integrals.

Examples of ample fields on $\mathcal{X}$ are the set $\{\emptyset, \mathcal{X}\}$, and the set of all subsets of $\mathcal{X}$, also called the *power set* of $\mathcal{X}$:

$$\wp(\mathcal{X}) := \{A \colon A \subseteq \mathcal{X}\}.$$

Any finite field is also an ample field. It is well-known that there is a correspondence between ample fields and partitions; see Wang [89], and Theorem 3.51 on p. 88.

Obviously, any ample field is a σ-field, and any σ-field is a field. If $\mathcal{X}$ is finite, then any field on $\mathcal{X}$ is also a σ-field and an ample field. If $\mathcal{X}$ is countable, then any σ-field on $\mathcal{X}$ is also an ample field.

Next, to each event A in a field $\mathcal{F}$ we attach a real number $\mu(A)$ that

measures our belief in A; typically, it represents a chance, a betting rate, or a price. To be a probability charge, μ should satisfy a number of additional properties.

Definition 3.16. Let $\mathcal{F}$ be a field on $\mathcal{X}$. A *charge* μ *on* $\mathcal{F}$, also called *finitely additive measure*, *finitely additive set function*, or *additive set function*, is an $\mathbb{R}^*$-valued mapping on $\mathcal{F}$ that assumes at most one of the values $+\infty$ and $-\infty$, and satisfies

(i) $\mu(\emptyset) = 0$, and

(ii) $\mu(A \cup B) = \mu(A) + \mu(B)$ whenever $A, B \in \mathcal{F}$ and $A \cap B = \emptyset$.

A charge μ on $\mathcal{F}$ is called a *probability charge* if additionally

(iii) $\mu(\mathcal{X}) = 1$, and

(iv) $\mu(A) \geq 0$ for any $A \in \mathcal{F}$.

The set of probability charges on $\mathcal{F}$ is denoted by $\mathcal{P}(\mathcal{F})$. Finally, a charge μ is said to be *σ-additive* if additionally $\mathcal{F}$ is a σ-field and

(v) for any sequence A_n of pairwise disjoint sets in $\mathcal{F}$, the limit

$$\lim_{n\to+\infty} \sum_{i=1}^{n} \mu(A_i)$$

exists in $\mathbb{R}^*$ and is equal to $\mu\left(\bigcup_{n\in\mathbb{N}} A_n\right)$ (and hence, the limit is independent of the order of the sequence).

A σ-additive charge is simply called a *measure* and a σ-additive probability charge is simply called a *probability measure*.

Note that in the literature, there are more general definitions of a measure: for instance, Bhaskara Rao and Bhaskara Rao [9, Definition 2.3.1, p. 47] allow a measure to be defined on a field that is not a σ-field, Halmos [40, Section 7, p. 30] allows a measure to be defined on a ring (*i.e.*, a collection of sets closed under union and difference), and also König [51, Chapter I, Section 2, p. 14] has a different definition of a measure, allowing it to be defined on what he calls a σ-oval. However, usually, measures are assumed to be defined on a σ-field, see for instance Kallenberg [48, Chapter 1, p. 8] and Schechter [70, Section 11.37, p. 288], and this is also the assumption we make. Observe

that probability charges do not assume the values $+\infty$ or $-\infty$; they are $[0,1]$-valued.

Perhaps the most commonly used charge is the *Lebesgue measure* λ on $\mathcal{B}(\mathbb{R})$, which is defined as the unique σ-additive measure on $\mathcal{B}(\mathbb{R})$ such that $\lambda((a,b)) = \lambda([a,b)) = \lambda((a,b]) = \lambda([a,b]) = b - a$ for any $a, b \in \mathbb{R}$ $(a \leq b)$: it measures the length of intervals of $\mathbb{R}$. Vitali [82] proved that there is no σ-additive measure on all of $\wp(\mathbb{R})$ that has this property—if we accept the axiom of choice; also see for instance Schechter [70, Section 21.22, p. 558]. This is also one of the reasons why we need to introduce the Borel σ-field. Sometimes the Lebesgue measure is defined on larger σ-fields, see for instance Halmos [40, Section 15]. We have found at least three different Lebesgue measures.[1] However, we prefer to take the simplest definition and *always* assume the Lebesgue measure to be defined only on the Borel σ-field of $\mathbb{R}$.

As another example, consider the *total variation* $\left|\mu\right|$ of a charge μ on a field $\mathcal{F}$, which is defined as

$$\left|\mu\right|(A) = \sup \sum_{i=1}^{n} \left|\mu(A_i)\right|$$

for any $A \in \mathcal{F}$, where the supremum is taken over all finite partitions $\{A_1, \ldots, A_n\} \subseteq \mathcal{F}$ of A. It is easy to check that $\left|\mu\right|$ is a charge. When μ is positive (that is, $\mu(A) \geq 0$ for all $A \in \mathcal{F}$, which holds for instance for probability charges and also for the Lebesgue measure), $\left|\mu\right|$ is equal to μ.

Identifying events with indicator gambles, observe that any real-valued mapping μ on $\mathcal{F}$ can be identified with a lower probability, an upper probability, or a probability. In doing this, $\mathcal{F}$ need not even be a field. Let's emphasise the difference between what we call probabilities, and probability charges. Probabilities, defined in the last paragraph of Section 3.3.1, are previsions whose domain only contains indicator gambles along with their negations—without any further restrictions. Probability charges, on the other hand, are additive, positive, and normed real-valued mappings defined on a field. So, as mathematical objects, they are quite different from probabilities. Through

[1] The Lebesgue measure defined on the Borel σ-field is sometimes also called the Borel-Lebesgue measure; see Schechter [70, Section 21.19]. The completion of the Borel-Lebesgue measure is usually also called the Lebesgue measure; see Halmos [40]. Finally, the Carathéodory extension of the Borel-Lebesgue measure to the σ-field of λ^*-measurable sets—again, see Halmos [40]—might as well be called the Lebesgue measure. The Carathéodory extension also agrees with the linear extension, as described in Section 4.3.4 on p. 116 ff.

the identification below, we shall establish a link between probabilities and probability charges in Theorem 3.21.

Definition 3.17. Let μ be any real-valued mapping defined on a class of subsets of $\mathcal{X}$. The *lower probability induced by* μ is defined as the lower prevision $\underline{\mathbf{P}}_\mu$ that maps all gambles I_A for $A \in \operatorname{dom} \mu$ to $\mu(A)$, the *upper probability induced by* μ the upper prevision $\overline{\mathbf{P}}_\mu$ that maps all gambles I_A for all $A \in \operatorname{dom} \mu$ to $\mu(A)$, and the *probability induced by* μ the prevision $\mathbf{P}_\mu$ that maps all gambles I_A for $A \in \mathcal{F}$ to $\mu(A)$ and all gambles $-I_A$ for $A \in \mathcal{F}$ to $-\mu(A)$:

$$\left.\begin{aligned} \underline{\mathbf{P}}_\mu(I_A) &:= \mu(A) \\ \overline{\mathbf{P}}_\mu(I_A) &:= \mu(A) \\ \mathbf{P}_\mu(I_A) = -\mathbf{P}_\mu(-I_A) &:= \mu(A) \end{aligned}\right\} \text{ for all } A \in \operatorname{dom} \mu.$$

A natural question is: are these (lower/upper) probabilities coherent? If not, do they at least avoid sure loss? In general, this question is hard to answer. In case the domain of μ is a field, we shall establish in Theorem 3.21 that $\mathbf{P}_\mu$ is a coherent probability if and only if μ is a probability charge. Hence, by Lemma 3.9, in such a case also $\underline{\mathbf{P}}_\mu$ is a coherent lower probability and $\overline{\mathbf{P}}_\mu$ is a coherent upper probability. The proof invokes an extension of $\mathbf{P}_\mu$ to the linear span of $\{I_A : A \in \mathcal{F}\}$, called the *Dunford integral*. This integral was introduced as an integral of vector-valued functions with respect to measures by Dunford [31, p. 443, Sect. 3], and extended to an integral of vector-valued functions with respect to charges by Dunford and Schwartz [30, Part I, Chapter III, Definition 2.17, p. 112]; also see Bhaskara Rao and Bhaskara Rao [9, Chapter 4] for a detailed study of the Dunford integral on scalar-valued functions. For establishing the coherence of probability charges in Theorem 3.21, we only need the Dunford integral defined on $\mathcal{F}$-simple (see Definition 3.18 below) gambles; see Dunford and Schwartz [30, Part I, Chapter III, Definition 2.13, p. 108]. As discussed by Dunford and Schwartz [30, Part I, Chapter III, Section 2, pp. 101–119], the Dunford integral is first defined on $\mathcal{F}$-simple gambles—this is the main reason why we call the integral defined below the Dunford integral—and then extended to more general functions through Cauchy sequences; this method is due to Dunford [31, Lemma 6, p. 444], and will be discussed in Section 4.3.8 on p. 161 ff., as it also forms a possible basis to extend coherent lower previsions to functions of X that are possibly unbounded; this is the subject of Chapter 5. Note that, on $\mathcal{F}$-simple gambles,

the Dunford integral coincides with the *S-integral* introduced by Hildebrandt [42, Sect. 1(f), p. 869] as an integral associated with charges. We shall discuss the S-integral in Section 4.3.5. As Hildebrandt [42, Sect. 1(f), p. 869] notes, 'it is possible to define the Lebesgue integral [with respect to a charge] by the Lebesgue process' and for $\mathcal{F}$-simple gambles (or more generally, $\mathcal{F}$-measurable gambles), 'obviously [the Lebesgue integral] exists', and the '[S-integral] exists also in this case and agrees with the [Lebesgue integral]'. Concluding, the integral of Definition 3.19 coincides with just about *any* integral found in the literature for charges and $\mathcal{F}$-simple gambles. Other extensions of probability charges will be discussed in Sections 4.3.2 and 4.3.4. So, what are $\mathcal{F}$-simple gambles?

Definition 3.18. Let $\mathcal{F}$ be a field on $\mathcal{X}$. A gamble f on X is called *$\mathcal{F}$-simple* if it belongs to the linear span of $I_{\mathcal{F}}$, that is, if

$$f = \sum_{i=1}^{n} a_i I_{A_i},$$

for some $n \in \mathbb{N}$, $a_1, \ldots, a_n$ in $\mathbb{R}$ and $A_1, \ldots, A_n$ in $\mathcal{F}$. The sum $\sum_{i=1}^{n} a_i I_{A_i}$ is called a *representation* of f. The set of $\mathcal{F}$-simple gambles is denoted by $\mathrm{span}(\mathcal{F})$.

Note that $\mathrm{span}(\mathcal{F})$ is a simplified notation for $\mathrm{span}(I_{\mathcal{F}})$, the linear span of all indicators of elements of $\mathcal{F}$.

Definition 3.19. Let $\mathcal{F}$ be a field on $\mathcal{X}$ and let μ be a charge on $\mathcal{F}$. An $\mathcal{F}$-simple gamble f is called *Dunford integrable* with respect to μ if it has a representation $f = \sum_{i=1}^{n} a_i I_{A_i}$ such that $\left|\mu\right|(A_i) < +\infty$ for all $i \in \{1, \ldots, n\}$. The *Dunford integral* of f with respect to μ is then defined as

$$\mathrm{D}\int f \,\mathrm{d}\mu := \sum_{i=1}^{n} a_i \mu(A_i). \tag{3.6}$$

Proof of unambiguity. Consider the linear space of gambles spanned by

$$\mathcal{K} := \{I_A : A \in \mathcal{F},\ \left|\mu\right|(A) < +\infty\}.$$

Note that $\mathrm{span}(\mathcal{K})$ is exactly the set of Dunford integrable $\mathcal{F}$-simple gambles. Define the mapping $\psi(I_A) := \mu(A)$ on $\mathcal{K}$. Now note that ψ satisfies the

condition

$$\sum_{i=1}^{n} \alpha_i I_{A_i} = 0 \implies \sum_{i=1}^{n} \alpha_i \psi(I_{A_i}) = 0 \tag{3.7}$$

for every $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$ and $I_{A_1}, \ldots, I_{A_n} \in \mathcal{K}$. Indeed, suppose $\sum_{i=1}^{n} \alpha_i I_{A_i} = 0$. Let $\mathcal{B} \subseteq \mathcal{F}$ be a finite partition of $\mathcal{X}$ such that for every $i \in \{1, \ldots, n\}$ it holds that A_i is a union of elements of $\mathcal{B}$. Then,

$$0 = \sum_{i=1}^{n} \alpha_i I_{A_i} = \sum_{i=1}^{n} \alpha_i \sum_{\substack{B \in \mathcal{B} \\ B \subseteq A_i}} I_B = \sum_{B \in \mathcal{B}} \left(\sum_{\substack{i \in \{1,\ldots,n\} \\ B \subseteq A_i}} \alpha_i \right) I_B$$

and since $\{I_B \colon B \in \mathcal{B}\}$ is linearly independent this means that

$$\sum_{\substack{i \in \{1,\ldots,n\} \\ B \subseteq A_i}} \alpha_i = 0$$

for every $B \in \mathcal{B}$. In particular,

$$\begin{aligned} \sum_{i=1}^{n} \alpha_i \psi(I_{A_i}) &= \sum_{i=1}^{n} \alpha_i \psi\left(\sum_{\substack{B \in \mathcal{B} \\ B \subseteq A_i}} I_B \right) = \sum_{i=1}^{n} \alpha_i \sum_{\substack{B \in \mathcal{B} \\ B \subseteq A_i}} \psi(I_B) \\ &= \sum_{B \in \mathcal{B}} \left(\sum_{\substack{i \in \{1,\ldots,n\} \\ B \subseteq A_i}} \alpha_i \right) \psi(I_B) = 0, \end{aligned}$$

where we used additivity of μ. Hence, Eq. (3.7) holds.

By a linear extension theorem (Schechter [70, Proposition 11.10]) it follows from Eq. (3.7) that ψ has a unique linear extension Ψ to $\operatorname{span}(\mathcal{K})$ given by

$$\Psi\left(\sum_{i=1}^{n} \alpha_i I_{A_i} \right) = \sum_{i=1}^{n} \alpha_i \psi(I_{A_i}). \tag{3.8}$$

But this means that Ψ is the Dunford integral. In other words, the Dunford integral is uniquely determined by (3.6). □

For $A \in \mathcal{F}$, we say that an $\mathcal{F}$-simple gamble f is Dunford integrable over A with respect to μ whenever $I_A f$ is Dunford integrable with respect to μ. In such a case, we call $\mathrm{D}\int I_A f \, \mathrm{d}\mu$ the Dunford integral of f over A with respect

to μ, and we also write

$$\mathrm{D}\int_A f \,\mathrm{d}\mu := \mathrm{D}\int I_A f \,\mathrm{d}\mu. \tag{3.9}$$

Lemma 3.20. *Let $\mathcal{F}$ be a field on $\mathcal{X}$ and let μ be a probability charge on $\mathcal{F}$. Then all $\mathcal{F}$-simple gambles are Dunford integrable with respect to μ, and the mapping $\mathbf{E}_\mu^{\mathrm{span}(\mathcal{F})}$ defined by*

$$\mathbf{E}_\mu^{\mathrm{span}(\mathcal{F})}(f) := \mathrm{D}\int f \,\mathrm{d}\mu \tag{3.10}$$

for any $\mathcal{F}$-simple gamble f, defines a linear prevision on $\mathrm{span}(\mathcal{F})$.

Proof. Dunford integrability of all $\mathcal{F}$-simple gambles follows from $|\mu| = \mu$ for probability charges μ. Note that $\mathbf{E}_\mu^{\mathrm{span}(\mathcal{F})}$ is self-conjugate, this easily follows from its definition, and hence, $\mathbf{E}_\mu^{\mathrm{span}(\mathcal{F})}$ is a prevision. Now, simply check the conditions of Theorem 3.7. □

We now arrive at the main result of this section: probability charges are coherent probabilities. Obviously, by Lemma 3.9, this implies that they are also coherent either as lower probabilities or upper probabilities.

Theorem 3.21. *Let $\mathcal{F}$ be a field. Let μ be any real-valued function defined on $\mathcal{F}$. Then $\mathbf{P}_\mu$ is coherent if and only if μ is a probability charge.*

Proof. Let μ be a probability charge defined on a field $\mathcal{F}$. Observe that the probability μ is the restriction of $\mathbf{E}_\mu^{\mathrm{span}(\mathcal{F})}$ which is a linear prevision by Lemma 3.20. Now apply Lemma 3.9 to arrive at the desired result.

Conversely, let $\mathbf{P}_\mu$ be coherent. Observe that $\mathbf{P}_\mu$ is a linear probability on $I_\mathcal{F} \cup -I_\mathcal{F}$. From $\mu(A) = \mathbf{P}_\mu(A)$ for all $A \in \mathcal{F}$, it is very easy to show that μ satisfies the properties of a probability charge simply by applying the linearity of $\mathbf{P}_\mu$. □

Equivalently, it holds that $\mathbf{P}_\mu$ avoids sure loss if and only if μ is a probability charge.

We have already shown that the linear prevision $\mathbf{E}_\mu^{\mathrm{span}(\mathcal{F})}$ is the unique linear extension of the probability $\mathbf{P}_\mu$ to the set of $\mathcal{F}$-simple gambles. We shall prove further on that $\mathbf{E}_\mu^{\mathrm{span}(\mathcal{F})}$ is actually the *only* coherent extension of $\mathbf{P}_\mu$ to $\mathrm{span}(\mathcal{F})$ (this will follow from a stronger result; see Proposition 4.28 on p. 112). In fact, it is also the only coherent extension of the lower probability $\underline{\mathbf{P}}_\mu$, and the only coherent extension of the upper probability $\overline{\mathbf{P}}_\mu$, to $\mathrm{span}(\mathcal{F})$.

In this sense, additivity implies self-conjugacy (see Corollary 4.33). In order to come up with event-based models, (lower or upper) probabilities, that are not self-conjugate, we need to weaken additivity. A few ways of doing this are described in Sections 3.5.4, 3.5.5, 3.5.7, 3.5.8, 3.5.9, and 3.5.10. But first, we fix the minimal properties a set function should satisfy, to be interpreted as a coherent (lower/upper) probability.

3.5.3 Set Functions

Definition

Let, for now, μ be an arbitrary set function, *i.e.*, an arbitrary real-valued mapping defined on an arbitrary collection $\mathcal{A}$ of subsets of $\mathcal{X}$, which we shall interpret either as a lower probability, an upper probability, or a probability; for simplicity, also assume that $\emptyset \in \mathcal{A}$ and $\mathcal{X} \in \mathcal{A}$. Then, *independently of how we interpret* μ, it follows from Theorem 3.5 that, if either $\underline{\mathbf{P}}_\mu$, $\overline{\mathbf{P}}_\mu$, or $\mathbf{P}_\mu$ is coherent, then $\mu(\emptyset) = 0$, $\mu(\mathcal{X}) = 1$, and $\mu(A) \leq \mu(B)$ for all A and B in $\mathcal{A}$ such that $A \subseteq B$. Explicitly:

- if $\underline{\mathbf{P}}_\mu$ is coherent, then $\underline{\mathbf{P}}_\mu(I_\emptyset) = 0$, $\underline{\mathbf{P}}_\mu(I_\mathcal{X}) = 1$, and $A \subseteq B \implies \underline{\mathbf{P}}_\mu(I_A) \leq \underline{\mathbf{P}}_\mu(I_B)$, for all A and B in $\mathcal{A}$,
- if $\overline{\mathbf{P}}_\mu$ is coherent, then $\overline{\mathbf{P}}_\mu(I_\emptyset) = 0$, $\overline{\mathbf{P}}_\mu(I_\mathcal{X}) = 1$, and $A \subseteq B \implies \overline{\mathbf{P}}_\mu(I_A) \leq \overline{\mathbf{P}}_\mu(I_B)$, for all A and B in $\mathcal{A}$, and
- if $\mathbf{P}_\mu$ is coherent, then $\mathbf{P}_\mu(I_\emptyset) = 0$, $\mathbf{P}_\mu(I_\mathcal{X}) = 1$, and $A \subseteq B \implies \mathbf{P}_\mu(I_A) \leq \mathbf{P}_\mu(I_B)$, for all A and B in $\mathcal{A}$.

Since we are only interested in set functions that induce either coherent lower probabilities, coherent upper probabilities, or coherent probabilities, it is consistent to include the above conditions into our definition of a set function. Why exactly take—only—these conditions? Indeed, they are necessary, but not sufficient for μ to induce a coherent lower probability, a coherent upper probability, or a coherent probability. Nevertheless,

- they are simple,
- they are independent of whether we interpret μ as a coherent lower probability, as a coherent upper probability, or as a coherent probability, and

- for all the specialised types of set functions we shall study further on, it suffices to impose only one additional property—such as nesting, 2-monotonicity, minimum preservation, *etc.*—for μ to induce a coherent lower probability, a coherent upper probability, or a coherent probability.

Definition 3.22. A *set function* is a real-valued mapping μ, defined on a collection $\mathcal{A}$ of subsets of $\mathcal{X}$, such that

(i) $\emptyset \in \mathcal{A}$ and $\mathcal{X} \in \mathcal{A}$,

(ii) $\mu(\emptyset) = 0$ and $\mu(\mathcal{X}) = 1$, and

(iii) $A \subseteq B \implies \mu(A) \leq \mu(B)$ for any A and B in $\mathcal{A}$.

The set $\mathcal{A}$ is called the *domain of* μ and is denoted by $\operatorname{dom} \mu$.

So, from now on, all set functions are assumed to satisfy the conditions of the above definition. This deviates from the terminology used in the literature: usually, set functions are only assumed to be non-negative, and zero on the empty set. But, for the purpose of studying the coherence of event-based models, Definition 3.22 simplifies our study a lot. Note that probability charges are set functions: since they induce coherent probabilities, they satisfy the above conditions; however, we shall not call all charges set functions: for instance, we shall not call the Lebesgue measure a set function, because $\lambda(\mathbb{R}) = +\infty$. In this work, this will not form any obstacle whatsoever.

Dual Set Functions

Defining lower and upper previsions in Section 3.3.1, we have seen that from any lower prevision $\underline{P}$ we can infer a conjugate upper prevision $\overline{P}$ on $\operatorname{dom} \overline{P} = -\operatorname{dom} \underline{P}$ which represents the same behavioural dispositions.

Let ν be a real-valued set function, and assume that we interpret it as a lower probability: we consider the lower probability $\underline{P} := \underline{\mathbf{P}}_\nu$ induced by ν—note that we will usually denote by ν any set function that is meant to be interpreted as a lower probability. The conjugate upper prevision $\overline{P}$ of $\underline{P}$ can be defined by $\overline{P}(-I_A) := -\nu(A)$ for all $A \in \operatorname{dom} \nu$, and unfortunately, $\overline{P}$ is not defined on a set of indicator gambles: $\overline{P}$ is not an upper probability induced by a set function π—we shall usually denote by π any set function that is meant to be interpreted as an upper probability. Hence, there seems to

be no correspondence between lower and upper probabilities similar to the correspondence between lower and upper previsions.

However, through a simple transformation, we can infer, through coherence, an upper probability from $\overline{P}$: by Theorem 3.5, it should hold that

$$\overline{P}(I_{\complement A}) = \overline{P}(1 - I_A) = 1 - \underline{P}(I_A)$$

whenever $\underline{P}$ is coherent. This suggests the following definition.

Definition 3.23. Let ν be a real-valued set function. The *dual* of ν is the real-valued set function π defined on $\left\{\complement A \colon A \in \operatorname{dom} \nu\right\}$ by

$$\pi(\complement A) := 1 - \nu(A),$$

for any $A \in \operatorname{dom} \nu$. If $\operatorname{dom} \nu$ is closed under complementation and $\nu(\complement A) = 1 - \nu(A)$ for all $A \in \operatorname{dom} \nu$, then ν is called *self-dual*.

There is a close relationship between conjugate upper previsions and dual set functions, as we shall see in Proposition 3.24 and Proposition 4.10. Note that fields are closed under complementation. We have already seen an example of self-dual set functions on a field: in case ν is a probability charge, its dual π is equal to ν, whence, probability charges are self-dual set functions.

The following proposition says that avoiding sure loss and coherence are preserved for the dual set function. In Proposition 4.10 on p. 99, we shall establish a stronger result: ν models the same behavioural dispositions as its dual π.

Proposition 3.24. *Let ν be a real-valued set function, and let π be its dual. Then $\underline{\mathbf{P}}_\nu$ avoids sure loss if and only if $\overline{\mathbf{P}}_\pi$ avoids sure loss, and $\underline{\mathbf{P}}_\nu$ is coherent if and only if $\overline{\mathbf{P}}_\pi$ is coherent. If ν is self-dual, then $\underline{\mathbf{P}}_\nu$ is coherent if and only if $\overline{\mathbf{P}}_\nu$ is coherent, if and only if $\mathbf{P}_\nu$ is coherent.*

Proof. Immediate from the definitions of avoiding sure loss and coherence. except for proving that, whenever ν is self-dual, coherence of $\underline{\mathbf{P}}_\nu$ is equivalent to coherence of $\mathbf{P}_\nu$. Since the proof is so much shorter using very simple result proved in Section 4.1, we shall cheat, and use that result.

By Lemma 3.9, if $\mathbf{P}_\nu$ is coherent, then $\underline{\mathbf{P}}_\nu$ is coherent too. Conversely, assume that $\underline{\mathbf{P}}_\nu$ is coherent. The easiest way to proceed, is to assume that $\underline{\mathbf{P}}_\nu$

has a coherent extension $\underline{P}$ to $I_A \cup -I_A$; this will be established in Section 4.1. If we can prove that this coherent extension $\underline{P}$ is equal to $\mathbf{P}_\nu$—and hence, that it is unique—then we have also established the coherence of $\mathbf{P}_\nu$. Since $\underline{P}$ is an extension of $\underline{\mathbf{P}}_\nu$, it holds that $\underline{P}(I_A) = \underline{\mathbf{P}}_\nu(I_A) = \nu(A) = \mathbf{P}_\nu(I_A)$. Since, additionally, ν is self-dual, it holds that $\underline{P}(-I_A) = \underline{P}(1 - I_A) - 1 = \underline{\mathbf{P}}_\nu(I_{\complement A}) - 1 = \nu(\complement A) - 1 = -\nu(A) = \mathbf{P}_\nu(-I_A)$. □

3.5.4 Nested Set Functions

A common way to model uncertainty about a random variable, is to identify a nested collection of sets, and to attach, to each event A in this collection, a real number $\mu(A)$ that measures our belief in A; we shall identify μ with a lower probability $\underline{\mathbf{P}}_\mu$, an upper probability $\overline{\mathbf{P}}_\mu$, or a probability $\mathbf{P}_\mu$. These set functions, which are far simpler than probability charges, have no specific name in the literature. I call them nested set functions, because they are defined on a collection of nested sets.

As lower probabilities, or upper probabilities, they arise naturally when modelling linguistic uncertainty; see Walley and De Cooman [88]. As probabilities, they arise as cumulative distribution functions; see Section 3.5.10.

In Theorem 3.27 below, we prove that a nested set function μ always induces a coherent probability $\mathbf{P}_\mu$, and hence, also a coherent lower probability $\underline{\mathbf{P}}_\mu$ and a coherent upper probability $\overline{\mathbf{P}}_\mu$.

Definition 3.25. A collection $\mathcal{A}$ of subsets of $\mathcal{X}$ is called *nested* whenever for any two elements A and B of $\mathcal{A}$ either $A \subseteq B$ or $B \subseteq A$.

Definition 3.26. A set function is called *nested* if its domain is a nested collection of sets.

Recall from Definition 3.22 that we assume set functions μ to be defined on at least $\emptyset$ and $\mathcal{X}$: this will simplify the condition for coherence and the expression for natural extension, and it guarantees that dom μ is a bounded chain with respect to $\subseteq$—a chain is bounded if it has a minimal and a maximal element; such nested set functions have been studied in the literature in the context of non-additive set functions, see for instance De Cooman and Aeyels [21] and Denneberg [28].

Also recall that we assume set functions to be monotone: $\mu(A) \leq \mu(B)$ whenever $A \subseteq B$, with minimum value $\mu(\emptyset) = 0$ and maximum value $\mu(\mathcal{X}) =$

1; otherwise, they induce neither coherent lower probabilities, nor coherent upper probabilities, nor coherent probabilities. Hence, a nested set function is an order preserving mapping from a chain of subsets of $\mathcal{X}$ to the unit interval $[0, 1]$.

The following theorem follows from Denneberg [28, Proposition 2.10]; we give an alternative proof below.

Theorem 3.27. *Let μ be a nested set function. Then $\mathbf{P}_\mu$, $\underline{\mathbf{P}}_\mu$ and $\overline{\mathbf{P}}_\mu$ are coherent.*

Proof. Indeed, by Lemma 3.9, it suffices to show that $\mathbf{P}_\mu$ is coherent. Denneberg [28, Proposition 2.10] showed that μ is the restriction of a probability charge. Hence, by Theorem 3.21 and Lemma 3.9, $\mathbf{P}_\mu$ is coherent.

Alternatively, define the set of gambles $\mathcal{K} := \{I_A : A \in \operatorname{dom}\mu\}$. Consider the set of gambles $\operatorname{span}(\mathcal{K})$ spanned by $\mathcal{K}$, this is a linear subspace of $\mathcal{L}(X)$. It is easy to show that $\mathcal{K}$ constitutes a basis for the linear space $\operatorname{span}(\mathcal{K})$. By a linear extension theorem (Schechter [70, Proposition 11.10]) we can uniquely define a linear mapping P on $\operatorname{span}(\mathcal{K})$ such that

$$P(I_A) = \mu(A)$$

for all $A \in \operatorname{dom}\mu$. P is an extension of the probability $\mathbf{P}_\mu$: to establish coherence, it suffices to show that P is a linear prevision. We check the conditions of Theorem 3.7. Clearly $P(f+g) = P(f)+P(g)$ for all $f, g \in \operatorname{span}(\mathcal{K})$ since it is a linear mapping. We only need to show that $P(f) \geq \inf f$ for any $f \in \operatorname{span}(\mathcal{K})$. Indeed, for any n in $\mathbb{N}$, $\alpha_1, \ldots, \alpha_n$ in $\mathbb{R}$, and $S_1 \subseteq S_2 \subseteq \cdots \subseteq S_n = \mathcal{X}$ in $\operatorname{dom}\mu$, we easily see that

$$\inf_{x\in\mathcal{X}}\left[\sum_{i=1}^{n}\alpha_i I_{S_i}(x)\right] = \min_{i=1}^{n}\left[\left(\sum_{j=i}^{n}\alpha_j\right)\right]$$

which is dominated by the following convex mixture—where we use $\mu(S_1) + \sum_{i=2}^{n}[\mu(S_i) - \mu(S_{i-1})] = \mu(S_n) = \mu(\mathcal{X}) = 1$, $\mu(S_1) \geq 0$, $\mu(S_i) - \mu(S_{i-1}) \geq 0$,

$$\leq \left(\sum_{j=1}^{n}\alpha_j\right)\mu(S_1) + \sum_{i=2}^{n}\left(\sum_{j=i}^{n}\alpha_j\right)[\mu(S_i) - \mu(S_{i-1})]$$

by $\mu(S_1) = P(I_{S_1})$, $\mu(S_i) - \mu(S_{i-1}) = P(I_{S_i \setminus S_{i-1}})$, and the linearity of P,

$$= P\left(\left(\sum_{j=1}^{n} \alpha_j\right) I_{S_1} + \sum_{i=2}^{n} \left(\sum_{j=i}^{n} \alpha_j\right) I_{S_i \setminus S_{i-1}}\right)$$
$$= P\left(\sum_{i=1}^{n} \alpha_i I_{S_i}\right).$$

□

3.5.5 2-Monotone and 2-Alternating Set Functions

The definition of 2-monotone and 2-alternating set functions below can be found for instance in Choquet [11, p. 131, and Chapter II, Section 8, p. 155].

Definition 3.28. A set function ν defined on a field $\mathcal{F}$ is called *2-monotone* if for any A and B in $\mathcal{F}$,

$$\nu(A \cup B) + \nu(A \cap B) \geq \nu(A) + \nu(B).$$

A set function π defined on $\mathcal{F}$ is called *2-alternating* if its dual is 2-monotone, or equivalently, if for any A and B in $\mathcal{F}$,

$$\pi(A \cup B) + \pi(A \cap B) \leq \pi(A) + \pi(B).$$

The property $\pi(A \cup B) + \pi(A \cap B) \leq \pi(A) + \pi(B)$ is sometimes also called *strong sub-additivity*; see for instance Choquet [11, p. 132].

As Walley [85, Section 6, pp. 51–52] notes, there are no clear behavioural arguments, similar to the behavioural arguments in favour of coherence, explaining 2-monotonicity. Certainly, not all coherent lower probabilities correspond to a 2-monotone set function. Nevertheless, 2-monotone set functions are mathematically very convenient, as we shall see in Section 4.3.10, and arise naturally in a number of important cases, as we shall see Section 4.3.4: these are the main reasons for studying 2-monotonicity.

In case $A \cap B = \emptyset$ it holds that $\nu(A \cup B) \geq \nu(A) + \nu(B)$ if ν is 2-monotone, and $\pi(A \cup B) \leq \pi(A) + \pi(B)$ if π is 2-alternating. This indicates that 2-monotone set functions are to be identified with lower probabilities and 2-alternating set functions, which are duals of 2-monotone set functions, with

upper probabilities; see Theorem 3.5(v) and Proposition 3.24. When are these lower and upper probabilities coherent?

The lower probability $\underline{\mathbf{P}}_\nu$ induced by a 2-monotone set function ν has an extension, known as the Choquet integral. This integral was introduced by Choquet [11, Section 48.1, p. 265] with respect to so-called *capacities*; we shall not study capacities in their full generality. Let it suffice to note that 2-monotone set functions are 2-monotone capacities when we equip $\mathcal{X}$ with any topology such that all elements of $\operatorname{dom}\nu$ are open: our definition of 2-monotone set function is then a special case of Choquet's [11, Chapter III, Section 15.2, p. 174] definition of 2-monotone capacities. Below we give a definition of the Choquet integral with respect to 2-monotone set functions, for $\mathcal{F}$-simple gambles. Our definition relies on the Dunford integral for $\mathcal{B}(\mathbb{R})$-simple gambles, which we introduced before—recall that $\mathcal{B}(\mathbb{R})$ is the Borel field on $\mathbb{R}$. Usually, the Choquet integral is defined through the Riemann integral, see for instance Janssen [45, Section 1.5.4.1], or the Riemann-Stieltjes integral, see for instance Walley [85, Section 6]—although some authors prefer to use other types of integration in their definition of the Choquet integral, as is the case for instance in Denneberg [28, Chapter 5]. In all cases investigated, the definition for $\mathcal{F}$-simple gambles, given below in terms of the Dunford integral, coincides with other definitions found in the literature.

Definition 3.29. Let $\mathcal{F}$ be a field on $\mathcal{X}$ and let ν be a 2-monotone set function on $\mathcal{F}$. Let f be any $\mathcal{F}$-simple gamble. Then the *lower decreasing distribution function* of f with respect to ν is defined by

$$G_{*\nu,f}(a) := \nu(\{x \in \mathcal{X} \colon f(x) > a\})$$

for any $a \in \mathbb{R}$. The *Choquet integral* of f with respect to ν is defined as

$$\mathrm{C}\int f \,\mathrm{d}\nu := \inf f + \mathrm{D}\int_{[\inf f, \sup f]} G_{*\nu,f} \,\mathrm{d}\lambda$$

where the Dunford integral is taken with respect to the Lebesgue measure λ on $\mathcal{B}(\mathbb{R})$.

Proof of Dunford integrability of $G_{*\nu,f}$. First observe that $\{x \in \mathcal{X} \colon f(x) > a\}$ is in $\mathcal{F}$ for any $a \in \mathbb{R}$ since f is $\mathcal{F}$-simple. Also f only takes a finite number of

values, say $a_1 < \cdots < a_n$. We can write

$$G_{*\nu,f} = I_{(-\infty,a_1)} + \sum_{i=1}^{n-1} G_{*\nu,f}(a_i) I_{[a_i,a_{i+1})} \tag{3.11}$$

hence, $G_{*\nu,f}$ is $\mathcal{B}(\mathbb{R})$-simple. From the above expression, it is easy to see that $G_{*\nu,f}$ is Dunford integrable over $[\inf f, \sup f] = [a_1, a_n]$ with respect to the Lebesgue measure. □

It is convenient to write the Dunford integral in the above definition of the Choquet integral as a sum. We can always write an $\mathcal{F}$-simple function as $f = \sum_{i=1}^{n} a_i I_{A_i}$ such that $a_1 < \cdots < a_n$ and such that $A_1, \ldots, A_n$ constitutes a partition of $\mathcal{F}$. We can write, using (3.11),

$$\begin{aligned}
\mathrm{C}\int f \,\mathrm{d}\nu &= a_1 + \sum_{i=1}^{n-1} (a_{i+1} - a_i) G_{*\nu,f}(a_i) \\
&= a_1 \left[1 - G_{*\nu,f}(a_1)\right] + \sum_{i=2}^{n-1} a_i \left[G_{*\nu,f}(a_{i-1}) - G_{*\nu,f}(a_i)\right] + a_n G_{*\nu,f}(a_{n-1}) \\
&= \sum_{i=1}^{n} a_i \left[\nu(\cup_{j=i}^{n} A_j) - \nu(\cup_{j=i+1}^{n} A_j)\right].
\end{aligned}$$

The last two expressions are especially interesting because they show how the Choquet integral of f is a convex combination of the values of f, with coefficients that depend in a non-trivial way on the shape of f. If we write the gamble f as $b_0 + \sum_{i=1}^{n} b_i I_{B_i}$ with b_0 in $\mathbb{R}$, $b_1, \ldots, b_n$ in $\mathbb{R}$ and strictly positive, and $\mathcal{X} \supsetneq B_1 \supsetneq B_2 \supsetneq \cdots \supsetneq B_n \supsetneq \emptyset$, then we can also write the above equality as

$$\begin{aligned}
\mathrm{C}\int f \,\mathrm{d}\nu &= b_0 + \sum_{i=1}^{n} b_i G_{*\nu,f}\left(\sum_{j=0}^{i-1} b_j\right) \\
&= b_0 + \sum_{i=1}^{n} b_i \nu(B_i),
\end{aligned} \tag{3.12}$$

which is a well-known identity (see for instance, Janssen [45]). The following result is due to Walley [85, Corollary 6.2, p. 55].

Lemma 3.30. *Let $\mathcal{F}$ be a field on $\mathcal{X}$ and let ν be a 2-monotone set function on $\mathcal{F}$.*

Define the lower prevision $\underline{\mathbf{E}}_\nu^{\mathrm{span}(\mathcal{F})}$ *by*

$$\underline{\mathbf{E}}_\nu^{\mathrm{span}(\mathcal{F})}(f) := \mathrm{C}\int f \,\mathrm{d}\nu \tag{3.13}$$

for any $\mathcal{F}$*-simple gamble* f*. Then* $\underline{\mathbf{E}}_\nu^{\mathrm{span}(\mathcal{F})}$ *is a coherent lower prevision on* $\mathrm{span}(\mathcal{F})$*.*

Proof. Simply check the conditions of Theorem 3.6. Conditions (i) and (ii) are easy to check. Condition (iii) is more difficult to prove; see for instance Choquet [11, p. 287], Walley [85, Lemma 6.3, p. 54], or Denneberg [28, Theorem 6.3, p. 75] for proofs. □

We now arrive at the main result: 2-monotone set functions induce coherent lower probabilities. It was proved by Walley [85, Corollary 6.3, p. 55].

Theorem 3.31. *Any 2-monotone set function induces a coherent lower probability.*

Proof. Let ν be a 2-monotone set function defined on a field $\mathcal{F}$. Observe that, for instance by Eq. (3.12), $\underline{\mathbf{P}}_\nu$ is the restriction of $\underline{\mathbf{E}}_\nu^{\mathrm{span}(\mathcal{F})}$ to the set $I_\mathcal{F}$ of indicators of elements of $\mathcal{F}$. But $\underline{\mathbf{E}}_\nu^{\mathrm{span}(\mathcal{F})}$ is a coherent prevision by Lemma 3.30. Now apply Lemma 3.9 to arrive at the desired result. □

3.5.6 2-Monotone Lower Previsions

Let's briefly generalise the notion of 2-monotonicity to lower previsions; this will turn out to be useful later on. The definition below is an instance of a general definition given by Choquet [11, Chapter III, Definition 13.1, p. 170, and Section 14.1, p. 171]

Definition 3.32. A lower prevision $\underline{P}$ defined on a lattice of gambles on X is called *2-monotone* if for all gambles f and g in $\mathrm{dom}\,\underline{P}$ it holds that

(i) $f \geq g \implies \underline{P}(f) \geq \underline{P}(g)$, and

(ii) $\underline{P}(f \vee g) + \underline{P}(f \wedge g) \geq \underline{P}(f) + \underline{P}(g)$.

Proposition 3.33. *Let* S *be any non-empty subset of* $\mathcal{X}$*. The vacuous lower prevision with respect to* S*, i.e.,* $\underline{P}_S$*, is a 2-monotone coherent lower prevision.*

Proof. The proposition can be proved easily by direct verification of Definition 3.32: let f and g be gambles in $\mathrm{dom}\,\underline{P}$. Clearly, by the coherence of

$\underline{P}_S$ (Theorem 3.5(iv)) it holds that $\underline{P}_S(f) \geq \underline{P}_S(g)$ whenever $f \geq g$, and since $f \vee g \geq f$ and $f \vee g \geq g$, also $\underline{P}_S(f \vee g) \geq \max\{\underline{P}_S(f), \underline{P}_S(g)\}$. Moreover,

$$\underline{P}_S(f \wedge g) = \inf_{x \in S} \min\{f(x), g(x)\} = \min\left\{\inf_{x \in S} f(x), \inf_{x \in S} g(x)\right\} = \min\{\underline{P}_S(f), \underline{P}_S(g)\}.$$

So, we can also conclude that

$$\begin{aligned}\underline{P}_S(f \vee g) + \underline{P}_S(f \wedge g) &\geq \max\{\underline{P}_S(f), \underline{P}_S(g)\} + \min\{\underline{P}_S(f), \underline{P}_S(g)\} \\ &= \underline{P}_S(f) + \underline{P}_S(g),\end{aligned}$$

which establishes the proposition. □

Proposition 3.34. *Any convex combination of 2-monotone lower previsions defined on a common domain, is 2-monotone.*

Proof. Let $\underline{P}_1, \ldots, \underline{P}_n$ be a finite family of 2-monotone lower previsions defined on a common domain $\mathcal{K}$, and let $\lambda_1, \ldots, \lambda_n$ be non-negative reals such that $\sum_{i=1}^n \lambda_i = 1$. For any f and g in $\mathcal{K}$, it holds that

$$\begin{aligned}\sum_{i=1}^n \lambda_i \underline{P}_i(f \vee g) + \sum_{i=1}^n \lambda_i \underline{P}_i(f \wedge g) &= \sum_{i=1}^n \lambda_i \left[\underline{P}_i(f \vee g) + \underline{P}_i(f \wedge g)\right] \\ &\geq \sum_{i=1}^n \lambda_i \left[\underline{P}_i(f) + \underline{P}_i(g)\right] \\ &= \sum_{i=1}^n \lambda_i \underline{P}_i(f) + \sum_{i=1}^n \lambda_i \underline{P}_i(g),\end{aligned}$$

and, whenever $f \geq g$, it obviously follows that $\sum_{i=1}^n \lambda_i \underline{P}_i(f) \geq \sum_{i=1}^n \lambda_i \underline{P}_i(g)$; so $\sum_{i=1}^n \lambda_i \underline{P}_i$ is 2-monotone as well. □

Proposition 3.35. *Let $\underline{P}_\alpha$ be a net of 2-monotone lower previsions defined on a common domain $\mathcal{K} \subseteq \mathcal{L}(X)$. If $\underline{P}(f) := \lim_\alpha \underline{P}_\alpha(f)$ exists and is real for all f in $\mathcal{K}$, then this point-wise limit $\underline{P}$ is a 2-monotone lower prevision on $\mathcal{K}$.*

Proof. For any f and g in $\mathcal{K}$, it holds that

$$\begin{aligned}\lim_\alpha \underline{P}_\alpha(f \vee g) + \lim_\alpha \underline{P}_\alpha(f \wedge g) &= \lim_\alpha \left[\underline{P}_\alpha(f \vee g) + \underline{P}_\alpha(f \wedge g)\right] \\ &\geq \lim_\alpha \left[\underline{P}_\alpha(f) + \underline{P}_\alpha(g)\right] = \lim_\alpha \underline{P}_\alpha(f) + \lim_\alpha \underline{P}_\alpha(g),\end{aligned}$$

since all limits exist and are real. Clearly, whenever $f \geq g$, it also follows that $\lim_\alpha \underline{P}_\alpha(f) \geq \lim_\alpha \underline{P}_\alpha(g)$ So, $\lim_\alpha \underline{P}_\alpha$ is a 2-monotone lower prevision on $\mathcal{K}$. □

3.5.7 Completely Monotone Set Functions and Belief Functions

Belief functions were introduced by Shafer [74, 75]. They are a special case of so-called completely monotone set functions. Note that $|J|$ denotes the cardinality of the finite set J, that is, the number of elements in J. The definition below can be found in Choquet [11, Chapter III, Definition 13.1, p. 170, and Section 14.1, p. 171]; Choquet's definition is actually far more general, and extends to mappings from any commutative semi-group to any commutative group, i.e., not only mappings from a field to $\mathbb{R}$. Also note that n-monotone set functions are n-monotone capacities when we equip $\mathcal{X}$ with any topology such that all elements of dom ν are open: our definition of n-monotone set functions is then a special case of Choquet's [11, Chapter III, Section 15.2, p. 174] definition of n-monotone capacities. (Note that, by our definition of a set function ν, it already holds that $\nu(A) \geq \nu(B)$ whenever $A \supseteq B$, this is why this condition is not included in the definition below.)

Definition 3.36. Let $\mathcal{F}$ be a field on $\mathcal{X}$, and let ν be a set function defined on $\mathcal{F}$. Let $n \in \mathbb{N}$, $n \geq 2$. Then ν is called *monotone of order n*, or *n-monotone*, if for any $A_1, \ldots, A_n \in \mathcal{F}$, it holds that

$$\nu\left(\bigcup_{i=1}^{n} A_i\right) \geq \sum_{\emptyset \neq J \subseteq \{1,2,\ldots,n\}} (-1)^{|J|+1} \nu\left(\bigcap_{i \in J} A_i\right). \tag{3.14}$$

If ν is n-monotone for all $n \geq 2$, then ν is called *completely monotone*, *monotone of order infinity*, or *∞-monotone*. If, additionally, $\mathcal{X}$ is a finite set and $\mathcal{F}$ is the power set of $\mathcal{X}$, then ν is called a *belief function*. A set function π on $\mathcal{F}$ is called *n-alternating* if its dual is n-monotone, or equivalently, if for any $A_1, \ldots, A_n \in \mathcal{F}$, it holds that

$$\pi\left(\bigcap_{i=1}^{n} A_i\right) \leq \sum_{\emptyset \neq J \subseteq \{1,2,\ldots,n\}} (-1)^{|J|+1} \pi\left(\bigcup_{i \in J} A_i\right). \tag{3.15}$$

If π is n-alternating for all $n \geq 2$, or equivalently, if the dual of π is completely monotone, then π is called *completely alternating*. The dual of a belief function is called a *plausibility function*.

It is easily verified that 2-monotonicity as defined in Definition 3.36 is equivalent to the 2-monotonicity as defined in Definition 3.28.

Shafer [75, Section 1, pp. 827–828] calls completely monotone set functions also belief functions, and calls completely alternating set functions also upper probability functions; in his definition, the domain of belief functions and upper probability functions does not even need to be a field. However, in the literature, belief and plausibility functions are usually understood to be defined on the power set of a finite set, following Shafer's [74, Chapter 2, Section 2, p. 38] original definition.

It is instructive to compare the definition of completely monotone set functions with the definition of probability charges, and to observe that complete monotonicity generalises a well-known combinatorial identity: for a probability charge μ on $\mathcal{F}$ it holds that

$$\mu\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{\emptyset \neq J \subseteq \{1,2,\ldots,n\}} (-1)^{|J|+1} \mu\left(\bigcap_{i \in J} A_i\right), \tag{3.16}$$

for any $n \in \mathbb{N}$, $n \geq 2$ and $A_1, \ldots, A_n \in \mathcal{F}$; for instance, see De Finetti [27, Vol. I, Sect. 3.8.3, p. 101]. This equality is also known as the *sieve formula* or the *inclusion-exclusion principle*; see for instance Aigner [1, 4.24(i), p. 158]. Hence, probability charges are also completely monotone set functions, but not every completely monotone set function corresponds to a probability charge. For example, as we shall see in Proposition 3.42 below, the vacuous lower prevision, restricted to $I_{\wp(\mathcal{X})}$, is a completely monotone set function that is not a probability charge. Choquet [11, Section 14.5, p. 173–174] notes that probability charges are the only set functions, defined on a field, which are both 2-monotone and 2-alternating, and proves that consequently, they must be n-monotone and n-alternating for any $n \in \mathbb{N}^*$, $n \geq 2$; he does not rely on the sieve formula to prove this.

Let's sum up a few obvious facts. Denote by $\mathbb{N}^*$ the set $\mathbb{N} \cup \{\infty\}$.

Proposition 3.37. *Let $n \in \mathbb{N}^*$, $n \geq 2$. If a set function ν is n-monotone, then for any $m \in \mathbb{N}^*$, $n \geq m \geq 2$, ν is also m-monotone.*

Proof. Immediate from Definition 3.36. □

Proposition 3.38. *Let $n \in \mathbb{N}^*$, $n \geq 2$. Any n-monotone set function ν induces a coherent lower probability $\underline{\mathbf{P}}_\nu$.*

Proof. Immediate from Proposition 3.37 and Theorem 3.31. □

The following proposition says that n-monotonicity is preserved under convex combinations and limits.

Proposition 3.39. *Let $n \in \mathbb{N}^*$, $n \geq 2$. The following statements hold.*

(i) Any convex combination of n-monotone set functions, defined on a common field, is n-monotone.

(ii) If the point-wise limit of a net of n-monotone set functions defined on a common field exists, then this limit is n-monotone.

Proof. We prove the statements for $n \in \mathbb{N}$, $n \geq 2$. The proof for $n = \infty$ is then immediate.

(i). It is easily verified that the convex combination of set functions is again a set function. We are left to check the condition for n-monotonicity. Let $\alpha_1, \ldots, \alpha_m$ be non-negative real numbers such that $\sum_{j=1}^m \alpha_j = 1$, let $\nu_1, \ldots, \nu_m$ be n-monotone set functions defined on a common field $\mathcal{F}$, and define $\nu(A) := \sum_{j=1}^m \alpha_j \nu_j(A)$ for every $A \in \mathcal{F}$. Then, for any $A_1, \ldots, A_n \in \mathcal{F}$, it holds that

$$\begin{aligned}
\nu\left(\bigcup_{i=1}^n A_i\right) &= \sum_{j=1}^n \alpha_j \nu_j \left(\bigcup_{i=1}^n A_i\right) \\
&\geq \sum_{j=1}^n \alpha_j \sum_{\emptyset \neq J \subseteq \{1,2,\ldots,n\}} (-1)^{|J|+1} \nu_j \left(\bigcap_{i \in J} A_i\right). \\
&= \sum_{j=1}^n \sum_{\emptyset \neq J \subseteq \{1,2,\ldots,n\}} (-1)^{|J|+1} \alpha_j \nu_j \left(\bigcap_{i \in J} A_i\right). \\
&= \sum_{\emptyset \neq J \subseteq \{1,2,\ldots,n\}} (-1)^{|J|+1} \sum_{j=1}^n \alpha_j \nu_j \left(\bigcap_{i \in J} A_i\right). \\
&= \sum_{\emptyset \neq J \subseteq \{1,2,\ldots,n\}} (-1)^{|J|+1} \nu \left(\bigcap_{i \in J} A_i\right),
\end{aligned}$$

hence, ν is n-monotone.

(ii). It is easily verified that the point-wise limit of a net of set functions, if this limit exists, is again a set function. It remains to check the condition for n-monotonicity. Let ν_α be a net of n-monotone set functions. Then

$$\begin{aligned}\lim_\alpha \nu_\alpha\left(\bigcup_{i=1}^n A_i\right) &\geq \lim_\alpha \sum_{\emptyset\neq J\subseteq\{1,2,\ldots,n\}} (-1)^{|J|+1}\nu_\alpha\left(\bigcap_{i\in J} A_i\right)\\ &= \sum_{\emptyset\neq J\subseteq\{1,2,\ldots,n\}} (-1)^{|J|+1}\lim_\alpha \nu_\alpha\left(\bigcap_{i\in J} A_i\right),\end{aligned}$$

hence, also $\lim_\alpha \nu_\alpha$ is an n-monotone set function. □

There's another way to construct n-monotone set functions, due to Choquet [11, Chapter V, Section 24.3, p. 198]. We shall use it a few times.

Definition 3.40. Let $\mathcal{F}$ be a field on $\mathcal{X}$ and $\mathcal{G}$ a field on $\mathcal{Y}$. A mapping $r\colon \mathcal{F} \to \mathcal{G}$ is called a *meet-homomorphism* or a *$\cap$-homomorphism from $\mathcal{F}$ to $\mathcal{G}$* if

(i) $r(\emptyset) = \emptyset$ and $r(\mathcal{X}) = \mathcal{Y}$, and

(ii) $r(A \cap B) = r(A) \cap r(B)$ for every A and B in $\mathcal{F}$.

Note that every $\cap$-homomorphism is monotone, *i.e.*, if $C \subseteq D$ for C and D in $\mathcal{F}$, then $r(C) \subseteq r(D)$, since $r(C) = r(C \cap D) = r(C) \cap r(D)$.

Lemma 3.41. *Let $n \in \mathbb{N}^*$, $n \geq 2$. Let $\mathcal{F}$ be a field on $\mathcal{X}$, let $\mathcal{G}$ be a field on $\mathcal{Y}$, let r be a $\cap$-homomorphism from $\mathcal{F}$ to $\mathcal{G}$, and let κ be an n-monotone set function on $\mathcal{G}$. Then $\nu := \kappa \circ r$ is an n-monotone set function on $\mathcal{F}$.*

Proof. Choquet's [11, Chapter V, Section 23.2, p. 197, and Section 24.3, p. 198] proof is rather short; it's essentially only a remark of two lines. For the sake of completeness, let's work out the details. We shall prove the statement for finite n; the case $n = \infty$ is then immediate.

It is easily shown that $\nu(\emptyset) = 0$, and $\nu(\mathcal{X}) = 1$, and that ν is monotone, *i.e.*, $\nu(A) \leq \nu(B)$ whenever $A \subseteq B$ for A and B in $\mathcal{F}$ (use the monotonicity of r and κ).

Now, for any $A_1, \ldots, A_n \in \mathcal{F}$, it holds that

$$\begin{aligned}\sum_{\emptyset \neq J \subseteq \{1,2,\ldots,n\}} (-1)^{|J|+1} \nu\left(\bigcap_{i \in J} A_i\right) &= \sum_{\emptyset \neq J \subseteq \{1,2,\ldots,n\}} (-1)^{|J|+1} \kappa\left(r\left(\bigcap_{i \in J} A_i\right)\right) \\ &= \sum_{\emptyset \neq J \subseteq \{1,2,\ldots,n\}} (-1)^{|J|+1} \kappa\left(\bigcap_{i \in J} r(A_i)\right)\end{aligned}$$

and since κ is n-monotone,

$$\leq \kappa\left(\bigcup_{i=1}^{n} r(A_i)\right)$$

and, since a $\cap$-homomorphism is monotone, it holds that $r(A_j) \subseteq r(\bigcup_{i=1}^{n} A_i)$ for all $j \in \{1, \ldots, n\}$, and hence, $\bigcup_{i=1}^{n} r(A_i) \subseteq r(\bigcup_{i=1}^{n} A_i)$. So, since also κ is monotone,

$$\leq \kappa\left(r\left(\bigcup_{i=1}^{n} A_i\right)\right) = \nu\left(\bigcup_{i=1}^{n} A_i\right).$$

This establishes the lemma. □

We can now easily prove the following proposition, which is strongly related to a very similar result by Choquet [11, Chapter V, Section 26.2, p. 205], which gives a particular example of a completely monotone set function; we shall need this when introducing S-integrals and Riemann-Stieltjes integrals.

Proposition 3.42. *Let S be any non-empty subset of $\mathcal{X}$. The set function ν defined by $\nu(A) := \underline{P}_S(I_A)$ is a completely monotone set function.*

Proof. Immediately from Choquet [11, Chapter V, Section 24.3, p. 198], Shafer [75, Section 2, p. 830, ll. 1–4], or Lemma 3.41, once observed that $\nu = \mu \circ r$, with μ a probability charge (*i.e.*, a completely monotone set function) on $\{\emptyset, \mathcal{X}\}$ defined by $\mu(\emptyset) := 0$ and $\mu(\mathcal{X}) := 1$, and r a $\cap$-homomorphism from $\wp(\mathcal{X})$ to $\{\emptyset, \mathcal{X}\}$, defined by

$$r(A) := \begin{cases} \mathcal{X}, & \text{if } S \subseteq A, \\ \emptyset, & \text{otherwise.} \end{cases}$$

for every $A \subseteq \mathcal{X}$. Indeed, r is a $\cap$-homomorphism, since $r(\emptyset) = \emptyset$, $r(\mathcal{X}) = \mathcal{X}$,

and

$$r(A \cap B) = \begin{cases} \mathcal{X}, & \text{if } S \subseteq A \cap B \\ \emptyset, & \text{otherwise} \end{cases} = \begin{cases} \mathcal{X}, & \text{if } S \subseteq A \text{ and } S \subseteq B \\ \emptyset, & \text{otherwise} \end{cases} = r(A) \cap r(B),$$

and also,

$$(\mu \circ r)(A) = \mu(r(A)) = \begin{cases} 1, & \text{if } S \subseteq A, \\ 0, & \text{otherwise,} \end{cases}$$

which is equal to $\nu(A)$. This establishes the lemma. □

What makes belief functions so special is that they can be uniquely characterised through convex combinations of vacuous lower previsions, *i.e.*, convex combinations of completely monotone set functions of the type of Proposition 3.42. This was proved by Shafer [74, Theorem 2.1 and Theorem 2.2] using simple combinatorics (the sieve formula and Möbius inversion); Choquet [11, Section 45.1, pp. 258–259] established a more general result, *i.e.*, for a more general class of completely monotone set functions, using geometry; also see Shafer [75, p. 830, ll. 1–4 and Theorem 2.1] for a straightforward extension of Choquet's result.

Theorem 3.43. *Assume that $\mathcal{X}$ is finite. Let ν be any real-valued mapping defined on the power set of $\mathcal{X}$. Then ν is a belief function if and only if there exists a mapping $m\colon \wp(\mathcal{X}) \to [0,1]$ that satisfies $m(\emptyset) = 0$, $\sum_{A \subseteq \mathcal{X}} m(A) = 1$, and*

$$\nu(A) = \sum_{\emptyset \neq B \subseteq \mathcal{X}} m(B) \underline{P}_B(I_A). \tag{3.17}$$

A mapping m satisfying the conditions described in Theorem 3.43 is called a *basic probability assignment*. It is well-known that the correspondence between basic probability assignments and belief functions is onto and one-to-one; see Shafer [74, Section 3, p. 39]. Transforming a belief function into its basic probability assignment is sometimes also called *Möbius inversion* or *the inverse Möbius transform*:

$$m(A) := \sum_{B \subseteq A} (-1)^{|A \setminus B|} \nu(B); \tag{3.18}$$

see Shafer [74, Theorem 2.2, p. 39].

Eq. (3.17) suggests the following coherent extension of $\underline{\mathbf{P}}_\nu$:

$$\underline{\mathbf{E}}_\nu(f) := \sum_{\emptyset \neq B \subseteq \mathcal{X}} m(B)\underline{P}_B(f), \qquad \text{for any gamble } f \text{ on } X. \tag{3.19}$$

This coherent lower prevision agrees with the Choquet integral with respect to ν, defined on p. 74 for $\mathcal{F}$-simple gambles—in this case $\mathcal{X}$ is a finite set and $\mathcal{F}$ is the power set of $\mathcal{X}$, all gambles are $\mathcal{F}$-simple. Thus, for belief functions, Eq. (3.19) can be used to calculate the Choquet integral. This result is apparently due to Walley [86, Note 2 of Section 3.2, p. 502].

Theorem 3.44. *Assume that $\mathcal{X}$ is finite, and let ν be a belief function on $\wp(\mathcal{X})$ corresponding to a basic probability assignment m. For any gamble f on X it holds that*

$$\mathrm{C}\int f \,\mathrm{d}\nu = \sum_{\emptyset \neq B \subseteq \mathcal{X}} m(B)\underline{P}_B(f).$$

3.5.8 Minimum and Maximum Preserving Set Functions

Minimum and maximum preserving set functions are a nice example of completely monotone and completely alternating set functions. On the power set of a finite set, they are belief functions. More generally, on finite fields, they are equivalent to (normed) necessity and possibility measures, defined further on in Section 3.5.9. For modelling so-called linguistic uncertainty through nested set functions, they are actually more reasonable than necessity and possibility measures; see Theorem 4.36(ii) on p. 117; also see Walley and De Cooman [88, p. 19].

Definition 3.45. Let $\mathcal{F}$ be a field. A set function ν defined on $\mathcal{F}$ is called *minimum preserving* if, for any A and B in $\mathcal{F}$,

$$\nu(A \cap B) = \nu(A) \wedge \nu(B).$$

A set function π defined on $\mathcal{F}$ is called *maximum preserving* if its dual is minimum preserving, or equivalently, if for any A and B in $\mathcal{F}$,

$$\pi(A \cup B) = \pi(A) \vee \pi(B).$$

As with 2-monotonicity, minimum preservation does not have a clear behavioural interpretation. Actually, minimum preservation, as a rationality

constraint, is clearly too strong in general: from De Cooman and Aeyels [21, Proposition 15&16], we infer that if $\mathcal{X}$ is finite, then the only coherent minimum preserving lower previsions defined on all gambles on X are vacuous lower previsions.

Using results from Section 3.5.5 about 2-monotone set functions, it is easy to prove that minimum preserving set functions induce coherent lower probabilities. As the proof is short, it is given below.

Theorem 3.46. *Any minimum preserving set function is 2-monotone, and hence, induces a coherent lower probability. Similarly, any maximum preserving set function is 2-alternating, and hence, induces a coherent upper probability.*

Proof. Let ν be a 2-monotone set function, and let $A, B \in \operatorname{dom} \nu$. Since ν is a set function, it is monotone, and hence, $\nu(A \cup B) \geq \nu(A)$ and $\nu(A \cup B) \geq \nu(B)$. Therefore, $\nu(A \cup B) \geq \nu(A) \vee \nu(B)$. Since, additionally, ν is minimum preserving,

$$\nu(A \cup B) + \nu(A \cap B) \geq \nu(A) \vee \nu(B) + \nu(A) \wedge \nu(B) = \nu(A) + \nu(B).$$

Now apply Theorem 3.31 and Proposition 3.24. □

Nguyen [62, Theorem 1, pp. 363–364] proved that minimum preserving set functions are completely monotone. So, when they are defined on the power set of a finite set, we can invoke Eq. (3.18) and Theorem 3.44 to calculate their Choquet integral.

Proposition 3.47. *Any minimum preserving set function is completely monotone.*

Corollary 3.48. *Any minimum preserving set function defined on the power set of a finite set is a belief function.*

3.5.9 Necessity and Possibility Measures

Zadeh [95] introduced possibility measures, aimed at modelling *linguistic uncertainty*, for instance, inferring from "tomorrow, it will rain a lot in Ghent" something about the amount it will rain tomorrow in Ghent. Dubois and Prade [29] introduced necessity measures, which are dual possibility measures. De Cooman [17, 18, 19] generalised possibility and necessity measures to complete lattices—linguistic variables typically assume values

in a complete lattice—and discussed the formal analogies between possibility theory and probability theory. Because of their interpretation as upper and lower probabilities, I shall only consider real-valued possibility and necessity measures.

Walley and De Cooman [88] modelled statements of the type "subject is predicate" as lower previsions, and obtained possibility measures (or more general, maximum preserving upper probabilities) only when the predicate is 'monotonic' in the random variable of interest. This means the predicate must be attached to a numerical scale on $\mathcal{X}$ (Walley and De Cooman [88, p. 13, ll. 19–21]):

> [...] monotonic predicates are the predicates q for which there is an underlying numerical scale which measures the degree of q-ness of [the possible outcomes x of X,]

or, more generally, that the predicate is attached to a complete preorder (a complete, transitive, and reflexive relation) on $\mathcal{X}$ (Walley and De Cooman [88, p. 14, Assumption 1]):

> [we assume that] there are degrees of q-ness, and every pair of elements in the possibility space [$\mathcal{X}$] can be compared according to how well they satisfy the property q, i.e. according to their degrees of q-ness.

For example (Walley and De Cooman [88, p. 13, ll. 2–10]),

> [...] suppose that we are trying to identify the man who committed a particular crime from a group of suspects, who make up the possibility space [$\mathcal{X}$]. An eyewitness describes the criminal as 'tall'. How should we model the resulting uncertainty about which suspect is the criminal? [...] there is a natural ordering of [$\mathcal{X}$] according to the height of the suspects, and 'tall' is increasing for an uncertain state which represents the *height* of the criminal.

So, as a behavioural uncertainty model, it appears that possibility measures can only model a very particular—but common—type of linguistic uncertainty.

Perhaps the easiest way to introduce necessity and possibility measures, is to take a look at Definition 3.45 on p. 84, and to note that

$$\nu\left(\bigcap_{A\in\mathcal{A}} A\right) = \min_{A\in\mathcal{A}} \nu(A) \text{ and } \pi\left(\bigcup_{A\in\mathcal{A}} A\right) = \max_{A\in\mathcal{A}} \pi(A)$$

for any non-empty finite subset $\mathcal{A}$ of $\mathcal{F}$. This suggests a stronger version of Definition 3.45 if ν and π are defined on an ample field—a field closed under arbitrary intersection and arbitrary union.

Definition 3.49. Let $\mathcal{F}$ be an ample field. A set function N defined on $\mathcal{F}$ is said to be an *infimum preserving set function*, or a *necessity measure*, if for any non-empty $\mathcal{A} \subseteq \mathcal{F}$,

$$N\left(\bigcap_{A\in\mathcal{A}} A\right) = \inf_{A\in\mathcal{A}} N(A).$$

A set function Π defined on $\mathcal{F}$ is said to be a *supremum preserving set function*, or a *possibility measure*, if its dual is infimum preserving, or equivalently, if for any non-empty $\mathcal{A} \subseteq \mathcal{F}$,

$$\Pi\left(\bigcup_{A\in\mathcal{A}} A\right) = \sup_{A\in\mathcal{A}} \Pi(A).$$

Note that, since possibility measures Π, as defined above, are set functions, it holds that $\Pi(\mathcal{X}) = 1$. In the literature, possibility measures for which $\Pi(\mathcal{X}) = 1$ are called *normed* possibility measures. Similarly, necessity measures, as defined above, are called normed necessity measures in the literature. ***We shall only consider normed possibility and necessity measures, and simply call them possibility and necessity measures.***

Before, we have denoted set functions by lower case letters, such as μ for set functions that induce coherent probabilities, ν for set functions that induce coherent lower probabilities, and π for set functions that induce coherent upper probabilities. In the literature, necessity and possibility measures are denoted by upper case letters Π and N, and the lower case letters ν and π are reserved for the necessity distribution and the possibility distribution; we follow this convention.

One important advantage of supremum preserving set functions, not shared by maximum preserving set functions in general, is their representability as a real-valued mapping on $\mathcal{X}$. This relies on a special property of ample fields, not shared by fields or σ-fields in general. First, we need to define atoms; see for instance Wang [89].

Definition 3.50. Let $\mathcal{F}$ be a field on $\mathcal{X}$, and let $x \in \mathcal{X}$. Then a set A in $\mathcal{F}$ is called an *atom* of $\mathcal{F}$ if

(i) $A \neq \emptyset$, and

(ii) if $B \in \mathcal{F}$ and $B \subseteq A$ then either $B = \emptyset$ or $B = A$.

The set of atoms of $\mathcal{F}$ will be denoted by $\mathbb{A}(\mathcal{F})$.

The following theorem is a concise summary of results by Wang [89]. As the proof is short, it is given below, for the sake of completeness.

Theorem 3.51. *A field $\mathcal{F}$ on $\mathcal{X}$ is an ample field if and only if there is a partition $\mathcal{B}$ of $\mathcal{X}$ such that every element of $\mathcal{F}$ is an arbitrary union of elements of $\mathcal{B}$:*

$$\mathcal{F} = \left\{ \bigcup_{A \in \mathcal{A}} A \colon \mathcal{A} \subseteq \mathcal{B} \right\}. \tag{3.20}$$

If so, this partition $\mathcal{B}$ is unique, and is the set $\mathbb{A}(\mathcal{F})$ of atoms of $\mathcal{F}$.

Proof. First, we show that, for any field $\mathcal{F}$, the atoms of $\mathcal{F}$ are pair-wise disjoint. Suppose A and B are atoms of $\mathcal{F}$. Clearly, $A \cap B \in \mathcal{F}$, $A \cap B \subseteq A$, and $A \cap B \subseteq B$. So, by the definition of atom, either $A \cap B = \emptyset$ or $A \cap B = A$, and either $A \cap B = \emptyset$ or $A \cap B = B$. Hence, indeed, either $A \cap B = \emptyset$ or $A \cap B = A = B$.

"if". If $\mathcal{F}$ is generated by a partition $\mathcal{B}$ through arbitrary union, then $\mathcal{F}$ is closed under arbitrary union, and hence, $\mathcal{F}$ is an ample field. Let's show that $\mathcal{B} = \mathbb{A}(\mathcal{F})$. Let $B \in \mathcal{B}$. Suppose that $A \in \mathcal{F}$ and $A \subseteq B$. Since $A \in \mathcal{F}$, there is a subset $\mathcal{A}$ of $\mathcal{B}$ such that $A = \cup_{C \in \mathcal{A}} C$. Since $\mathcal{B}$ is a partition, $\mathcal{A} \subseteq \mathcal{B}$, and $A \subseteq B$, it must hold that $\mathcal{A} \subseteq \{B\}$. Hence, either $\mathcal{A} = \emptyset$, in which case $A = \emptyset$, or $\mathcal{A} = \{B\}$, in which case $A = B$. So, indeed, $B \in \mathbb{A}(\mathcal{F})$. We proved that $\mathcal{B} \subseteq \mathbb{A}(\mathcal{F})$. But, all elements of $\mathbb{A}(\mathcal{F})$ are pair-wise disjoint, so, it must hold that $\mathcal{B} = \mathbb{A}(\mathcal{F})$.

"only if". Suppose $\mathcal{F}$ is an ample field. We must prove that $\mathbb{A}(\mathcal{F})$ is the only partition of $\mathcal{X}$ for which Eq. (3.20) is satisfied. First, we show that $\mathbb{A}(\mathcal{F})$ is a partition of $\mathcal{X}$. Since $\mathcal{F}$ is an ample field, this implies that Eq. (3.20) is satisfied for the partition $\mathbb{A}(\mathcal{F})$. Then, by the "if"-part, $\mathbb{A}(\mathcal{F})$ is the only partition of $\mathcal{X}$ such that Eq. (3.20) is satisfied.

So, is $\mathbb{A}(\mathcal{F})$ a partition of $\mathcal{X}$? Above, we showed that all atoms of $\mathcal{F}$ are pair-wise disjoint. It remains to prove that $\bigcup_{A \in \mathbb{A}(\mathcal{F})} A = \mathcal{X}$. Define for any $x \in \mathcal{X}$ the set

$$[x]_{\mathcal{F}} := \bigcap_{A \in \mathcal{F},\, x \in A} A.$$

Then, $[x]_{\mathcal{F}}$ is an atom of $\mathcal{F}$. Indeed, since $x \in [x]_{\mathcal{F}}$, it is non-empty, and since $\mathcal{F}$ is closed under arbitrary intersection—since it is an ample field—also $[x]_{\mathcal{F}} \in \mathcal{F}$. If $B \in \mathcal{F}$ and $B \subseteq [x]_{\mathcal{F}}$, then we have the following cases.

(a) $x \in B$. Then $[x]_{\mathcal{F}} \cap B = [x]_{\mathcal{F}}$ by definition of $[x]_{\mathcal{F}}$, and hence, $B \supseteq [x]_{\mathcal{F}}$. Together with $B \subseteq [x]_{\mathcal{F}}$, it follows that $B = [x]_{\mathcal{F}}$.

(b) $x \notin B$. Then $(\complement[x]_{\mathcal{F}}) \cup B = \complement[x]_{\mathcal{F}}$, and hence, $B \subseteq \complement[x]_{\mathcal{F}}$. Together with $B \subseteq [x]_{\mathcal{F}}$, it follows that $B = \emptyset$.

So, $[x]_{\mathcal{F}}$ is indeed an atom of $\mathcal{F}$. Since $x \in [x]_{\mathcal{F}}$ for all $x \in \mathcal{X}$, we find that

$$\mathcal{X} \supseteq \bigcup_{A \in \mathbb{A}(\mathcal{F})} A \supseteq \bigcup_{x \in \mathcal{X}} [x]_{\mathcal{F}} \supseteq \mathcal{X},$$

which establishes the desired equality. □

So, if $\mathcal{F}$ is an ample field, the set $\mathbb{A}(\mathcal{F})$ of atoms of $\mathcal{F}$ is the finest partition of $\mathcal{X}$ whose elements belong to $\mathcal{F}$, and any element of $\mathcal{F}$ is an arbitrary union of elements of this partition. Conversely, if $\mathcal{F}$ is generated through arbitrary union of elements of a partition, then $\mathcal{F}$ is an ample field whose atoms are given by the generating partition. This establishes an isomorphism between partitions and ample fields. For example, the ample field $\wp(\mathcal{X})$ corresponds to the partition $\{\{x\} \colon x \in \mathcal{X}\}$ of all singletons of $\mathcal{X}$. The ample field $\{\emptyset, \mathcal{X}\}$ corresponds to the partition $\{\mathcal{X}\}$.

An important consequence of Theorem 3.51 is that, if $\mathcal{F}$ is an ample field, then every singleton of $\mathcal{X}$ is contained in an atom of $\mathcal{F}$. As in the proof, we shall denote this atom by $[x]_{\mathcal{F}}$.

What happens if we restrict a possibility measure defined on an ample field $\mathcal{F}$ to the corresponding partition $\mathbb{A}(\mathcal{F})$? As we shall prove shortly, a possibility measure is uniquely determined by its restriction to atoms, and a necessity measure by its restriction to complements of atoms.

Definition 3.52. Let $\mathcal{F}$ be an ample field. Let Π be a possibility measure on $\mathcal{F}$. The gamble

$$\pi(x) := \Pi([x]_{\mathcal{F}}), \text{ for all } x \in \mathcal{X},$$

is called the *possibility distribution induced by* Π. For a necessity measure N on $\mathcal{F}$, the gamble

$$\nu(x) := N(\complement[x]_{\mathcal{F}}), \text{ for all } x \in \mathcal{X},$$

is called the *necessity distribution induced by* N.

The next theorem and corollary is a summary of results by De Cooman [17]. Again, since the proofs are short, they are given here for the sake of completeness. The notion of $\mathcal{F}$-measurability will be defined in Section 4.3.2; see Definition 4.25 on p. 109. The only thing we need to know now, is that in case $\mathcal{F}$ is an ample field, a gamble f is $\mathcal{F}$-measurable if and only if it is constant on the atoms of $\mathcal{F}$; see Proposition 4.27 on p. 111.

Theorem 3.53. *Let $\mathcal{F}$ be an ample field. A set function N defined on $\mathcal{F}$ is a necessity measure if and only if there is an $\mathcal{F}$-measurable gamble ν on $\mathcal{X}$ such that* $\inf \nu = 0$, $\sup \nu \leq 1$ *and*

$$N(A) = \inf_{x \in \complement A} \nu(x), \tag{3.21}$$

for any A in $\mathcal{F} \setminus \{\mathcal{X}\}$, and $N(\mathcal{X}) = 1$. If so, then ν is the necessity distribution induced by N. Similarly, a set function Π defined on $\mathcal{F}$ is a possibility measure if and only if there is an $\mathcal{F}$-measurable gamble π on $\mathcal{X}$ such that $\inf \pi \geq 0$, $\sup \pi = 1$, *and*

$$\Pi(A) = \sup_{x \in A} \pi(x), \tag{3.22}$$

for any A in $\mathcal{F} \setminus \{\emptyset\}$, and $\Pi(\emptyset) = 0$. If so, then π is the possibility distribution induced by Π.

Proof. Immediate from Theorem 3.51. Let's fill in the details. First of all, it suffices to prove only the first part of the theorem, concerning necessity measures. The part about possibility measures follows then from their dual necessity measures.

"if". Let ν be an $\mathcal{F}$-measurable mapping on $\mathcal{X}$ such that $\inf \nu = 0$, $\sup \nu \leq 1$ and suppose that Eq. (3.21) holds. The properties $N(\emptyset) = 0$, $N(\mathcal{X}) = 1$ and $N(A) \leq N(B)$ for all A and B in $\mathcal{F}$ such that $A \subseteq B$, are immediate. Let $\mathcal{A}$ be a subset of $\mathcal{F} \setminus \{\mathcal{X}\}$. Then

$$\begin{aligned} N\left(\bigcap_{A \in \mathcal{A}} A\right) &= \inf\left\{\nu(x) \colon x \in \complement\left(\bigcap_{A \in \mathcal{A}} A\right)\right\} \\ &= \inf\left\{\nu(x) \colon x \in \bigcup_{A \in \mathcal{A}} \complement A\right\} \\ &= \inf\left\{\nu(x) \colon \exists A \in \mathcal{A} \colon x \in \complement A\right\} \end{aligned}$$

$$
\begin{aligned}
&= \inf\left(\bigcup_{A\in\mathcal{A}}\left\{\nu(x)\colon x\in \complement A\right\}\right)\\
&= \inf\left\{\inf\left\{\nu(x)\colon x\in \complement A\right\} : A\in\mathcal{A}\right\}\\
&= \inf_{A\in\mathcal{A}} N(A),
\end{aligned}
$$

and hence, N is an infimum preserving set function; it is a necessity measure. From the equality Eq. (3.21) it also easily follows that, for any $x \in \mathcal{X}$,

$$
N(\complement[x]_{\mathcal{F}}) = \inf_{y\in\complement\complement[x]_{\mathcal{F}}} \nu(y) = \inf_{y\in[x]_{\mathcal{F}}} \nu(y) = \nu(x),
$$

since ν is $\mathcal{F}$-measurable, and using Proposition 4.27. Hence, ν coincides with the necessity distribution induced by N.

"only if". Suppose that N is a necessity measure on $\mathcal{F}$. Then the necessity distribution ν induced by N, satisfies Eq. (3.21). Indeed, since N is infimum preserving, and $A = \bigcap_{B\in\mathbb{A}(\mathcal{F}),\, B\cap\complement A\neq\emptyset} \complement B$ by Theorem 3.51 applied on $\complement A$, it follows that

$$
N(A) = \inf_{\substack{B\in\mathbb{A}(\mathcal{F})\\ B\cap\complement A\neq\emptyset}} N(\complement B) = \inf_{x\in\complement A} N(\complement[x]_{\mathcal{F}}) = \inf_{x\in\complement A} \nu(x)
$$

for any $A \in \mathcal{F} \setminus \{\mathcal{X}\}$. From this equality, it also follows that $\sup \nu = \sup_{x\in\mathcal{X}} N(\complement[x]_{\mathcal{F}}) \leq 1$, and

$$
\inf_{x\in\mathcal{X}} \nu(x) = \inf_{B\in\mathbb{A}(\mathcal{F})} N(\complement B) = N\left(\bigcap_{B\in\mathbb{A}(\mathcal{F})} \complement B\right) = N\left(\complement \bigcup_{B\in\mathbb{A}(\mathcal{F})} B\right) = N(\complement\mathcal{X}) = 0,
$$

applying Theorem 3.51 to get $\bigcup_{B\in\mathbb{A}(\mathcal{F})} B = \mathcal{X}$. □

The following corollary is similar to a well-known result in probability theory, namely, that there is an onto and one-to-one correspondence between (sufficiently regular) probability density functions on $[a, b]$, and (sufficiently regular) probability measures on $[a, b]$; see De Cooman [17].

Definition 3.54. A *necessity distribution* ν on X is a gamble on X such that $0 = \inf \nu \leq \sup \nu \leq 1$. A *possibility distribution* π on X is a gamble on X such that $0 \leq \inf \pi \leq \sup \pi = 1$.

Corollary 3.55. *Let $\mathcal{F}$ be an ample field. There is an onto and one-to-one correspon-*

dence between necessity measures on $\mathcal{F}$ and $\mathcal{F}$-measurable necessity distributions, and an onto and one-to-one correspondence between possibility measures on $\mathcal{F}$ and $\mathcal{F}$-measurable possibility distributions.

Proof. Immediate from Theorem 3.53, and the fact that, through Eq. (3.21), a necessity measure N on $\mathcal{F}$ uniquely determines an $\mathcal{F}$-measurable necessity distribution—namely, its induced necessity distribution—and any $\mathcal{F}$-measurable necessity distribution ν uniquely determines a necessity measure N on $\mathcal{F}$: indeed, suppose $N(A) \neq M(A)$ for some $A \in \mathcal{F}$, where M is a necessity measure that also satisfies Eq. (3.21) with necessity distribution ν. Then

$$\inf_{x \in \complement A} \nu(x) = N(A) \neq M(A) = \inf_{x \in \complement A} \nu(x)$$

We have arrived at a contradiction. Hence, N is unique.

The proof for possibility measures now follows from their dual necessity measures. □

3.5.10 Cumulative Distribution Functions and P-Boxes

A probability box, or p-box, models uncertainty about a real-valued random variable X through bounds on the cumulative distribution function of X; we refer to Ferson, Kreinovich, Ginzburg, Myers, and Sentz [33] for an in depth discussion of this model. Within our framework, we shall view a p-box as a lower prevision defined on a particular set of indicator gambles and their negations.

We shall be mostly concerned with p-boxes modelling uncertainty about bounded random variables, *i.e.*, gambles. Therefore, we restrict our study to p-boxes defined on compact intervals only.

Definition 3.56. Let $\mathcal{X} = [a, b]$ be a compact interval in $\mathbb{R}$. An ordered pair (F_*, F^*) of real-valued mappings on $\mathcal{X}$ is called a *p-box* on $\mathcal{X}$, and the lower prevision $\underline{\mathbf{P}}_{(F_*,F^*)}$ defined by

$$\underline{\mathbf{P}}_{(F_*,F^*)}(I_{[a,x]}) = F_*(x) \text{ and } \underline{\mathbf{P}}_{(F_*,F^*)}(-I_{[a,x]}) = -F^*(x) \text{ for all } x \in \mathcal{X}, \tag{3.23}$$

is called the *lower prevision induced by* (F_*, F^*).

The identification of p-boxes (F_*, F^*) with $\underline{\mathbf{P}}_{(F_*,F^*)}$ means that we interpret $F_*(x)$ as a lower probability and $F^*(x)$ as an upper probability for the event

$[a, x]$. So, p-boxes are nothing but a specification of a lower and an upper cumulative distribution function.

Through the above identifications, we may also translate our lower prevision behavioural concepts to p-boxes. We say that a p-box avoids sure loss if its corresponding lower prevision avoids sure loss, etc. It is easy to show exactly when a p-box is coherent. Let's first consider the simpler case $F_* = F^*$.

Definition 3.57. Let $\mathcal{X} = [a, b]$ be a compact interval in $\mathbb{R}$. A *cumulative distribution function* on $\mathcal{X}$ is a self-conjugate p-box (F, F) on $\mathcal{X}$.

A cumulative distribution function F (we shall write simply F instead of (F, F) for cumulative distribution functions) induces a probability defined by $\mathbf{P}_F := \underline{\mathbf{P}}_{(F,F)}$. It satisfies $\mathbf{P}_F(I_{[a,x]}) = -\mathbf{P}_F(-I_{[a,x]}) = F(x)$. When is $\mathbf{P}_F$ coherent?

Lemma 3.58. *Let $\mathcal{X} = [a, b]$ be a compact interval in $\mathbb{R}$ and let F be cumulative distribution function F on $\mathcal{X}$. Then $\mathbf{P}_F$ is coherent if and only if*

(i) $x \leq y \implies F(x) \leq F(y)$ *for all* $x, y \in \mathcal{X}$, *and*

(ii) $0 \leq F(a) \leq F(b) = 1$.

Proof. "if". Immediate from Theorem 3.27.

"only if". Immediate from the properties of coherence, Theorem 3.5. □

Theorem 3.59. *Let $\mathcal{X} = [a, b]$ be a compact interval in $\mathbb{R}$, and let (F_*, F^*) be a p-box on $\mathcal{X}$. Then $\underline{\mathbf{P}}_{(F_*,F^*)}$ is coherent if and only if both $\mathbf{P}_{F_*}$ and $\mathbf{P}_{F^*}$ are coherent, and additionally $F_*(x) \leq F^*(x)$ for all $x \in \mathcal{X}$:*

(i) $x \leq y \implies F_*(x) \leq F_*(y)$ *and* $F^*(x) \leq F^*(y)$, *for all* $x, y \in \mathcal{X}$

(ii) $0 \leq F_*(x) \leq F^*(x) \leq 1 = F_*(b) = F^*(b)$, *for all* $x \in \mathcal{X}$

Proof. "only if". The conditions for coherence of $\mathbf{P}_{F_*}$ and $\mathbf{P}_{F^*}$ (described in Lemma 3.58) and the condition $F_*(x) \leq F^*(x)$ for all $x \in \mathcal{X}$ are easily derived from the coherence of $\underline{\mathbf{P}}_{(F_*,F^*)}$.

"if". If both $\mathbf{P}_{F_*}$ and $\mathbf{P}_{F^*}$ are coherent, then, by Lemma 3.11, so must be their lower envelope

$$\underline{P}(I_{[a,x]}) = \min\{\mathbf{P}_{F_*}(I_{[a,x]}), \mathbf{P}_{F^*}(I_{[a,x]})\} = \min\{F_*(x), F^*(x)\} = F_*(x), \text{ and}$$
$$\underline{P}(-I_{[a,x]}) = \min\{\mathbf{P}_{F_*}(-I_{[a,x]}), \mathbf{P}_{F^*}(-I_{[a,x]})\} = \min\{-F_*(x), -F^*(x)\} = -F^*(x),$$

where we used $F_*(x) \leq F^*(x)$. But $\underline{P}$ is exactly equal to $\underline{\mathbf{P}}_{(F_*,F^*)}$. Hence, $\underline{\mathbf{P}}_{(F_*,F^*)}$ is coherent. □

Chapter 4

Inference

We now turn to the following simple inference problem. Suppose we have a lower prevision $\underline{P}$, defined on a subset of the set of all gambles on X. Can we infer from $\underline{P}$ something about the lower prevision of a gamble that is not in the domain of $\underline{P}$? More generally, can we extend $\underline{P}$ to a coherent lower prevision defined on a larger domain? Walley [86] proved that this is possible if $\underline{P}$ avoids sure loss, in which case he proved there is point-wise smallest coherent lower prevision in the set of all coherent extensions of $\underline{P}$. This is only one of the many ways to define natural extension; we refer to Walley [86] on this subject.

Walley [86] also demonstrated how natural extension encompasses many other extension methods known from the literature: the Lebesgue integral on the closed unit interval, the Choquet integral with respect to 2-monotone set functions, inner and outer measures, Bayes's rule, *etc*. We shall discuss these results, and generalise some of them.

4.1 Natural Extension

Let $\underline{P}$ be any lower prevision, and let $\mathcal{K}$ be a set of gambles that includes $\operatorname{dom}\underline{P}$. Let's carefully sum up the properties which our extension $\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{K}}$ of $\underline{P}$ to $\mathcal{K}$ should satisfy. Note that Walley only discusses the case $\mathcal{K} = \mathcal{L}(X)$; as we shall see in Corollary 4.4, all other cases follow from $\mathcal{K} = \mathcal{L}(X)$. We shall denote $\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{L}(X)}$ by $\underline{\mathbf{E}}_{\underline{P}}$.

First of all, any behavioural disposition expressed by $\underline{P}$ should also be

expressed by $\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{K}}$. Hence, any transaction implied by $\underline{P}$, such as buying a gamble f for a price s, should also be implied by $\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{K}}$. In such a case, we say that $\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{K}}$ is a behavioural extension of $\underline{P}$:

Definition 4.1. We say that a lower prevision $\underline{Q}$ is a *behavioural extension* of a lower prevision $\underline{P}$ if $\operatorname{dom}\underline{P} \subseteq \operatorname{dom}\underline{Q}$ and $\underline{P}(f) \leq \underline{Q}(f)$ for any gamble $f \in \operatorname{dom}\underline{P}$.

Thus, $\operatorname{dom}\underline{P} \subseteq \mathcal{K}$ and $\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{K}}(f) \geq \underline{P}(f)$ for all f in $\operatorname{dom}\underline{P}$. Secondly, $\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{K}}$ must be coherent, as argued in Section 3.4.2. Last but not least, we want the buying prices $\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{K}}$ to be as low as possible: any coherent behavioural extension of $\underline{P}$ to $\mathcal{K}$ must also be a behavioural extension of $\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{K}}$. This can only be the case when $\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{K}}$ is the point-wise smallest coherent behavioural extension of $\underline{P}$ to $\mathcal{K}$. Because of this property, $\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{K}}$ is sometimes also called the *least committal extension* of $\underline{P}$; see for instance Walley [85, p. 28].

Definition 4.2. Let $\underline{P}$ be a lower prevision, and let $\operatorname{dom}\underline{P} \subseteq \mathcal{K} \subseteq \mathcal{L}(X)$. The point-wise smallest coherent behavioural extension of $\underline{P}$ to $\mathcal{K}$, if it exists, is called the *natural extension of $\underline{P}$ to $\mathcal{K}$*, and it is denoted by $\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{K}}$. The natural extension of $\underline{P}$ to $\mathcal{L}(X)$ is simply called the *natural extension of $\underline{P}$*, and is denoted by $\underline{\mathbf{E}}_{\underline{P}}$.

The main contribution of the following theorem, again due to Walley [86, Chapter 3], is that avoiding sure loss of $\underline{P}$ is necessary and sufficient for the existence of its natural extension $\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{K}}$. It also gives an explicit expression for $\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{K}}$, and a number of criteria for checking avoiding sure loss.

Theorem 4.3. *Let $\underline{P}$ be a lower prevision, and let* $\operatorname{dom}\underline{P} \subseteq \mathcal{K} \subseteq \mathcal{L}(X)$. *Define the $\mathcal{L}(X)$–$\mathbb{R}^*$ map $\underline{E}$ by*

$$\underline{E}(f) := \sup\Bigg\{\alpha + \sum_{i=1}^{n} \lambda_i \underline{P}(f_i) : \alpha \in \mathbb{R},\ n \in \mathbb{N},\ \lambda_1, \ldots, \lambda_n \geq 0,\ f_1, \ldots, f_n \in \operatorname{dom}\underline{P},\ (\forall x \in \mathcal{X})\Big(\alpha + \sum_{i=1}^{n} \lambda_i f_i(x) \leq f(x)\Big)\Bigg\} \quad (4.1)$$

for any gamble $f \in \mathcal{L}(X)$. The following conditions are equivalent.

(i) $\underline{E}(f) < +\infty$ *for some gamble $f \in \mathcal{L}(X)$.*

(ii) $\underline{E}(f) < +\infty$ *for all gambles* $f \in \mathcal{L}(X)$.

(iii) $\underline{E}$ *is a coherent lower prevision.*

(iv) The natural extension of $\underline{P}$ *to* $\mathcal{K}$ *exists and is equal to* $\underline{E}$ *restricted to* $\mathcal{K}$.

(v) $\underline{P}$ *has at least one coherent behavioural extension.*

(vi) $\underline{P}$ *has at least one behavioural extension that avoids sure loss.*

(vii) $\underline{P}$ *avoids sure loss.*

The following corollary, whose proof is immediate from the above theorem, tells us that $\underline{\mathbf{E}}_{\underline{P}}$ uniquely determines the natural extension $\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{K}}$ of $\underline{P}$ to any domain $\mathcal{K}$ that includes $\operatorname{dom}\underline{P}$.

Corollary 4.4. *Let* $\underline{P}$ *be a lower prevision, and let* $\operatorname{dom}\underline{P} \subseteq \mathcal{K} \subseteq \mathcal{L}(X)$. *Then* $\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{K}}$ *exists if and only if* $\underline{\mathbf{E}}_{\underline{P}}$ *exists, and in such a case*

$$\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{K}}(f) = \underline{\mathbf{E}}_{\underline{P}}(f) \text{ for all } f \in \mathcal{K}.$$

So, from now on, in proofs, we can focus our attention on the natural extension $\underline{\mathbf{E}}_{\underline{P}}$ of $\underline{P}$ to the set of all gambles on X. An alternative, and simpler expression for natural extension is obtained when $\underline{P}$ is defined on a linear space and is coherent. This theorem is also due to Walley [86, Definition 3.1.1 and Theorem 3.1.2, pp. 122–124]; the proof is very short, we provide it for the sake of completeness.

Theorem 4.5. *Let* $\underline{P}$ *be any coherent lower prevision defined on a linear space, and let* $\operatorname{dom}\underline{P} \subseteq \mathcal{K} \subseteq \mathcal{L}(X)$. *Then the natural extension of* $\underline{P}$ *to* $\mathcal{K}$ *exists, and for any gamble* $f \in \mathcal{K}$,

$$\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{K}}(f) := \sup\{a + \underline{P}(g) \colon a \in \mathbb{R},\ g \in \operatorname{dom}\underline{P},\ a + g \leq f\}. \tag{4.2}$$

for any gamble $f \in \mathcal{K}$.

Proof. Look at Eq. (4.1) and note that

$$\sum_{i=1}^{n} \lambda_i \underline{P}(f_i) \leq \underline{P}\left(\sum_{i=1}^{n} \lambda_i f_i\right). \tag{4.3}$$

Since we are looking for the supremum, we can replace $\sum_{i=1}^{n} \lambda_i \underline{P}(f_i)$ by $\underline{P}(g)$ with $g = \sum_{i=1}^{n} \lambda_i f_i$. □

$\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{K}}$ coincides with $\underline{P}$ on $\operatorname{dom}\underline{P}$ if $\underline{P}$ is coherent; this follows directly from the definition of natural extension in case $\operatorname{dom}\underline{P} = \mathcal{K}$. We refer to Walley [86, Theorem 3.1.2(d)] for a proof.

Proposition 4.6. *Let $\underline{P}$ be a lower prevision, and let $\operatorname{dom}\underline{P} \subseteq \mathcal{K} \subseteq \mathcal{L}(X)$. If $\underline{P}$ is coherent then $\underline{P}$ and $\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{K}}$ coincide on $\operatorname{dom}\underline{P}$.*

Let's end with few nice results about natural extension, not due to Walley.

Proposition 4.7. *Let $\underline{P}$ and $\underline{Q}$ be lower previsions on X that avoid sure loss. If $\underline{Q}$ is a behavioural extension of $\underline{P}$, then $\underline{\mathbf{E}}_{\underline{Q}}$ is a behavioural extension of $\underline{\mathbf{E}}_{\underline{P}}$ too: $\underline{\mathbf{E}}_{\underline{Q}}(f) \geq \underline{\mathbf{E}}_{\underline{P}}(f)$ for every gamble f on X.*

Proof. If $\underline{Q}$ is a behavioural extension of $\underline{P}$, then any coherent behavioural extension of $\underline{Q}$ is also a coherent behavioural extension of $\underline{P}$; now apply the definition of natural extension: it is the point-wise smallest coherent behavioural extension to the set of all gambles on X. □

Proposition 4.8. *Let $\underline{P}$ be a lower prevision that avoids sure loss. Let $\underline{Q}$ be any coherent behavioural extension of $\underline{P}$. Then $\underline{P}$ is equivalent to $\underline{Q}$ (that is, $\underline{\mathbf{E}}_{\underline{P}} = \underline{\mathbf{E}}_{\underline{Q}}$) if and only if $\underline{Q}$ and $\underline{\mathbf{E}}_{\underline{P}}$ coincide on $\operatorname{dom}\underline{Q}$.*

Proof. Note that since $\underline{P}$ avoids sure loss and $\underline{Q}$ is coherent, both $\underline{\mathbf{E}}_{\underline{P}}$ and $\underline{\mathbf{E}}_{\underline{Q}}$ exist.

"if". Since $\underline{Q}$ is a behavioural extension of $\underline{P}$, any behavioural extension of $\underline{Q}$ is also a behavioural extension of $\underline{P}$. Hence, $\underline{\mathbf{E}}_{\underline{Q}} \geq \underline{\mathbf{E}}_{\underline{P}}$. To prove the converse inequality, let $\underline{R}$ be any coherent behavioural extension of $\underline{P}$ to the set of all gambles on X. The claim is established if we can show that $\underline{R}$ is also a behavioural extension of $\underline{Q}$. Indeed, $\underline{R} \geq \underline{\mathbf{E}}_{\underline{P}}$ by definition of natural extension. Since $\underline{Q} = \underline{\mathbf{E}}_{\underline{P}}$ on $\operatorname{dom}\underline{Q}$ it follows that also $\underline{R} \geq \underline{Q}$ on $\operatorname{dom}\underline{Q}$, which means that $\underline{R}$ is a behavioural extension of $\underline{Q}$.

"only if". Suppose $\underline{\mathbf{E}}_{\underline{P}} = \underline{\mathbf{E}}_{\underline{Q}}$. Since $\underline{Q}$ is coherent, it follows from Proposition 4.6 that $\underline{Q}$ and $\underline{\mathbf{E}}_{\underline{Q}}$ coincide on $\operatorname{dom}\underline{Q}$, and hence, $\underline{\mathbf{E}}_{\underline{P}}$ and $\underline{Q}$ coincide on $\operatorname{dom}\underline{Q}$. □

Corollary 4.9. *Let $\underline{P}$ be a lower prevision that avoids sure loss, and let $\operatorname{dom}\underline{P} \subseteq \mathcal{J} \subseteq \mathcal{K} \subseteq \mathcal{L}(X)$. Then*

$$\underline{\mathbf{E}}_{\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{J}}}^{\mathcal{K}}(f) = \underline{\mathbf{E}}_{\underline{P}}^{\mathcal{K}}(f), \text{ for all } f \in \mathcal{K}, \text{ and } \underline{\mathbf{E}}_{\underline{P}}^{\mathcal{J}}(f) = \underline{\mathbf{E}}_{\underline{P}}^{\mathcal{K}}(f), \text{ for all } f \in \mathcal{J}.$$

Proof. By Corollary 4.4, $\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{J}}$ coincides with $\underline{\mathbf{E}}_{\underline{P}}$ on $\mathcal{J}$. Hence, by Proposition 4.8, $\underline{\mathbf{E}}_{\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{J}}}$ is equal to $\underline{\mathbf{E}}_{\underline{P}}$. Therefore, also $\underline{\mathbf{E}}_{\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{J}}}^{\mathcal{K}}$ is equal to $\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{K}}$, again by Corollary 4.4. The other equality is immediate from Corollary 4.4. □

The first equality of Corollary 4.9 could be called *transitivity of natural extension*: let $\underline{P}$, $\underline{Q}$, and $\underline{R}$ be three coherent lower previsions on X, and assume that $\operatorname{dom}\underline{P} \subseteq \operatorname{dom}\underline{Q} \subseteq \operatorname{dom}\underline{R}$; if $\underline{Q}$ is the natural extension of $\underline{P}$ to $\operatorname{dom}\underline{Q}$, and $\underline{R}$ is the natural extension of $\underline{Q}$ to $\operatorname{dom}\underline{R}$, then $\underline{R}$ is the natural extension of $\underline{P}$ to $\operatorname{dom}\underline{R}$.

The following proposition generalises Proposition 3.24 on p. 70. Recall that the dual of a set function ν is defined as $\pi(\complement A) := 1 - \nu(A)$ for any $A \in \operatorname{dom}\nu$.

Proposition 4.10. *Let ν be a set function defined on a collection $\mathcal{A}$ of subsets of $\mathcal{X}$, and let π be its dual. Then $\underline{\mathbf{P}}_\nu$ and $\overline{\mathbf{P}}_\pi$ are equivalent: $\underline{\mathbf{P}}_\nu$ avoids sure loss if and only if $\overline{\mathbf{P}}_\pi$ avoids sure loss, and in such a case, $\underline{\mathbf{E}}_{\underline{\mathbf{P}}_\nu} = \underline{\mathbf{E}}_{\overline{\mathbf{P}}_\pi}$.*

Proof. Immediate from Theorem 4.3. □

4.2 Linear Extension and Integration

In the following sections we shall study how integration with respect to probability charges and other event-based uncertainty structures can be obtained through natural extension, and *vice versa*. Let's start with some general considerations about integration. Many of the integrals we know are linear functionals that can be written as (a linear combination of)[1] linear previsions defined on a linear space of gambles. So, it seems natural to me to define integration as a kind of linear natural extension for lower previsions. The idea of defining integrability and integrals through natural extension is new, and we shall explore its relation with some of the more common integrals further on.

Definition 4.11. Let $\underline{P}$ be a lower prevision that avoids sure loss. Then the *linear extension* $\mathbf{E}_{\underline{P}}$ of $\underline{P}$ is defined as the natural extension $\underline{\mathbf{E}}_{\underline{P}}$ restricted to the

[1] For instance, by the Jordan decomposition theorem, any bounded charge is a linear combination of two probability charges; see Bhaskara Rao and Bhaskara Rao [9, Theorem 2.5.3]. Therefore, as long as integrals are linear, we can reduce integration with respect to bounded charges to integration with respect to probability charges. For bounded positive charges this comes down to renormalisation. Renormalisation will be extensively used further on.

domain where it is self-conjugate, that is, to $\{f \in \mathcal{L}(X)\colon \underline{\mathbf{E}}_{\underline{P}}(f) = \overline{\mathbf{E}}_{\underline{P}}(f)\}$. We say that f is *$\underline{P}$-integrable* if $f \in \operatorname{dom} \mathbf{E}_{\underline{P}}$, and we shall call $\mathbf{E}_{\underline{P}}(f)$ the *integral of f with respect to $\underline{P}$*, or simply the *$\underline{P}$-integral*.

Since $\mathbf{E}_{\underline{P}}$ is a restriction of a coherent lower prevision, namely $\underline{\mathbf{E}}_{\underline{P}}$, and since $\mathbf{E}_{\underline{P}}$ is self-conjugate by its definition, it follows that it is a linear prevision.

Proposition 4.12. *$\mathbf{E}_{\underline{P}}$ is a linear prevision.*

We emphasise again that the $\underline{P}$-integral introduced here is a linear integral. Not all integrals are linear. The most important non-linear integral is probably the Choquet integral; this integral is usually not a $\underline{P}$-integral. The Choquet integral is the only non-linear integral used in this work.

The linear extension is in a sense the *unique* coherent behavioural extension, as explained in the following proposition.

Proposition 4.13. *Let $\underline{P}$ be a lower prevision that avoids sure loss. Then for any coherent behavioural extension $\underline{Q}$ of $\underline{P}$, it holds that $\mathbf{E}_{\underline{P}}(f) = \underline{Q}(f)$ whenever $f \in \operatorname{dom} \mathbf{E}_{\underline{P}} \cap \operatorname{dom} \underline{Q}$.*

Proof. Let $f \in \operatorname{dom} \mathbf{E}_{\underline{P}} \cap \operatorname{dom} \underline{Q}$. Simply note that the linear extension coincides with the natural extension on its domain, and the natural extension is the point-wise smallest coherent behavioural extension of $\underline{P}$. In particular,

$$\mathbf{E}_{\underline{P}}(f) = \underline{\mathbf{E}}_{\underline{P}}(f) \leq \underline{\mathbf{E}}_{\underline{Q}}(f) \leq -\underline{\mathbf{E}}_{\underline{Q}}(-f) \leq -\underline{\mathbf{E}}_{\underline{P}}(-f) = -\mathbf{E}_{\underline{P}}(-f) = \mathbf{E}_{\underline{P}}(f),$$

hence, $\mathbf{E}_{\underline{P}}(f) = \underline{\mathbf{E}}_{\underline{Q}}(f)$. Now, by coherence of $\underline{Q}$ and Proposition 4.6 we have that $\underline{Q}(f) = \underline{\mathbf{E}}_{\underline{Q}}(f)$. Hence, $\mathbf{E}_{\underline{P}}(f) = \underline{Q}(f)$. □

For previsions we don't only have uniqueness, but we can even establish equivalence of P and $\mathbf{E}_P$. Below, we prove an even more general statement; compare with Proposition 4.8.

Proposition 4.14. *Let P be a linear prevision. Then for any coherent behavioural extension $\underline{Q}$ of P such that $\operatorname{dom} \underline{Q} \subseteq \operatorname{dom} \mathbf{E}_P$ it holds that $\underline{\mathbf{E}}_{\underline{Q}} = \mathbf{E}_P$, that is, $\underline{Q}$ is equivalent to P.*

Proof. Since any behavioural extension of $\underline{Q}$ must also be a behavioural extension of P, it clearly holds that $\underline{\mathbf{E}}_{\underline{Q}} \geq \mathbf{E}_P$. The converse is established if we can show that any behavioural extension $\underline{R}$ of P, defined on $\mathcal{L}(X)$, is also a

behavioural extension of $\underline{Q}$. Indeed, by Proposition 4.13 we easily find that $\mathbf{E}_P = \underline{R}$ on $\operatorname{dom}\mathbf{E}_P$, and $\mathbf{E}_P = \underline{Q}$ on $\operatorname{dom}\underline{Q}$. Hence, $\underline{Q} = \underline{R}$ on $\operatorname{dom}\underline{Q}$. But this means that $\underline{R}$ is a behavioural extension of $\underline{Q}$. □

The following is an immediate consequence of Proposition 4.7 on p. 98.

Proposition 4.15. *Let $\underline{P}$ and $\underline{Q}$ be lower previsions on X that avoid sure loss. If $\underline{Q}$ is a behavioural extension of $\underline{P}$, then $\mathbf{E}_{\underline{Q}}$ is an extension of $\mathbf{E}_{\underline{P}}$: it holds that* $\operatorname{dom}\mathbf{E}_{\underline{P}} \subseteq \operatorname{dom}\mathbf{E}_{\underline{Q}}$ *and* $\mathbf{E}_{\underline{P}}(f) = \mathbf{E}_{\underline{Q}}(f)$ *for every f in* $\operatorname{dom}\mathbf{E}_{\underline{P}}$.

Proof. By Proposition 4.7 on p. 98, we already have that $\underline{\mathbf{E}}_{\underline{Q}}$ is a behavioural extension of $\underline{\mathbf{E}}_{\underline{P}}$. Consequently, for any gamble f in $\operatorname{dom}\mathbf{E}_{\underline{P}}$ it holds that

$$\mathbf{E}_{\underline{P}}(f) = \underline{\mathbf{E}}_{\underline{P}}(f) \leq \underline{\mathbf{E}}_{\underline{Q}}(f) \leq \overline{\mathbf{E}}_{\underline{Q}}(f) \leq \overline{\mathbf{E}}_{\underline{P}}(f) = \mathbf{E}_{\underline{P}}(f),$$

so f belongs to $\operatorname{dom}\mathbf{E}_{\underline{Q}}$ and $\mathbf{E}_{\underline{Q}}(f) = \mathbf{E}_{\underline{P}}(f)$. □

The $\underline{P}$-integrable gambles interact in additive way with other gambles.

Proposition 4.16. *Let $\underline{P}$ be a lower prevision on X that avoids sure loss. For any pair of gambles f and g on X of which at least one is $\underline{P}$-integrable it holds that*

$$\underline{\mathbf{E}}_{\underline{P}}(f+g) = \underline{\mathbf{E}}_{\underline{P}}(f) + \underline{\mathbf{E}}_{\underline{P}}(g) \quad \text{and} \quad \overline{\mathbf{E}}_{\underline{P}}(f+g) = \overline{\mathbf{E}}_{\underline{P}}(f) + \overline{\mathbf{E}}_{\underline{P}}(g).$$

Proof. Suppose for instance f is $\underline{P}$-integrable, that is, $\underline{\mathbf{E}}_{\underline{P}}(f) = \overline{\mathbf{E}}_{\underline{P}}(f)$. Then the first equality follows from

$$\underline{\mathbf{E}}_{\underline{P}}(f) + \underline{\mathbf{E}}_{\underline{P}}(g) \leq \underline{\mathbf{E}}_{\underline{P}}(f+g) \leq \overline{\mathbf{E}}_{\underline{P}}(f) + \underline{\mathbf{E}}_{\underline{P}}(g) = \underline{\mathbf{E}}_{\underline{P}}(f) + \underline{\mathbf{E}}_{\underline{P}}(g),$$

where we used the coherence of $\underline{\mathbf{E}}_{\underline{P}}$ and Theorem 3.5(v). For upper previsions the proof is similar. □

The linear extension is a natural generalisation to gambles of the Jordan extension (see Denneberg [28, p. 29]) in measure theory; the Jordan extension will be defined in Section 4.3.4. Indeed, restricting the self-conjugacy condition to indicators, we recover the condition for Jordan measurability. The following corollary tells us that linear extension is also very similar to the construction of the Carathéodory extension (see Denneberg [28, p. 24]) in measure theory; the Carathéodory extension will also be defined in Section 4.3.4. However, the condition below, restricted to indicators, is not

similar to the condition for Carathéodory measurability—the equivalence is true, but it is not immediate.

Corollary 4.17. *Let $\underline{P}$ be a lower prevision that avoids sure loss. A gamble f on X is $\underline{P}$-integrable if and only if*

$$\underline{\mathbf{E}}_{\underline{P}}(f+g) = \underline{\mathbf{E}}_{\underline{P}}(f) + \underline{\mathbf{E}}_{\underline{P}}(g), \quad \text{for all } g \in \mathscr{L}(X).$$

Proof. "only if". Immediate from Proposition 4.16.

"if". Take $g = -f$. Then $0 = \underline{\mathbf{E}}_{\underline{P}}(0) = \underline{\mathbf{E}}_{\underline{P}}(f) + \underline{\mathbf{E}}_{\underline{P}}(-f) = \underline{\mathbf{E}}_{\underline{P}}(f) - \overline{\mathbf{E}}_{\underline{P}}(f)$. □

The following proposition gives a lower bound on the domain of the linear extension. Recall that $\mathbb{R}(X)$ denotes the set of constant gambles on X. The relation $\underline{P}(f) = \sup f$ in the proposition below is related to so-called null-sets and null-gambles. With respect to a probability measure μ, a subset A of $\mathcal{X}$ is called a null-set if its outer measure $\mu^*(A)$ is zero; we shall see in Section 4.3.4 that the outer measure corresponds to a coherent upper prevision. Hence, $\mu^*(A)$ translates into $\overline{P}(I_A) = 0$, or equivalently, $\underline{P}(-I_A) = \sup[-I_A]$. The proposition below says that null-sets belong to the domain of the linear extension—it is not hard to imagine how this also extends to null-gambles f, which satisfy $\overline{P}(|f|) = 0$. Null-sets and null-gambles will be introduced in Section 5.3.1 on p. 225 ff. and Section 5.3.2 on p. 230 ff.

Recall the definition of 2-monotonicity for lower previsions; see Definition 3.32 on p. 76.

Proposition 4.18. *Let $\underline{P}$ be a lower prevision that avoids sure loss. The following statements hold.*

(i) If both f and $-f$ are in $\operatorname{dom}\underline{P}$, and $\underline{P}(f) = -\underline{P}(-f)$, then $f \in \operatorname{dom}\underline{\mathbf{E}}_{\underline{P}}$ and $\underline{P}(f) = \underline{\mathbf{E}}_{\underline{P}}(f)$.

(ii) If $f \in \operatorname{dom}\underline{P}$ and $\underline{P}(f) = \sup f$, then $f \in \operatorname{dom}\underline{\mathbf{E}}_{\underline{P}}$ and $\underline{P}(f) = \underline{\mathbf{E}}_{\underline{P}}(f)$.

(iii) If $|f| \in \operatorname{dom}\overline{P}$ and $\overline{P}(|f|) = 0$, then any gamble g such that $|g| \leq |f|$, belongs to $\operatorname{dom}\underline{\mathbf{E}}_{\underline{P}}$, and $\underline{\mathbf{E}}_{\underline{P}}(g) = 0$. In particular, $\underline{\mathbf{E}}_{\underline{P}}(f) = \underline{\mathbf{E}}_{\underline{P}}(|f|) = \underline{\mathbf{E}}_{\underline{P}}(-|f|) = 0$.

(iv) $\mathbb{R}(X) \subseteq \operatorname{dom}\underline{\mathbf{E}}_{\underline{P}}$.

(v) $\operatorname{dom}\underline{\mathbf{E}}_{\underline{P}}$ is a uniformly closed linear space.

(vi) If $\underline{\mathbf{E}}_{\underline{P}}$ is 2-monotone then $\operatorname{dom}\underline{\mathbf{E}}_{\underline{P}}$ is a uniformly closed linear lattice.

Proof. (i). If both f and $-f$ are in $\operatorname{dom}\underline{P}$ and $\underline{P}(f) = -\underline{P}(-f)$ then, by the definition of natural extension as the point-wise smallest coherent behavioural extension of $\underline{P}$,

$$\overline{P}(f) = \underline{P}(f) \leq \underline{\mathbf{E}}_{\underline{P}}(f) \leq \overline{\mathbf{E}}_{\underline{P}}(f) \leq \overline{P}(f);$$

recall that $-\underline{P}(-f) = \overline{P}(f)$ by definition. Hence, $\underline{P}(f) = \underline{\mathbf{E}}_{\underline{P}}(f) = \overline{\mathbf{E}}_{\underline{P}}(f) = \mathbf{E}_{\underline{P}}(f)$.

(ii). If $f \in \operatorname{dom}\underline{P}$ and $\underline{P}(f) = \sup f$, then

$$\sup f = \underline{P}(f) \leq \underline{\mathbf{E}}_{\underline{P}}(f) \leq \overline{\mathbf{E}}_{\underline{P}}(f) \leq \sup f,$$

and hence, also $\underline{P}(f) = \underline{\mathbf{E}}_{\underline{P}}(f) = \overline{\mathbf{E}}_{\underline{P}}(f) = \mathbf{E}_{\underline{P}}(f)$.

(iii). If $\overline{P}(|f|) = 0$, then, since $\overline{P}(|f|) \geq \inf|f| \geq 0$, it follows that $\inf|f| = 0$, or equivalently, $\sup -|f| = \underline{P}(-|f|)$. Now, apply (ii) to find that $-|f|$ belongs to $\operatorname{dom}\mathbf{E}_{\underline{P}}$, and $\mathbf{E}_{\underline{P}}(-|f|) = 0$, and hence, $\mathbf{E}_{\underline{P}}(|f|) = -\mathbf{E}_{\underline{P}}(-|f|) = 0$ too. For any gamble g on X such that $|g| \leq |f|$, we find that

$$0 = \mathbf{E}_{\underline{P}}(-|f|) \leq \underline{\mathbf{E}}_{\underline{P}}(-|g|) \leq \underline{\mathbf{E}}_{\underline{P}}(g) \leq \overline{\mathbf{E}}_{\underline{P}}(g) \leq \overline{\mathbf{E}}_{\underline{P}}(|g|) \leq \mathbf{E}_{\underline{P}}(|f|) = 0.$$

Therefore, $\underline{\mathbf{E}}_{\underline{P}}(g) = \overline{\mathbf{E}}_{\underline{P}}(g) = 0$, so g belongs to $\operatorname{dom}\mathbf{E}_{\underline{P}}$, and $\mathbf{E}_{\underline{P}}(g) = 0$.

(iv). $\underline{\mathbf{E}}_{\underline{P}}$ is coherent, so $\underline{\mathbf{E}}_{\underline{P}}(a) = \overline{\mathbf{E}}_{\underline{P}}(a) = \mathbf{E}_{\underline{P}}(a) = a$ for any $a \in \mathbb{R}(X)$.

(v). Let $\alpha_1, \ldots, \alpha_n, \alpha'_1, \ldots, \alpha'_m$ be non-negative reals, and $h_1, \ldots, h_n, h'_1, \ldots, h'_m$ gambles in $\operatorname{dom}\mathbf{E}_{\underline{P}}$. By the coherence of the natural extension, and the self-conjugacy of the linear extension, it follows from Theorem 3.5(vi)&(v) on p. 55 that

$$\underline{\mathbf{E}}_{\underline{P}}\left(\sum_{i=1}^{n}\alpha_i h_i - \sum_{i=1}^{n}\alpha'_i h'_i\right) \geq \sum_{i=1}^{n}\alpha_i \underline{\mathbf{E}}_{\underline{P}}(h_i) + \sum_{i=1}^{m}\alpha'_i \underline{\mathbf{E}}_{\underline{P}}(-h'_i)$$

and since $\underline{\mathbf{E}}_{\underline{P}}(h_i) = \overline{\mathbf{E}}_{\underline{P}}(h_i)$, and also $\underline{\mathbf{E}}_{\underline{P}}(-h'_i) = \mathbf{E}_{\underline{P}}(-h'_i) = -\mathbf{E}_{\underline{P}}(h'_i) = -\underline{\mathbf{E}}_{\underline{P}}(h'_i) = \overline{\mathbf{E}}_{\underline{P}}(-h'_i)$,

$$= \sum_{i=1}^{n}\alpha_i \overline{\mathbf{E}}_{\underline{P}}(h_i) + \sum_{i=1}^{m}\alpha'_i \overline{\mathbf{E}}_{\underline{P}}(-h'_i) \geq \overline{\mathbf{E}}_{\underline{P}}\left(\sum_{i=1}^{n}\alpha_i h_i - \sum_{i=1}^{m}\alpha'_i h'_i\right)$$

Now use $\underline{\mathbf{E}}_{\underline{P}} \leq \overline{\mathbf{E}}_{\underline{P}}$. So, $\operatorname{dom}\mathbf{E}_{\underline{P}}$ is a linear space. It is uniformly closed, because of Lemma 3.14 on p. 59.

(vi). It suffices to show that $|f|$ belongs to $\operatorname{dom}\mathbf{E}_{\underline{P}}$ whenever f belongs to $\operatorname{dom}\mathbf{E}_{\underline{P}}$—indeed, $f \vee g = (f + g + |f - g|)/2$ and $f \wedge g = (f + g - |f - g|)/2$, and

by (v) we already have that dom $\mathbf{E}_{\underline{P}}$ is a uniformly closed linear space.

Observe that, by the 2-monotonicity of $\underline{\mathbf{E}}_{\underline{P}}$,

$$\underline{\mathbf{E}}_{\underline{P}}(f \vee -f) + \underline{\mathbf{E}}_{\underline{P}}(f \wedge -f) \geq \underline{\mathbf{E}}_{\underline{P}}(f) + \underline{\mathbf{E}}_{\underline{P}}(-f),$$

or equivalently, since $f \vee -f = |f|$ and $f \wedge -f = -|f|$, and $\underline{\mathbf{E}}_{\underline{P}}(-\bullet) = -\overline{\mathbf{E}}_{\underline{P}}(\bullet)$,

$$\overline{\mathbf{E}}_{\underline{P}}(f) - \underline{\mathbf{E}}_{\underline{P}}(f) \geq \overline{\mathbf{E}}_{\underline{P}}(|f|) - \underline{\mathbf{E}}_{\underline{P}}(|f|).$$

So, if f belongs to dom $\mathbf{E}_{\underline{P}}$, then the left hand side is zero. By the coherence of $\underline{\mathbf{E}}_{\underline{P}}$, the right hand side is non-negative, and hence, the inequality implies that the right hand side must be zero too. But this means that $|f|$ belongs to dom $\mathbf{E}_{\underline{P}}$. □

Note that for just about any notion of integrability in the literature, the set of integrable bounded functions is at least uniformly closed: for instance, Darboux [14, Théorème V (second one), p. 82] proves this result for Riemann integrability.

We now give some conditions under which $\underline{P}$-integrals of two gambles are equal.

Proposition 4.19. *Let $\underline{P}$ be a lower prevision on X that avoids sure loss, let f and g be gambles on X, and define $N := \{x \in \mathcal{X} : f(x) \neq g(x)\}$. The following statements hold.*

(i) If f is $\underline{P}$-integrable, and $\overline{\mathbf{E}}_{\underline{P}}(I_N) = 0$, then g is $\underline{P}$-integrable and $\mathbf{E}_{\underline{P}}(f) = \mathbf{E}_{\underline{P}}(g)$.

(ii) If f and g are $\underline{P}$-integrable, $\underline{\mathbf{E}}_{\underline{P}}(I_N) = 0$, and $\underline{\mathbf{E}}_{\underline{P}}$ is 2-monotone, then $\mathbf{E}_{\underline{P}}(f) = \mathbf{E}_{\underline{P}}(g)$.

Proof. Define $\lambda := \sup|f - g| \geq 0$, and note that $|f - g| \leq \lambda I_N$.

(i). By the coherence of $\underline{\mathbf{E}}_{\underline{P}}$, Theorem 3.5(iv)&(vi) on p. 55:

$$0 \leq \overline{\mathbf{E}}_{\underline{P}}(|f - g|) \leq \lambda \overline{\mathbf{E}}_{\underline{P}}(I_N) = 0.$$

Applying Theorem 3.5(ix), it follows that $\underline{\mathbf{E}}_{\underline{P}}(g) = \mathbf{E}_{\underline{P}}(f) = \overline{\mathbf{E}}_{\underline{P}}(g)$.

(ii). If $\underline{\mathbf{E}}_{\underline{P}}$ is 2-monotone, then by Proposition 4.18(vi), dom $\mathbf{E}_{\underline{P}}$ is a linear lattice. Therefore, if f and g are $\underline{P}$-integrable, then also $|f - g|$ is $\underline{P}$-integrable,

and hence, by the coherence of $\underline{\mathbf{E}}_{\underline{P}}$, Theorem 3.5(iv)&(vi) on p. 55:

$$0 \leq \overline{\mathbf{E}}_{\underline{P}}(|f-g|) = \underline{\mathbf{E}}_{\underline{P}}(|f-g|) \leq \lambda \underline{\mathbf{E}}_{\underline{P}}(I_N) = 0.$$

Again applying Theorem 3.5(ix), it follows that $\mathbf{E}_{\underline{P}}(g) = \mathbf{E}_{\underline{P}}(f)$. □

Let's end with a few obvious, but important, results.

Corollary 4.20. *Let P be a linear prevision. Then $\mathbf{E}_P$ is an extension of P.*

Proof. Immediate from Proposition 4.18(i). □

If $\operatorname{dom} \underline{P} \subseteq \mathcal{K} \subseteq \operatorname{dom} \mathbf{E}_{\underline{P}}$, then we denote $\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{K}}$ also by $\mathbf{E}_{\underline{P}}^{\mathcal{K}}$ to emphasise that it is a linear prevision.

Corollary 4.21. *Let P be a linear prevision, and let $\operatorname{dom} P \subseteq \mathcal{J} \subseteq \mathcal{K} \subseteq \operatorname{dom} \mathbf{E}_P$. Then*

$$\mathbf{E}^{\mathcal{K}}_{\mathbf{E}_P^{\mathcal{J}}}(f) = \mathbf{E}_P^{\mathcal{K}}(f) \text{ for all } f \in \mathcal{K}, \text{ and } \mathbf{E}_P^{\mathcal{J}}(f) = \mathbf{E}_P^{\mathcal{K}}(f) \text{ for all } f \in \mathcal{J}.$$

Proof. Immediate from Corollary 4.9, once observed that $\mathbf{E}_P^{\mathcal{K}} = \underline{\mathbf{E}}_P^{\mathcal{K}}$, $\mathbf{E}_P^{\mathcal{J}} = \underline{\mathbf{E}}_P^{\mathcal{J}}$, and $\mathbf{E}^{\mathcal{K}}_{\mathbf{E}_P^{\mathcal{J}}} = \underline{\mathbf{E}}^{\mathcal{K}}_{\mathbf{E}_P^{\mathcal{J}}}$, by Corollary 4.4. □

4.3 Examples of Natural and Linear Extension

4.3.1 Uncertainty Models and Equivalence

We have already seen many examples where mathematical constructs for modelling uncertainty—briefly, uncertainty models—are identified with lower previsions: a probability charge μ is identified with its induced probability $\mathbf{P}_\mu$; a nested set function μ is identified with its induced lower probability $\underline{\mathbf{P}}_\mu$, its induced upper probability $\overline{\mathbf{P}}_\mu$, or its induced probability $\mathbf{P}_\mu$; a 2-monotone set function ν is identified with its induced lower probability $\underline{\mathbf{P}}_\nu$, as are minimum preserving set functions and necessity measures; a cumulative distribution function F is identified with its induced probability $\mathbf{P}_F$; and a p-box (F_*, F^*) with its induced lower prevision $\underline{\mathbf{P}}_{(F_*,F^*)}$. Through this identification it was proved under what conditions these models are coherent: see Theorem 3.21 on p. 67, Proposition 3.24 on p. 70, Theorem 3.27 on p. 72, Theorem 3.31 on p. 76, Proposition 3.38 on p. 80, Theorem 3.46 on p. 85,

Lemma 3.58 on p. 93, and Theorem 3.59 on p. 93. In this section we formalise the idea of identifying mathematical constructs with lower previsions. This provides us with a unifying language for studying their relations.

Again, only because all these uncertainty models can be interpreted as lower previsions, it seems logical to describe all relations between them through the language of lower previsions. So, the definition below is not the most general one we can think of—without doubt there are uncertainty models not interpretable as lower previsions—but it is general enough for the purpose of this work (see for instance De Cooman [20] for a very general approach).

Definition 4.22. An *uncertainty structure* is a pair $(\mathcal{S}, \mathfrak{s})$, where $\mathcal{S}$ is a non-empty set, and $\mathfrak{s}$ maps each element of $\mathcal{S}$ to a lower prevision on X. Elements s of $\mathcal{S}$ are called *uncertainty models*. An uncertainty model s is said to *avoid sure loss* whenever $\mathfrak{s}(s)$ avoids sure loss; the set of uncertainty models that avoid sure loss is denoted by $\text{asl}(\mathcal{S})$. If s avoids sure loss, then $\underline{\mathbf{E}}_{\mathfrak{s}(s)}$ exists, is called the *natural extension* of s, and is also denoted by $\underline{\mathbf{E}}_s$ whenever $\mathfrak{s}$ is clear from the context. Similarly, $\mathbf{E}_{\mathfrak{s}(s)}$ is called the *linear extension* of s and is simply denoted by $\mathbf{E}_s$. A gamble f is called *s-integrable* whenever it is $\underline{\mathbf{E}}_s$-integrable and in such a case we call $\mathbf{E}_s(f)$ the *s-integral* of f. An uncertainty model s is said to be *coherent* whenever $\mathfrak{s}(s)$ is coherent, that is, whenever $\mathfrak{s}(s) = \underline{\mathbf{E}}_s$ on $\text{dom}\,\mathfrak{s}(s)$; the set of coherent models in $\mathcal{S}$ is denoted by $\text{coh}(\mathcal{S})$. An uncertainty structure must satisfy the following conditions which guarantee that the idea of natural extension pulls back to $\mathcal{S}$ in a very simple way.

(i) Every coherent uncertainty model $s \in \mathcal{S}$ is uniquely determined by the natural extension of $\mathfrak{s}(s)$. That is, for every coherent s, $s' \in \mathcal{S}$ it holds that

$$\underline{\mathbf{E}}_s = \underline{\mathbf{E}}_{s'} \iff s = s'. \tag{4.4}$$

So, $\mathfrak{s}$ must be one-to-one between $\text{coh}(\mathcal{S})$ and $\{\underline{\mathbf{E}}_s : s \in \text{coh}(\mathcal{S})\}$.

(ii) For every uncertainty model $s \in \mathcal{S}$ that avoids sure loss there is a coherent uncertainty model $s^* \in \text{coh}(\mathcal{S})$ such that $\underline{\mathbf{E}}_s = \underline{\mathbf{E}}_{s^*}$. So, $\{\underline{\mathbf{E}}_s : s \in \text{asl}(\mathcal{S})\} = \{\underline{\mathbf{E}}_s : s \in \text{coh}(\mathcal{S})\}$

For every $s \in \text{asl}(\mathcal{S})$ there is a unique $s^* \in \text{coh}(\mathcal{S})$ such that $\underline{\mathbf{E}}_s = \underline{\mathbf{E}}_{s^*}$ and this uncertainty model s^* is called the *natural extension in $\mathcal{S}$* of s.

Let's give a few examples of uncertainty structures.

- Let $\mathcal{K} \subseteq \mathcal{L}(X)$. Any set of lower previsions defined on a common domain $\mathcal{K}$ is an uncertainty structure when equipped with the identity map.

- Let $\mathcal{F}$ be a field. The set $\mathcal{P}(\mathcal{F})$ of probability charges on $\mathcal{F}$ equipped with $\mathbf{P}_\bullet$ is an uncertainty structure; see Definition 3.17 on p. 64: $\mathbf{P}_\mu$ is the induced probability of a probability charge μ. Indeed, all of its elements are coherent:

$$\text{coh}(\mathcal{P}(\mathcal{F})) = \text{asl}(\mathcal{P}(\mathcal{F})) = \mathcal{P}(\mathcal{F}). \tag{4.5}$$

  Hence, condition (ii) is satisfied. Condition (i) is also satisfied. The natural extension $\underline{\mathbf{E}}_\mu$ coincides with $\mathbf{E}_\mu^{\text{span}(\mathcal{F})}$ on $\text{span}(\mathcal{F})$. This will follow from the fact that every $\mathcal{F}$-simple gamble f is μ-integrable with μ-integral given by $\mathbf{E}_\mu^{\text{span}(\mathcal{F})}(f)$ (see for instance Proposition 4.28 on p. 112).

- Let $\mathcal{A}$ be a nested collection of subsets of $\mathcal{X}$, and suppose that $\emptyset \in \mathcal{A}$ and $\mathcal{X} \in \mathcal{A}$. The set of nested set functions on $\mathcal{A}$ equipped with either $\underline{\mathbf{P}}_\bullet$, $\overline{\mathbf{P}}_\bullet$, or $\mathbf{P}_\bullet$, are three different uncertainty structures.

- The set of 2-monotone set functions on a field $\mathcal{F}$ is an uncertainty structure when equipped with $\underline{\mathbf{P}}_\bullet$; once more see Definition 3.17 on p. 64: $\underline{\mathbf{P}}_\nu$ is the induced lower probability of a 2-monotone set function ν. Again, all of its elements are coherent so condition (ii) is trivially satisfied, as well as condition (i). The natural extension $\underline{\mathbf{E}}_\nu$ coincides with $\underline{\mathbf{E}}_\nu^{\text{span}(\mathcal{F})}$ on $\text{span}(\mathcal{F})$, as we shall prove further on. In case ν is a probability charge, it holds that $\underline{\mathbf{E}}_{\mathbf{P}_\nu} = \underline{\mathbf{E}}_{\underline{\mathbf{P}}_\nu}$.

- Let $\mathcal{F}$ be an ample field. The set of necessity measures on $\mathcal{F}$ is an uncertainty structure when equipped with $\underline{\mathbf{P}}_\bullet$. The set of possibility measures on $\mathcal{F}$ is an uncertainty structure when equipped with $\overline{\mathbf{P}}_\bullet$.

- P-boxes and cumulative distribution functions are uncertainty structures too when equipped with $\underline{\mathbf{P}}_{(\bullet,\bullet)}$.

Let's now use uncertainty structures in order to identify various relations between different uncertainty models.

Definition 4.23. Let $(\mathcal{S}, \mathfrak{s})$ and $(\mathcal{T}, \mathfrak{t})$ be two uncertainty structures, and let $s \in \mathcal{S}$ and $t \in \mathcal{T}$. We say that s is *equivalent* to t, and we write $s \sim t$, whenever

(i) both s and t incur sure loss, or

(ii) their natural extensions coincide: $\underline{\mathbf{E}}_s = \underline{\mathbf{E}}_t$.

Let $(\mathcal{S}, \mathfrak{s})$ and $(\mathcal{T}, \mathfrak{t})$ be two uncertainty structures. Then $\mathcal{S}$ is said to be *weaker* than $\mathcal{T}$ and we write $\mathcal{S} \preccurlyeq \mathcal{T}$ whenever for every $s \in \mathrm{coh}(\mathcal{S})$ there is a $t \in \mathrm{coh}(\mathcal{T})$ such that $s \sim t$. Finally, $\mathcal{S}$ and $\mathcal{T}$ are said to be *equivalent* and we write $\mathcal{S} \sim \mathcal{T}$ if each is weaker than the other one, that is, whenever $\mathcal{S} \preccurlyeq \mathcal{T}$ and $\mathcal{T} \preccurlyeq \mathcal{S}$.

Clearly, $\sim$ is the equivalence relation that corresponds to the partial semi-ordering $\preccurlyeq$ on the set of uncertainty structures. This partial semi-ordering has a unique maximal element (up to equivalence), namely, the set of all coherent lower previsions on $\mathcal{L}(X)$ equipped with the identity map. Every other uncertainty structure is equivalent to some subset of this set, and the partial semi-ordering corresponds to set inclusion.

Sometimes, we want to express the relation between uncertainty models restricting to only a subset of all gambles.

Definition 4.24. Let $\mathcal{K} \subseteq \mathcal{L}(X)$. Let $(\mathcal{S}, \mathfrak{s})$ and $(\mathcal{T}, \mathfrak{t})$ be two uncertainty structures, and let $s \in \mathcal{S}$ and $t \in \mathcal{T}$. We say that s is *equivalent* to t on $\mathcal{K}$, and we write $s \sim_{\mathcal{K}} t$, whenever

(i) both s and t incur sure loss, or

(ii) their natural extensions coincide on $\mathcal{K}$: $\underline{\mathbf{E}}_s(f) = \underline{\mathbf{E}}_t(f)$ for all gambles f in $\mathcal{K}$.

$\mathcal{S}$ is said to be *weaker* than $\mathcal{T}$ on $\mathcal{K}$, and we write $\mathcal{S} \preccurlyeq_{\mathcal{K}} \mathcal{T}$, whenever for every $s \in \mathrm{coh}(\mathcal{S})$ there is a $t \in \mathrm{coh}(\mathcal{T})$ such that $s \sim_{\mathcal{K}} t$. Finally, $\mathcal{S}$ and $\mathcal{T}$ are said to be *equivalent* on $\mathcal{K}$ and we write $\mathcal{S} \sim_{\mathcal{K}} \mathcal{T}$ if both are weaker on $\mathcal{K}$ than the other one, that is, whenever $\mathcal{S} \preccurlyeq_{\mathcal{K}} \mathcal{T}$ and $\mathcal{T} \preccurlyeq_{\mathcal{K}} \mathcal{S}$.

For example, by definition, any two uncertainty structures $(\mathcal{S}, \mathfrak{s})$ and $(\mathcal{T}, \mathfrak{t})$ are at least equivalent on the set $\mathbb{R}(X)$ of constant gambles on X.

4.3.2 Integration of Probability Charges

Let's now consider s-integrals for some special uncertainty structures. First of all, we take a look at the set of probability charges defined on a field $\mathcal{F}$ equipped with $\mathbf{P}_\bullet$. For a probability charge μ, the set of all $\mathbf{P}_\mu$-integrable gambles, the domain of $\mathbf{E}_{\mathbf{P}_\mu}$, is denoted by $\mathcal{L}_\mu(X)$. Using the notation introduced in Section 4.3.1, the set of all μ-integrable gambles, which is the domain of $\mathbf{E}_\mu$, is denoted by $\mathcal{L}_\mu(X)$.

What can we say about $\mathcal{L}_\mu(X)$? Applying Proposition 4.18(i)&(v) with $\underline{P}$ equal to $\mathbf{P}_\mu$, it follows that $\mathcal{L}_\mu(X)$ contains at least the uniform closure of span($\mathcal{F}$). Gambles in this uniform closure will be called *$\mathcal{F}$-measurable*, and the set of $\mathcal{F}$-measurable gambles is denoted by

$$\mathcal{L}_\mathcal{F}(X) := \text{cl}(\text{span}(\mathcal{F})). \tag{4.6}$$

This characterisation is equivalent to many other characterisations of measurability known in the literature; see for instance Greco [38], Bhaskara Rao and Bhaskara Rao [9], Walley [86], Denneberg [28], and Janssen [45]. Bhaskara Rao and Bhaskara Rao [9] called it $\mathcal{F}$-continuity. In Hildebrandt [42, Sect. 1(f), p. 869] and Walley [86, Section 3.2.1, p. 129], the definition of $\mathcal{F}$-measurability is stronger than our definition of $\mathcal{F}$-measurability given below, unless $\mathcal{F}$ is a σ-field. However, if $\mathcal{F}$ is a field, but not a σ-field, Hildebrandt's [42] and Walley's [86] set of $\mathcal{F}$-measurable gambles is not even a linear space; therefore, I prefer the more general definition. In case $\mathcal{F}$ is a σ-field, our definition reduces to the classical definition of $\mathcal{F}$-measurability, also called Borel-measurability, for gambles.

The characterisation (A) in the definition below is due to Greco [38], and the characterisation (B) is due to Janssen [45], who also established equivalence with Greco's definition. Below, we give a shorter proof.

Definition 4.25. Let $\mathcal{F}$ be a field on X and let f be a gamble on X. Then the following conditions are equivalent; if any (hence all) of them are satisfied, we say that f is *$\mathcal{F}$-measurable*.

(A) For any $a \in \mathbb{R}$ and any $\epsilon > 0$ there is an $A \in \mathcal{F}$ such that

$$\{x \in \mathcal{X} \colon f(x) \geq a\} \supseteq A \supseteq \{x \in \mathcal{X} \colon f(x) \geq a + \epsilon\}.$$

(B) There is a sequence f_n of $\mathcal{F}$-simple gambles that converges uniformly to f, that is, $\lim_{n\to\infty} \sup\left|f - f_n\right| = 0$.

The set of all $\mathcal{F}$-measurable gambles is denoted by $\mathcal{L}_{\mathcal{F}}(X)$.

Proof of equivalence. (A) $\Longrightarrow$ (B). Assume that condition (A) is satisfied. Let $\epsilon > 0$. Then (B) is established if we can find an $\mathcal{F}$-simple gamble g such that $\sup\left|f - g\right| \leq \epsilon$.

Let $a_0, \ldots, a_n$ be a finite sequence of real numbers such that $a_0 < \inf f$, $0 < a_{i+1} - a_i < \frac{\epsilon}{3}$ for $i \in \{0, \ldots, n-1\}$ and $\sup f < a_n$. Define $A_i = \{x \in \mathcal{X} \colon f(x) \geq a_i\}$ for $i \in \{0, \ldots, n\}$. By (A) there is a sequence $B_0, \ldots, B_{n-1}$ of members of $\mathcal{F}$ such that

$$A_0 \supseteq B_0 \supseteq A_1 \supseteq B_1 \cdots \supseteq A_{n-1} \supseteq B_{n-1} \supseteq A_n.$$

With $a_{-1} := 0$, define the $\mathcal{F}$-simple gamble

$$g := \sum_{i=0}^{n-1} (a_i - a_{i-1}) I_{B_i}.$$

We show that g has the desired property: $\left|f(x) - g(x)\right| < \epsilon$ for any $x \in \mathcal{X}$. Indeed, let $x \in \mathcal{X}$. First observe that by construction of $a_0, \ldots, a_n$ there is a unique $j \in \{1, \ldots, n\}$ such that $a_{j-1} \leq f(x) < a_j$. By the construction of the sequence $B_0, \ldots, B_{n-1}$ it holds that $I_{B_i}(x) = 1$ for $i < j$ and $I_{B_i}(x) = 0$ for $i > j$ (for $i = j$ both values are possible). We can conclude that

$$\begin{aligned}
\left|f(x) - g(x)\right| &= \left|f(x) - \sum_{i=0}^{n-1} (a_i - a_{i-1}) I_{B_i}\right| \\
&\leq \left|f(x) - \sum_{i=0}^{j-1} (a_i - a_{i-1})\right| + \left|a_j - a_{j-1}\right| \\
&= \left|f(x) - a_{j-1}\right| + \left|a_j - a_{j-1}\right| \\
&\leq \left|f(x) - a_j\right| + 2\left|a_j - a_{j-1}\right| \\
&< \epsilon,
\end{aligned}$$

which establishes the first part of the proof.

(B) $\Longrightarrow$ (A). Conversely, let $a \in \mathbb{R}$ and $\epsilon > 0$, and suppose there is a sequence of $\mathcal{F}$-simple gambles such that $\sup\left|f - f_n\right|$ converges to zero. Then

there is an $n_\epsilon \in \mathbb{N}$ such that $\sup\left|f - f_{n_\epsilon}\right| < \frac{\epsilon}{2}$. We find that

$$\{x \in \mathcal{X}\colon f(x) \geq a\} \supseteq \{x \in \mathcal{X}\colon f_{n_\epsilon}(x) \geq a + \tfrac{\epsilon}{2}\} \supseteq \{x \in \mathcal{X}\colon f(x) \geq a + \epsilon\}.$$

Now simply observe that $\{x \in \mathcal{X}\colon f_{n_\epsilon}(x) \geq a + \frac{\epsilon}{2}\}$ belongs to $\mathcal{F}$, since f_{n_ϵ} is $\mathcal{F}$-simple. □

We note that for a σ-field $\mathcal{F}$, $\mathcal{F}$-measurability is equivalent to the well-known notion of Borel-measurability, as we prove now; see for instance Kallenberg [48, Lemma 1.11, p. 7]. As the proof is short, it is given below.

Proposition 4.26. *Let $\mathcal{F}$ be a σ-field on X. A gamble f on X is $\mathcal{F}$-measurable if and only if for any $a \in \mathbb{R}$ the set $\{x \in \mathcal{X}\colon f(x) > a\}$ belongs to $\mathcal{F}$, or equivalently, if and only if for any $a \in \mathbb{R}$ the set $\{x \in \mathcal{X}\colon f(x) \geq a\}$ belongs to $\mathcal{F}$.*

Proof. "if". Simply take $A := \{x \in \mathcal{X}\colon f(x) > a\}$ or $A := \{x \in \mathcal{X}\colon f(x) \geq a\}$ in Definition 4.25(A).

"only if". Suppose f is *$\mathcal{F}$-measurable*. By Definition 4.25(A) there is a sequence A_n in $\mathcal{F}$ such that

$$\{x \in \mathcal{X}\colon f(x) \geq a + \tfrac{1}{n+1}\} \supseteq A_n \supseteq \{x \in \mathcal{X}\colon f(x) \geq a + \tfrac{2}{n+1}\}$$

Taking the countable union over $n \in \mathbb{N}$, we find that

$$\{x \in \mathcal{X}\colon f(x) > a\} \supseteq \cup_{n\in\mathbb{N}} A_n \supseteq \{x \in \mathcal{X}\colon f(x) > a\},$$

which means that $\{x \in \mathcal{X}\colon f(x) > a\} = \cup_{n\in\mathbb{N}} A_n$. Since $\mathcal{F}$ is a σ-field, it is closed under countable union, and hence, $\cup_{n\in\mathbb{N}} A_n$ belongs to $\mathcal{F}$. This establishes the proposition.

For the other part, construct the sequence A_n in $\mathcal{F}$ such that

$$\{x \in \mathcal{X}\colon f(x) \geq a - \tfrac{2}{n+1}\} \supseteq A_n \supseteq \{x \in \mathcal{X}\colon f(x) \geq a - \tfrac{1}{n+1}\},$$

and take countable intersection to arrive at the desired result. □

For ample fields—fields closed under arbitrary union—we have the following necessary and sufficient condition.

Proposition 4.27. *Let $\mathcal{F}$ be an ample field on X. A gamble f on X is $\mathcal{F}$-measurable if and only if for any $a \in \mathbb{R}$ the set $\{x \in \mathcal{X}\colon f(x) = a\}$ belongs to $\mathcal{F}$, or equivalently, if and only if f is constant on the atoms of $\mathcal{F}$.*

Proof. The two conditions are equivalent, since $\{x \in \mathcal{X}\colon f(x) = a\}$ belongs to $\mathcal{F}$ if and only if $\{x \in \mathcal{X}\colon f(x) = a\}$ is a union of atoms of $\mathcal{F}$, by Theorem 3.51.

"if". Suppose that for any $a \in \mathbb{R}$ the set $\{x \in \mathcal{X}\colon f(x) = a\}$ belongs to $\mathcal{F}$. Since

$$\{x \in \mathcal{X}\colon f(x) \geq b\} = \bigcup_{a \geq b} \{x \in \mathcal{X}\colon f(x) = a\}$$

for any $b \in \mathbb{R}$, and $\mathcal{F}$ is closed under arbitrary union, it follows that $\{x \in \mathcal{X}\colon f(x) \geq b\}$ belongs to $\mathcal{F}$ for any $b \in \mathbb{R}$. Now apply Proposition 4.26.

"only if". If f is $\mathcal{F}$-measurable, then, by Proposition 4.26, for any $a \in \mathbb{R}$, both $\{x \in \mathcal{X}\colon f \geq a\}$ and $\{x \in \mathcal{X}\colon f > a\}$ belong to $\mathcal{F}$. But, $\mathcal{F}$ is a field, and hence, it is closed under complementation and intersection. We find that also

$$\{x \in \mathcal{X}\colon f \geq a\} \cap \complement\{x \in \mathcal{X}\colon f > a\} = \{x \in \mathcal{X}\colon f = a\}$$

belongs to $\mathcal{F}$. □

Proposition 4.28. *Let $\mathcal{F}$ be a field on $\mathcal{X}$ and let μ be a probability charge on $\mathcal{F}$. Then any $\mathcal{F}$-measurable gamble is μ-integrable, that is,*

$$\operatorname{span}(\mathcal{F}) \subseteq \mathcal{L}_{\mathcal{F}}(X) \subseteq \mathcal{L}_{\mu}(X). \tag{4.7}$$

Proof. The inclusion $\operatorname{span}(\mathcal{F}) \subseteq \mathcal{L}_{\mathcal{F}}(X)$ follows easily from the definition of $\mathcal{F}$-measurability, Definition 4.25(B) on p. 109.

By Proposition 4.18(i) all gambles in $\operatorname{span}(\mathcal{F})$ are μ-integrable. By Proposition 4.18(v) all gambles in the uniform closure of this set are also μ-integrable. But, by Definition 4.25 these are exactly the $\mathcal{F}$-measurable gambles. □

4.3.3 Linear Previsions and Probability Charges

Through $\mathcal{F}$-measurability, we shall now establish a fairly general correspondence between linear previsions and probability charges; similar results were proved by for instance, Hildebrandt [42], Dunford and Schwartz [30, Chapter VI, p. 492 ff.], and Bhaskara Rao and Bhaskara Rao [9, Theorem 4.7.4, p. 135]. This will also be a first example of equivalent uncertainty structures.

Its impact is that, in Section 4.4, it will allow us to obtain the natural extension by lower and upper integrals, introduced further on in Section 4.3.5.

Definition 4.29. We can define a map from the set of linear previsions to the set of probability charges. Let P be a linear prevision, and let $\mathcal{F}$ be a field such that $I_{\mathcal{F}} \subseteq \operatorname{dom} P$. The restriction of P to $I_{\mathcal{F}}$ corresponds to a probability charge on $\mathcal{F}$:

$$\boldsymbol{\mu}_P^{\mathcal{F}}(A) := P(I_A) \text{ for any } A \in \mathcal{F}. \tag{4.8}$$

Conversely, we can also define a map from the set of probability charges to the set of linear previsions. Let μ be a probability charge defined on the field $\mathcal{F}$, and let $\mathcal{K}$ be any set of μ-integrable gambles: $\mathcal{K} \subseteq \mathcal{L}_\mu(X)$. Then the mapping $\mathbf{E}_\mu^{\mathcal{K}}$ defined by

$$\mathbf{E}_\mu^{\mathcal{K}}(f) := \mathbf{E}_\mu(f) \text{ for any } f \in \mathcal{K}, \tag{4.9}$$

is a linear prevision.

Proof that these maps are well-defined. Let P be a linear prevision, and let $\mathcal{F}$ be a field such that $I_{\mathcal{F}} \subseteq \operatorname{dom} P$. To see that $\boldsymbol{\mu}_P^{\mathcal{F}}$ is a probability charge, observe that the conditions of Definition 3.16 follow from the linearity of P.

Let μ be a probability charge on $\mathcal{F}$, and let $\mathcal{K} \subseteq \mathcal{L}_\mu(X)$. To see that $\mathbf{E}_\mu^{\mathcal{K}}$ is a linear prevision, observe that by its definition, it is a restriction of the linear extension $\mathbf{E}_\mu$, which is obviously a linear prevision. Therefore, by Lemma 3.9 also $\mathbf{E}_\mu^{\mathcal{K}}$ is a linear prevision. □

By Proposition 4.18(iv) ($\mathbb{R}(X) \subseteq \operatorname{dom} \mathbf{E}_P$) and Proposition 4.13, any linear prevision P has a unique coherent behavioural extension to a linear prevision on $\operatorname{dom} P \cup \mathbb{R}(X)$, through $P(a) := a$ for all $a \in \mathbb{R}(X)$. So, we can always assume that there is a field $\mathcal{F}$ such that $I_{\mathcal{F}} \subseteq \operatorname{dom} P$.

Recall that $\mathbf{E}_\mu$ is a short notation for $\mathbf{E}_{\mathbf{P}_\mu}$: the linear extension of the probability $\mathbf{P}_\mu$ induced by μ. Hence, $\mathbf{E}_\mu^{\mathcal{K}}$ is the linear extension of $\mathbf{P}_\mu$ restricted to $\mathcal{K}$. If $\operatorname{dom} \underline{P} \subseteq \mathcal{K}$, then $\mathbf{E}_\mu^{\mathcal{K}}$ is equal to $\mathbf{E}_{\mathbf{P}_\mu}^{\mathcal{K}}$, defined on p. 105.

Also note that for $\mathcal{K} = \operatorname{span}(\mathcal{F})$, the linear prevision $\mathbf{E}_\mu^{\operatorname{span}(\mathcal{F})}$, as defined in Definition 4.29 above by Eq. (4.9), coincides with the linear prevision $\mathbf{E}_\mu^{\operatorname{span}(\mathcal{F})} := \mathrm{D}\int \bullet \, \mathrm{d}\mu$, as defined in Lemma 3.20 on p. 67 by Eq. (3.10). Indeed, by Lemma 3.20, $\mathrm{D}\int \bullet \, \mathrm{d}\mu$, defined on the set $\operatorname{span}(\mathcal{F})$ of $\mathcal{F}$-simple gambles, is a coherent behavioural extension of $\mathbf{P}_\mu$. But, by Proposition 4.28, any

$\mathcal{F}$-simple gamble is μ-integrable: $\operatorname{span}(\mathcal{F}) \subseteq \mathcal{L}_\mu(X) = \operatorname{dom} \mathbf{E}_{\mathbf{P}_\mu}$. Hence, by Proposition 4.13, $\mathrm{D}\int \bullet \, \mathrm{d}\mu$ and $\mathbf{E}_{\mathbf{P}_\mu}$ coincide on $\operatorname{span}(\mathcal{F})$. Thus, $\mathrm{D}\int \bullet \, \mathrm{d}\mu = \mathbf{E}_{\mathbf{P}_\mu}^{\operatorname{span}(\mathcal{F})} = \mathbf{E}_\mu^{\operatorname{span}(\mathcal{F})}$: this shows that our notation is consistent.

We can actually extend the Dunford integral to a much larger set of gambles, and in Theorem 4.62 we shall prove that $\mathrm{D}\int \bullet \, \mathrm{d}\mu = \mathbf{E}_\mu$ for this extension.

The two maps given above constitute equivalences between particular sets of probability charges and sets of linear previsions; a more general version of the theorem below can be found in Bhaskara Rao and Bhaskara Rao [9, Theorem 4.7.4, p. 135].

Theorem 4.30. *Let $\mathcal{F}$ be a field. Then $\boldsymbol{\mu}_\bullet^{\mathcal{F}}$ and $\mathbf{E}_\bullet^{\mathcal{L}_\mathcal{F}(X)}$ are onto and one-to-one between the set of linear previsions on $\mathcal{L}_\mathcal{F}(X)$ and the set of probability charges on $\mathcal{F}$.*

Proof. Let P be any linear prevision on $\mathcal{L}_\mathcal{F}(X)$, and let μ be any probability charge on $\mathcal{F}$.

In order to establish that the maps are onto and one-to-one, it suffices to show that

$$\boldsymbol{\mu}^{\mathcal{F}}_{\mathbf{E}_\mu^{\mathcal{L}_\mathcal{F}(X)}} = \mu \text{ and } \mathbf{E}^{\mathcal{L}_\mathcal{F}(X)}_{\boldsymbol{\mu}_P^{\mathcal{F}}} = P.$$

For any $A \in \mathcal{F}$, $\mathbf{E}_\mu^{\mathcal{L}_\mathcal{F}(X)}(I_A) = \mathbf{E}_{\mathbf{P}_\mu}(I_A)$, and applying Corollary 4.20, $\mathbf{E}_{\mathbf{P}_\mu}(I_A) = \mathbf{P}_\mu(A) = \mu(A)$; hence, the first equality holds. To see that the second equality also holds, let f be any $\mathcal{F}$-measurable gamble. By Definition 4.25 f can be uniformly approximated by a sequence f_n of $\mathcal{F}$-simple gambles. For any such $\mathcal{F}$-simple gamble $f_n = \sum_{j=1}^{m_n} a_{n,j} I_{A_{n,j}}$, it holds that

$$\begin{aligned}\mathbf{E}^{\mathcal{L}_\mathcal{F}(X)}_{\boldsymbol{\mu}_P^{\mathcal{F}}}(f_n) = \mathbf{E}^{\operatorname{span}(\mathcal{F})}_{\boldsymbol{\mu}_P^{\mathcal{F}}}(f_n) &= \mathrm{D}\int f_n \, \mathrm{d}\boldsymbol{\mu}_P^{\mathcal{F}} \\ &= \sum_{j=1}^{m_n} a_{n,j} \boldsymbol{\mu}_P^{\mathcal{F}}(A_{n,j}) = \sum_{j=1}^{m_n} a_{n,j} P(I_{A_{n,j}}) = P(f_n).\end{aligned}$$

Let's explain each equality. The first equality follows from Proposition 4.28, by which $\operatorname{dom} P \subseteq \operatorname{span}(\mathcal{F}) \subseteq \mathcal{L}_\mathcal{F}(X) \subseteq \mathcal{L}_\mu(X)$, and Corollary 4.21, which gives the desired equality. The next equality is simply an application of the definition of $\mathbf{E}^{\operatorname{span}(\mathcal{F})}_{\boldsymbol{\mu}_P^{\mathcal{F}}}$ given in Lemma 3.20, which agrees with Definition 4.29 as we explained before. The next equality is a direct application of the definition of the Dunford integral for $\mathcal{F}$-simple gambles; see Definition 3.19

on p. 65. Next, we apply the definition of $\boldsymbol{\mu}_P^{\mathcal{F}}$; see Definition 4.29. Finally, we invoke the linearity of P.

Since both $\mathbf{E}_{\boldsymbol{\mu}_P^{\mathcal{F}}}^{\mathcal{L}_{\mathcal{F}}(X)}$ and P are coherent previsions, they are continuous with respect to the topology of uniform convergence, so

$$\mathbf{E}_{\boldsymbol{\mu}_P^{\mathcal{F}}}^{\mathcal{L}_{\mathcal{F}}(X)}(f) = \lim_{n\to+\infty} \mathbf{E}_{\boldsymbol{\mu}_P^{\mathcal{F}}}^{\mathcal{L}_{\mathcal{F}}(X)}(f_n) = \lim_{n\to+\infty} P(f_n) = P(f).$$

This establishes the claim. □

Corollary 4.31. *For a given field $\mathcal{F}$, the set of probability charges on $\mathcal{F}$ (equipped with $\mathbf{P}_\bullet$) is equivalent to the set of linear previsions on $\mathcal{L}_{\mathcal{F}}(X)$ (equipped with the identity map).*

Proof. Using Theorem 4.30 it is easy to check the conditions of Definition 4.23. □

Corollary 4.32. *$\boldsymbol{\mu}_\bullet^{\wp(\mathcal{X})}$ and $\mathbf{E}_\bullet^{\mathcal{L}(X)}$ are onto and one-to-one maps between probability charges on $\wp(\mathcal{X})$ and linear previsions on $\mathcal{L}(X)$. Hence, the set $\mathcal{P}(\wp(\mathcal{X}))$ of all probability charges on $\wp(\mathcal{X})$ is equivalent to the set $\mathcal{P}(X)$ of all linear previsions on $\mathcal{L}(X)$.*

Proof. See Theorem 4.30 and observe that $\mathcal{L}_{\wp(\mathcal{X})}(X) = \mathcal{L}(X)$ whenever μ is defined on $\wp(\mathcal{X})$. □

Finally, we obtain the following new characterisation of self-conjugacy, which I find quite surprising.

Corollary 4.33. *Let $\mathcal{F}$ be a field on $\mathcal{X}$ and $\underline{P}$ be a coherent lower prevision defined on a symmetric domain ($\operatorname{dom}\underline{P} = -\operatorname{dom}\underline{P}$) such that*

$$I_{\mathcal{F}} \subseteq \operatorname{dom}\underline{P} \subseteq \mathcal{L}_{\mathcal{F}}(X). \tag{4.10}$$

Then $\underline{P}$ is self-conjugate if and only if its restriction to $I_{\mathcal{F}}$ corresponds to a probability charge on $\mathcal{F}$, that is, if and only if $\boldsymbol{\mu}_{\underline{P}}^{\mathcal{F}}$ is a probability charge.

Proof. "only if". If $\underline{P}$ is self-conjugate, it must be a coherent prevision, in which case its restriction to $I_{\mathcal{F}}$ corresponds to $\boldsymbol{\mu}_{\underline{P}}^{\mathcal{F}}$, which is a probability charge by Definition 4.29.

"if". Since $\underline{P}$ is coherent, it coincides with its natural extension $\underline{\mathbf{E}}_{\underline{P}}$ on $\operatorname{dom}\underline{P}$, by Proposition 4.6. The corollary is therefore established if we can show that $\underline{\mathbf{E}}_{\underline{P}}$ is self-conjugate on $\mathcal{L}_{\mathcal{F}}(X)$.

First, note that since the restriction of $\underline{P}$ to $I_{\mathcal{F}}$ corresponds to a probability charge, for any A in $\mathcal{F}$ it holds in particular that $\underline{P}(I_A) + \underline{P}(I_{\complement A}) = \underline{P}(I_{\mathcal{X}}) = 1$, and hence, also $\underline{\mathbf{E}}_{\underline{P}}(I_A) + \underline{\mathbf{E}}_{\underline{P}}(I_{\complement A}) = 1$. But this means that, by Theorem 3.5(iii),

$$\underline{\mathbf{E}}_{\underline{P}}(I_A) = 1 - \underline{\mathbf{E}}_{\underline{P}}(I_{\complement A}) = 1 - \underline{\mathbf{E}}_{\underline{P}}(1 - I_A) = -\underline{\mathbf{E}}_{\underline{P}}(-I_A) = \overline{\mathbf{E}}_{\underline{P}}(I_A). \tag{4.11}$$

Thus, $\underline{\mathbf{E}}_{\underline{P}}$ restricted to $I_{\mathcal{F}} \cup -I_{\mathcal{F}}$, say Q, is a linear prevision: it is the restriction of a coherent lower prevision, and it is self-conjugate. By Theorem 3.21 on p. 67, it follows that the set function μ, defined by $\mu(A) := Q(A)$ for all $A \in \mathcal{F}$, is a probability charge; note that $\mathbf{P}_\mu = Q$. Its linear extension $\mathbf{E}_\mu = \mathbf{E}_Q$ is, by its definition, a linear prevision defined on the set of all μ-integrable gambles f. But, since $\underline{\mathbf{E}}_{\underline{P}}$ is also a coherent behavioural extension of Q, it must hold that $\mathbf{E}_Q(f) = \underline{\mathbf{E}}_{\underline{P}}(f)$ for all μ-integrable gambles f, by Proposition 4.13. But, by Proposition 4.28, $\mathcal{F}$-measurability implies μ-integrability, and so we also have that $\mathbf{E}_Q(f) = \underline{\mathbf{E}}_{\underline{P}}(f)$ for all $\mathcal{F}$-measurable gambles f. This means that $\underline{\mathbf{E}}_{\underline{P}}$ is self-conjugate on the set $\mathcal{L}_{\mathcal{F}}(X)$ of $\mathcal{F}$-measurable gambles. □

4.3.4 Inner and Outer Set Function, Completion, and Other Extensions

The following definition describes well-known extensions for probability charges, 2-monotone set functions, and nested set functions that can be characterised by means of natural extension.

Definition 4.34. Let μ be a set function defined on a collection $\mathcal{A}$ of subsets of $\mathcal{X}$. The *inner set function* and *outer set function* induced by μ are the set functions μ_* and μ^* defined for all $B \subseteq \mathcal{X}$ by

$$\mu_*(B) := \sup\{\mu(A) \colon A \in \mathcal{A}, A \subseteq B\},$$
$$\mu^*(B) := \inf\{\mu(A) \colon A \in \mathcal{A}, A \supseteq B\}.$$

Proposition 4.35. *Let ν be a set function defined on a collection $\mathcal{A}$ of subsets of $\mathcal{X}$, and let π be its dual. Then $\nu_*(A) = 1 - \pi^*(\complement A)$ and $\nu^*(A) = 1 - \pi_*(\complement A)$ for any $A \in \mathcal{A}$.*

Proof. Immediate from the definition of inner and outer set function. □

The next theorem summarises the most important properties of the inner and outer set function. Properties (i), (ii) and (vi) are due to De Cooman

and Aeyels [21]; their results are proved by means of nested set functions induced by a multi-valued map; for the sake of completeness, an alternative proof without reference to multi-valued maps is given below. The continuity condition below for necessity and possibility measures is sufficient, but not necessary; it is a special case of the (necessary and sufficient) continuity condition given by De Cooman and Aeyels [21]; this follows immediately from De Cooman and Aeyels [21, Section 6.2]. Property (iii) is due to Walley [86, Corollary 3.1.9, p. 127], (iv) is due to Choquet [11, Chapter IV, Lemma 18.3&18.4, pp. 185–186], and (v) summarises (iii) and (iv) for probability charges.

Recall that by Theorem 3.27 on p. 72, lower probabilities, upper probabilities, and probabilities induced by nested set functions are always coherent, and by Theorem 3.31 on p. 76, lower probabilities induced by 2-monotone set functions are always coherent.

Theorem 4.36. *The following statements hold.*

(i) Let ν and π be nested set functions. Then $\underline{\mathbf{P}}_{\nu_*}$ is coherent, and equal to $\underline{\mathbf{E}}_{\underline{\mathbf{P}}_\nu}$ on indicators:

$$\nu_*(A) = \underline{\mathbf{E}}_{\underline{\mathbf{P}}_\nu}(I_A), \quad \text{for all } A \subseteq \mathcal{X}.$$

Similarly, $\overline{\mathbf{P}}_{\pi^*}$ is coherent and equal to $\overline{\mathbf{E}}_{\overline{\mathbf{P}}_\pi}$ on indicators:

$$\pi^*(A) = \overline{\mathbf{E}}_{\overline{\mathbf{P}}_\pi}(I_A), \quad \text{for all } A \subseteq \mathcal{X}.$$

(ii) Let ν and π be nested set functions. Then ν_* is a minimum preserving set function. If, additionally, $\operatorname{dom}\nu$ is closed under arbitrary intersection and ν is continuous from above, i.e.,

$$\nu\left(\bigcap_{A\in\mathcal{A}} A\right) = \inf_{A\in\mathcal{A}} \nu(A), \text{ for any non-empty } \mathcal{A} \subseteq \operatorname{dom}\nu,$$

then ν_* is a necessity measure. Similarly, π^* is a maximum preserving set function. If, additionally, $\operatorname{dom}\pi$ is closed under arbitrary union and π is continuous from below, i.e.,

$$\pi\left(\bigcup_{A\in\mathcal{A}} A\right) = \sup_{A\in\mathcal{A}} \pi(A), \text{ for any non-empty } \mathcal{A} \subseteq \operatorname{dom}\pi,$$

then π^* is a possibility measure.

(iii) Let ν and π be set functions defined on a field. If $\underline{\mathbf{P}}_\nu$ is coherent, then $\underline{\mathbf{P}}_{\nu_*}$ is also coherent and equal to $\underline{\mathbf{E}}_{\underline{\mathbf{P}}_\nu}$ on indicators:

$$\nu_*(A) = \underline{\mathbf{E}}_{\underline{\mathbf{P}}_\nu}(I_A), \quad \text{for all } A \subseteq \mathcal{X}.$$

Similarly, if $\overline{\mathbf{P}}_\pi$ is coherent, then $\overline{\mathbf{P}}_{\pi^*}$ is also coherent and equal to $\overline{\mathbf{E}}_{\overline{\mathbf{P}}_\pi}$ on indicators:

$$\pi^*(A) = \overline{\mathbf{E}}_{\overline{\mathbf{P}}_\pi}(I_A), \quad \text{for all } A \subseteq \mathcal{X}.$$

(iv) Let $n \in \mathbb{N}^*$, $n \geq 2$. If ν is an n-monotone set function, then ν_* is also an n-monotone set function. If π is an n-alternating set function, then π^* is also an n-alternating set function.

(v) If μ is a probability charge, then μ_* is completely monotone, μ^* is completely alternating, and

$$\mu_*(A) = \underline{\mathbf{E}}_\mu(I_A), \text{ and } \mu^*(A) = \overline{\mathbf{E}}_\mu(I_A), \quad \text{for all } A \subseteq \mathcal{X}. \tag{4.12}$$

(vi) Let $\mathcal{F}$ be an ample field, let N be a necessity measure on $\mathcal{F}$ with necessity distribution ν, and let Π be a possibility measure on $\mathcal{F}$ with possibility distribution π. Then N_* is a necessity measure with necessity distribution ν and Π^* is a possibility measure with possibility distribution π: for any $A \subseteq \mathcal{X}$, $A \neq \emptyset$ and $A \neq \mathcal{X}$, it holds that

$$N_*(A) = \inf_{x \in \complement A} \nu(x) \quad \text{and} \quad \Pi^*(A) = \sup_{x \in A} \pi(x).$$

Proof. (i). It suffices to prove that $\nu_*(A) = \underline{\mathbf{E}}_{\underline{\mathbf{P}}_\nu}(I_A)$ for all $A \subseteq \mathcal{X}$. Indeed, since $\mathcal{X} \in \operatorname{dom} \nu$ and $\nu(\mathcal{X}) = 1$ by the coherence of $\underline{\mathbf{P}}_\nu$ (Theorem 3.27), it follows from Theorem 4.3 on p. 96 that

$$\underline{\mathbf{E}}_{\underline{\mathbf{P}}_\nu}(A) = \sup\Bigg\{ \sum_{i=1}^n \lambda_i \nu(S_i) \colon n \in \mathbb{N},\ \lambda_1, \ldots, \lambda_n \geq 0,\ S_1, \ldots, S_n \in \operatorname{dom} \nu,\ \sum_{i=1}^n \lambda_i I_{S_i} \leq I_A \Bigg\}.$$

Choosing $n = 1$ and $\lambda_1 = 1$, it immediately follows that

$$\underline{E}_{\underline{P}_\nu}(A) \geq \sup\{\nu(S)\colon S \in \operatorname{dom}\nu,\, S \subseteq A\} = \nu_*(A).$$

To see that the converse inequality holds too, it suffices to show that for any $n \in \mathbb{N}$, $\lambda_1, \ldots, \lambda_n \geq 0$, and $S_1, \ldots, S_n \in \operatorname{dom}\nu$ such that $\sum_{i=1}^n \lambda_i I_{S_i} \leq I_A$, there is an $S \in \operatorname{dom}\nu$ such that $S \subseteq A$ and $\sum_{i=1}^n \lambda_i \nu(S_i) \leq \nu(S)$.

If $\sum_{i=1}^n \lambda_i I_{S_i} = 0$, choose $S = \emptyset$, and use $\emptyset \in \operatorname{dom}\nu$ and $\nu(\emptyset) = 0$ (which hold by the coherence of $\underline{P}_\nu$). Otherwise, without loss of generality, we can assume that $\lambda_i > 0$ and $S_i \neq \emptyset$ for all $i \in \{1, \ldots, n\}$. Define the non-empty set

$$S := \bigcup_{i=1}^n S_i.$$

Since $\operatorname{dom}\nu$ is a nested collection of sets, it follows that $S \in \operatorname{dom}\nu$. If $x \in S$, then (since $\lambda_i > 0$ for all $i \in \{1, \ldots, n\}$ and $n \geq 1$)

$$0 < \sum_{i=1}^n \lambda_i I_{S_i}(x) \leq I_A(x);$$

this can only happen when also $I_A(x) = 1$. Hence, $S \subseteq A$. Finally, take any $x \in \cap_{i=1}^n S_i$—this set is non-empty since $\operatorname{dom}\nu$ is a nested collection, $S_i \neq \emptyset$ for all $i \in \{1, \ldots, n\}$, and $n \geq 1$. Observe that $x \in S$, and therefore also $x \in A$. We find that

$$1 = I_A(x) \geq \sum_{i=1}^n \lambda_i I_{S_i}(x) = \sum_{i=1}^n \lambda_i,$$

and hence,

$$\nu(S) \geq \sum_{i=1}^n \lambda_i \nu(S) \geq \sum_{i=1}^n \lambda_i \nu(S_i),$$

where we used $\nu(S) \geq \nu(S_i)$ for all $i \in \{1, \ldots, n\}$: indeed, $S \supseteq S_i$ by definition of S, and hence, $\nu(S) \geq \nu(S_i)$ by the coherence of $\underline{P}_\nu$; see Theorem 3.27. This establishes that $\nu_*(A) = \underline{E}_{\underline{P}_\nu}(I_A)$ for all $A \subseteq \mathcal{X}$.

To prove that $\pi^*(A) = \overline{E}_{\overline{P}_\pi}(I_A)$ for all $A \subseteq \mathcal{X}$, apply Proposition 4.35 and Proposition 4.10 on p. 99.

(ii). From (i) and Theorem 3.5(iv): $\nu_*(A \cap B) = \underline{E}_{\underline{P}_\nu}(A \cap B) \leq \underline{E}_{\underline{P}_\nu}(A) = \nu_*(A)$ and $\nu_*(A \cap B) = \underline{E}_{\underline{P}_\nu}(A \cap B) \leq \underline{E}_{\underline{P}_\nu}(B) = \nu_*(B)$, and hence, $\nu_*(A \cap B) \leq \nu_*(A) \wedge \nu_*(B)$. To establish the converse inequality, it suffices to show that, for any $C, D \in$

$\operatorname{dom} \nu$ such that $C \subseteq A$ and $D \subseteq B$, there is an $E \in \operatorname{dom} \nu$ such that $E \subseteq A \cap B$ and $\nu(E) \geq \nu(C) \wedge \nu(D)$. Indeed, take $E = C \cap D$, and observe that $E = C$ or $E = D$ since $\operatorname{dom} \nu$ is a nested collection. Hence,

$$\begin{aligned} \nu_*(A \cap B) &= \sup\{\nu(E) \colon E \in \operatorname{dom} \nu,\ E \subseteq A \cap B\} \\ &\geq \sup\{\nu(C) \wedge \nu(D) \colon C, D \in \operatorname{dom} \nu,\ C \subseteq A,\ D \subseteq B\} \end{aligned}$$

and since the conditions $C \subseteq A$ and $D \subseteq B$ are logically independent,

$$\begin{aligned} &= \sup\{\nu(C) \colon C \in \operatorname{dom} \nu,\ C \subseteq A\} \wedge \sup\{\nu(D) \colon D \in \operatorname{dom} \nu,\ D \subseteq B\} \\ &= \nu_*(A) \wedge \nu_*(B). \end{aligned}$$

Suppose that $\operatorname{dom} \nu$ is closed under arbitrary intersection, and ν is continuous from above. Let $\mathcal{A}$ be any collection of subsets of $\mathcal{X}$. Again, the inequality $\nu_*(\bigcap_{A \in \mathcal{A}} A) \leq \inf_{A \in \mathcal{A}} \nu_*(A)$ follows from (i) and Theorem 3.5(iv). To establish the converse inequality, it suffices to show that, whenever for all $A \in \mathcal{A}$ we have $C_A \subseteq \operatorname{dom} \nu$ such that $C_A \subseteq A$, there is an $E \in \operatorname{dom} \nu$ such that $E \subseteq \bigcap_{A \in \mathcal{A}} A$ and $\nu(E) \geq \inf_{A \in \mathcal{A}} \nu(C_A)$. Indeed, take $E = \bigcap_{A \in \mathcal{A}} C_A \colon E \in \operatorname{dom} \nu$ for some $A \in \mathcal{A}$ since $\operatorname{dom} \nu$ is a nested collection closed under arbitrary intersection, and $\nu(E) = \inf_{A \in \mathcal{A}} \nu(C_A)$ since ν is continuous from above. Now, proceed as above.

To prove that π^* is maximum preserving, or a possibility measure if π is continuous from below, apply Proposition 4.35 and Proposition 4.10 on p. 99.

(iii). See Walley [86, Corollary 3.1.9, p. 127].

(iv). We prove the n-monotone case: the n-alternating case then follows from Proposition 4.35 and Proposition 4.10 on p. 99.

Choquet [11, Chapter IV, Lemma 18.3, p. 186, ll. 6–9] gives a simple proof—note that Choquet's 'interior capacity' does not depend on any topology on $\mathcal{X}$, and coincides with our inner set function, but Choquet's 'exterior capacity' does depend on the topology on $\mathcal{X}$, and therefore has no relation with the outer set function in general; see Choquet [11, Chapter V, Section 15.2, p. 174]. The 2-monotone case can also be found in for instance Walley [85, Lemma 6.1]. Choquet's proof is rather brief, so let's fill in the details for the sake of completeness. We prove the case where n is finite; the case $n = \infty$ is then immediate.

Let $B_1, \ldots, B_n$ be subsets of $\mathcal{X}$. By the definition of the inner set function

ν_*, for any non-empty $J \subseteq \{1,\dots,n\}$, there is a set A_J in $\operatorname{dom}\nu$ such that $A_J \subseteq \bigcap_{i\in J} B_i$ and $\nu_*(\bigcap_{i\in J} B_i) \leq \nu(A_J) + \epsilon$. For every $i \in \{1,\dots,n\}$, define $A_i := \bigcup_{J\ni i} A_J$. Since $A_J \subseteq \bigcap_{i\in J} B_i$ for every non-empty $J \subseteq \{1,\dots,n\}$, it follows that

$$A_i = \bigcup_{J\ni i} A_J \subseteq \bigcup_{J\ni i} \bigcap_{i'\in J} B_{i'} \subseteq B_i$$

for all $i \in \{1,\dots,n\}$: if there is a $J \ni i$ (namely, $J = \{i\}$), such that $x \in B_{i'}$ for every $i' \in J$, then it must hold in particular that $x \in B_i$ (since $i \in J$). Also,

$$\bigcap_{i\in J} A_i = \bigcap_{i\in J} \bigcup_{J'\ni i} A_{J'} \supseteq A_J$$

for every non-empty $J \subseteq \{1,\dots,n\}$: if $x \in A_J$, then, obviously, for every $i \in J$, there is a $J' \ni i$ (namely, $J' = J$) such that $x \in A_{J'}$. Hence, for every $\epsilon > 0$, we find that

$$\nu_*\left(\bigcup_{i=1}^n B_i\right) \geq \nu_*\left(\bigcup_{i=1}^n A_i\right) = \nu\left(\bigcup_{i=1}^n A_i\right)$$

since $A_i \subseteq B_i$ and $A_i \in \operatorname{dom}\nu$ for all $i \in \{1,\dots,n\}$, and now using the n-monotonicity of ν,

$$\geq \sum_{\emptyset\neq J\subseteq\{1,2,\dots,n\}} (-1)^{|J|+1} \nu\left(\bigcap_{i\in J} A_i\right).$$

and since $\nu(\bigcap_{i\in J} A_i) \leq \nu_*(\bigcap_{i\in J} B_i) \leq \nu(A_J) + \epsilon \leq \nu(\bigcap_{i\in J} A_i) + \epsilon$ for all non-empty $J \subseteq \{1,\dots,n\}$, we may conclude that

$$\geq -(2^n - 1)\epsilon + \sum_{\emptyset\neq J\subseteq\{1,2,\dots,n\}} (-1)^{|J|+1} \nu_*\left(\bigcap_{i\in J} B_i\right).$$

Since this inequality holds for all $\epsilon > 0$, it follows that it must also hold for $\epsilon = 0$, and hence, ν_* is n-monotone.

(v). Immediate (iii) and (iv), once observed that a probability charge is a completely monotone and completely alternating set function (see Choquet [11, Section 14.5, pp. 173–174]). An alternative proof follows from application of Theorem 4.41 and Theorem 4.42, given further on.

(vi). By (iii), $N_*(A) = \underline{\mathbf{E}}_{\underline{\mathbf{P}}_N}(I_A)$ for all $A \subseteq \mathcal{X}$. De Cooman and Aeyels [21, Theorem 6] proved that $\underline{\mathbf{E}}_{\underline{\mathbf{P}}_N}(I_{\complement\{x\}}) = \nu(x)$. This establishes the claim. Alternatively, for any $A \subseteq \mathcal{X}$, $A \neq \mathcal{X}$, it holds that

$$\begin{aligned} N_*(A) &= \sup\{N(B)\colon B \in \mathcal{F},\, B \subseteq A\} \\ &= \sup\left\{\inf_{x\in\complement B} \nu(x)\colon B \in \mathcal{F},\, B \subseteq A\right\} \\ &= \sup\left\{\inf_{x\in B} \nu(x)\colon B \in \mathcal{F},\, \complement A \subseteq B\right\} \end{aligned}$$

and this supremum is reached for $B \in \mathcal{F}$ as small as possible such that $\complement A \subseteq B$, and by Theorem 3.51 on p. 88, this is $B = \cup_{x\in\complement A}[x]_{\mathcal{F}}$, hence,

$$= \inf_{x\in\complement A} \nu(x).$$

The proof for Π^* is similar. □

The above proposition about inner and outer extension has a nice consequence, which is again due to De Cooman and Aeyels [21, Section 7.1]; the corollary below is a special case of their result. We give an alternative proof below.

Corollary 4.37. *Let N be a necessity measure, and let v be its induced necessity distribution. Then $\underline{\mathbf{P}}_N$ is equivalent to the lower probability induced by the restriction of N to the dual cut sets of v, i.e., the lower probability $\underline{\mathbf{P}}_\nu$ induced by the nested set function*

$$\nu(\{x \in \mathcal{X}\colon v(x) \leq z\}) := N(\{x \in \mathcal{X}\colon v(x) \leq z\}), \text{ for all } z \in \mathbb{R}.$$

Similarly, let Π be a possibility measure, and let p be its induced possibility distribution. Then $\overline{\mathbf{P}}_\Pi$ is equivalent to the upper probability induced by the restriction of Π to the strict cut sets of p, i.e., the upper probability $\overline{\mathbf{P}}_\pi$ induced by the nested set function

$$\pi(\{x \in \mathcal{X}\colon p(x) > z\}) := \Pi(\{x \in \mathcal{X}\colon p(x) > z\}), \text{ for all } z \in \mathbb{R},$$

Proof. As it follows from Theorem 3.46 that N is coherent, it suffices to show that $N(A) = \nu_*(A)$ for all $A \in \operatorname{dom} N$, by Theorem 4.6 and Theorem 4.8.

Observe that ν is a nested set function, and $\operatorname{dom}\nu$ is closed under arbitrary intersection: for any $Z \subseteq \mathbb{R}$, it holds that

$$\bigcap_{z\in Z}\{x \in \mathcal{X}\colon N(x) \le z\} = \{x \in \mathcal{X}\colon N(x) \le \inf Z\};$$

and ν is continuous from above since N is a necessity measure—hence, ν_* is necessity measure by Theorem 4.36(ii). Is it equal to N? Yes, by Theorem 3.53, this holds if the necessity distribution v_* induced by ν_* is equal to the necessity distribution v induced by N. Indeed, for any $x \in \mathcal{X}$,

$$\begin{aligned} v_*(x) &= \nu_*(\complement\{x\}) \\ &= \sup\{\nu(S)\colon S \in \operatorname{dom}\nu,\ x \notin S\} \\ &= \sup\{N(\{y \in \mathcal{X}\colon v(y) \le z\})\colon z \in \mathbb{R},\ x \notin \{y \in \mathcal{X}\colon v(y) \le z\}\} \\ &= \sup\{N(\{y \in \mathcal{X}\colon v(y) \le z\})\colon z \in \mathbb{R},\ v(x) > z\} \\ &= \sup\{\inf\{v(y)\colon y \in \mathcal{X},\ v(y) > z\}\colon z \in \mathbb{R},\ v(x) > z\} \end{aligned}$$

and since $\inf\{v(y)\colon y \in \mathcal{X},\ v(y) > z\}$ is non-decreasing in z,

$$= \lim_{z \xrightarrow{<} v(x)} \inf\{v(y)\colon y \in \mathcal{X},\ v(y) > z\}$$

Since $\inf\{v(y)\colon y \in \mathcal{X},\ v(y) > z\} \ge z$, its limit for z to $v(x)$ is greater or equal than $v(x)$. Also, since $v(x) \in \{v(y)\colon y \in \mathcal{X},\ v(y) > z\}$ whenever $z < v(x)$, it follows that $\inf\{v(y)\colon y \in \mathcal{X},\ v(y) > z\} \le v(x)$ whenever $z < v(x)$. Hence, the limit is less or equal than $v(x)$. We conclude that $v_*(x) = v(x)$. □

In the literature, we find the following extensions of probability charges (see for instance Denneberg [28, pp. 24–29]). We shall investigate their relation to the linear extension we introduced before.

Definition 4.38. Let μ be a probability charge defined a the field $\mathcal{F}$. The *completion* of μ is the probability charge $\overline{\mu}$ defined by $\overline{\mu}(A \Delta N) := \mu(A)$ for any $A \in \mathcal{F}$ and $N \subseteq \mathcal{X}$, whenever there is an $M \in \mathcal{F}$ such that $N \subseteq M$ and $\mu(M) = 0$. The *Carathéodory field* of μ is defined as

$$C_\mu := \{A \subseteq \mathcal{X}\colon \mu^*(B) = \mu^*(B \cap A) + \mu^*(B \setminus A) \text{ for all } B \subseteq \mathcal{X}\},$$

and *Jordan field* of μ is defined as

$$\mathcal{J}_\mu := \{A \subseteq \mathcal{X} \colon \mu_*(A) = \mu^*(A)\}.$$

The *Carathéodory extension* of a probability charge μ is the unique extension of μ to a probability charge μ_C defined on C_μ, and the *Jordan extension* of a probability charge μ is the unique extension of μ to a probability charge $\mu_\mathcal{J}$ defined on $\mathcal{J}_\mu$.

Through Theorem 4.36, we can easily show that the probability $\mathbf{P}_{\mu_\mathcal{J}}$ induced by the Jordan extension $\mu_\mathcal{J}$ is equal to the linear extension $\mathbf{E}_\mu$ of $\mathbf{P}_\mu$ restricted to indicators. By Theorem 4.30, this establishes existence and uniqueness of the Jordan extension. Note that, as a result, $I_{\mathcal{J}_\mu} = \operatorname{dom} \mathbf{E}_\mu \cap I_{\wp(\mathcal{X})}$.

Proof of existence and uniqueness of the Carathéodory extension for probability measures can be found in for instance Kallenberg [40, Theorem 2.1, p. 24]. For probability charges, this follows from a result by Denneberg [28], who showed that Jordan and Carathéodory extensions coincide. Anyway, the Carathéodory extension of μ satisfies the following interesting property:

$$\mu^*(A \cup B) + \mu^*(A \cap B) = \mu_C(A) + \mu^*(B)$$

for any $A \in C_\mu$ and $B \subseteq \mathcal{X}$; note that $\mu_C(A) = \mu^*(A) = \mu_*(A)$. In case $A \cap B = \emptyset$, the above equation follows from Proposition 4.16 and Theorem 4.36 above. A set $A \in C_\mu$ is also called μ^**-measurable* (see Halmos [40, Section 11]) or *Carathéodory measurable* (see Denneberg [28, Chapter 2, p. 24]). Following Hildebrandt [43, Chapter V, Theorem 2.8], Carathéodory measurability has also a connection with Riemann integrability. Similarly, a set in $\mathcal{J}_\mu$ is called *Jordan measurable*.

Taking $B = \mathcal{X}$ in the definition of C_μ makes clear that the Jordan field includes the Carathéodory field. Once observed that μ_* is a 2-monotone set function, it follows easily that the Jordan field is actually equal to the Carathéodory field; see Denneberg [28, Proposition 2.9]. Therefore, the Jordan extension is equal to the Carathéodory extension.

It is easy to show that the completion of a probability charge μ is also a probability charge that agrees with the linear extension of μ, but in general, the domain of the completion is smaller than the domain of the Carathéodory or Jordan extension. Let's also mention that sets in the domain of the comple-

tion of the Lebesgue measure are called *Lebesgue measurable*. So, measurability is legion, and all want their personal soldier? Anyway, let's summarise:

Theorem 4.39. *Let $\mathcal{F}$ be a field on $\mathcal{X}$ and let μ be a probability charge. Then, for any $A \subseteq \mathcal{X}$, $A \in \operatorname{dom} \overline{\mu}$ implies $I_A \in \operatorname{dom} \mathbf{E}_\mu$, and*

$$\mathbf{E}_\mu(I_A) = \overline{\mu}(A) \text{ for any } A \in \operatorname{dom} \overline{\mu},$$

The linear extension of μ restricted to events, the Carathéodory extension of μ, and the Jordan extension of μ coincide: for any $A \subseteq \mathcal{X}$ it holds that

$$I_A \in \operatorname{dom} \mathbf{E}_\mu \iff A \in \mathcal{C}_\mu \iff A \in \mathcal{J}_\mu$$

and, if A satisfies any of these equivalent conditions,

$$\mathbf{E}_\mu(I_A) = \mu_{\mathcal{C}}(A) = \mu_{\mathcal{J}}(A).$$

Proof. The completion $\overline{\mu}$ agrees on its domain with the linear extension of μ. Indeed, let $A \in \operatorname{dom} \overline{\mu}$. We must show that $\overline{\mu}(A) = \mu_*(A) = \mu^*(A)$. Since $\mu_*(A) \leq \mu^*(A)$, it suffices to show that $\mu_*(A) \geq \overline{\mu}(A) \geq \mu^*(A)$. Since $A \in \operatorname{dom} \overline{\mu}$, there are $B \in \mathcal{F}$, $M \in \mathcal{F}$, and $N \subseteq M$, such that $A = B \Delta N$, $\mu(M) = 0$, and $\overline{\mu}(A) = \mu(B)$; see Definition 4.38. Hence,

$$\mu_*(B\Delta N) \geq \left|\mu_*(B) - \mu_*(N)\right| = \mu_*(B) = \mu(B) = \mu^*(B) + \mu^*(N) \geq \mu^*(B\Delta N),$$

where we used $\mu_*(N) \leq \mu^*(N) \leq \mu^*(M) = \mu(M) = 0$, the coherence of the natural extension $\underline{\mathbf{E}}_\mu$ (Theorem 3.5), and Theorem 4.36(v)—for the first inequality, use $I_{B\Delta N} = |I_B - I_N|$, and for the second inequality, use $I_B + I_N \geq I_{B\Delta N}$. Therefore, $\mu_*(A) \geq \overline{\mu}(A) \geq \mu^*(A)$.

It follows from Eq. (4.12) that $\mu_*(A) = \mu^*(A)$ if and only if $\underline{\mathbf{E}}_\mu(I_A) = \overline{\mathbf{E}}_\mu(I_A)$, and therefore, the Jordan extension corresponds to the linear extension of μ restricted to indicators. Denneberg [28, Proposition 2.9], has proved that the Jordan extension coincides with the Carathéodory extension. Let's give a shorter version of the proof.

Clearly, if $A \in \mathcal{C}_\mu$ then $A \in \mathcal{J}_\mu$; simply take $B = \mathcal{X}$ in the definition of the Carathéodory field. Hence, $\mathcal{C}_\mu \subseteq \mathcal{J}_\mu$. Conversely, assume that $A \in \mathcal{J}_\mu$. By Theorem 4.36, μ^* induces a coherent upper prevision, and hence, μ^* is

sub-additive, so

$$\mu^*(B) \leq \mu^*(B \cap A) + \mu^*(B \setminus A).$$

To prove the converse inequality, observe that μ^* is 2-monotone, again by Theorem 4.36. Hence,

$$\mu^*(B) \geq \mu^*(B \cap A) + \mu^*(B \cup A) - \mu^*(A).$$

Again, apply Theorem 4.36, and use Proposition 4.16, to see that

$$\mu^*(B \cup A) - \mu^*(A) = \overline{\mathbf{E}}_\mu(I_{B\cup A}) - \overline{\mathbf{E}}_\mu(I_A) = \overline{\mathbf{E}}_\mu(I_{B\cup A} - I_A) = \overline{\mathbf{E}}_\mu(I_{B\setminus A}) = \mu^*(B \setminus A),$$

where we used the fact that A belongs to the Jordan field, $\underline{\mathbf{E}}_\mu(I_A) = \mu_*(A) = \mu^*(A) = \overline{\mathbf{E}}_\mu(I_A)$, to apply Proposition 4.16. Hence, $A \in C_\mu$. □

For the purpose of this work, the most important observation is that the inner set functions induced by nested set functions, possibility measures, and probability charges, are 2-monotone set functions.

4.3.5 The S-Integral

One of the simplest kinds of integrals on charges one can think of is the S-integral. Of all integrals we shall discuss, this integral is most closely related to the idea of natural extension, as we shall prove shortly in Theorem 4.42.

The S-integral was suggested by Moore and Smith [57, Section 5, p. 114, ll. 10–13] to provide a conceptually simpler definition of the Lebesgue integral. It was then defined by Kolmogoroff [50, Zweites Kapitel, §2, p. 663, Nr. 12] for arbitrary functions, and again by Hildebrandt [42, Sect. 1(f), p. 869] for bounded functions. Gould [37, Definition 4.3, p. 201, and Definition 6.1& Theorem 6.2, p. 213] extended the S-integral to unbounded functions and charges that assume values in a Banach space—incidentally, Gould aimed at a generalisation of the Dunford integral. In this section, we shall only consider the S-integral of bounded functions, *i.e.*, gambles, with respect to real-valued bounded positive charges. For this case, Bhaskara Rao and Bhaskara Rao [9, Section 4.5] have defined the S-integral through a lower and an upper S-integral; equivalence is immediate from Gould [37, Theorem 4.7(c)], in case of real-valued positive charges. Hence, Bhaskara Rao's and Bhaskara Rao's construction of Hildebrandt's S-integral—restricted to positive charges and

bounded functions—is similar to Darboux's [14] construction of Riemann's [66] integral, as we shall discuss in Section 4.3.6. With respect to the Lebesgue measure (or the Lebesgue-Stieltjes measure, which will be introduced in Definition 4.47), the S-integral is also called the Y-integral, or the Young-Stieltjes integral. It was introduced by Young [94] and is extensively discussed by Hildebrandt [43, Chapter VII, Section 3]. The S-integral is a straightforward generalisation of the Young-Stieltjes integral. For gambles, the Young-Stieltjes integral provides us with an alternative to, and, in my opinion, also a simpler formulation, of the Lebesgue-Stieltjes integral; see Hildebrandt [43]. In particular, it does not involve any measurability conditions. In any case, the Young-Stieltjes integral and the Lebesgue-Stieltjes integral agree on a very large class of gambles; see Hildebrandt [43, Chapter VII, Theorem 3.9].

Hildebrandt [42] claims that the S-integral is of the **S**tieltjes type, whence the term **S**-integral. But, as we shall prove in Theorem 4.53, not all Riemann-Stieltjes integrals are representable by an S-integral: the S-integral is therefore not really of the Stieltjes type. Nevertheless, we shall keep the term S-integral.

However, the S-integral does generalise the Riemann integral, as well as a large class of Riemann-Stieltjes integrals. Apparently, this is a new result: in the vast literature on integration, I have found only very little material covering this topic, perhaps because the connection requires integration with respect to charges on fields of $\mathbb{R}$ that are not σ-fields. Let's summarise what I believe is known:

- If a gamble is Riemann-integrable, then it is S-integrable with respect to the Lebesgue measure restricted to the field generated by $\{[a, x] : x \in \mathcal{X}\}$ (which we denote by $\mathcal{F}_{[]}(\mathcal{X})$); see Hildebrandt [42, p. 870, ll. 1–6]—we shall establish the converse of this claim in Corollary 4.54.

- De Finetti [27, Vol. I, Sections 6.2–6.4, pp. 222–241] has argued that '[the lower and upper Riemann-Stieltjes integral] expresses *all that one can obtain from F*, that is, distributional knowledge, [...]' (De Finetti [27, Vol. I, p. 235, ll. 14–17]. This suggests that the lower and upper Riemann-Stieltjes integral should coincide with the lower and upper S-integral induced by a Riemann-Stieltjes charge; we shall prove that this claim is approximately correct, if we follow Darboux's approach to Riemann-Stieltjes integration (see Section 4.3.6 on p. 132 ff.).

So, we shall take some time for an extensive discussion of the relation between

the S-integral and the Riemann-Stieltjes integral; the results are summarised in Theorem 4.52 on p. 145, Theorem 4.53 on p. 151, and Corollary 4.54 on p. 156. Why do we do this? Because, as a result, we shall be able to

- provide new characterisations of the Choquet integral described in Section 4.3.10, and
- characterise the natural extension of cumulative distribution functions and p-boxes as Riemann-Stieltjes integrals.

This only adds to the importance of the S-integral.

The S-integral, and hence, as we shall see shortly, natural extension, provides us with a tool to study all of the above-mentioned integrals and their connections. Conversely, and certainly of more importance for our purposes here, all of these integrals provide us with a tool to calculate natural extension itself, not only for probability charges, but for any lower prevision that avoids sure loss. This will be extensively discussed in Section 4.4.

Let $\mathbb{P}(\mathcal{F})$ denote the set of all finite partitions of $\mathcal{X}$ whose elements belong to the field $\mathcal{F}$. We define a relation $\leq$ on $\mathbb{P}(\mathcal{F})$: say that $\mathcal{A} \leq \mathcal{B}$ whenever $\mathcal{B}$ is a *refinement* of $\mathcal{A}$, *i.e.*, whenever every element of $\mathcal{B}$ is a subset of an element of $\mathcal{A}$. It follows easily that

(i) $\leq$ is *reflexive*: every finite partition is a refinement of itself,

(ii) $\leq$ is *transitive*: if a finite partition $\mathcal{A}$ refines a finite partition $\mathcal{B}$, and $\mathcal{B}$ refines a finite partition $\mathcal{C}$, then $\mathcal{A}$ also refines $\mathcal{C}$, and

(iii) $\leq$ satisfies the *composition property*: since $\mathcal{F}$ is a field, every two finite partitions in $\mathcal{F}$ have a common finite refinement in $\mathcal{F}$, *i.e.*, for every $\mathcal{A}$ and $\mathcal{B}$ in $\mathbb{P}(\mathcal{F})$ there is a $\mathcal{C}$ in $\mathbb{P}(\mathcal{F})$ such that $\mathcal{A} \leq \mathcal{C}$ and $\mathcal{B} \leq \mathcal{C}$.

A set equipped with a relation that is transitive, reflexive, and that satisfies the composition property, is called a *directed set*. Hence, $\mathbb{P}(\mathcal{F})$ is a directed set with respect to $\leq$.

As a consequence, we can take the so-called Moore-Smith limit over $\mathbb{P}(\mathcal{F})$; see Moore and Smith [57, Section I, p. 103]: for every *net* α on $\mathbb{P}(\mathcal{F})$, *i.e.*, every map $\alpha\colon \mathbb{P}(\mathcal{F}) \to \mathbb{R}$, the *Moore-Smith limit* of α, if it exists, is the unique real number a such that for every $\epsilon > 0$, there is an $\mathcal{A}_\epsilon \in \mathbb{P}(\mathcal{F})$ such that $|\alpha(\mathcal{A}) - a| < \epsilon$ for all $\mathcal{A} \geq \mathcal{A}_\epsilon$; we shall denote the Moore-Smith limit of a net α on $\mathbb{P}(\mathcal{F})$ by $\lim_{\mathcal{B}\in\mathbb{P}(\mathcal{F})} \alpha(\mathcal{B})$. The Moore-Smith limit is a natural generalisation

of the limit of sequences, and this limit is used in the definition of the S-integral.

Definition 4.40. Let $\mathcal{F}$ be a field on $\mathcal{X}$ and let μ be a bounded positive charge on $\mathcal{F}$. A gamble f on X is called *S-integrable* with respect to μ if and only if the *lower and upper S-integral* of f

$$\mathrm{S}\underline{\int} f \,\mathrm{d}\mu := \lim_{\mathcal{B}\in\mathbb{P}(\mathcal{F})} \sum_{B\in\mathcal{B}} \underline{P}_B(f)\mu(B) = \sup_{\mathcal{B}\in\mathbb{P}(\mathcal{F})} \sum_{B\in\mathcal{B}} \underline{P}_B(f)\mu(B), \tag{4.13}$$

$$\mathrm{S}\overline{\int} f \,\mathrm{d}\mu := \lim_{\mathcal{B}\in\mathbb{P}(\mathcal{F})} \sum_{B\in\mathcal{B}} \overline{P}_B(f)\mu(B) = \inf_{\mathcal{B}\in\mathbb{P}(\mathcal{F})} \sum_{B\in\mathcal{B}} \overline{P}_B(f)\mu(B) \tag{4.14}$$

coincide. In such a case, the *S-integral* of f with respect to μ is defined as

$$\mathrm{S}\int f \,\mathrm{d}\mu := \mathrm{S}\underline{\int} f \,\mathrm{d}\mu = \mathrm{S}\overline{\int} f \,\mathrm{d}\mu. \tag{4.15}$$

Proof of existence of the Moore-Smith limits. Observe that $\sum_{B\in\mathcal{B}} \underline{P}_B(f)\mu(B)$ is a non-decreasing net over $\mathcal{B} \in \mathbb{P}(\mathcal{F})$, and is bounded from above by $\mu(\mathcal{X}) \sup f$. Hence, it converges and its Moore-Smith limit over $\mathbb{P}(\mathcal{F})$ coincides with the supremum over $\mathbb{P}(\mathcal{F})$. Proof of existence of the upper S-integral is similar. □

In case μ is a probability charge, recall that any finite convex combination and any point-wise limit of coherent lower previsions is coherent, and note that the lower S-integral with respect to μ is the point-wise limit of a net of convex combinations of coherent lower previsions; in fact, it's a point-wise limit of the natural extensions of belief functions, introduced in Section 3.5.7. Hence, up to normalisation, the lower S-integral is a coherent lower prevision, and the upper S-integral is its conjugate.

Theorem 4.41. *Let μ be a probability charge. Then the lower S-integral with respect to μ is a 2-monotone coherent lower prevision. The lower S-integral, restricted to $I_{\wp(\mathcal{X})}$, corresponds to a completely monotone set function.*

Proof. By Definition 4.40, the lower S-integral is the point-wise limit of a net of convex combinations of vacuous lower previsions. But, by Proposition 3.33 on p. 76, vacuous lower previsions are 2-monotone and coherent, and by Lemma 3.10 on p. 57, Lemma 3.12 on p. 58, Proposition 3.34 on p. 77, and Proposition 3.35 on p. 77, limits and convex combinations of 2-monotone

coherent lower previsions are 2-monotone and coherent. Therefore, the lower S-integral must be a 2-monotone coherent lower prevision.

Apply Proposition 3.39 on p. 80 and Proposition 3.42 on p. 82 to see why the lower S-integral, restricted to $I_{\wp(\mathcal{X})}$, corresponds to a completely monotone set function. □

The following theorem yields an even stronger result.

Theorem 4.42. *Let $\mathcal{F}$ be a field on $\mathcal{X}$ and let μ be a probability charge on $\mathcal{F}$. For any gamble f on X it holds that*

$$\underline{\mathrm{S}\int} f \,\mathrm{d}\mu = \underline{\mathbf{E}}_\mu(f), \qquad \overline{\mathrm{S}\int} f \,\mathrm{d}\mu = \overline{\mathbf{E}}_\mu(f). \tag{4.16}$$

Proof. We start with a simple observation. For every finite subset $\mathcal{A} \Subset \mathcal{F}$ and $\lambda_A \in \mathbb{R}$ (for all $A \in \mathcal{A}$) there are a finite partition $\mathcal{B}_\mathcal{A} \in \mathbb{P}(\mathcal{F})$ and a $\kappa_B \in \mathbb{R}$ (for all $B \in \mathcal{B}_\mathcal{A}$) such that

$$\sum_{A\in\mathcal{A}} \lambda_A I_A = \sum_{B\in\mathcal{B}_\mathcal{A}} \kappa_B I_B.$$

If $\sum_{B\in\mathcal{B}_\mathcal{A}} \kappa_B I_B \leq f$, then for any $B \in \mathcal{B}_\mathcal{A}$, it holds that $\kappa_B \leq f(x)$ for all $x \in B$, and hence, $\kappa_B \leq \inf_{x\in B} f(x) = \underline{P}_B(f)$. So, $\sum_{A\in\mathcal{A}} \lambda_A I_A \leq f$ implies that

$$\sum_{A\in\mathcal{A}} \lambda_A I_A = \sum_{B\in\mathcal{B}_\mathcal{A}} \kappa_B I_B \leq \sum_{B\in\mathcal{B}_\mathcal{A}} \underline{P}_B(f) I_B \leq f, \text{ and } \sum_{A\in\mathcal{A}} \lambda_A \mu(A) \leq \sum_{B\in\mathcal{B}_\mathcal{A}} \underline{P}_B(f)\mu(B). \tag{4.17}$$

Therefore, by definition of the lower S-integral,

$$\underline{\mathrm{S}\int} f \,\mathrm{d}\mu = \sup_{\mathcal{B}\in\mathbb{P}(\mathcal{F})} \sum_{B\in\mathcal{B}} \underline{P}_B(f)\mu(B),$$

and, taking the supremum over a larger set,

$$\leq \sup_{\mathcal{A}\Subset\mathcal{F}} \left\{ \sum_{A\in\mathcal{A}} \lambda_A \mu(A) \colon A \in \mathcal{F},\ \lambda_A \in \mathbb{R},\ \sum_{A\in\mathcal{A}} \lambda_A I_A \leq f \right\} = \underline{\mathbf{E}}_\mu(f)$$

where the equality with $\underline{\mathbf{E}}_\mu$ follows from the fact that $\mathcal{X} \in \mathcal{F}$, so we can omit the constant in Eq. (4.1), and from the self-conjugacy of $\mathbf{P}_\mu$, so we can allow the coëfficients λ_A to be any real number in Eq. (4.1), instead of only non-negative reals. Now, invoking Eq. (4.17),

$$\leq \sup_{\mathcal{A} \Subset \mathcal{F}} \sum_{B \in \mathcal{B}_\mathcal{A}} \underline{P}_B(f)\mu(B)$$

and since $\{\mathcal{B}_\mathcal{A} \colon \mathcal{A} \Subset \mathcal{F}\}$ is a subset of $\mathbb{P}(\mathcal{F})$,

$$\leq \mathrm{S} \underline{\int} f \,\mathrm{d}\mu.$$

Equality of the upper S-integral follows by conjugacy. □

Theorem 4.43. *Let $\mathcal{F}$ be a field on $\mathcal{X}$ and let μ be a probability charge on $\mathcal{F}$. Then $\mathrm{S} \int \bullet \,\mathrm{d}\mu = \mathbf{E}_\mu$, that is, a gamble f on X is μ-integrable if and only if it is S-integrable with respect to μ and in such a case*

$$\mathrm{S} \int f \,\mathrm{d}\mu = \mathbf{E}_\mu(f). \tag{4.18}$$

Proof. Immediately from Theorem 4.42. □

If μ is defined on an ample field, the lower S-integral can be obtained from the S-integral as follows. This is similar to a result by Aeyels and De Cooman [21], which was proved in the context of natural extension of a possibility measure.

Theorem 4.44. *Let $\mathcal{F}$ be an ample field on $\mathcal{X}$ and let μ be a probability charge on $\mathcal{F}$. Then, for any gamble f on X it holds that*

$$\underline{\mathbf{E}}_\mu(f) = \mathrm{S} \underline{\int} f \,\mathrm{d}\mu = \mathrm{S} \int [f]^\downarrow_\mathcal{F} \,\mathrm{d}\mu = \mathbf{E}_\mu([f]^\downarrow_\mathcal{F}), \tag{4.19}$$

where the gamble $[f]^\downarrow_\mathcal{F}$ is defined as

$$[f]^\downarrow_\mathcal{F}(x) := \underline{P}_{[x]_\mathcal{F}}(f) = \inf_{y \in [x]_\mathcal{F}} f(y)$$

for any x in $\mathcal{X}$.

Proof. From Theorem 3.51 it follows that $\underline{P}_A(f) = \underline{P}_A([f]^{\downarrow}_{\mathcal{F}})$ for any $A \in \mathcal{F}$. Now, apply the definition of the lower S-integral to obtain that $\underline{\mathrm{S}\int} f \,\mathrm{d}\mu = \underline{\mathrm{S}\int}[f]^{\downarrow}_{\mathcal{F}} \,\mathrm{d}\mu$. But, $[f]^{\downarrow}_{\mathcal{F}}$ is constant on the atoms of $\mathcal{F}$ by its definition, and hence, by Proposition 4.27, it is $\mathcal{F}$-measurable, and finally, applying Proposition 4.28, we find that $[f]^{\downarrow}_{\mathcal{F}}$ is μ-integrable. Now, apply Theorem 4.43. □

For a general bounded positive charge μ, the lower S-integral satisfies

$$\underline{\mathrm{S}\int} f \,\mathrm{d}\mu = \mu(\mathcal{X})\,\underline{\mathrm{S}\int} f \,\mathrm{d}\tfrac{\mu}{\mu(\mathcal{X})} = \mu(\mathcal{X})\underline{\mathbf{E}}_{\frac{\mu}{\mu(\mathcal{X})}}(f),$$

assuming that $\mu(\mathcal{X}) > 0$. If we call the right hand side the *natural extension of* μ, and if we say that a gamble is *μ-integrable* whenever it is $\frac{\mu}{\mu(\mathcal{X})}$-integrable, then Theorem 4.42, Theorem 4.43, and Theorem 4.44 remain valid for all bounded positive charges.

4.3.6 The Riemann and the Riemann-Stieltjes Integral

Consider a compact interval $\mathcal{X} = [a, b]$ in $\mathbb{R}$ and the Lebesgue measure λ on the Borel field $\mathcal{B}(\mathcal{X})$. The Riemann integral, as defined by Darboux [14, Section II, p. 65], is very much like the S-integral with respect to λ, except that it takes a Moore-Smith limit over the set $\mathbb{S}(\mathcal{X})$ of *subdivisions of* $\mathcal{X}$,

$$\mathbb{S}(\mathcal{X}) := \left\{\left\{[a, x_1],\ [x_1, x_2], \ldots, [x_n, b]\right\} \colon a \leq x_1 \leq \cdots \leq x_n \leq b\right\}, \tag{4.20}$$

instead of over the set of finite partitions $\mathbb{P}(\mathcal{B}(\mathcal{X}))$ on the Borel field $\mathcal{B}(\mathcal{X})$. Indeed, the set $\mathbb{S}(\mathcal{X})$ of subdivisions of $\mathcal{X}$ constitutes a directed set: for any $\mathcal{S}, \mathcal{T} \in \mathbb{S}(\mathcal{X})$, either we define $\mathcal{S} \leq_{|\bullet|} \mathcal{T}$ if $|\mathcal{S}| \leq |\mathcal{T}|$, where $|\mathcal{S}| := \max_{S \in \mathcal{S}} \lambda(S)$, or we define $\mathcal{S} \leq \mathcal{T}$ if $\mathcal{S}$ is a refinement of $\mathcal{T}$, *i.e.*, if for any $S \in \mathcal{S}$ there is a $T \in \mathcal{T}$ such that $T \subseteq S$. Both $\leq$ and $\leq_{|\bullet|}$ are transitive, reflexive, and satisfy the composition property: the set $\mathbb{S}(\mathcal{X})$ constitutes a directed set with respect to both $\leq$ and $\leq_{|\bullet|}$.

Hence, we have two ways to take the Moore-Smith limit [57, p. 103] over $\mathbb{S}(\mathcal{X})$ to calculate lower and upper integrals. Darboux [14, p. 69, l. 23–p. 70, l. 13] showed that it does not matter whether the Moore-Smith limit is taken with respect to $\leq$ or with respect to $\leq_{|\bullet|}$. We, however, shall take the $\leq$-limit, because that is more like the S-integral, and hence, like natural

extension. Note that the original definition of the Riemann integral, by Riemann [66, Section 4, p. 102], involves the $\leq_{|\bullet|}$-limit, and is not formulated through lower and upper integrals. For an extensive discussion of these matters, also see for instance Hildebrandt's book [43]: Chapter II, pp. 27–32, in particular Definition 2.1, Definition 2.2, Theorem 3.2 and Theorem 3.10. For our purposes, it is convenient to take the result of Darboux's extension [14, Section II, pp. 64–71] of Riemann's analysis [66, Sections 4–6, pp. 102–108] as our definition of the Riemann integral, because lower and upper integrals obtained by the Moore-Smith limit with respect to $\leq$ are more closely linked to the lower and upper S-integral, and hence, to natural extension.

Definition 4.45. Let $\mathcal{X} = [a, b]$ be a compact interval in $\mathbb{R}$, and let λ denote the Lebesgue measure on the Borel field $\mathcal{B}(\mathcal{X})$. A gamble f on X is called *Riemann integrable* if and only if the *lower and upper Riemann integral* of f

$$\mathrm{R}\underline{\int_a^b} f(x)\,\mathrm{d}x := \lim_{\mathcal{S}\in\mathbb{S}(\mathcal{X})} \sum_{S\in\mathcal{S}} \underline{P}_S(f)\lambda(S) = \sup_{\mathcal{S}\in\mathbb{S}(\mathcal{X})} \sum_{S\in\mathcal{S}} \underline{P}_S(f)\lambda(S), \tag{4.21}$$

$$\mathrm{R}\overline{\int_a^b} f(x)\,\mathrm{d}x := \lim_{\mathcal{S}\in\mathbb{S}(\mathcal{X})} \sum_{S\in\mathcal{S}} \overline{P}_S(f)\lambda(S) = \inf_{\mathcal{S}\in\mathbb{S}(\mathcal{X})} \sum_{S\in\mathcal{S}} \overline{P}_S(f)\lambda(S) \tag{4.22}$$

coincide. In such a case, the *Riemann integral* of f is defined as

$$\mathrm{R}\int_a^b f(x)\,\mathrm{d}x := \mathrm{R}\underline{\int_a^b} f(x)\,\mathrm{d}x = \mathrm{R}\overline{\int_a^b} f(x)\,\mathrm{d}x. \tag{4.23}$$

The set of Riemann integrable gambles on X is denoted by $\mathcal{L}_{\mathrm{dX}}(X)$.

Proof of existence of the Moore-Smith limits. Similar to the proof for the lower and upper S-integral. □

Note that the lower and upper Riemann integrals are sometimes called the lower and upper Darboux integrals. Again, we have the following interesting property:

Theorem 4.46. *Let $\mathcal{X} = [0, 1]$. The lower Riemann integral on $\mathcal{X}$ is a 2-monotone coherent lower prevision. The lower Riemann integral on $\mathcal{X}$, restricted to $I_{\wp(\mathcal{X})}$, corresponds to a completely monotone set function.*

Proof. Similar to the proof of Theorem 4.41 on p. 129. □

Obviously, Theorem 4.46 extends to arbitrary compact intervals $\mathcal{X} = [a, b]$ in $\mathbb{R}$, after renormalisation.

The Riemann integral can be easily generalised as follows. Subdivisions consist of closed intervals $[x, y]$ with $a \leq x \leq y \leq b$. We know that the Lebesgue measure is the unique (σ-additive) measure on $\mathscr{B}(\mathcal{X})$ such that $\lambda([x, y]) = y - x$ for all $a \leq x \leq y \leq b$. Now, instead, fix a real-valued non-decreasing bounded function F on $\mathcal{X}$ (such functions have very nice properties: in particular $F(x+) := \lim_{\epsilon>0} F(x + \epsilon)$ exists for every $a \leq x < b$ and $F(x-) := \lim_{\epsilon>0} F(x - \epsilon)$ exists for every $a < x \leq b$, and F is continuous except at countably many points of $[a, b]$; see for instance Schechter [70, Proposition 19.22]). Then, as we shall prove below, there is a charge ρ_F on a field that contains all closed intervals, such that $\rho_F([x, y]) = F(y) - F(x)$ for all $a \leq x \leq y \leq b$. If we replace the Lebesgue measure λ in the definition of the Riemann integral by this charge ρ_F then we obtain exactly the Riemann-Stieltjes integral defined below. These integrals will play an important rôle in the natural extension of cumulative distribution functions and p-boxes. Before proceeding with the definition, a few remarks are necessary to motivate why we choose our particular approach.

First of all, contrary to the Riemann integral, the Riemann-Stieltjes integral defined by the (norm based) $\leq_{|\bullet|}$-limit is not equivalent to the Riemann-Stieltjes integral defined by the (refinement based) $\leq$-limit unless F is continuous; see for instance Hildebrandt [43, Chapter II, Theorem 10.9]. As mentioned before, we prefer to take the $\leq$-limit over $\mathbb{S}(\mathcal{X})$ because this construction of the limit is closer to the definition of the S-integral, and hence, to the idea of natural extension. Another reason is that this construction yields a strictly larger set of integrable gambles whenever F is not continuous.

If F is continuous, then there actually exists a unique measure λ_F on the Borel field $\mathscr{B}(\mathcal{X})$ such that $\lambda_F([x, y]) = F(y) - F(x)$ for all $a \leq x \leq y \leq b$. This measure (or, its completion) is called the Lebesgue-Stieltjes measure induced by F; see Halmos [40, Section 15.9]. It can be defined through the Lebesgue measure as $\lambda_F(A) := \lambda(F(A))$ for any $A \in \mathscr{B}(\mathcal{X})$—the continuous image of a Borel set again is a Borel set. But, if F is not continuous, there is no such measure: σ-additivity cannot hold. Indeed, suppose for instance that $F(y+) > F(y)$, with $F(y+) := \lim_{\epsilon>0} F(y + \epsilon) > F(y)$.

$$F(y+) - F(x) = \inf_{n \in \mathbb{N}} F\left(y + \tfrac{1}{n}\right) - F(x) = \inf_{n \in \mathbb{N}} \rho_F\left(\left[x, y + \tfrac{1}{n}\right]\right)$$

$$> \rho_F\left(\cap_{n\in\mathbb{N}}\left[x, y+\tfrac{1}{n}\right]\right) = F(y) - F(x),$$

and a similar strict inequality in case $F(x-) < F(x)$, with $F(x-) := \lim_{\epsilon>0} F(x-\epsilon)$. So, either we have to give up σ-additivity, or we have to give up that $\rho_F([x,y]) = F(y) - F(x)$ for every closed interval $[x,y]$. Since we need that $\rho_F([x,y]) = F(y) - F(x)$ in our definition of the Riemann-Stieltjes integral, we have to give up σ-additivity for ρ_F.

Note that, for instance, Hildebrandt [43, Chapter V, Sections 4–6] introduces a σ-additive Lebesgue-Stieltjes measure λ_F even if F is not continuous (and uses this measure λ_F to define the Young-Stieltjes integral as the lower S-integral with respect to λ_F; see Hildebrandt [43, Chapter VII, Definition 3.3]). However, the equality $\lambda_F([x,y]) = F(x) - F(y)$ does not hold for all closed intervals if F is not continuous; instead, it satisfies $\lambda_F([x,y]) = F(y+) - F(x-)$. The same holds for the Lebesgue-Stieltjes integral introduced by for instance Kallenberg [48, Proposition 2.14, p. 31]. Therefore, these Lebesgue-Stieltjes measures are not attractive candidates for studying the relation between the S-integral and the Riemann-Stieltjes integral—it is not clear how to obtain the Riemann-Stieltjes integral using these Lebesgue-Stieltjes measures. Moreover, why insist on a measure if a charge can do the job more efficiently?

These are some of the complications that we should be aware of. Note however, that ρ_F is uniquely determined on $\mathcal{F}_{[]}(\mathcal{X})$, that is the smallest field generated by the set of closed (or open) intervals in $\mathcal{X}$; this follows from a result by Bhaskara Rao and Bhaskara Rao [9, Theorem 3.2.5, p. 65]. Below, we invoke natural extension to obtain the same result. For our purposes, this charge will be sufficiently general as a replacement for the Lebesgue measure in the Riemann-Stieltjes integral if we do not want to assume F to be continuous. Of course, by the Hahn-Banach theorem, ρ_F can be extended to $\mathcal{B}(\mathcal{X})$ (see for instance Schechter [70, Section 29.32, HB26]), but in general this extension is not unique.

Definition 4.47. Let $\mathcal{X} = [a,b]$ be a compact interval of $\mathbb{R}$, and let F be a real-valued non-decreasing bounded function on $\mathcal{X}$. Define the *Riemann-Stieltjes charge* ρ_F as the unique charge on $\mathcal{F}_{[]}(\mathcal{X})$ such that $\rho_F([x,y]) = F(y) - F(x)$ for every $a \le x \le y \le b$. If F is continuous, define the *Lebesgue-Stieltjes measure* λ_F as the unique σ-additive extension of ρ_F to the Borel σ-field $\mathcal{B}(\mathcal{X})$.

Proof of existence and uniqueness. Again we can demonstrate the power of nat-

ural extension. Assume that $F(b) - F(a) = 1$. Define the probability P by

$$P(I_{[x,y]}) = -P(-I_{[x,y]}) = F(y) - F(x)$$

for every $a \leq x \leq y \leq b$. If this probability avoids sure loss, then there is only one linear prevision Q defined on $\mathrm{cl}(\mathrm{span}(\{I_{[x,y]} \colon a \leq x \leq y \leq b\})) = \mathrm{cl}(\mathrm{span}(\mathcal{F}_{[]}(\mathcal{X}))) = \mathcal{L}_{\mathcal{F}_{[]}(\mathcal{X})}(X)$ that is a behavioural extension of P, through Proposition 4.18(v). But by Theorem 4.30 there is only one charge on $\mathcal{F}_{[]}(\mathcal{X})$ whose linear extension to $\mathcal{L}_{\mathcal{F}_{[]}(\mathcal{X})}(X)$ is Q. This is exactly ρ_F.

To see that P avoids sure loss, consider the following coherent behavioural extension of P to $\mathcal{F}_{[]}(\mathcal{X})$,

$$\mu_F(A) := \sum_{I \in \mathcal{I}_A} F(\sup I) - F(\inf I), \tag{4.24}$$

for any $A \in \mathcal{F}_{[]}(\mathcal{X})$, where $\mathcal{I}_A$ is the (unique and finite) smallest set of disjoint intervals that make up A. It is follows immediately that μ_F is a bounded positive charge, and hence, a probability charge, which extends P. Uniqueness and existence of ρ_F follow as explained above. Note that as a consequence, μ_F is actually equal to ρ_F.

If $F(b) - F(a) \neq 1$ and $F(b) > F(a)$, define $G(x) := \frac{F(x)}{F(b)-F(a)}$ for all $x \in \mathcal{X}$. Since $G(b) - G(a) = 1$, there is a unique probability charge ρ_G on $\mathcal{F}_{[]}(\mathcal{X})$ such that $\rho_G([x, y]) = \frac{F(y)-F(x)}{F(b)-F(a)}$ for all $a \leq x \leq y \leq b$, and hence, $\rho_F := [F(b) - F(a)]\rho_G$ is the unique charge such that $\rho_F([x, y]) = F(y) - F(x)$ for all $a \leq x \leq y \leq b$.

If $F(b) = F(a)$, then $\rho_F([x, y]) = F(y) - F(x) = 0$ for all $a \leq x \leq y \leq b$, extends uniquely to $\rho_F(A) := 0$ for all $A \in \mathcal{F}_{[]}(\mathcal{X})$—by the monotonicity of ρ_F, $\rho_F(\emptyset) = 0$, and $\rho_F(\mathcal{X}) = 0$.

In case F is continuous, $\lambda_F(A) := \lambda(F(A))$ for any $A \in \mathcal{B}(\mathcal{X})$ identifies a measure with the desired properties, whose uniqueness is similar to the uniqueness of the Lebesgue measure on $\mathcal{B}(\mathcal{X})$; see for instance Halmos [40, Section 16.9]. □

Unlike the Riemann-Stieltjes charge we have just defined, the Lebesgue-Stieltjes measure is only defined given that F is continuous. Remark that in case F is continuous ρ_F is usually not equivalent to λ_F: it may happen that $\rho_{F*}(A) < \lambda_{F*}(A)$ for some $A \subseteq \mathcal{X}$. For instance, assume $a < b$, let F be the identity map on $\mathcal{X}$, and let $A = \{x \in \mathcal{X} \colon x \text{ not rational}\}$. Then $\rho_{F*}(A) = 0$ but $\lambda_{F*}(A) = \lambda_F(A) = \lambda(F(A)) = \lambda(A) = b - a > 0$. Hence, by Theorem 4.36(v)

on p. 117, the natural extension of ρ_F does not coincide with the natural extension of λ_F. Of course, ρ_F is equivalent to λ_F on $\mathcal{F}_{[]}(\mathcal{X})$ (the set A in the above example is chosen such that it does not belong to $\mathcal{F}_{[]}(\mathcal{X})$). As a consequence, in the definition below ρ_F can be replaced by λ_F whenever F is continuous.

Definition 4.48. Let $\mathcal{X} = [a, b]$ be a compact interval in $\mathbb{R}$, let F be a real-valued non-decreasing bounded function on $\mathcal{X}$. A gamble f on X is called *Riemann-Stieltjes integrable* with respect to F if and only if the *lower and upper Riemann-Stieltjes integral* of f

$$\text{R-S}\underline{\int_a^b} f(x)\,\mathrm{d}F(x) := \lim_{\mathcal{S}\in\mathbb{S}(\mathcal{X})} \sum_{S\in\mathcal{S}} \underline{P}_S(f)\rho_F(S) = \sup_{\mathcal{S}\in\mathbb{S}(\mathcal{X})} \sum_{S\in\mathcal{S}} \underline{P}_S(f)\rho_F(S), \tag{4.25}$$

$$\text{R-S}\overline{\int_a^b} f(x)\,\mathrm{d}F(x) := \lim_{\mathcal{S}\in\mathbb{S}(\mathcal{X})} \sum_{S\in\mathcal{S}} \overline{P}_S(f)\rho_F(S) = \inf_{\mathcal{S}\in\mathbb{S}(\mathcal{X})} \sum_{S\in\mathcal{S}} \overline{P}_S(f)\rho_F(S) \tag{4.26}$$

coincide. In such a case, the *Riemann-Stieltjes integral* of f is defined as

$$\text{R-S}\int_a^b f(x)\,\mathrm{d}F(x) := \text{R-S}\underline{\int_a^b} f(x)\,\mathrm{d}F(x) = \text{R-S}\overline{\int_a^b} f(x)\,\mathrm{d}F(x). \tag{4.27}$$

The set of Riemann-Stieltjes integrable gambles on X is denoted by $\mathcal{L}_{\mathrm{d}F}(X)$.

Proof of existence of the Moore-Smith limits. Similar to the proof for the lower and upper S-integral. □

The Riemann integral is just the Riemann-Stieltjes integral with respect to the identity map on $\mathcal{X}$. So in the following, we can concentrate our study on the Riemann-Stieltjes integral. Again, we have the following interesting property:

Theorem 4.49. *Let $\mathcal{X} = [a, b]$ be a compact interval in $\mathbb{R}$, and let F be a real-valued non-decreasing bounded function on $\mathcal{X}$ such that $F(b) - F(a) = 1$. The lower Riemann-Stieltjes integral with respect to F is a 2-monotone coherent lower prevision. Restricted to $I_{\wp(\mathcal{X})}$, the lower Riemann-Stieltjes integral with respect to F corresponds to a completely monotone set function.*

Proof. Similar to the proof of Theorem 4.41 on p. 129. □

Theorem 4.49 extends to arbitrary real-valued non-decreasing bounded functions F on $\mathcal{X}$ after renormalisation.

Since F is non-decreasing and bounded, the Riemann-Stieltjes charge ρ_F is guaranteed to be a bounded positive charge. Consequently we can define an S-integral with respect to ρ_F. In the rest of this section, we shall investigate the relation between the Riemann-Stieltjes integral and the S-integral with respect to ρ_F, and with respect to other charges related to ρ_F.

In case $F(b) - F(a) = 1$, ρ_F is a probability charge, and the lower Riemann-Stieltjes integral is the point-wise limit of a net of convex combinations of coherent lower previsions. Hence, up to normalisation, the lower Riemann-Stieltjes integral is a coherent lower prevision, and the upper Riemann-Stieltjes integral is its conjugate. One of the many consequences of this simple observation is that we can restrict our study to the lower Riemann-Stieltjes integral only, since the upper Riemann-Stieltjes integral follows uniquely from it.

In order to link the lower Riemann-Stieltjes integral to the lower S-integral we must somehow be able to convert subdivisions into partitions. Obvious candidates are finite partitions, whose elements belong to a field $\mathcal{F}$ that contains a sufficiently large number of intervals, for instance all closed intervals $[x, y]$ for $a \leq x \leq y \leq b$. Further on, we shall prove that we can replace subdivisions $\mathbb{S}(\mathcal{X})$ by partitions $\mathbb{P}(\mathcal{F})$, as long as every element of $\mathbb{P}(\mathcal{F})$ can be approximated by a sequence in $\mathbb{S}(\mathcal{X})$ and vice versa. If such fields $\mathcal{F}$ exist, then this suggests yet another way to construct the Riemann-Stieltjes integral, directly by means of the S-integral. Theorem 4.52 characterises all fields $\mathcal{F}$ for which we can realise the above-mentioned approximation.

First, we need some results about Riemann-Stieltjes integrable indicator gambles. It is convenient to say that a set A is Riemann-Stieltjes integrable whenever I_A is Riemann-Stieltjes integrable. Roughly stated, the following lemma connects the Riemann-Stieltjes lower and upper integral of sets to the inner and outer set functions induced by ρ_F, and states that a set is Riemann-Stieltjes integrable if and only if the parts, that can neither be contained in nor excluded by closed intervals, are sufficiently small in Riemann-Stieltjes charge.

Recall that a real-valued function F on $\mathbb{R}$ is said to be *continuous on* $A \subseteq \mathbb{R}$ whenever, for every $\epsilon > 0$ and every $x \in A$, there is a $\delta_{\epsilon,x} > 0$ such that, for every $y \in A$, $|x - y| < \delta_{\epsilon,x}$ implies that $|F(y) - F(x)| < \epsilon$; in particular, if F is

continuous at every point of A, then F is continuous on A, but the converse only holds if A is open.

Lemma 4.50. *Let $\mathcal{X} = [a, b]$ be a compact interval in $\mathbb{R}$, and let F be a real-valued non-decreasing bounded function on $\mathcal{X}$. Let $A \subseteq \mathcal{X}$. The following propositions hold.*

(i) It holds that

$$\text{R-S}\underline{\int}_a^b I_A(x)\,\mathrm{d}F(x) \leq \rho_{F*}(A),$$

with equality if F is continuous on $B \cap A$, where $B \in \mathcal{F}_{[]}(\mathcal{X})$ is some set such that $\mathrm{cl}(A \setminus B) \subseteq A$. Similarly,

$$\text{R-S}\overline{\int}_a^b I_A(x)\,\mathrm{d}F(x) \geq \rho_F^*(A),$$

with equality if F is continuous on $B \setminus A$, where $B \in \mathcal{F}_{[]}(\mathcal{X})$ is some set such that $A \subseteq \mathrm{int}(A \cup B)$.

(ii) If F is continuous on some $B \in \mathcal{F}_{[]}(\mathcal{X})$ such that $\mathrm{cl}(A \setminus B) \subseteq A \subseteq \mathrm{int}(A \cup B)$ then

$$\text{R-S}\underline{\int}_a^b I_A(x)\,\mathrm{d}F(x) = \rho_{F*}(A) \quad \textit{and} \quad \text{R-S}\overline{\int}_a^b I_A(x)\,\mathrm{d}F(x) = \rho_F^*(A)$$

(iii) The set A is Riemann-Stieltjes integrable if and only if

$$\lim_{\mathcal{S} \in \mathbb{S}(\mathcal{X})} \sum_{\substack{S \in \mathcal{S} \\ S \not\subseteq A \text{ and } S \not\subseteq \complement A}} \rho_F(S) = 0. \tag{4.28}$$

Hence, in such a case

$$\text{R-S}\int_a^b I_A(x)\,\mathrm{d}F(x) = \rho_{F*}(A) = \rho_F^*(A). \tag{4.29}$$

Proof. Throughout the proof, we shall assume that $F(b) - F(a) = 1$. The general case follows simply by renormalisation.

Observe that $\overline{P}_S(I_A) = 1$ if and only if $S \not\subseteq \complement A$, and $\underline{P}_S(I_A) = 1$ if and only

if $S \subseteq A$. This will be repeatedly used. Also observe that

$$\text{R-S}\underline{\int}_a^b I_A(x)\,\mathrm{d}F(x) = 1 - \text{R-S}\overline{\int}_a^b I_{\complement A}(x)\,\mathrm{d}F(x) \quad \text{and} \quad \rho_{F*}(A) = 1 - \rho_F^*(\complement A).$$

This effectively halves the number of things we need to prove.

(i) To show that $\text{R-S}\underline{\int}_a^b I_A(x)\,\mathrm{d}F(x) \leq \rho_{F*}(A)$ we shall prove that for every subdivision $\mathcal{S} \in \mathbb{S}(\mathcal{X})$ we can find a $B_{\mathcal{S}} \in \mathcal{F}_{[]}(\mathcal{X})$ such that $B_{\mathcal{S}} \subseteq A$ and

$$\sum_{S\in\mathcal{S}} \underline{P}_S(I_A)\rho_F(S) = \sum_{\substack{S\in\mathcal{S}\\ S\subseteq A}} \rho_F(S) \leq \rho_F(B_{\mathcal{S}})$$

In such a case, it will follow that

$$\sum_{S\in\mathcal{S}} \underline{P}_S(I_A)\rho_F(S) \leq \sup\{\rho_F(B)\colon B \in \mathcal{F}_{[]}(\mathcal{X}),\, B \subseteq A\} = \rho_{F*}(A),$$

for any subdivision $\mathcal{S}$, and hence,

$$\sup_{\mathcal{S}\in\mathbb{S}(\mathcal{X})} \sum_{S\in\mathcal{S}} \underline{P}_S(I_A)\rho_F(S) = \text{R-S}\underline{\int}_a^b I_A(x)\,\mathrm{d}F(x) \leq \rho_{F*}(A).$$

Indeed, take $B_{\mathcal{S}} = \underset{\substack{S\in\mathcal{S}\\ S\subseteq A}}{\cup} S$. Since ρ_F is zero on finite sets and the sets $S \in \mathcal{S}$ only overlap on a finite set, it follows from the additivity of the charge ρ_F that

$$\rho_F(B_{\mathcal{S}}) = \rho_F\left(\bigcup_{\substack{S\in\mathcal{S}\\ S\subseteq A}} S\right) = \sum_{\substack{S\in\mathcal{S}\\ S\subseteq A}} \rho_F(S),$$

which implies the desired inequality.

To establish $\text{R-S}\overline{\int}_a^b I_A(x)\,\mathrm{d}F(x) \geq \rho_{F*}(A)$ in case F is continuous on $A \cap C$ for some $C \in \mathcal{F}_{[]}(\mathcal{X})$ such that $\mathrm{cl}(A \setminus C) \subseteq A$, we shall show that for every $B \in \mathcal{F}_{[]}(\mathcal{X})$ such that $B \subseteq A$ we can find a $K_B \geq 0$ (that may depend on B) such that for every $\epsilon > 0$ we can find a subdivision $\mathcal{S}_\epsilon^B \in \mathbb{S}(\mathcal{X})$ such that

$$\sum_{S\in\mathcal{S}_\epsilon^B} \underline{P}_S(I_A)\rho_F(S) = \sum_{\substack{S\in\mathcal{S}_\epsilon^B\\ S\subseteq A}} \rho_F(S) \geq \rho_F(B) - K_B\epsilon$$

We shall then have that

$$\sup_{\mathcal{S}\in\mathbb{S}(\mathcal{S})}\sum_{\substack{S\in\mathcal{S}\\S\subseteq A}}\rho_F(S)=\text{R-S}\underline{\int_a^b} I_A(x)\,\mathrm{d}F(x)\geq\rho_F(B)-K_B\epsilon$$

for any $B\in\mathcal{F}_{[]}(\mathcal{X})$ such that $B\subseteq A$ and any $\epsilon>0$, and hence also for $\epsilon=0$, and therefore

$$\text{R-S}\underline{\int_a^b} I_A(x)\,\mathrm{d}F(x)\geq\sup\{\rho_F(B)\colon B\in\mathcal{F}_{[]}(\mathcal{X}),\,B\subseteq A\}=\rho_{F*}(A).$$

Indeed, fix $\epsilon>0$ and $B\in\mathcal{F}_{[]}(\mathcal{X})$, $B\subseteq A$. Since F is continuous on $A\cap C$, for every $x\in A\cap C$, there is a $\delta_{x,\epsilon}>0$ such that, for every $y\in A\cap C$, $\left|F(x)-F(y)\right|<\epsilon$ whenever $\left|x-y\right|<\delta_{x,\epsilon}$. Since both B and $C\in\mathcal{F}_{[]}(\mathcal{X})$, also $B\setminus C$ and $B\cap C\in\mathcal{F}_{[]}(\mathcal{X})$. Hence, these two disjoint sets, which jointly make up B, both must be a finite union of disjoint non-empty intervals. Let $\mathcal{I}^B$ be a set of disjoint non-empty intervals that make up $B\setminus C$ and let $\mathcal{J}^B$ be a set of disjoint non-empty intervals that make up $B\cap C$, or briefly, $\mathcal{I}^B$ is an interval partition of $B\setminus C$ and $\mathcal{J}^B$ is an interval partition of $B\cap C$ (it may happen that $\mathcal{I}^B$ or $\mathcal{J}^B$ are empty).

We are now going to construct a subdivision $\mathcal{S}_\epsilon^B\in\mathbb{S}(\mathcal{X})$, based on the interval partition $\mathcal{I}^B$ of $B\setminus C$ and the interval partition $\mathcal{J}^B$ of $B\cap C$. Note that the elements of $\mathcal{S}_\epsilon^B\in\mathbb{S}(\mathcal{X})$ must be closed intervals.

For each non-empty interval $I\in\mathcal{I}^B$ define the closed interval $S^I:=\mathrm{cl}(I)$. Since $B\subseteq A$ also $I\subseteq A$ and therefore it holds that

$$S^I=\mathrm{cl}(I)=\mathrm{cl}(I\setminus C)\subseteq\mathrm{cl}(A\setminus C)\subseteq A,$$

where we invoked the special property of C. Obviously also $\rho_F(S^I)=\rho_F(I)$ since they have exactly the same extreme points.

Since $\mathcal{J}^B$ makes up $B\cap C$ and $B\subseteq A$ obviously also $J\subseteq A\cap C$ for all $J\in\mathcal{J}^B$. Unfortunately, J is not necessarily closed, and so it may happen that $\mathrm{cl}\,J\not\subseteq A\cap C$. However, we can choose for every J a non-empty *closed* interval (possibly a singleton) $S_\epsilon^J\subseteq J$ such that

$$\left|\overline{J}-\overline{S}_\epsilon^J\right|<\delta_{\overline{J},\epsilon}\text{ and }\left|\underline{S}_\epsilon^J-\underline{J}\right|<\delta_{\underline{J},\epsilon},\tag{4.30}$$

where $\underline{*}$ and $\overline{*}$ denote the lower and upper end-point (that is, infimum and

supremum) of an interval $*$. Obviously, since $S^J_\epsilon \subseteq J \subseteq A \cap C$ for all $J \in \mathcal{J}^B$ the extreme points of all intervals J and S^J_ϵ belong to $A \cap C$. By the continuity of F on $A \cap C$ and Eq. (4.30) it follows that

$$\left|F(\bar{J}) - F(\overline{S}^J_\epsilon)\right| < \epsilon \text{ and } \left|F(\underline{S}^J_\epsilon) - F(\underline{J})\right| < \epsilon.$$

But this means that

$$\left|\rho_F(J) - \rho_F(S^J_\epsilon)\right| = F(\bar{J}) - F(\overline{S}^J_\epsilon) + F(\underline{S}^J_\epsilon) - F(\underline{J}) = \left|F(\bar{J}) - F(\overline{S}^J_\epsilon)\right| + \left|F(\underline{S}^J_\epsilon) - F(\underline{J})\right| < 2\epsilon.$$

Now clearly all closed intervals S^I for $I \in \mathcal{I}^B$ and S^J_ϵ for $J \in \mathcal{J}^B$ are disjoint, up to a possible finite overlap. Hence, there is a subdivision $\mathcal{S}^B_\epsilon \in \mathbb{S}(\mathcal{X})$ such that

$$\mathcal{S}^B_\epsilon \supseteq \{S^I \colon I \in \mathcal{I}^B\} \cup \{S^J_\epsilon \colon J \in \mathcal{J}^B\}.$$

For this subdivision

$$\begin{aligned}\sum_{\substack{S \in \mathcal{S}^B_\epsilon \\ S \subseteq A}} \rho_F(S) &\geq \sum_{I \in \mathcal{I}^B} \rho_F(S^I) + \sum_{J \in \mathcal{J}^B} \rho_F(S^J_\epsilon) \geq \sum_{I \in \mathcal{I}^B} \rho_F(I) + \sum_{J \in \mathcal{J}^B} [\rho_F(J) - 2\epsilon] \\ &= \rho_F(B \setminus C) + \rho_F(B \cap C) - 2\left|\mathcal{J}^B\right|\epsilon = \rho_F(B) - 2\left|\mathcal{J}^B\right|\epsilon,\end{aligned}$$

and so the desired inequality is satisfied for $K_B = 2\left|\mathcal{J}^B\right| \geq 0$.

(ii) Immediately from (i).

(iii) By definition, I_A is Riemann-Stieltjes integrable if and only if

$$\lim_{\mathcal{S} \in \mathbb{S}(\mathcal{X})} \sum_{S \in \mathcal{S}} \overline{P}_S(I_A)\rho_F(S) = \lim_{\mathcal{S} \in \mathbb{S}(\mathcal{X})} \sum_{S \in \mathcal{S}} \underline{P}_S(I_A)\rho_F(S)$$

Since both Moore-Smith limits exist and are real, this is equivalent to

$$\begin{aligned}\lim_{\mathcal{S} \in \mathbb{S}(\mathcal{X})} \left(\sum_{S \in \mathcal{S}} \overline{P}_S(I_A)\rho_F(S) - \sum_{S \in \mathcal{S}} \underline{P}_S(I_A)\rho_F(S) \right) \\ = \lim_{\mathcal{S} \in \mathbb{S}(\mathcal{X})} \sum_{S \in \mathcal{S}} \left(\overline{P}_S(I_A) - \underline{P}_S(I_A)\right)\rho_F(S) = 0\end{aligned}$$

But $\overline{P}_S(I_A) - \underline{P}_S(I_A) = 1$ if and only if $S \not\subseteq \complement A$ and $S \not\subseteq A$, and in all other cases the difference is zero since $\overline{P}_S \geq \underline{P}_S$. This shows that A is Riemann-Stieltjes integrable if and only if Eq. (4.28) holds. □

It follows from this lemma that, if F is continuous at the extreme points x and y (that is, on a neighbourhood of each of those points), then the intervals $[x, y]$, $[x, y)$, $(x, y]$ and (x, y) are Riemann-Stieltjes integrable and their Riemann-Stieltjes integrals are all equal to $F(y) - F(x)$. Beware if F is not continuous at some point x: for instance, the singleton $\{y\}$ is then not Riemann-Stieltjes integrable, even though $\rho_{F*}(\{y\}) = \rho_F^*(\{y\}) = \rho_F(\{y\}) = 0$.

Let's calculate the lower and upper Riemann-Stieltjes integral of intervals in the general case, not assuming continuity of F, and derive Riemann-Stieltjes integrability conditions for intervals from those expressions. Let $a \leq x \leq y \leq b$. For closed intervals $[x, y]$ (such as singletons), we get, after some manipulation, directly from Definition 4.48 that

$$\text{R-S}\underline{\int_a^b} I_{[x,y]}(z)\,\mathrm{d}F(z) = F(y) - F(x) = \rho_F([x, y]), \tag{4.31}$$

$$\text{R-S}\overline{\int_a^b} I_{[x,y]}(z)\,\mathrm{d}F(z) = F(y+) - F(x-), \tag{4.32}$$

if we agree to let $F(a-) := F(a)$ and $F(b+) := F(b)$. Similarly,

$$\text{R-S}\underline{\int_a^b} I_{(x,y)}(z)\,\mathrm{d}F(z) = F(y-) - F(x+), \tag{4.33}$$

$$\text{R-S}\overline{\int_a^b} I_{(x,y)}(z)\,\mathrm{d}F(z) = F(y) - F(x) = \rho_F((x, y)), \tag{4.34}$$

again, if we agree to let $F(a-) := F(a)$ and $F(b+) := F(b)$. For the other intervals, we have:

$$\text{R-S}\underline{\int_a^b} I_{[x,y)}(z)\,\mathrm{d}F(z) = F(y-) - F(x), \tag{4.35}$$

$$\text{R-S}\overline{\int_a^b} I_{[x,y)}(z)\,\mathrm{d}F(z) = F(y) - F(x-), \tag{4.36}$$

and

$$\text{R-S}\underline{\int_a^b} I_{(x,y]}(z)\,\mathrm{d}F(z) = F(y) - F(x+), \tag{4.37}$$

$$\text{R-S}\overline{\int_a^b} I_{(x,y]}(z)\,\mathrm{d}F(z) = F(y+) - F(x). \tag{4.38}$$

if we once more agree to let $F(a-) := F(a)$ and $F(b+) := F(b)$. Observe that the **lo**wer Riemann-Stieltjes integral coincides with the Riemann-Stieltjes charge ρ_F on c**lo**sed intervals, and the **upp**er Riemann-Stieltjes integral coincides with ρ_F on **op**en intervals.

The countable set $A = \{a + \frac{1}{n} : n \in \mathbb{N}\}$ is Riemann-Stieltjes integrable if F is continuous at every point of A and on $[a, a + \epsilon]$ for some $\epsilon > 0$. In case F is only continuous at every point of A, but not at a itself, then its lower and upper Riemann-Stieltjes integral are given by 0 and $F(a+) - F(a)$. But, even if F is continuous everywhere, not every countable set (and hence, not every Borel set) is Riemann-Stieltjes integrable: for instance, the set $Q = \{x \in [0,1] : x \text{ rational}\}$ is not Riemann integrable. Indeed, because Q is dense in $[0,1]$ every closed interval $S \subseteq [0,1]$ with Lebesgue measure $\lambda(S) > 0$ has the property that $S \not\subseteq Q$ and $S \not\subseteq \complement Q$. The last example clearly shows that the collection of all Riemann-Stieltjes integrable sets is usually not a σ-field. Perhaps surprisingly, it does constitute a field. This is a consequence of Proposition 4.18(i)&(v) on p. 102 given that the lower Riemann-Stieltjes integral has all the properties of a coherent lower prevision after renormalisation. Below, we give an alternative proof. Hildebrandt [43, Chapter V, Theorem 2.6] gives a proof in case F is the identity map, that is, for Riemann integrability.

Proposition 4.51. *Let $\mathcal{X} = [a,b]$ be a compact interval in $\mathbb{R}$ and let F be a real-valued non-decreasing bounded function on $\mathcal{X}$. The collection of all subsets of $\mathcal{X}$ that are Riemann-Stieltjes integrable with respect to F is a field.*

Proof. By Lemma 4.50 the set of all Riemann-Stieltjes integrable sets is closed under complementation, as Eq. (4.28) clearly implies, and contains the empty set. It remains to show that it is also closed under (finite) union. Let A and $B \subseteq \mathcal{X}$ be Riemann-Stieltjes integrable. We must show that $A \cup B$ is Riemann-Stieltjes integrable. The following implication holds:

$$\Big(S \not\subseteq A \cup B \text{ and } S \not\subseteq \complement(A \cup B)\Big) \implies \Big((S \not\subseteq A \text{ and } S \not\subseteq \complement A) \text{ or } (S \not\subseteq B \text{ and } S \not\subseteq \complement B)\Big)$$

Indeed, if $S \not\subseteq A \cup B$ then $S \not\subseteq A$ and $S \not\subseteq B$, and if $S \not\subseteq \complement(A \cup B)$ then $S \not\subseteq \complement A$ or $S \not\subseteq \complement B$. From this implication, it is easy to see that for any subdivision

$\mathcal{S} \in \mathbb{S}(\mathcal{X})$

$$\sum_{\substack{S\in\mathcal{S} \\ S\not\subseteq A\cup B \text{ and } S\not\subseteq (A\cup B)}} \rho_F(S) \leq \sum_{\substack{S\in\mathcal{S} \\ S\not\subseteq A \text{ and } S\not\subseteq A \\ \text{or} \\ S\not\subseteq B \text{ and } S\not\subseteq B}} \rho_F(S)$$

$$\leq \sum_{\substack{S\in\mathcal{S} \\ S\not\subseteq A \text{ and } S\not\subseteq A}} \rho_F(S) + \sum_{\substack{S\in\mathcal{S} \\ S\not\subseteq B \text{ and } S\not\subseteq B}} \rho_F(S).$$

Now take the Moore-Smith limit over $\mathcal{S} \in \mathbb{S}(\mathcal{X})$ and again apply Eq. (4.28) to see that $A \cup B$ is Riemann-Stieltjes integrable. □

Let's now present a theorem that characterises all charges that are, through Theorem 4.42, equivalent to a given lower Riemann-Stieltjes integral, and hence, by conjugacy, also equivalent to the corresponding upper Riemann-Stieltjes integral.

Theorem 4.52. *Let $\mathcal{X} = [a, b]$ be a compact interval in $\mathbb{R}$ and let F be a real-valued non-decreasing bounded function on $\mathcal{X}$. Let $\mathcal{F}$ be any field on $\mathcal{X}$ and let μ be any bounded positive charge on $\mathcal{F}$. Then the lower Riemann-Stieltjes integral with respect to F is equal to the natural extension of μ, that is,*

$$\text{R-S}\underline{\int_a^b} f(x)\,\mathrm{d}F(x) = \underline{\mathrm{E}}_\mu(f) = \text{S}\underline{\int} f\,\mathrm{d}\mu \quad \textit{for all } f \in \mathcal{L}(\mathcal{X}), \tag{4.39}$$

if and only if

(i) every $A \in \mathcal{F}$ is Riemann-Stieltjes integrable and

$$\mu(A) = \text{R-S}\int_a^b I_A(x)\,\mathrm{d}F(x), \textit{ and}$$

(ii) $\mathcal{F}$ is dense in $\mathcal{F}_{[]}(\mathcal{X})$ in the sense that for every closed interval S of $\mathcal{X}$ and every $\epsilon > 0$ there is at least one element $A_\epsilon^S \in \mathcal{F}$ such that

$$A_\epsilon^S \subseteq S \quad \textit{and} \quad \text{R-S}\underline{\int_a^b} I_{A_\epsilon^S \Delta S}(x)\,\mathrm{d}F(x) < \epsilon,$$

or equivalently, for every open interval T of $\mathcal{X}$ and every $\epsilon > 0$ there is at least

one element $B_\epsilon^T \in \mathcal{F}$ *such that*

$$B_\epsilon^T \supseteq T \quad \textit{and} \quad \text{R-S}\overline{\int_a^b} I_{B_\epsilon^T \Delta T}(x)\,\mathrm{d}F(x) < \epsilon.$$

The two inequalities above can also be written as $\left|\mu(A_\epsilon^S) - \rho_F(S)\right| < \epsilon$ resp. $\left|\mu(B_\epsilon^T) - \rho_F(T)\right| < \epsilon$. This is a consequence of Proposition 4.16 on p. 101 along with some properties of the lower and upper Riemann-Stieltjes integral we have just demonstrated; see details in the proof below.

Proof. As usual, we shall assume that $F(b) - F(a) = 1$ throughout the proof, as in that case, $\text{R-S}\underline{\int_a^b} \bullet\,\mathrm{d}F$ is a coherent lower prevision on $\mathcal{L}(X)$. The general case follows after renormalisation. In the marginal case that $F(a) = F(b)$, the proof is immediate.

"if". Fix any gamble f on X. We first prove that, under conditions (i) and (ii), $\text{R-S}\overline{\int_a^b} f(x)\,\mathrm{d}F(x) \leq \text{S}\overline{\int} f\,\mathrm{d}\mu$.

Let $\mathcal{S} \in \mathbb{S}(\mathcal{X})$ be any subdivision of $\mathcal{X}$. We shall prove that we can find a $K_\mathcal{S} \geq 0$, which may depend on $\mathcal{S}$, such that for every $\epsilon > 0$ we can find a partition $\mathcal{B}_\epsilon \in \mathbb{P}(\mathcal{F})$ such that

$$\sum_{S\in\mathcal{S}} \underline{P}_S(f)\rho_F(S) \leq K_\mathcal{S}\epsilon + \sum_{B\in\mathcal{B}_\epsilon} \underline{P}_B(f)\mu(B).$$

If we can do this for every $\epsilon > 0$, then of course also

$$\sum_{S\in\mathcal{S}} \underline{P}_S(f)\rho_F(S) \leq \text{S}\underline{\int} f\,\mathrm{d}\mu,$$

for every subdivision $\mathcal{S} \in \mathbb{S}(\mathcal{X})$, and hence, $\text{R-S}\underline{\int_a^b} f(x)\,\mathrm{d}F(x) \leq \text{S}\overline{\int} f\,\mathrm{d}\mu$.

Fix therefore $\epsilon > 0$. $\mathcal{S}$ is a finite set of closed intervals, so by condition (ii), for each $S \in \mathcal{S}$ there is an $A_\epsilon^S \in \mathcal{F}$ such that $A_\epsilon^S \subseteq S$ and, applying the coherence of $\text{R-S}\underline{\int_a^b} \bullet\,\mathrm{d}F(x)$, $I_\mathcal{F} \subseteq \operatorname{dom} \text{R-S}\underline{\int_a^b} \bullet\,\mathrm{d}F(x)$ (this is condition (i)), Proposition 4.16 on p. 101, the fact that the lower Riemann-Stieltjes integral is equal to the Riemann-Stieltjes charge for closed intervals, $A_\epsilon^S \Delta S = S \setminus A_\epsilon^S$, and $A_\epsilon^S \in \mathcal{F}$,

$$\text{R-S}\underline{\int_a^b} I_{A_\epsilon^S \Delta S}(x)\,\mathrm{d}F(x) = \text{R-S}\underline{\int_a^b} \left[I_S(x) - I_{A_\epsilon^S}(x)\right]\,\mathrm{d}F(x)$$

$$= \text{R-S}\underline{\int_a^b} I_S(x)\,\mathrm{d}F(x) - \text{R-S}\int_a^b I_{A_\epsilon^S}(x)\,\mathrm{d}F(x)$$
$$= \rho_F(S) - \mu(A_\epsilon^S) = \left|\rho_F(S) - \mu(A_\epsilon^S)\right| < \epsilon. \quad (4.40)$$

The finite family $\mathcal{A}_\epsilon := \{A_\epsilon^S : S \in \mathcal{S}\} \subseteq \mathcal{F}$ does not necessarily constitute a partition of $\mathcal{X}$, but, since $\mathcal{F}$ is a field, we do have a finite partition $\mathcal{B}_\epsilon \in \mathbb{P}(\mathcal{F})$ such that each $A_\epsilon^S \in \mathcal{A}_\epsilon$ is a unique (finite) union of elements of $\mathcal{B}_\epsilon$. For instance, take for $\mathcal{B}_\epsilon$ the atoms of the field generated by $\mathcal{A}_\epsilon$ (finite fields are ample fields: Theorem 3.51 on p. 88 applies).

Define $\mathcal{B}_\epsilon^S := \{B \in \mathcal{B}_\epsilon : B \subseteq A_\epsilon^S\}$ for all $S \in \mathcal{S}$. This set identifies all elements of $\mathcal{B}_\epsilon$ which make up A_ϵ^S; note that $\mathcal{B}_\epsilon^S$ can be empty, which happens exactly if A_ϵ^S is the empty set. $\mathcal{B}_\epsilon$ is a finite partition, so for every $S \in \mathcal{S}$ it holds that

$$\epsilon > \left|\rho_F(S) - \mu(A_\epsilon^S)\right| = \left|\rho_F(S) - \mu\left(\cup_{B\in\mathcal{B}_\epsilon^S} B\right)\right| = \left|\rho_F(S) - \sum_{B\in\mathcal{B}_\epsilon^S} \mu(B)\right|$$

and therefore,

$$\sum_{S\in\mathcal{S}} \underline{P}_S(f)\rho_F(S) \leq \sum_{S\in\mathcal{S}} \epsilon \underline{P}_S(f) + \sum_{S\in\mathcal{S}} \sum_{B\in\mathcal{B}_\epsilon^S} \underline{P}_S(f)\mu(B)$$

and since $\underline{P}_S(f) \leq \sup|f|$,

$$\leq |\mathcal{S}| \sup|f|\epsilon + \sum_{S\in\mathcal{S}} \sum_{B\in\mathcal{B}_\epsilon^S} \underline{P}_S(f)\mu(B)$$

and since $B \subseteq A_\epsilon^S \subseteq S$ whenever $B \in \mathcal{B}_\epsilon^S$, it holds that $\underline{P}_S \leq \underline{P}_B$ whenever $B \in \mathcal{B}_\epsilon^S$, so

$$\leq |\mathcal{S}| \sup|f|\epsilon + \sum_{S\in\mathcal{S}} \sum_{B\in\mathcal{B}_\epsilon^S} \underline{P}_B(f)\mu(B),$$

The only two problems left are that some $B \in \mathcal{B}_\epsilon$ might be counted more than once in the above expression, and that some $B \in \mathcal{B}$ might not be counted at all. Fortunately, in both cases $\mu(B)$ is at most linear in ϵ.

If some $B \in \mathcal{B}$ is counted more than once, then it must be that $B \in \mathcal{B}_\epsilon^S \cap \mathcal{B}_\epsilon^T$ for some $S \neq T$ in $\mathcal{S}$. This conditions happens only if $B \subseteq A_\epsilon^S \cap A_\epsilon^T$. But, the

elements of $\mathcal{S}$ overlap only on a finite set. Therefore, since the elements of $\mathcal{A}_\epsilon$ are such that $A_\epsilon^S \subseteq S$ for all $S \in \mathcal{S}$, they can also only overlap on a finite set. Hence, if $B \in \mathcal{B}_\epsilon^S \cap \mathcal{B}_\epsilon^T$ then B must a finite, Riemann-Stieltjes integrable set (not every finite set is Riemann-Stieltjes integrable but B is by construction). Hence, $\mu(B) = \text{R-S}\int_a^b I_B(x)\,\mathrm{d}F(x) \leq \rho_{F*}(B) = 0$, and therefore $\mu(B) = 0$ holds.

On the other hand, if $B \in \mathcal{B}_\epsilon$ is not counted at all, then it must be that $B \notin \mathcal{B}_\epsilon^S$ for all $S \in \mathcal{S}$. But,

$$\sum_{\substack{B\in\mathcal{B}\\ \forall S\in\mathcal{S}:\, B\notin\mathcal{B}_\epsilon^S}} \mu(B) = 1 - \sum_{\substack{B\in\mathcal{B}\\ \exists S\in\mathcal{S}:\, B\in\mathcal{B}_\epsilon^S}} \mu(B) = 1 - \sum_{S\in\mathcal{S}}\sum_{B\in\mathcal{B}_\epsilon^S} \mu(B)$$
$$< 1 - \sum_{S\in\mathcal{S}} \left[\rho_F(S) - \epsilon\right] = |\mathcal{S}|\,\epsilon$$

and hence, if $B \notin \mathcal{B}_\epsilon^S$ for all $S \in \mathcal{S}$ then it can only be that $\mu(B) < |\mathcal{S}|\,\epsilon$.

So, in all cases we find that

$$\sum_{S\in\mathcal{S}}\sum_{B\in\mathcal{B}_\epsilon^S} \underline{P}_B(f)\mu(B) \leq |\mathcal{S}| \sup|f|\epsilon + \sum_{B\in\mathcal{B}} \underline{P}_B(f)\mu(B),$$

and therefore, with $K_{\mathcal{S}} = 2|\mathcal{S}| \sup|f|\epsilon$, the desired inequality is established. Hence, so is $\text{R-S}\underline{\int}_a^b f(x)\,\mathrm{d}F(x) \leq \text{S}\underline{\int} f\,\mathrm{d}\mu$.

Let's now turn to the converse inequality: we prove that, under conditions (i) and (ii), also $\text{R-S}\underline{\int}_a^b f(x)\,\mathrm{d}F(x) \geq \text{S}\underline{\int} f\,\mathrm{d}\mu$.

First, assume that f is a non-negative gamble on X; this simplifies the proof considerably. Let $\mathcal{B} \in \mathbb{P}(\mathcal{F})$ be a finite partition of $\mathcal{X}$. We shall show that we can find a $K_{\mathcal{B},f} \geq 0$, that may depend on $\mathcal{B}$ and on f, such that for every $\epsilon > 0$ sufficiently small, we can find a subdivision $\mathcal{S}_\epsilon \in \mathbb{S}(\mathcal{X})$ such that

$$\sum_{B\in\mathcal{B}} \underline{P}_B(f)\mu(B) \leq K_{\mathcal{B},f}\epsilon + \sum_{S\in\mathcal{S}_\epsilon} \underline{P}_S(f)\rho_F(S).$$

If we can do this for every $\epsilon > 0$ sufficiently small, then

$$\sum_{B\in\mathcal{B}} \underline{P}_B(f)\mu(B) \leq \text{R-S}\underline{\int}_a^b f(x)\,\mathrm{d}F(x),$$

for every $\mathcal{B} \in \mathbb{P}(\mathcal{X})$, and hence, $\text{S}\underline{\int} f\,\mathrm{d}\mu \leq \text{R-S}\underline{\int}_a^b f(x)\,\mathrm{d}F(x)$.

Indeed, fix $\epsilon > 0$, and observe that, by (i), every $B \in \mathcal{B}$ is Riemann-

Stieltjes integrable with respect to F, with Riemann-Stieltjes integral $\mu(B)$. By the definition of Riemann-Stieltjes integrability and Lemma 4.50, there is a subdivision $\mathcal{S}_\epsilon^B \in \mathbb{S}(\mathcal{X})$ such that

$$\sum_{\substack{S\in\mathcal{S}_\epsilon^B \\ S\subseteq B}} \rho_F(S) + \epsilon > \mu(B); \tag{4.41}$$

the left-hand side follows by direct application of the definition of the lower Riemann-Stieltjes integral. Since $\mathbb{S}(\mathcal{X})$ is a directed set and $\mathcal{B}$ is finite, there is a subdivision $\mathcal{S}_\epsilon$ that refines every $\mathcal{S}_\epsilon^B \in \mathbb{S}(\mathcal{X})$ for all $B \in \mathcal{B}$. The above inequality remains valid when replacing all subdivisions by $\mathcal{S}_\epsilon$. Hence,

$$\begin{aligned}
\sum_{B\in\mathcal{B}} \underline{P}_B(f)\mu(B) &\leq \sum_{B\in\mathcal{B}} \underline{P}_B(f)\left(\sum_{\substack{S\in\mathcal{S}_\epsilon \\ S\subseteq B}} \rho_F(S) + \epsilon\right) \\
&\leq |\mathcal{B}|\sup[f]\epsilon + \sum_{B\in\mathcal{B}} \underline{P}_B(f)\left(\sum_{\substack{S\in\mathcal{S}_\epsilon \\ S\subseteq B}} \rho_F(S)\right) \\
&= |\mathcal{B}|\sup[f]\epsilon + \sum_{S\in\mathcal{S}_\epsilon}\left(\sum_{\substack{B\in\mathcal{B} \\ S\subseteq B}} \underline{P}_B(f)\right)\rho_F(S)
\end{aligned}$$

Fix $S \in \mathcal{S}_\epsilon$. Since $\mathcal{B}$ is a partition, either there is exactly one B_0 for which $S \subseteq B_0$, in which case $\sum_{\substack{B\in\mathcal{B} \\ S\subseteq B}} \underline{P}_B(f) = \underline{P}_{B_0}(f) \leq \underline{P}_S(f)$, or there is no B for which $S \subseteq B$, in which case $\sum_{\substack{B\in\mathcal{B} \\ S\subseteq B}} \underline{P}_B(f) = 0 \leq \underline{P}_S(f)$. So,

$$\sum_{S\in\mathcal{S}_\epsilon}\left(\sum_{\substack{B\in\mathcal{B} \\ S\subseteq B}} \underline{P}_B(f)\right)\rho_F(S) \leq \sum_{S\in\mathcal{S}_\epsilon} \underline{P}_S(f)\rho_F(S)$$

The desired inequality follows, with $K_{\mathcal{B},f} = |\mathcal{B}|\sup[f]$, for any non-negative gamble f.

Since $\text{R-S}\underline{\int}_a^b [\alpha + f(x)]\,\mathrm{d}F(x) = \alpha + \text{R-S}\underline{\int}_a^b f(x)\,\mathrm{d}F(x)$ for any $\alpha \in \mathbb{R}$, and similarly for the lower S-integral—both are coherent lower previsions—the inequality $\text{S}\underline{\int} f\,\mathrm{d}\mu \leq \text{R-S}\underline{\int}_a^b f(x)\,\mathrm{d}F(x)$ follows for all gambles f.

"only if". Conversely, assume that the lower Riemann-Stieltjes integral

and the lower S-integral are equal for all gambles f on X. We prove that conditions (i) and (ii) hold.

(i). First, we must show that all elements of $\mathcal{F}$ are Riemann-Stieltjes integrable. We prove this through contraposition.

Suppose $\mathcal{F}$ has bad element A, namely, one that is not Riemann-Stieltjes integrable. Consider the indicator gamble I_A. Obviously, this gamble is $\mathcal{F}$-measurable, and hence, it is S-integrable with respect to μ: its S-integral is simply given by $\mu(A)$. But it is not Riemann-Stieltjes integrable. Hence, the lower and upper Riemann-Stieltjes integral are not equal, and so certainly at least one of them must be different from the S-integral $\mu(A)$. So, we find that $\text{R-S}\underline{\int}_a^b I_A \,\mathrm{d}F \neq \text{S}\underline{\int} I_A \,\mathrm{d}\mu$ or $\text{R-S}\underline{\int}_a^b -I_A \,\mathrm{d}F \neq \text{S}\underline{\int} -I_A \,\mathrm{d}\mu$. We have arrived at a contradiction.

Now we have that all elements of $\mathcal{F}$ must be Riemann-Stieltjes integrable, consider again $A \in \mathcal{F}$. The S-integral of I_A with respect to μ is obviously equal to $\mu(A)$. By assumption, this must be equal to the Riemann-Stieltjes integral of I_A. Hence, $\mu(A) = \text{R-S}\int I_A(x)\,\mathrm{d}F(x)$ holds for all elements A of $\mathcal{F}$.

Finally, suppose there is a closed interval S and an $\epsilon > 0$ such that for every $A \in \mathcal{F}$, $A \not\subseteq S$ or $\left|\mu(A) - \rho_F(S)\right| \geq \epsilon$, or equivalently, if $A \subseteq S$ then $\left|\mu(A) - \rho_F(S)\right| \geq \epsilon$. Consider the gamble I_S. Since S is a closed interval, the lower Riemann-Stieltjes integral of I_S with respect to F is equal to $\rho_F(S)$. But the lower S-integral of I_S with respect to μ is bounded away from $\rho_F(S)$ by at least ϵ. Indeed,

$$\left|\text{R-S}\underline{\int}_a^b I_S(x)\,\mathrm{d}F(x) - \text{S}\underline{\int} I_S \,\mathrm{d}\mu\right| = \left|\rho_F(S) - \sup_{\mathcal{B}\in\mathbb{P}(\mathcal{F})} \sum_{A\in\mathcal{B}} \underline{P}_A(I_S)\mu(A)\right|$$
$$= \left|\rho_F(S) - \sup_{\mathcal{B}\in\mathbb{P}(\mathcal{F})} \sum_{\substack{A\in\mathcal{B}\\ A\subseteq S}} \mu(A)\right|$$

and, since μ is additive, for any $\mathcal{B}$ in $\mathbb{P}(\mathcal{F})$ there is an A' in $\mathcal{F}$, namely, $A' := \cup_{A\in\mathcal{B}, A\subseteq S}$ such that $\sum_{A\in\mathcal{B}, A\subseteq S} \mu(A) = \mu(A')$, and conversely, for every A' in $\mathcal{F}$ there is a partition $\mathcal{B}$ in $\mathbb{P}(\mathcal{F})$, namely, $\mathcal{B} := \{A', \complement A'\}$, such that $\mu(A') = \sum_{A\in\mathcal{B}, A\subseteq S} \mu(A)$. So,

$$= \left| \rho_F(S) - \sup_{\substack{A\in\mathcal{F}\\ A\subseteq S}} \mu(A) \right|$$

$$= \left| \inf_{\substack{A\in\mathcal{F}\\ A\subseteq S}} [\rho_F(S) - \mu(A)] \right|$$

and since, as we demonstrated before, for each $A \in \mathcal{F}$ such that $A \subseteq S$, we have that $\rho_F(S) - \mu(A) = \text{R-S}\underline{\int}_a^b I_{A\Delta S}(x)\,\mathrm{d}F(x) \geq 0$ (see Eq. (4.40) on p. 147),

$$= \inf_{\substack{A\in\mathcal{F}\\ A\subseteq S}} [\rho_F(S) - \mu(A)] \geq \epsilon.$$

Again we have arrived at a contradiction. □

Condition (i) saying that every element of $\mathcal{F}$ must be Riemann-Stieltjes integrable and that the charge of each of these sets $A \in \mathcal{F}$ must be equal to its Riemann-Stieltjes integral, is obvious. What is striking is that we only have to add that $\mathcal{F}$ is dense in $\mathcal{F}_{[]}(\mathcal{X})$, as in condition (ii). This weak assumption is sufficient for the linear extension of μ to coincide with the Riemann-Stieltjes integral, and the natural extension of μ to coincide with the lower Riemann-Stieltjes integral. Of course, it has still not been proved that there actually exist such μ. That is the subject of the following important theorem, which gives necessary and sufficient conditions on F for the existence of charges μ that satisfy the conditions of Theorem 4.52. Loosely speaking, it says that at every point F should be continuous from at least one side. In many practical cases this is satisfied.

Perhaps it is instructive to note that by Theorem 4.52(i) and Lemma 4.50 the only candidates for μ are restrictions of ρ_{F*} or ρ_F^* to fields of Riemann-Stieltjes integrable sets. The theorem suggests a very simple restriction.

Theorem 4.53. *Let $\mathcal{X} = [a, b]$ be a compact interval in $\mathbb{R}$ and let F be a real-valued non-decreasing bounded function on $\mathcal{X}$. Then there is a field $\mathcal{F}$ on $\mathcal{X}$ and a bounded positive charge μ on $\mathcal{F}$, such that the lower Riemann-Stieltjes integral with respect to F is equal to the lower S-integral with respect to μ if and only if for every $a < x < b$ either $F(x+) = F(x)$ or $F(x-) = F(x)$. In that case, such a charge μ is given by the restriction of ρ_F to the field of those sets in $\mathcal{F}_{[]}(\mathcal{X})$ that are Riemann-Stieltjes integrable with respect to F.*

Proof. "if". Suppose F satisfies the one-sided continuity condition at every point $a < x < b$. Define the charge μ as suggested, that is, the restriction of ρ_F to the field of Riemann-Stieltjes integrable elements, with respect to F, in $\mathcal{F}_{[]}(\mathcal{X})$. We check that μ satisfies the conditions of Theorem 4.52. Every element of $\mathcal{F}$ is Riemann-Stieltjes integrable, and by Lemma 4.50, $\rho_F(A) = \text{R-S}\int I_A(x)\,\mathrm{d}F(x)$ for every Riemann-Stieltjes integrable set $A \in \mathcal{F}_{[]}(\mathcal{X})$: so condition (i) is satisfied.

Now comes a subtle part. We shall need the following result: if F is continuous at x from the right, then for every $\epsilon > 0$, there is an $x_\epsilon > x$ such that F is continuous at x_ϵ (from both sides) and $F(x_\epsilon) - F(x) < \epsilon$. Indeed, if F is continuous at x from the right, then for every $\epsilon > 0$ there is a δ_ϵ such that $F(z) - F(x) < \epsilon$ whenever $x < z < x + \delta_\epsilon$. But, since F is non-decreasing, there are only a countable number of points at which F is not continuous (see for instance Schechter [70, Proposition 19.22]). Hence, since $(x, x + \delta_\epsilon)$ is uncountable, there must be an $x_\epsilon \in (x, x + \delta_\epsilon)$ at which F is continuous: this x_ϵ satisfies $F(x_\epsilon) - F(x) < \epsilon$ as required.

Similarly, we see that if F is continuous at x from the left, then for every $\epsilon > 0$, there is an $x_\epsilon < x$ such that F is continuous at x_ϵ and $F(x) - F(x_\epsilon) < \epsilon$. Now, we are in a comfortable situation to check condition (ii) of Theorem 4.52.

Let $\epsilon > 0$, and let $S = [x, y]$ be any closed interval in $\mathcal{X}$. Define $A^S_\epsilon = [x_\epsilon, y_\epsilon]$ as follows. If $x = a$ or if F is continuous at $x > a$ from the left, take $x_\epsilon := x$; otherwise, F must be continuous at x from the right and we can take $x_\epsilon > x$ such that $F(x_\epsilon) - F(x) < \epsilon$, and (by the above result) F is continuous at x_ϵ. In both cases it holds that $F(x_\epsilon) - F(x) < \epsilon$, and either F is continuous at x_ϵ from the left or $x_\epsilon = a$.

Similarly, if $y = b$ or if F is continuous at $y < b$ from the right, take $y_\epsilon := y$; otherwise, F must be continuous at y from the left and we can take $y_\epsilon < y$ such that $F(y) - F(y_\epsilon) < \epsilon$, and F is continuous at y_ϵ. Again, in both cases it holds that $F(y) - F(y_\epsilon) < \epsilon$, and either F is continuous at y_ϵ from the right or $y_\epsilon = b$.

Hence, it holds that $A^S_\epsilon \subseteq S$ by definition of x_ϵ and y_ϵ. Also, by Eqs. (4.31) and (4.32) on p. 143, A^S_ϵ is Riemann-Stieltjes integrable since either $x_\epsilon = a$ or F is continuous at x_ϵ from the left, and either $y_\epsilon = b$ or F is continuous at y_ϵ from the right. Finally,

$$\left|\mu(A^S_\epsilon) - \rho_F(S)\right| = \left|\rho_F(A^S_\epsilon) - \rho_F(S)\right| = F(x_\epsilon) - F(x) + F(y) - F(y_\epsilon) < 2\epsilon,$$

since μ is a restriction of ρ_F. By Eq. (4.40) on p. 147, this establishes condition (ii) of Theorem 4.52.

"only if". Suppose there is an $a < x_0 < b$ such that $F(x_0-) < F(x_0) < F(x_0+)$, and assume *ex absurdo* that there is a bounded positive charge ν defined on some field $\mathcal{F}$ on $\mathcal{X}$ such that the lower Riemann-Stieltjes integral with respect to F is equal to the lower S-integral with respect to ν. By Theorem 4.52(i), we already have that $\mathcal{F}$ contains only Riemann-Stieltjes integrable sets, and $\mathrm{S}\int I_A \,\mathrm{d}\nu = \nu(A) = \text{R-S}\int_a^b I_A \,\mathrm{d}F$ for all $A \in \mathcal{F}$. So, consider for instance the closed interval $A = [x_0, b]$, which is not Riemann-Stieltjes integrable by Eqs. (4.31)&(4.32) on p. 143 (and the fact that $F(x_0-) \neq F(x_0)$), and therefore does not belong to $\mathcal{F}$. We have proved before that

$$\text{R-S}\underline{\int}_a^b I_A(x)\,\mathrm{d}F(x) = F(b) - F(x_0);$$

see Eq. (4.31) on p. 143. We shall obtain a contradiction by showing that $\mathrm{S}\underline{\int} I_A \,\mathrm{d}\nu$ is not equal to $F(b) - F(x_0)$.

Define μ as the restriction of ρ_{F*} to the field $\mathcal{H}$ of Riemann-Stieltjes integrable sets; note that $\mathcal{F} \subseteq \mathcal{H}$ since $\mathcal{F}$ contains only Riemann-Stieltjes integrable sets. By Lemma 4.50, $\mu(B) = \rho_{F*}(B) = \text{R-S}\int_a^b I_B \,\mathrm{d}F$ for all Riemann-Stieltjes integrable sets B, and hence, also for all $B \in \mathcal{F}$, since $\mathcal{F} \subseteq \mathcal{H}$. Therefore, since ν satisfies Theorem 4.52(i), $\nu(B) = \mu(B)$ for all $B \in \mathcal{F}$: $\mathbf{P}_\mu$ is a behavioural extension of $\mathbf{P}_\nu$. Hence, by Proposition 4.7 on p. 98 and Theorem 4.42 on p. 130,

$$\mathrm{S}\underline{\int} I_A \,\mathrm{d}\nu \leq \mathrm{S}\underline{\int} I_A \,\mathrm{d}\mu = \underline{\mathbf{E}}_\mu(I_A) = \mu_*(A) = \sup_{B\in\mathcal{H},\, B\subseteq A} \mu(B)$$

where we also invoked Theorem 4.36(v) on p. 117. Since, as we shall prove below, for every $B \in \mathcal{H}$, *i.e.*, every Riemann-Stieltjes integrable set B, such that $x_0 \in B$, there is an $\epsilon > 0$ such that $[x_0 - \epsilon, x_0] \subseteq B$, it follows that if $B \in \mathcal{H}$ and $B \subseteq A = [x_0, b]$, then $x_0 \notin B$, *i.e.*, $B \subseteq (x_0, b]$:

$$= \sup_{B\in\mathcal{H},\, B\subseteq(x_0,b]} \mu(B)$$

Observe that $B \in \mathcal{H}$ and $B \subseteq (x_0, b]$ imply that $\mu(B) = \text{R-S}\int_a^b I_B(x)\,\mathrm{d}F(x) \leq \text{R-S}\underline{\int}_a^b I_{(x_0,b]}(x)\,\mathrm{d}F(x) = F(b) - F(x_0+)$, where we used Eq. (4.37) on p. 143, so

$$\begin{aligned} &\leq F(b) - F(x_0+) \\ &< F(b) - F(x_0) = \text{R-S}\underline{\int}_a^b I_A(x)\,\mathrm{d}F(x). \end{aligned}$$

We have arrived at a contradiction.

It still remains to prove that if B is Riemann-Stieltjes integrable and $x_0 \in B$, then there must be an $\epsilon > 0$ such that $[x_0 - \epsilon, x_0] \subseteq B$; recall that $F(x_0-) < F(x_0) < F(x_0+)$.

Indeed, assume *ex absurdo* that $[x_0 - \epsilon, x_0] \not\subseteq B$ for all $\epsilon > 0$. Obviously, $[x_0, x_0 + \delta] \not\subseteq \bigcap B$ for all $\delta > 0$, since $x_0 \notin \bigcap B$. Now, since B is Riemann-Stieltjes integrable, it follows by Lemma 4.50(iii) that

$$0 = \lim_{\mathcal{S} \in \mathbb{S}(\mathcal{X})} \sum_{\substack{S \in \mathcal{S} \\ S \not\subseteq B \text{ and } S \not\subseteq \bigcap B}} \rho_F(S)$$

and by restricting the sum to closed intervals that contain x_0,

$$\geq \lim_{\epsilon > 0, \delta > 0} [\rho_F([x_0, x_0 - \epsilon]) + \rho_F([x_0, x_0 + \delta])] = F(x_0+) - F(x_0-) > 0,$$

a contradiction. □

Whenever there are points at which F is discontinuous from both sides, the lower Riemann-Stieltjes integral does not have a lower S-integral representation. It can therefore not be written as the natural extension of some bounded positive charge. In such a case, we could try to do one of the following.

(i) For some applications, it may be good enough to have only equivalence with respect to a restricted set of gambles.

(ii) We could try and find a sequence F_n approximating F such that the sequence $\text{R-S}\underline{\int}_a^b f(x)\,\mathrm{d}F_n(x)$ converges (in limit inferior) to $\text{R-S}\underline{\int}_a^b f(x)\,\mathrm{d}F(x)$ for all gambles f on X. If possible there will be also a sequence of

bounded positive charges μ_n such that $\mathrm{S}\underline{\int} f \,\mathrm{d}\mu_n$ converges (in limit inferior) to $\text{R-S}\underline{\int}_a^b f(x)\,\mathrm{d}F(x)$ by Theorem 4.53.

(iii) Perhaps $\text{R-S}\underline{\int}_a^b \bullet\,\mathrm{d}F(x)$ is equivalent to a more general uncertainty structure. For instance, we know that the lower Riemann-Stieltjes integral is the point-wise limit of the natural extension of belief functions (up to normalisation, as always). Hence, we could try to find a belief function, or completely monotone set function, as a representation for the Riemann-Stieltjes integral. However, in many cases we do have a probability charge representation, so belief functions and completely monotone set functions probably are an overkill.

(iv) In a similar spirit we could try establishing equivalence to a lower envelope or a convex mixture of lower S-integrals with respect to ρ_F and related charges.

Let's end this section on Riemann-Stieltjes integrals with some remarks. First of all, the restriction of ρ_{F*} suggested by Theorem 4.53 is usually far from being restricted to the smallest possible field.

In case F is continuous, other choices that satisfy the conditions of Theorem 4.52 are for instance the smallest field that contains $\{[a,x]\colon x \in \mathcal{X}\}$. We shall denote it by $\mathcal{F}_{[)}(\mathcal{X})$. It consists of all finite unions of intervals of the type $[a,x]$ and $(x,y]$ for $a \leq x < y \leq b$. The field of all complements of elements of $\mathcal{F}_{[)}(\mathcal{X})$ will be denoted by $\mathcal{F}_{(]}(\mathcal{X})$, and this one contains all finite unions of intervals of the type $[x,y)$ and $[y,b]$ for $a \leq x < y \leq b$. The smallest field that contains both of these fields is exactly $\mathcal{F}_{[]}(\mathcal{X})$ which contains all finite unions of intervals (this choice is suggested in Theorem 4.53). The fields $\mathcal{F}_{[)}(\mathcal{X})$ and $\mathcal{F}_{(]}(\mathcal{X})$ have the benefit over $\mathcal{F}_{[]}(\mathcal{X})$ that they have fewer elements.

Still in case F is continuous, we can consider even smaller fields, even fields with only a countable number of elements. For instance, let A be any dense subset of $\mathcal{X}$. Then any field generated by only $\{[a,x]\colon x \in A\}$ will do. If A is the set of rational numbers in $\mathcal{X}$, then this field is countable. Contrast the partitions generated by these fields with the subdivisions of $\mathcal{X}$ defined in Eq. (4.20).

In case F is the identity map, Theorem 4.52 says that *the natural extension of the Lebesgue measure restricted to any of these fields will give you the lower Riemann integral*, not only for Riemann integrable gambles, but for any gamble. If it

had not been proved, such statement would be hardly believable.

Corollary 4.54. *Let $\mathcal{X} = [a, b]$ be a compact interval in $\mathbb{R}$. The set of bounded positive charges that are equivalent to the lower Riemann integral on X is exactly given by all restrictions of the Lebesgue measure λ on $\mathcal{B}(\mathcal{X})$ to a field $\mathcal{F}$ containing only Riemann integrable sets and satisfying the condition that $\mathcal{F}$ is dense (as in Theorem 4.52(ii)).*

In case F is only continuous from at least one side at every point, it mostly suffices to restrict any of those fields to those sets that are Riemann-Stieltjes integrable, as suggested by Theorem 4.53. Note that one of the reasons why this is possible is because, independently of F, the number of intervals that are not Riemann-Stieltjes integrable is at most countable. Theorem 4.53 exactly proves that the remaining set of intervals is sufficiently dense in $\mathcal{F}_{[]}(\mathcal{X})$.

One more important consequence of Theorem 4.52, is that the Riemann-Stieltjes charge really lives up to its name if F is continuous, and moreover, that we might as well take the S-integral with respect to the Lebesgue-Stieltjes measure for calculating the Riemann-Stieltjes integral (but not for the lower and upper Riemann-Stieltjes integral).

Proposition 4.55. *Let $\mathcal{X} = [a, b]$ be a compact interval in $\mathbb{R}$, and let F be a real-valued non-decreasing bounded function on $\mathcal{X}$. Assume that F is continuous. Then for any gamble f on X*

$$\text{R-S}\underline{\int_a^b} f(x)\,\mathrm{d}F(x) = \text{S}\underline{\int} f\,\mathrm{d}\rho_F.$$

Hence, f is Riemann-Stieltjes integrable with respect to F if and only if it is S-integrable with respect to the Riemann-Stieltjes charge ρ_F. In such a case, f is also S-integrable with respect to the Lebesgue-Stieltjes measure λ_F and

$$\text{R-S}\int_a^b f(x)\,\mathrm{d}F(x) = \text{S}\int f\,\mathrm{d}\rho_F = \text{S}\int f\,\mathrm{d}\lambda_F.$$

Proof. The equality $\text{R-S}\underline{\int_a^b} \bullet\,\mathrm{d}F(x) = \text{S}\underline{\int} \bullet\,\mathrm{d}\rho_F$ follows from Theorem 4.53; note that, since F is continuous, all sets in $\overline{\mathcal{F}}_{[]}(\mathcal{X})$ are Riemann-Stieltjes integrable with respect to F (see Eqs. (4.31)–(4.38) on pp. 143–143). It remains to show that $\text{S}\int \bullet\,\mathrm{d}\rho_F = \text{S}\int \bullet\,\mathrm{d}\lambda_F$ on $\operatorname{dom}\left(\text{S}\int \bullet\,\mathrm{d}\rho_F\right)$, *i.e.*, that $\text{S}\int \bullet\,\mathrm{d}\lambda_F$ is an extension of $\text{S}\int \bullet\,\mathrm{d}\rho_F$.

By definition of the Lebesgue-Stieltjes measure (see Definition 4.47 on p. 135), λ_F is the unique σ-additive extension of ρ_F to the Borel σ-field $\mathcal{B}(\mathcal{X})$; hence, λ_F coincides with ρ_F on $\operatorname{dom}\rho_F = \mathcal{F}_{[]}(\mathcal{X})$. Consequently, $\mathbf{P}_{\lambda_F}$ is a behavioural extension of $\mathbf{P}_{\rho_F}$. So, by Proposition 4.15 on p. 101, $\mathbf{E}_{\mathbf{P}_{\lambda_F}} = \mathbf{E}_{\lambda_F}$ is an extension of $\mathbf{E}_{\mathbf{P}_{\rho_F}} = \mathbf{E}_{\rho_F}$: $\operatorname{dom}\mathbf{E}_{\rho_F} \subseteq \operatorname{dom}\mathbf{E}_{\lambda_F}$ and $\mathbf{E}_{\rho_F}(f) = \mathbf{E}_{\lambda_F}(f)$ for every gamble f in $\operatorname{dom}\mathbf{E}_{\rho_F}$. But, the S-integral coincides with the linear extension by Theorem 4.43 on p. 131; this yields the desired equality. □

To see why not every gamble that is S-integrable with respect to the Lebesgue measure is also Riemann integrable, consider the set

$$A := \{x \in \mathcal{X} \colon x \text{ is rational}\}. \tag{4.42}$$

Obviously, I_A is S-integrable with respect to the Lebesgue measure λ on the Borel σ-field $\mathcal{B}(\mathcal{X})$. Indeed, since A is countable it belongs to $\mathcal{B}(\mathcal{X})$, so I_A is $\mathcal{B}(\mathcal{X})$-measurable (it is even $\mathcal{B}(\mathcal{X})$-simple). Now apply Proposition 4.28 and Theorem 4.43 to find that I_A is S-integrable with respect to λ.

But it is well-known that I_A is not Riemann integrable. Indeed, since the rational numbers are dense in $\mathbb{R}$ it holds that $\underline{P}_B(I_A) = 0$ and $\overline{P}_B(I_A) = 1$ for every interval B of $\mathcal{X}$ that is larger than a singleton, and therefore also for every element B of a partition in $\mathbb{P}(\mathcal{F}_{[]}(\mathcal{X}))$ (except for finite sets, but these sets have charge zero so they do not contribute to the lower and upper integral). Hence, the lower Riemann integral of I_A is equal to zero, whereas the upper Riemann integral of I_A is equal to one, which means that I_A is not Riemann integrable.

The following characterisation of Riemann integrability, given for the sake of completeness, is apparently due to Lebesgue. Note that $\operatorname{dom}\overline{\lambda}$, the domain of the completion of the Lebesgue measure, constitutes a σ-field. Sets in the domain of $\overline{\lambda}$ are called *Lebesgue measurable sets*; see Halmos [40, Section 15].

Theorem 4.56. *Let $\mathcal{X} = [a, b]$ be a compact interval in $\mathbb{R}$. A gamble f on X is Riemann integrable if and only if it is $\operatorname{dom}\overline{\lambda}$-measurable and continuous almost everywhere, that is,*

$$\lambda^*(\{x \in [a, b] \colon f \text{ not continuous at } x\}) = 0.$$

Proof. See for instance Schechter [70, Theorem 24.46]. □

In the light of our previous results, I find this quite surprising. Indeed, let $\mathcal{F}$ be any field such that the conditions of Theorem 4.52 are satisfied; this is a subset of $\operatorname{dom}\overline{\lambda}$. Then Riemann integrability turns out to lie in between $\mathcal{F}$-measurability and $\operatorname{dom}\overline{\lambda}$-measurability:

$$\mathcal{L}_{\mathcal{F}}(X) \subseteq \mathcal{L}_{\mathrm{d}X}(X) \subseteq \mathcal{L}_{\operatorname{dom}\overline{\lambda}}(X).$$

From Theorem 4.56, we might also want to remember that $\mathcal{F}$-measurability implies continuity almost everywhere, for any field $\mathcal{F}$ that satisfies the conditions of Theorem 4.52.

4.3.7 Natural Extension of Cumulative Distribution Functions

In this section we shall be concerned with the natural extension of cumulative distribution functions F, which are self-conjugate p-boxes (F, F). Using the results of the previous section about Riemann-Stieltjes integrals and its equivalence to restrictions of the Riemann-Stieltjes charge to a particular class of fields, we shall easily establish necessary and sufficient conditions for the natural extension of a cumulative distribution function to be equal to the lower Riemann-Stieltjes integral with respect to F.

Let's first characterise the natural extension of F in terms of the lower S-integral.

Definition 4.57. Let $\mathcal{X} = [a, b]$ be a compact interval of $\mathbb{R}$, and let F be a coherent cumulative distribution function on $\mathcal{X}$. The unique probability charge μ_F on $\mathcal{F}_{\langle\rangle}(\mathcal{X})$ that is equivalent to F, *i.e.*, such that $\mu_F([a, x]) = F(x)$ for all $x \in \mathcal{X}$, is called the *probability charge induced by* F.

Proof of existence and uniqueness. Consider the linear extension $\mathbf{E}_F$ of F. By definition, this is the linear extension of the coherent probability $\mathbf{P}_F$ induced by F: for all $x \in \mathcal{X}$, $\mathbf{P}_F(I_{[a,x]}) = -\mathbf{P}_F(-I_{[a,x]}) = F(x)$. Through Proposition 4.18(i)&(v) on p. 102, $\mathbf{E}_F$ is defined on at least $\operatorname{cl}(\operatorname{span}(\{I_{[a,x]} \colon x \in \mathcal{X}\})) = \mathcal{L}_{\mathcal{F}_{\langle\rangle}(\mathcal{X})}(X)$, since $\mathcal{F}_{\langle\rangle}(\mathcal{X})$ is the field generated by $\{[a, x] \colon x \in \mathcal{X}\}$. So we can define the charge

$$\mu_F(A) := \mathbf{E}_F(I_A)$$

for all sets A in the field $\mathcal{F}_{\langle\rangle}(\mathcal{X})$ (which is the field generated by the collection

$\{[a, x] : x \in \mathcal{X}\}$).

Clearly $\mathbf{P}_{\mu_F}$ is a coherent behavioural extension of $\mathbf{P}_F$ and $\operatorname{dom}\mathbf{P}_{\mu_F} \subseteq \operatorname{dom}\mathbf{E}_F$. Hence, by Proposition 4.14 on p. 100, it follows that $\mathbf{P}_{\mu_F}$ is equivalent to $\mathbf{P}_F$, or equivalently, μ_F is equivalent to F. It's uniqueness follows from Proposition 4.13 on p. 100: every charge κ on $\mathcal{F}_{\{\}}(\mathcal{X})$ such that $\kappa([a, x]) = F(x)$ for all $x \in \mathcal{X}$, induces a coherent behavioural extension of $\mathbf{P}_F$. Hence, $\mathbf{P}_\kappa$ coincides with $\mathbf{E}_F$ on $\operatorname{dom}\mathbf{E}_F \cap \operatorname{dom}\mathbf{P}_\kappa$. But, $\operatorname{dom}\mathbf{P}_\kappa = \operatorname{dom}\mathbf{P}_{\mu_F} \subseteq \operatorname{dom}\mathbf{E}_F$, so, it follows that $\kappa(A) = \mathbf{E}_F(I_A) = \mu_F(A)$ for all $A \in \mathcal{F}_{\{\}}(\mathcal{X})$: κ must be equal to μ_F. □

Let us emphasise that μ_F is *not* the Riemann-Stieltjes charge ρ_F (but, it is true that ρ_F is a behavioural extension of μ_F). In general, μ_F is not even equivalent to ρ_F—unless F is continuous. Since μ_F is equivalent to F by definition, and the lower S-integral with respect to μ_F is the natural extension of μ_F as stated in Theorem 4.43 on p. 131, we immediately have that $\underline{\mathrm{S}\int} \bullet \,\mathrm{d}\mu_F$ is the natural extension of $\mathbf{P}_F$. Alternatively, note that $\mathbf{P}_{\mu_F}$ is the natural extension of $\mathbf{P}_F$ to $I_{\mathcal{F}_{\{\}}(\mathcal{X})} \cup -I_{\mathcal{F}_{\{\}}(\mathcal{X})}$, and $\mathrm{S}\underline{\int} \bullet \,\mathrm{d}\mu_F$ is the natural extension of $\mathbf{P}_{\mu_F}$; therefore, by Corollary 4.9 on p. 98, $\mathrm{S}\underline{\int} \bullet \,\mathrm{d}\mu_F$ is also the natural extension of $\mathbf{P}_F$.

Theorem 4.58. *Let $\mathcal{X} = [a, b]$ be a compact interval of $\mathbb{R}$, and let F be a coherent cumulative distribution function on $\mathcal{X}$, inducing the probability charge μ_F. Then*

$$\underline{\mathbf{E}}_F(f) = \mathrm{S}\underline{\int} f \,\mathrm{d}\mu_F \qquad \text{for any gamble } f \text{ on } X.$$

The natural extension $\underline{\mathbf{E}}_F$ for a singleton $\{x\}$ with $a < x \leq b$ is given by

$$\underline{\mathbf{E}}_F(I_{\{x\}}) = \sup_{\epsilon \geq 0,\, A \in \mathcal{F}_{\{\}}(\mathcal{X}),\, A \subseteq \{x\}} \mu_F(A) = \mu_F(\emptyset) = 0,$$
$$\overline{\mathbf{E}}_F(I_{\{x\}}) = \inf_{\epsilon \geq 0,\, (x-\epsilon, x] \supseteq \{x\}} \mu_F((x - \epsilon, x]) = F(x) - F(x-),$$

and for $x = a$ we have $\underline{\mathbf{E}}_F(I_{\{a\}}) = \overline{\mathbf{E}}_F(I_{\{a\}}) = F(a)$. These expressions are easily obtained by invoking the inner and outer set functions (defined in Section 4.3.4) with respect to μ_F as a means of calculating the natural extension of μ_F, and hence, F, to events. Taking into account linearity, *i.e.*, Proposition 4.16 on

p. 101, we have for $a < x \leq y \leq b$ that

$$\underline{\mathbf{E}}_F(I_{[x,y]}) = \mathbf{E}_F(I_{[a,y]}) - \mathbf{E}_F(I_{[a,x]}) + \underline{\mathbf{E}}_F(\{x\}) = F(y) - F(x) \text{ and}$$
$$\overline{\mathbf{E}}_F(I_{[x,y]}) = \mathbf{E}_F(I_{[a,y]}) - \mathbf{E}_F(I_{[a,x]}) + \overline{\mathbf{E}}_F(\{x\}) = F(y) - F(x-),$$

and similarly,

$$\underline{\mathbf{E}}_F(I_{(x,y)}) = \mathbf{E}_F(I_{[a,y]}) - \mathbf{E}_F(I_{[a,x]}) - \overline{\mathbf{E}}_F(\{y\}) = F(y-) - F(x) \text{ and}$$
$$\overline{\mathbf{E}}_F(I_{(x,y)}) = \mathbf{E}_F(I_{[a,y]}) - \mathbf{E}_F(I_{[a,x]}) - \underline{\mathbf{E}}_F(\{x\}) = F(y) - F(x).$$

Expressions for other intervals can be obtained in a similar way. We are now in a comfortable position to connect the natural extension of F to the lower Riemann-Stieltjes integral.

Theorem 4.59. *Let $\mathcal{X} = [a, b]$ be a compact interval in $\mathbb{R}$ and let F be a coherent cumulative distribution function on $\mathcal{X}$. Then*

$$\underline{\mathbf{E}}_F(f) = \text{R-S}\underline{\int_a^b} f(x)\,\mathrm{d}F(x) \quad \text{for all gambles } f \text{ on } X, \tag{4.43}$$

if and only if $F(a) = 0$ and F is continuous from the right at every point $x \in [a, b)$.

Proof. "if". Let μ_F be the probability charge induced by F. By Theorem 4.58, it suffices to check the conditions of Theorem 4.52.

Consider $\mathcal{A} := \{[a, x] \colon x \in \mathcal{X}\}$; $\operatorname{dom} \mu_F = \mathcal{F}_{[]}(\mathcal{X})$ is the field generated by $\mathcal{A}$. Every element of $\mathcal{A}$ is Riemann-Stieltjes integrable. Indeed, using Eqs. (4.31)&(4.32) on p. 143, for every $a \leq y < b$ it holds that

$$\begin{aligned}\text{R-S}\underline{\int_a^b} I_{[a,y]}(x)\,\mathrm{d}F(x) &= F(y) - F(a) \\ &= F(y+) - F(a) = \text{R-S}\overline{\int_a^b} I_{[a,y]}(x)\,\mathrm{d}F(x),\end{aligned}$$

and therefore, since $F(a) = 0$, indeed $\mu_F([a, y]) = F(y) = \text{R-S}\int_a^b I_{[a,y]}(x)\,\mathrm{d}F(x)$. For $y = b$ we have

$$\text{R-S}\underline{\int_a^b} I_{[a,b]}(x)\,\mathrm{d}F(x) = F(b) - F(a) = \text{R-S}\overline{\int_a^b} I_{[a,b]}(x)\,\mathrm{d}F(x),$$

so $[a, b]$ is Riemann-Stieltjes integrable too and, again since $F(a) = 0$, it follows that $\mu_F([a, b]) = F(b) = \text{R-S}\int_a^b I_{[a,b]}(x)\,\mathrm{d}F(x)$. Hence, all sets in $\mathcal{A}$ are Riemann-Stieltjes integrable. By Proposition 4.51 it follows that all sets in the field generated by $\mathcal{A}$ are Riemann-Stieltjes integrable: all sets in $\mathcal{F}_{[]}(\mathcal{X}) = \operatorname{dom}\mu_F$ are Riemann-Stieltjes integrable.

Let $S = [x, y] \subseteq \mathcal{X}$ and $\epsilon > 0$. Take $A_\epsilon^S = (x, y]$ in $\mathcal{F}_{[]}(\mathcal{X})$. Then $A_\epsilon^S \Delta S = \{x\}$, and hence, $\text{R-S}\underline{\int}_a^b I_{A\Delta S}(z)\,\mathrm{d}F(z) = 0 < \epsilon$.

So, all conditions of Theorem 4.52 are satisfied. Therefore,

$$\underline{\mathbf{E}}_F(f) = \underline{\mathbf{E}}_{\mu_F}(f) = \text{R-S}\underline{\int}_a^b f(x)\,\mathrm{d}F(x)$$

for any gamble f on X.

"only if". We are given that $\underline{\mathbf{E}}_F(f) = \text{R-S}\underline{\int}_a^b f(x)\,\mathrm{d}F(x)$ for all gambles f on X. Take $f = I_{[a,b]}$. It follows that

$$1 = \underline{\mathbf{E}}_F(I_{[a,b]}) = \text{R-S}\underline{\int}_a^b I_{[a,b]}\,\mathrm{d}F(x) = F(b) - F(a) = 1 - F(a),$$

and hence, $F(a) = 0$.

Now let $x \in (a, b)$ and consider the gamble $I_{\{x\}}$. Then

$$F(x) - F(x-) = \overline{\mathbf{E}}_F(I_{\{x\}}) = \text{R-S}\overline{\int}_a^b I_{\{x\}}(z)\,\mathrm{d}F(z) = F(x+) - F(x-).$$

So, $F(x) = F(x+)$, and hence, F must be continuous from the right at every $x \in (a, b)$. For $x = a$ we have

$$0 = F(a) = \mu_F(\{a\}) = \overline{\mathbf{E}}_\mu(I_{\{a\}}) = \text{R-S}\overline{\int}_a^b I_{\{a\}}(z)\,\mathrm{d}F(z) = F(a+) - F(a),$$

and hence, also at $x = a$, F must be continuous from the right. □

4.3.8 The Dunford Integral

Another way to integrate a charge, introduced by Dunford [31] for measures, and extended by Dunford and Schwartz [30] to charges, is to start from an integral defined on simple gambles only, and try to approximate other gambles by a Cauchy sequence of simple ones. This is the core idea behind

the definition of the Dunford integral. Note that we use the Dunford integral for $\mathcal{F}$-simple gambles already defined in Definition 3.19 on p. 65. A nice thing about its definition is that it applies not only to gambles, but to all functions $f: \mathcal{X} \to \mathbb{R}$—in this sense it is more general than the S-integral introduced in Section 4.3.5. The definition relies on the outer set function μ^* induced by a probability charge μ; see Definition 4.34 on p. 116

Definition 4.60. Let $\mathcal{F}$ be a field on $\mathcal{X}$ and let μ be a probability charge on $\mathcal{F}$. Then a function $f: \mathcal{X} \to \mathbb{R}$ is called *Dunford integrable* with respect to μ if and only if there is a sequence of $\mathcal{F}$-simple gambles f_n such that

(i) $\lim_{n,m\to+\infty} \mathrm{D}\int \left|f_n - f_m\right| \,\mathrm{d}\mu \to 0$, and

(ii) for any $\epsilon > 0$ it holds that $\lim_{n\to+\infty} \mu^*\left(\left\{x \in \mathcal{X}\colon\ \left|f(x) - f_n(x)\right| > \epsilon\right\}\right) \to 0$.

In such a case, the *Dunford integral* of f with respect to μ is defined as

$$\mathrm{D}\int f \,\mathrm{d}\mu := \lim_{n\to+\infty} \mathrm{D}\int f_n \,\mathrm{d}\mu, \tag{4.44}$$

where we should note that the limit on the right hand side is independent of the sequence f_n satisfying the two conditions above. Such sequence is called a *determining sequence* for f.

In order to show that Dunford integrals and μ-integrals coincide on gambles, we shall need the following lemma.

Lemma 4.61. *Let $\underline{P}$ be any coherent lower prevision on $\mathcal{L}(X)$. Let f_n be a bounded sequence of gambles on X, that is, there is an $\alpha \in \mathbb{R}$ such that $\sup_{n\in\mathbb{N}} \sup\left|f_n\right| \leq \alpha$. Then the following statements are equivalent.*

(i) $\lim_{n\to+\infty} \overline{P}\left(\left|f_n\right|\right) = 0$.

(ii) For any $\epsilon > 0$ it holds that $\lim_{n\to+\infty} \overline{P}\left(\left\{x \in \mathcal{X}\colon\ \left|f_n(x)\right| > \epsilon\right\}\right) = 0$.

Proof. Define $A_{n,\epsilon} := \left\{x \in \mathcal{X}\colon\ \left|f_n(x)\right| > \epsilon\right\}$.

(i) $\Longrightarrow$ (ii). Let $\epsilon > 0$ and $\delta > 0$ such that $\delta < \epsilon$. Then there is an $N_\delta \in \mathbb{N}$ such that $\overline{P}\left(\left|f_n\right|\right) < \delta^2$ for every $n \geq N_\delta$. Since $\epsilon I_{A_{n,\epsilon}} \leq \left|f_n\right| I_{A_{n,\epsilon}}$, it follows from the coherence of $\underline{P}$ that

$$\overline{P}\left(A_{n,\epsilon}\right) \leq \frac{\overline{P}\left(\left|f_n\right| A_{n,\epsilon}\right)}{\epsilon} \leq \frac{\overline{P}\left(\left|f_n\right|\right)}{\epsilon} < \frac{\delta^2}{\epsilon} < \delta \tag{4.45}$$

for every $n \geq N_\delta$. So $\lim_{n\to+\infty} \overline{P}(A_{n,\epsilon}) = 0$.

(ii) $\Longrightarrow$ (i). Let $\epsilon > 0$. Then there is an $N_\epsilon \in \mathbb{N}$ such that $\overline{P}(A_{n,\epsilon}) < \frac{\epsilon}{\alpha+1}$ for every $n \geq N_\epsilon$. It follows from the coherence of $\underline{P}$ that

$$\overline{P}\left(\left|f_n\right|\right) \leq \overline{P}\left(\left|f_n\right| A_{n,\epsilon}\right) + \overline{P}\left(\left|f_n\right| \complement A_{n,\epsilon}\right) < \alpha\frac{\epsilon}{\alpha+1} + \epsilon < 2\epsilon \tag{4.46}$$

for every $n \geq N_\epsilon$. □

A rather long proof for the equivalence of Dunford integrability and S-integrability for gambles, and a proof for the equality of the corresponding integrals on gambles, was given by Bhaskara Rao and Bhaskara Rao [9, Theorem 4.5.7 and Proposition 4.5.8]. We give a much shorter and conceptually simpler proof by repeated application of Lemma 4.61.

Theorem 4.62. *Let $\mathcal{F}$ be a field on $\mathcal{X}$ and let μ be a probability charge on $\mathcal{F}$. Let f be any gamble. Then the following conditions are equivalent.*

(i) f is Dunford integrable with respect to μ.

(ii) f is S-integrable with respect to μ.

(iii) f is μ-integrable.

If any (and hence all) of these conditions are satisfied, then

$$\mathrm{D}\int f\,\mathrm{d}\mu = \mathrm{S}\int f\,\mathrm{d}\mu = \mathbf{E}_\mu(f). \tag{4.47}$$

Proof. Equivalence of S-integrability with respect to μ and μ-integrability, and equality of the corresponding integrals, has already been established in Theorem 4.43 on p. 131.

For $\mathcal{F}$-simple gambles, equality of the Dunford integral and natural extension, and hence, equality of the Dunford integral and the S-integral, was noted in Section 4.3.3, p. 112 ff. Equality of the S-integral and the Dunford integral for $\mathcal{F}$-simple gambles is also immediate from their respective definitions; see Definition 3.19 on p. 65 and Definition 4.40 on p. 129. To prove the general case, we shall extensively use the equality of the S-integral and the Dunford integral for $\mathcal{F}$-simple gambles.

Suppose that f is S-integrable with respect to μ. Then, for any $n \in \mathbb{N}$, $n > 0$, there is a finite partition $\mathcal{B}_n \in \mathbb{P}(\mathcal{F})$ such that

$$\sum_{B\in\mathcal{B}_n} \overline{P}_B(f)\mu(B) - \sum_{B\in\mathcal{B}_n} \underline{P}_B(f)\mu(B) < \frac{1}{n}.$$

Define the gambles $g_n := \sum_{B\in\mathcal{B}_n} \overline{P}_B(f)I_B$ and $h_n := \sum_{B\in\mathcal{B}_n} \underline{P}_B(f)I_B$, then, since g_n and h_n are $\mathcal{F}$-simple, the above condition can be written as $\mathrm{S}\int (g_n - h_n)\,\mathrm{d}\mu < \frac{1}{n}$, or, since $g_n \geq h_n$, also as $\mathrm{S}\int \left|g_n - h_n\right|\,\mathrm{d}\mu < \frac{1}{n}$. But $g_n \geq f \geq h_n$, and hence $\left|g_n - h_n\right| \geq \left|g_n - f\right|$. In particular, we find that, given that the lower S-integral is coherent,

$$\mathrm{S}\overline{\int} \left|g_n - f\right|\,\mathrm{d}\mu \leq \mathrm{S}\overline{\int} \left|g_n - h_n\right|\,\mathrm{d}\mu = \mathrm{S}\int (g_n - h_n)\,\mathrm{d}\mu < \frac{1}{n}. \tag{4.48}$$

Since, again by the coherence of the lower S-integral, $\mathrm{S}\int \left|g_n - g_m\right|\,\mathrm{d}\mu \leq \mathrm{S}\overline{\int} \left|g_n - f\right|\,\mathrm{d}\mu + \mathrm{S}\overline{\int} \left|f - g_m\right|\,\mathrm{d}\mu$, this implies that $\mathrm{S}\int \left|g_n - g_m\right|\,\mathrm{d}\mu$ converges to zero. But $g_n - g_m$ is $\mathcal{F}$-simple, so $\mathrm{S}\int \left|g_n - g_m\right|\,\mathrm{d}\mu = \mathrm{D}\int \left|g_n - g_m\right|\,\mathrm{d}\mu$, and so $\mathrm{D}\int \left|g_n - g_m\right|\,\mathrm{d}\mu$ must converge to zero as well. Also observe that $\sup_{n\in\mathbb{N}} \sup\left|f - g_n\right| \leq 2\sup\left|f\right|$, and hence, Lemma 4.61 applies on Eq. (4.48): for any $\epsilon > 0$ it holds that $\mathrm{S}\overline{\int} \left\{x \in \mathcal{X}\colon \left|f(x) - g_n(x)\right| > \epsilon\right\}\,\mathrm{d}\mu$ converges to zero. Now apply Theorem 4.36(v) on p. 117 and Theorem 4.42 on p. 130 to see that $\mu^*\left(\left\{x \in \mathcal{X}\colon \left|f(x) - g_n(x)\right| > \epsilon\right\}\right)$ converges to zero. We have demonstrated that f is Dunford integrable with respect to μ. Again by Eq. (4.48), and the coherence of the lower S-integral (Theorem 3.5(xii) on p. 55) it holds that

$$\mathrm{S}\int f\,\mathrm{d}\mu = \lim_{n\to+\infty} \mathrm{S}\int g_n\,\mathrm{d}\mu = \mathrm{D}\int f\,\mathrm{d}\mu,$$

so the corresponding integrals are equal as well.

Conversely, assume that f is Dunford integrable with respect to μ. Then there is a sequence f_n of $\mathcal{F}$-simple gambles such that for every $\epsilon > 0$ the sequence $\mu^*\left(\left\{x \in \mathcal{X}\colon \left|f(x) - f_n(x)\right| > \epsilon\right\}\right)$ converges to zero. In particular, there is an $\mathcal{F}$-simple gamble f_ϵ such that

$$\mu^*\left(\left\{x \in \mathcal{X}\colon \left|f(x) - f_\epsilon(x)\right| > \epsilon\right\}\right) < \epsilon.$$

Define the set $A_\epsilon := \left\{x \in \mathcal{X}\colon \left|f(x) - f_\epsilon(x)\right| > \epsilon\right\}$, so $\mu^*(A_\epsilon) = \inf\{\mu(B)\colon A_\epsilon \subseteq B \in$

$\mathcal{F}\} < \epsilon$. This means that there must also be a $B_\epsilon \in \mathcal{F}$ such that $A_\epsilon \subseteq B_\epsilon$ but still $\mu(B_\epsilon) < \epsilon$. Since f_ϵ is $\mathcal{F}$-simple there is a finite partition $\mathcal{A}_\epsilon \in \mathbb{P}(\mathcal{F})$ on whose elements f_ϵ is constant. Define the finite partition

$$\mathcal{B}_\epsilon := \{B_\epsilon\} \cup \{A \cap \complement B_\epsilon \colon A \in \mathcal{A}_\epsilon\}.$$

Fix $A \in \mathcal{A}_\epsilon$. Observe that for any $x \in A \cap \complement B_\epsilon$ it holds in particular that $x \notin A_\epsilon$, and hence $\left|f(x) - f_\epsilon(x)\right| \leq \epsilon$. Since actually f_ϵ is constant on $A \cap \complement B_\epsilon$, we also find that, since $\underline{P}_{A\cap\complement B_\epsilon}$ is coherent,

$$\begin{aligned}\overline{P}_{A\cap\complement B_\epsilon}(f) - \underline{P}_{A\cap\complement B_\epsilon}(f) &= \overline{P}_{A\cap\complement B_\epsilon}(f - f_\epsilon) - \underline{P}_{A\cap\complement B_\epsilon}(f - f_\epsilon) \\ &\leq 2\overline{P}_{A\cap\complement B_\epsilon}(|f - f_\epsilon|) \leq 2\epsilon.\end{aligned}$$

Look at the definition of the S-integral, and define the $\mathcal{F}$-simple functions $g_\epsilon := \sum_{B\in\mathcal{B}_\epsilon} \overline{P}_B(f) I_B$ and $h_\epsilon := \sum_{B\in\mathcal{B}_\epsilon} \underline{P}_B(f) I_B$. If we can show that $\mathrm{S}\int (g_\epsilon - h_\epsilon)\,\mathrm{d}\mu \leq K\epsilon$ for some $K > 0$ which may depend on f, then we have proved that f is S-integrable. Indeed, in such a case, since $g_\epsilon \geq f \geq h_\epsilon$, and by the coherence of the lower S-integral,

$$\mathrm{S}\int g_\epsilon\,\mathrm{d}\mu \geq \mathrm{S}\overline{\int} f\,\mathrm{d}\mu \geq \mathrm{S}\underline{\int} f\,\mathrm{d}\mu \geq \mathrm{S}\int h_\epsilon\,\mathrm{d}\mu \geq \mathrm{S}\int g_\epsilon\,\mathrm{d}\mu + K\epsilon,$$

for any $\epsilon > 0$, and hence, in the limit for ϵ to zero, we recover that $\mathrm{S}\overline{\int} f\,\mathrm{d}\mu = \mathrm{S}\underline{\int} f\,\mathrm{d}\mu$: f is S-integrable.

So, let's show that $\mathrm{S}\int (g_\epsilon - h_\epsilon)\,\mathrm{d}\mu \leq K\epsilon$ for some $K > 0$:

$$\begin{aligned}\mathrm{S}\int (g_\epsilon - h_\epsilon)\,\mathrm{d}\mu &= \sum_{B\in\mathcal{B}_\epsilon} \left(\overline{P}_B(f) - \underline{P}_B(f)\right)\mu(B) \\ &= \left(\overline{P}_{B_\epsilon}(f) - \underline{P}_{B_\epsilon}(f)\right)\mu(B_\epsilon) \\ &\quad + \sum_{A\in\mathcal{A}_\epsilon} \left(\overline{P}_{A\cap\complement B_\epsilon}(f) - \underline{P}_{A\cap\complement B_\epsilon}(f)\right)\mu(A \cap \complement B_\epsilon) \\ &\leq 2\epsilon \sup|f| + 2\epsilon \sum_{A\in\mathcal{A}_\epsilon} \mu(A \cap \complement B_\epsilon) \\ &\leq 2(\sup|f| + 1)\epsilon\end{aligned}$$

and hence, the desired inequality is satisfied for $K = 2(\sup|f|+1)$. This means

that f is S-integrable, with S-integral $\mathrm{S}\int f\,\mathrm{d}\mu = \lim_{\epsilon>0}\mathrm{S}\int g_\epsilon\,\mathrm{d}\mu$. From this equality, which can also be written as $\lim_{\epsilon>0}\mathrm{S}\int|g_\epsilon - f|\,\mathrm{d}\mu = 0$, we can infer two things. Firstly:

$$\begin{aligned}\lim_{n,m\to+\infty}\mathrm{D}\int\left|g_{\frac{1}{n}} - g_{\frac{1}{m}}\right|\mathrm{d}\mu &= \lim_{n,m\to+\infty}\mathrm{S}\int\left|g_{\frac{1}{n}} - g_{\frac{1}{m}}\right|\mathrm{d}\mu \\ &\leq \lim_{n\to+\infty}\mathrm{S}\int\left|g_{\frac{1}{n}} - f\right|\mathrm{d}\mu + \lim_{m\to+\infty}\mathrm{S}\int\left|f - g_{\frac{1}{m}}\right|\mathrm{d}\mu = 0,\end{aligned}$$

and secondly, noting that $\sup_{n\in\mathbb{N}}\left|f - g_{\frac{1}{n}}\right| \leq 2\sup|f|$, and once more applying Lemma 4.61: for any $\epsilon > 0$ it holds that $\mathrm{S}\overline{\int}\left\{x\in\mathcal{X}\colon \left|f(x) - g_{\frac{1}{n}}(x)\right| > \epsilon\right\}\mathrm{d}\mu$ converges to zero. Now apply Theorem 4.36(v) on p. 117 and Theorem 4.42 on p. 130 to see that for any $\epsilon > 0$ also $\mu^*\left(\left\{x\in\mathcal{X}\colon \left|f(x) - g_{\frac{1}{n}}(x)\right| > \epsilon\right\}\right)$ converges to zero. Hence, $g_{\frac{1}{n}}$ is a determining sequence for f, so

$$\mathrm{D}\int f\,\mathrm{d}\mu = \lim_{n\to+\infty}\mathrm{D}\int g_{\frac{1}{n}}\,\mathrm{d}\mu = \lim_{n\to+\infty}\mathrm{S}\int g_{\frac{1}{n}}\,\mathrm{d}\mu = \mathrm{S}\int f\,\mathrm{d}\mu.$$

So, f is not only S-integrable, also its Dunford integral is equal to its S-integral. This concludes the proof. □

We shall need the following results on Dunford integration further on, we refer to Bhaskara Rao and Bhaskara Rao [9] for a proof.

Lemma 4.63. *Let $\mathcal{F}$ be a field on $\mathcal{X}$ and let μ be a probability charge on $\mathcal{F}$. Let f be a Dunford integrable random quantity, and let f_n be a determining sequence for f. Then* $\lim_{n\to\infty}\mathrm{D}\int|f - f_n|\,\mathrm{d}\mu = 0$.

Theorem 4.64. *Let $\mathcal{F}$ be a field on $\mathcal{X}$ and let μ be a probability charge on $\mathcal{F}$. Let f be a random quantity on $\mathcal{X}$, and let f_n be a sequence of Dunford integrable random quantities. If*

(i) $\lim_{n,m\to+\infty}\mathrm{D}\int\left|f_n - f_m\right|\mathrm{d}\mu = 0$, *and*

(ii) for any $\epsilon > 0$ it holds that $\lim_{n\to+\infty}\mu^*\left(\left\{x\in\mathcal{X}\colon \left|f(x) - f_n(x)\right| > \epsilon\right\}\right) = 0$,

then f is Dunford integrable, and $\mathrm{D}\int f\,\mathrm{d}\mu = \lim_{n\to\infty}\mathrm{D}\int f_n\,\mathrm{d}\mu$.

4.3.9 Intermezzo: More Properties of the Riemann-Stieltjes Integral

The results derived in this section will be used in Section 4.3.10. Most of them are well-known properties, except for Proposition 4.69 and Proposition 4.70. In Proposition 4.69, we shall give a formula for partial integration of lower and upper Riemann-Stieltjes integrals, *i.e.*, without imposing any integrability, continuity, or other regularity conditions, except for monotonicity. In Proposition 4.70, we shall generalise a well-known result about modifying the integrand without changing the value of the Riemann-Stieltjes integral.

Proposition 4.65. *Let $\mathscr{X} = [a, b]$ be a compact interval in $\mathbb{R}$, let F be a real-valued non-decreasing bounded function on $\mathscr{X}$. Then*

$$\text{R-S}\underline{\int_a^y} f(x)\,\mathrm{d}F(x) + \text{R-S}\underline{\int_y^b} f(x)\,\mathrm{d}F(x) = \text{R-S}\underline{\int_a^b} f(x)\,\mathrm{d}F(x).$$

for any gamble f on X and $y \in \mathscr{X}$.

Proof. Immediately from Definition 4.48, and the observation that for every subdivision $\mathcal{S}$ of $[a, b]$, there is a subdivision of $[a, b]$ that refines both $\mathcal{S}$ and $\{[a, y], [y, b]\}$. □

Proposition 4.66. *Let $\mathscr{X} = [a, b]$ be a compact interval in $\mathbb{R}$, let F be a real-valued non-decreasing bounded function on $\mathscr{X}$. The lower and upper Riemann-Stieltjes integral are uniformly continuous on $\mathscr{L}(X)$ with respect to the supremum norm.*

Proof. This follows from the coherence of the lower Riemann-Stieltjes integral and Theorem 3.5(xiii). □

Proposition 4.67. *Let $\mathscr{X} = [a, b]$ be a compact interval in $\mathbb{R}$, and let F be a real-valued non-decreasing bounded function on $\mathscr{X}$. The set of Riemann-Stieltjes integrable gambles with respect to F is a uniformly closed linear lattice, on which the Riemann-Stieltjes integral is a positive linear functional.*

Proof. Immediately from Theorem 4.49 on p. 137 and Proposition 4.18(vi) on p. 102. □

Proposition 4.68. *Let $\mathscr{X} = [a, b]$ be a compact interval in $\mathbb{R}$, and let f be a non-decreasing or non-increasing gamble on X. Then f is Riemann integrable.*

Proof. Suppose f is non-decreasing. For each $\epsilon > 0$, select a subdivision $\mathcal{S}_\epsilon \in \mathbb{S}(\mathcal{X})$ such that $\lambda(S) < \epsilon$ for each $S \in \mathcal{S}_\epsilon$. We find that

$$\sum_{S\in\mathcal{S}_\epsilon} \left[\overline{P}_S(f) - \underline{P}_S(f)\right] \lambda(S) < \epsilon \sum_{S\in\mathcal{S}_\epsilon} \left[\overline{P}_S(f) - \underline{P}_S(f)\right] = \epsilon \left(\sup f - \inf f\right),$$

where the last equality holds because f is non-decreasing. We find that

$$\begin{aligned} \overline{\mathrm{R}\int_a^b} f(x)\,\mathrm{d}x - \underline{\mathrm{R}\int_a^b} f(x)\,\mathrm{d}x &= \inf_{\mathcal{S}\in\mathbb{S}(\mathcal{X})} \sum_{S\in\mathcal{S}} \left[\overline{P}_S(f) - \underline{P}_S(f)\right] \lambda(S) \\ &\leq \inf_{\epsilon>0} \sum_{S\in\mathcal{S}_\epsilon} \left[\overline{P}_S(f) - \underline{P}_S(f)\right] \lambda(S) \\ &\leq \inf_{\epsilon>0} \epsilon \left(\sup f - \inf f\right) = 0. \end{aligned}$$

Therefore, f is Riemann integrable.

If f is non-increasing, then $-f$ is non-decreasing, and hence, Riemann integrable. From Proposition 4.67 it follows that f is also Riemann integrable. □

Proposition 4.69. *Let $\mathcal{X} = [a, b]$ be a compact interval in $\mathbb{R}$, and let f and g be non-decreasing gambles on X. Then*

$$\begin{aligned} &\underline{\text{R-S}\int_a^b} f(x)\,\mathrm{d}g(x) + \overline{\text{R-S}\int_a^b} g(x)\,\mathrm{d}f(x) \\ &\qquad = f(b)g(b) - f(a)g(a). \\ &\qquad = \overline{\text{R-S}\int_a^b} f(x)\,\mathrm{d}g(x) + \underline{\text{R-S}\int_a^b} g(x)\,\mathrm{d}f(x) \end{aligned}$$

Hence, f is Riemann-Stieltjes integrable with respect to g if and only if g is Riemann-Stieltjes integrable with respect to f, and in such a case

$$\text{R-S}\int_a^b f(x)\,\mathrm{d}g(x) + \text{R-S}\int_a^b g(x)\,\mathrm{d}f(x) = f(b)g(b) - f(a)g(a).$$

Proof. Since f and g are non-decreasing, for any closed interval S in $\mathcal{X}$ it holds that $\underline{P}_S(f) = f(\min S)$ and $\overline{P}_S(g) = g(\max S)$. We find

$$
\begin{aligned}
&\text{R-S}\underline{\int_a^b} f(x)\,\mathrm{d}g(x) + \text{R-S}\overline{\int_a^b} g(x)\,\mathrm{d}f(x) \\
&\qquad = \lim_{\mathcal{S}\in\mathbb{S}(\mathcal{X})} \sum_{S\in\mathcal{S}} \left[\underline{P}_S(f)\rho_g(S) + \overline{P}_S(g)\rho_f(S)\right] \\
&\qquad = \lim_{\mathcal{S}\in\mathbb{S}(\mathcal{X})} \sum_{S\in\mathcal{S}} \Big\{f(\min S)[g(\max S) - g(\min S)] \\
&\qquad\qquad\qquad + g(\max S)[f(\max S) - f(\min S)]\Big\} \\
&\qquad = \lim_{\mathcal{S}\in\mathbb{S}(\mathcal{X})} \sum_{S\in\mathcal{S}} \{-f(\min S)g(\min S) + g(\max S)f(\max S)\} \\
&\qquad = -f(a)g(a) + f(b)g(b).
\end{aligned}
$$

□

It is well-known that if two gambles f and g are Riemann integrable, and differ at most on a countable set, or more general, on a set of outer Lebesgue measure zero, then their Riemann integrals are equal; for instance, see Darboux [14, Section IV, p. 75, ll. 12–20] in case f and g differ only on a finite set, see Hildebrandt [43, Chapter II, Theorem 15.9, p. 74] for a generalisation to Riemann-Stieltjes integrals, and requiring only that f and g differ at most on a set of measure zero (with respect to what could be called the outer Riemann-Stieltjes measure).

We give a very short proof of a stronger version of this property: if both f and g are Riemann-Stieltjes integrable, then their Riemann-Stieltjes integrals are equal whenever they differ only on a set whose indicator has zero *lower* Riemann-Stieltjes integral. Indeed, surprisingly, and perhaps counter-intuitively, we only need the lower Riemann-Stieltjes integral of the indicator to be zero. For instance: if two Riemann integrable gambles f and g differ only on a set of inner Lebesgue measure zero, then their Riemann integrals are equal.

Proposition 4.70. *Let $\mathcal{X} = [a, b]$ be a compact interval of $\mathbb{R}$, let F be a bounded non-decreasing real-valued function on $\mathcal{X}$, and let f and g be gambles on X that are Riemann-Stieltjes integrable with respect to F. If $f = g$ except on a set of zero lower Riemann-Stieltjes integral, i.e., if $f(x) = g(x)$ for all $x \in \mathcal{X} \setminus N$, where N is some*

subset of $\mathcal{X}$ *such that* R-S$\underline{\int}_a^b I_N(x)\,\mathrm{d}F(x) = 0$, *then*

$$\text{R-S}\int_a^b f(x)\,\mathrm{d}F(x) = \text{R-S}\int_a^b g(x)\,\mathrm{d}F(x).$$

Proof. Immediately from the 2-monotonicity of the lower Riemann-Stieltjes integral (Theorem 4.49 on p. 137) and Proposition 4.19(ii) on p. 104. □

Enrique Miranda (personal communication) has proved that, if f and g are Riemann integrable gambles that are equal except on $N \subseteq \mathcal{X}$, then $\mathrm{R}\underline{\int}_a^b I_N(x)\,\mathrm{d}x = 0$ if and only if $\lambda^*(N) = 0$; this tells us that the above condition reduces to a well-known condition in case of Riemann integration:

Corollary 4.71. *Let* $\mathcal{X} = [a, b]$ *be a compact interval of* $\mathbb{R}$, *and let* f *and* g *be gambles on* X *that are Riemann integrable. Let* $N = \{x \in \mathcal{X} : f(x) \neq g(x)\}$. *Then* $\mathrm{R}\underline{\int}_a^b I_N(x)\,\mathrm{d}x = 0$ *if and only if the outer Lebesgue measure of* N *is zero, and in that case,* $\mathrm{R}\int_a^b f(x)\,\mathrm{d}x = \mathrm{R}\int_a^b g(x)\,\mathrm{d}x$.

Proof. Equality of the Riemann integrals follows from Proposition 4.70. It remains to show the equivalence of the two conditions. Let us denote by ρ is the restriction of the Lebesgue measure to $\mathcal{F}_{[]}(\mathcal{X})$.

"if". If the outer Lebesgue measure of N is zero, then so must be the lower Riemann integral of I_N, since

$$\lambda^*(N) \geq \lambda_*(N) \geq \rho_*(N) = \mathrm{R}\underline{\int}_a^b I_N(x)\,\mathrm{d}x,$$

where we used Lemma 4.50(i) on p. 139 in the final step.

"only if". Since f and g are Riemann-integrable, so is $|f - g|$ (by Proposition 4.67), and since $\mathrm{R}\underline{\int}_a^b I_N(x)\,\mathrm{d}x = 0$ it also follows that $\mathrm{R}\int_a^b |f(x) - g(x)|\,\mathrm{d}x = 0$. Indeed,

$$0 \leq \mathrm{R}\int_a^b |f(x) - g(x)|\,\mathrm{d}x \leq \sup|f - g|\,\mathrm{R}\underline{\int}_a^b I_N(x)\,\mathrm{d}x = 0$$

By Theorem 4.56, it follows in fact that $|f - g|$ is $\mathrm{dom}\,\overline{\lambda}$-measurable, and so there is a sequence h_n of non-negative $\mathrm{dom}\,\overline{\lambda}$-simple gambles that converges

uniformly to $|f - g|$ from below. But, for every n in $\mathbb{N}$,

$$0 \leq \mathrm{S}\int h_n \,\mathrm{d}\overline{\lambda} \leq \mathrm{S}\overline{\int} |f-g| \,\mathrm{d}\overline{\lambda} \leq \mathrm{R}\overline{\int_a^b} |f(x)-g(x)| \,\mathrm{d}x = \mathrm{R}\int_a^b |f(x)-g(x)| \,\mathrm{d}x = 0,$$

so $\mathrm{S}\int h_n \,\mathrm{d}\overline{\lambda} = 0$. Now, fix n in $\mathbb{N}$. Since h_n is $\operatorname{dom}\overline{\lambda}$-simple, there is an $\epsilon > 0$ such that, for any x in $\mathcal{X}$, $h_n(x) > 0$ if and only if $h_n(x) > \epsilon$. Therefore,

$$0 = \mathrm{S}\int h_n \,\mathrm{d}\overline{\lambda} \geq \epsilon\overline{\lambda}(\{x \in \mathcal{X} \colon h_n(x) > \epsilon\}) \geq 0,$$

which in turn implies that $\overline{\lambda}(\{x \in \mathcal{X} \colon h_n(x) > 0\}$ must be zero as well. But, since the non-negative sequence of gambles h_n converges uniformly to the non-negative gamble $|f - g|$ from below,

$$\{x \in \mathcal{X} \colon f(x) > 0\} = \bigcup_{n\in\mathbb{N}} \{x \in \mathcal{X} \colon h_n(x) > 0\},$$

and so, since $\overline{\lambda}$ is a measure, it follows that $\overline{\lambda}(\{x \in \mathcal{X} \colon f(x) > 0\}) = 0$ as well. But, λ^* is an extension of $\overline{\lambda}$, so the equivalence is established. □

4.3.10 Natural Extension of 2-Monotone Set Functions: The Choquet Integral

Let us now proceed with a general definition of the Choquet integral, due to Choquet [11, Section 48.1, p. 265], and introduced before as a behavioural extension of 2-monotone set functions to simple gambles, and investigate its relation to natural extension. The most general, and I think also the most elegant way of extending the Choquet integral with respect to 2-monotone set functions for arbitrary gambles, is suggested by Walley; see Walley [85, Section 6, p. 52, ll. 22–24] and Walley [86, Note 2 of Section 3.2, p. 502]: he observes that through natural extension we don't need to impose any measurability conditions—compare for instance with Greco's [38] upper ν-measurability; a definition is given below.

Walley's argument is as follows. First, we extend, through natural extension, the 2-monotone set function ν defined on a field $\mathcal{F}$ to a set function

ν_* defined on the power set $\wp(\mathcal{X})$. This set function ν_* is 2-monotone; see Theorem 4.36(iv). Now, generalising a result by Walley [85], we can show that, for gambles, Greco's [38] Choquet integral with respect to ν is equal, on its domain of upper ν-measurable gambles, to the natural extension of $\underline{\mathbf{P}}_\nu$. Hence, since ν_* is equal, on its domain, to the natural extension of ν, also the Choquet integral with respect to ν_* is the natural extension of $\underline{\mathbf{P}}_\nu$ to upper ν_*-measurable gambles; this is a consequence of Proposition 4.8. But, every gamble on X is upper ν_*-measurable, since ν_* is defined on the σ-field $\wp(\mathcal{X})$ and every gamble on X is $\wp(\mathcal{X})$-measurable. So, the Choquet integral with respect to ν_* is the natural extension of $\underline{\mathbf{P}}_\nu$ to the set of all gambles. The details of the proof will be given below.

In conclusion, for 2-monotone measures, it makes perfect sense to take Greco's [38] Choquet integral with respect to ν_* also as the Choquet integral with respect to ν, unless you like unnecessarily complicated measurability conditions. Of course, as described in great detail by Denneberg [28], Greco's [38] approach is more general because it allows for a Choquet integral with respect to an arbitrary set function, not just 2-monotone set functions defined on a field. However, Walley [85, Corollary 6.2] proved that the Choquet integral with respect to a set function ν—defined on the power set, positive, and $\nu(\emptyset) = 0$—corresponds to a coherent lower prevision if and only if ν is 2-monotone; in a completely different context, this result was first proved by Choquet [11, Theorem on p. 289], and a very general formulation of this result can be found in Denneberg [28, Chapter 6]. Therefore, in our study of the interplay between integration and natural extension, it is not exactly clear what is the rôle of the Choquet integral with respect to set functions that are not 2-monotone. However, we shall not concern ourselves with this question any further.

In the definition below, the Choquet integral expressed as Riemann and Riemann-Stieltjes integrals are well-known results. Our contribution is also an expression of the Choquet integral in terms of a Dunford and an S-integral with respect to the Lebesgue measure—or any sufficiently dense restriction of the Lebesgue measure, as suggested by Theorem 4.52. These are simple consequences of Proposition 4.55 on p. 156 and Theorem 4.62 on p. 163

Definition 4.72. Let $\mathcal{F}$ be a field on $\mathcal{X}$ and let ν be a 2-monotone set function on $\mathcal{F}$. Let f be any gamble on X. Let $G_{*\nu,f}$ be the *lower decreasing distribution function* of f with respect to ν, and $F^*_{\nu,f}$ the *upper distribution function* of f

with respect to ν, that is,

$$G_{*\nu,f}(z) := \nu_*(\{x \in \mathcal{X} \colon f(x) > z\}), \text{ and}$$
$$F^*_{\nu,f}(z) := \pi^*(\{x \in \mathcal{X} \colon f(x) \leq z\}) = 1 - G_{*\nu,f}(z),$$

for any $z \in \mathbb{R}$, where π denotes the dual 2-alternating set function induced by ν. The *Choquet integral* of f with respect to ν is defined as

$$\begin{aligned}
\mathrm{C}\int f \,\mathrm{d}\nu &:= \inf f + \mathrm{D}\int_{[\inf f, \sup f]} G_{*\nu,f} \,\mathrm{d}\lambda, \\
&= \inf f + \mathrm{S}\int_{[\inf f, \sup f]} G_{*\nu,f} \,\mathrm{d}\lambda, \\
&= \inf f + \mathrm{R}\int_{\inf f}^{\sup f} G_{*\nu,f}(z) \,\mathrm{d}z, \\
&= \text{R-S}\int_{\inf f - \epsilon}^{\sup f} z \,\mathrm{d}F^*_{\nu,f}(z), \\
&= \mathrm{R}\int_{0 \wedge \inf f}^{0} [G_{*\nu,f}(z) - 1] \,\mathrm{d}z + \mathrm{R}\int_{0}^{0 \vee \sup f} G_{*\nu,f}(z) \,\mathrm{d}z, \\
&= \mathrm{R}\int_0^1 \check{G}_{*\nu,f}(t) \,\mathrm{d}t,
\end{aligned}$$

where the Dunford integral and the S-integral are taken with respect to the Lebesgue measure λ on $\mathcal{B}(\mathbb{R})$, $\epsilon > 0$ is arbitrary, and $\check{G}_{*\nu,f}$ is the *pseudo-inverse* of $G_{*\nu,f}$ which is defined as

$$\check{G}_{*\nu,f}(t) := \inf f \vee \sup\{z \in [\inf f, \sup f] \colon G_{*\nu,f}(z) > t\}, \qquad \text{for all } t \in [0,1].$$

Proof of existence and equality of all integrals. Since $G_{*\nu,f}$ is non-increasing it is Riemann integrable on $[\inf f, \sup f]$ by Proposition 4.68. The Riemann integral of $G_{*\nu,f}$ can be converted into an S-integral using Proposition 4.55 on p. 156. The S-integral is equal to the Dunford integral by Theorem 4.62 on p. 163. The Riemann integral converts into a Riemann-Stieltjes integral through Proposition 4.69 on p. 168:

$$\inf f + \mathrm{R}\int_{\inf f}^{\sup f} G_{*\nu,f}(z) \,\mathrm{d}z$$

$$= \inf f - \mathrm{R} \int_{\inf f}^{\sup f} \left[-G_{*\nu,f}(z)\right] \mathrm{d}z$$

and since $-G_{*\nu,f}$ is non-decreasing, we can apply Proposition 4.69,

$$= \inf f - \Bigg\{ - \text{R-S} \int_{\inf f}^{\sup f} z \, \mathrm{d}\left[-G_{*\nu,f}(z)\right] + \sup f \left(-G_{*\nu,f}(\sup f)\right) - \inf f \left(-G_{*\nu,f}(\inf f)\right) \Bigg\}$$

and since $-G_{*\nu,f} = F^*_{\nu,f} - 1$, we can replace the Riemann-Stieltjes integral with respect to $-G_{*\nu,f}$ with a Riemann-Stieltjes integral with respect to $F^*_{\nu,f}$. Also, $G_{*\nu,f}(\sup f) = \nu_*(\emptyset) = 0$, hence,

$$\begin{aligned} &= \inf f - \left\{ - \text{R-S} \int_{\inf f}^{\sup f} z \, \mathrm{d}F^*_{\nu,f}(z) + (\inf f) G_{*\nu,f}(\inf f) \right\} \\ &= \inf f \left(1 - G_{*\nu,f}(\inf f)\right) + \text{R-S} \int_{\inf f}^{\sup f} z \, \mathrm{d}F^*_{\nu,f}(z) \\ &= (\inf f) F^*_{\nu,f}(\inf f) + \text{R-S} \int_{\inf f}^{\sup f} z \, \mathrm{d}F^*_{\nu,f}(z) \end{aligned}$$

Now, observe that, since $F^*_{\nu,f}$ is non-decreasing and again applying Proposition 4.69,

$$\begin{aligned} \text{R-S} \int_{\inf f - \epsilon}^{\inf f} z \, \mathrm{d}F^*_{\nu,f}(z) &= - \mathrm{R} \int_{\inf f - \epsilon}^{\inf f} F^*_{\nu,f}(z) \, \mathrm{d}z \\ &\quad + (\inf f) F^*_{\nu,f}(\inf f) - (\inf f - \epsilon) F^*_{\nu,f}(\inf f - \epsilon) \\ &= (\inf f) F^*_{\nu,f}(\inf f). \end{aligned}$$

and apply Proposition 4.65.

Next, observe that

$$\begin{aligned} \inf f &= 0 \wedge \inf f + 0 \vee \inf f \\ &= \mathrm{R} \int_{0 \wedge \inf f}^{0} (-1) \, \mathrm{d}z + \mathrm{R} \int_{0}^{0 \vee \inf f} (+1) \, \mathrm{d}z \end{aligned}$$

and since $G_{*\nu,f}(z) = \nu_*(\mathcal{X}) = 1$ whenever $0 \leq x < \inf f$,

$$= \mathrm{R}\int_{0\wedge\inf f}^{0} (-1)\,\mathrm{d}z + \mathrm{R}\int_{0}^{0\vee\inf f} G_{*\nu,f}(z)\,\mathrm{d}z,$$

and, by Proposition 4.65,

$$\mathrm{R}\int_{\inf f}^{\sup f} G_{*\nu,f}(z)\,\mathrm{d}z = \mathrm{R}\int_{0\wedge\inf f}^{0} G_{*\nu,f}(z)\,\mathrm{d}z + \mathrm{R}\int_{0\vee\inf f}^{\sup f} G_{*\nu,f}(z)\,\mathrm{d}z,$$

Hence, by Proposition 4.65 and Proposition 4.67,

$$\begin{aligned}
&\inf f + \mathrm{R}\int_{\inf f}^{\sup f} G_{*\nu,f}(z)\,\mathrm{d}z \\
&= \mathrm{R}\int_{0\wedge\inf f}^{0} (-1)\,\mathrm{d}z + \mathrm{R}\int_{0}^{0\vee\inf f} G_{*\nu,f}(z)\,\mathrm{d}z \\
&\quad + \mathrm{R}\int_{0\wedge\inf f}^{0} G_{*\nu,f}(z)\,\mathrm{d}z + \mathrm{R}\int_{0\vee\inf f}^{\sup f} G_{*\nu,f}(z)\,\mathrm{d}z \\
&= \mathrm{R}\int_{0\wedge\inf f}^{0} [G_{*\nu,f}(z) - 1]\,\mathrm{d}z + \mathrm{R}\int_{0}^{0\vee\sup f} G_{*\nu,f}(z)\,\mathrm{d}z.
\end{aligned}$$

For the equality involving the Riemann integral over $\check{G}_{*\nu,f}$ we refer to Denneberg [28, Proposition 1.2, Lemma 1.3 and Corollary 1.5]. □

To see why we need to integrate from $\inf f - \epsilon$ when expressing the Choquet integral as a Riemann-Stieltjes integral, let ν be any 2-monotone set function and consider a constant gamble $a \in \mathbb{R}(X)$. Then

$$F^*_{\nu,a}(z) = \pi^*(\{x \in \mathcal{X} \colon a \leq z\}) = \begin{cases} 0, & \text{if } z < a, \\ 1, & \text{if } z \geq a. \end{cases}$$

Clearly, $\inf a = \sup a = a$, and hence,

$$\text{R-S}\int_{a}^{a} z\,\mathrm{d}F^*_{\nu,a}(z) = 0,$$

which is not the Choquet integral of the constant gamble a. But,

$$\text{R-S}\underline{\int_{a-\epsilon}^{a}} z\,\mathrm{d}F^*_{\nu,a}(z) = \sup_{0<\delta<\epsilon}\left[0 + \underline{P}_{[a-\epsilon+\delta,a]}(z) \times 1\right] = a,$$

and similar for the upper Riemann-Stieltjes integral,

$$\text{R-S}\overline{\int_{a-\epsilon}^{a}} z\,\mathrm{d}F^*_{\nu,a}(z) = \inf_{0<\delta<\epsilon}\left[0 + \overline{P}_{[a-\epsilon+\delta,a]}(z) \times 1\right] = a.$$

In the literature, when the Choquet integral is defined as a Riemann-Stieltjes integral, the Riemann-Stieltjes integral is over $\mathbb{R}$ instead of over $[\inf f - \epsilon, \sup f]$. This agrees with our definition since

$$\lim_{c\to-\infty} \text{R-S}\int_{c}^{\inf f-\epsilon} z\,\mathrm{d}F^*_{\nu,f}(z) = 0, \text{ and } \lim_{d\to+\infty} \text{R-S}\int_{\sup f}^{d} z\,\mathrm{d}F^*_{\nu,f}(z) = 0.$$

Now apply Proposition 4.65 to see that

$$\text{R-S}\int_{-\infty}^{+\infty} z\,\mathrm{d}F^*_{\nu,f}(z) := \lim_{\substack{c\to-\infty\\ d\to+\infty}} \text{R-S}\int_{c}^{d} z\,\mathrm{d}F^*_{\nu,f}(z) = \text{R-S}\int_{\inf f-\epsilon}^{\sup f} z\,\mathrm{d}F^*_{\nu,f}(z).$$

A similar argument shows that

$$\begin{aligned}\mathrm{R}\int_{-\infty}^{0} [G_{*\nu,f}(z) - 1]\,\mathrm{d}z + \mathrm{R}\int_{0}^{+\infty} G_{*\nu,f}(z)\,\mathrm{d}z \\ = \mathrm{R}\int_{0\wedge\inf f}^{0} [G_{*\nu,f}(z) - 1]\,\mathrm{d}z + \mathrm{R}\int_{0}^{0\vee\sup f} G_{*\nu,f}(z)\,\mathrm{d}z.\end{aligned}$$

Our definition of lower decreasing distribution function extends and simplifies the definition of the decreasing distribution function for upper ν-measurable gambles in case ν is a 2-monotone set function. As a consequence, our definition of the Choquet integral extends the definition of the Choquet integral for upper ν-measurable gambles. The proof of this claim, given below for the sake of completeness, is immediate.

Definition 4.73. Let ν be a 2-monotone set function defined on a field $\mathcal{F}$. A gamble f is called *upper ν-measurable* if there is an at most countable set

$N \subseteq \mathbb{R}$ such that

$$\nu_*(\{x \in \mathcal{X} \colon f(x) > z\}) = \nu^*(\{x \in \mathcal{X} \colon f(x) > z\})$$

for all $z \in \mathbb{R} \setminus N$; note that the left hand side is $G_{*\nu,f}(z)$. If f is upper ν-measurable, then any real-valued function equal to $G_{*\nu,f}$ except on an at most countable set is called a *decreasing distribution function* of f with respect to ν.

The definition of upper ν-measurability relies on both the inner set function ν_* and the outer set function ν^* induced by ν. Recall that ν_* is 2-monotone, as stated in Theorem 4.36(iv), but usually ν^* will not be 2-monotone.

Proposition 4.74. *Let ν be a 2-monotone set function defined on a field $\mathcal{F}$. Let f be an upper ν-measurable function. Then $G_{*\nu,f}$ is a decreasing distribution function of f with respect to ν.*

Proof. $G_{*\nu,f}$ is equal to $G_{*\nu,f}$ except on the empty set, which is an at most countable subset of $\mathbb{R}$. Now, apply the definition of decreasing distribution function. □

Corollary 4.75. *Let ν be a 2-monotone set function defined on a field $\mathcal{F}$. Let f be an upper ν-measurable function, and let $G_{\nu,f}$ be a decreasing distribution function of f with respect to ν. Then*

$$\mathrm{C}\int f \,\mathrm{d}\nu = \mathrm{R}\int_{0\wedge\inf f}^{0} [G_{\nu,f}(z) - 1]\,\mathrm{d}z + \mathrm{R}\int_{0}^{0\vee\sup f} G_{\nu,f}(z)\,\mathrm{d}z,$$

Proof. Simply observe that the Riemann integral over $G_{\nu,f}$ is equal to the Riemann integral over $G_{*\nu,f}$, since they are equal except on an at most countable subset of $\mathbb{R}$: every countable subset of $\mathbb{R}$ has Lebesgue measure zero, and therefore, the lower Riemann integral of its indicator is zero too, so Proposition 4.70 applies. □

So, our definition of the Choquet integral is a generalisation of the Choquet integral found in Greco [38] and Denneberg [28] for upper ν-measurable gambles, at least, when ν is 2-monotone. Again, the Choquet integral of set functions that are not 2-monotone may not give us the natural extension; this was established by Walley [85], Corollary 6.2.

The previously defined notions of lower decreasing distribution function and Choquet integral of Definition 3.29 on p. 74, which deals with $\mathcal{F}$-simple

gambles only, are compatible with Definition 4.72. Indeed, let ν be a 2-monotone set function defined on a field $\mathcal{F}$, and let f be an $\mathcal{F}$-simple gamble. For any $z \in \mathbb{R}$,

$$G_{*\nu,f}(z) = \nu_*(\{x \in \mathcal{X} \colon f(x) > z\}) = \nu(\{x \in \mathcal{X} \colon f(x) > z\}),$$

since $\{x \in \mathcal{X} \colon f(x) > z\}$ is in $\mathcal{F}$ for any $z \in \mathbb{R}$, ν_* is the natural extension of ν and ν is coherent; see Theorem 4.36 on p. 117. This is establishes equivalence of the two definitions, for $\mathcal{F}$-simple gambles.

Let's now proceed with the main result of this section, which is due to Walley [85]: the Choquet integral with respect to a 2-monotone set function ν is equal to the natural extension of ν. Below, we give more direct proof—in contradistinction to Walley, we do not rely on duality results (*i.e.*, results of Section 4.4).

Theorem 4.76. *Let $\mathcal{F}$ be a field on $\mathcal{X}$ and let ν be a 2-monotone set function on $\mathcal{F}$. For any gamble f on X it holds that*

$$\underline{\mathbf{E}}_\nu(f) = \mathrm{C}\int f \,\mathrm{d}\nu = \inf f + \mathrm{R}\int_{\inf f}^{\sup f} G_{*\nu,f}(z)\,\mathrm{d}z.$$

Proof. Let f be a $\wp(\mathcal{X})$-simple gamble on X. We can write the gamble f as $b_0 + \sum_{i=1}^n b_i I_{B_i}$ with b_0 in $\mathbb{R}$, $b_1, \ldots, b_n$ in $\mathbb{R}$ and strictly positive, and $\mathcal{X} \supsetneq B_1 \supsetneq B_2 \supsetneq \cdots \supsetneq B_n \supsetneq \emptyset$. It follows easily from Definition 4.72, Theorem 4.36 and Proposition 4.8 that $G_{*\nu,f} = G_{*\nu_*,f}$, since $(\nu_*)_* = \nu_*$. Therefore, also $\mathrm{C}\int f \,\mathrm{d}\nu = \mathrm{C}\int f \,\mathrm{d}\nu_*$. Hence,

$$\mathrm{C}\int f \,\mathrm{d}\nu = \mathrm{C}\int f \,\mathrm{d}\nu_* = b_0 + \sum_{i=1}^n b_i \nu_*(B_i),$$

where the last equality was established in Eq. (3.12) on p. 75. From the coherence of $\underline{\mathbf{E}}_\nu$ and Theorem 4.36, it follows that

$$\underline{\mathbf{E}}_\nu\left(b_0 + \sum_{i=1}^n b_i I_{B_i}\right) \geq b_0 + \sum_{i=1}^n b_i \underline{\mathbf{E}}_\nu(I_{B_i}) = b_0 + \sum_{i=1}^n b_i \nu_*(B_i),$$

hence,

$$\underline{\mathbf{E}}_\nu(f) \geq \mathrm{C}\int f \,\mathrm{d}\nu.$$

It was established in Lemma 3.30 that $\mathrm{C}\int \bullet \,\mathrm{d}\nu_*$ is a coherent lower prevision on the set of $\wp(\mathcal{X})$-simple gambles. Obviously, $\mathrm{C}\int \bullet \,\mathrm{d}\nu_*$ is a behavioural extension of ν_*. Hence, by the definition of natural extension as the point-wise smallest behavioural extension,

$$\mathrm{C}\int f\,\mathrm{d}\nu = \mathrm{C}\int f\,\mathrm{d}\nu_* \geq \underline{\mathbf{E}}_{\nu_*}(f) = \underline{\mathbf{E}}_{\nu}(f),$$

where we again applied Theorem 4.36 and Proposition 4.8.

We conclude that $\mathrm{C}\int \bullet \,\mathrm{d}\nu = \underline{\mathbf{E}}_{\nu}$ on the set of $\wp(\mathcal{X})$-simple gambles. But, with respect to the topology of uniform convergence on $\mathcal{L}(X)$, this set is dense in $\mathcal{L}(X)$, so if we can show that both lower previsions $\underline{\mathbf{E}}_{\nu}$ and $\mathrm{C}\int \bullet \,\mathrm{d}\nu$ are continuous on $\mathcal{L}(X)$, then $\mathrm{C}\int \bullet \,\mathrm{d}\nu = \underline{\mathbf{E}}_{\nu}$ on the set of all gambles on X, and the theorem is established.

Indeed, $\underline{\mathbf{E}}_{\nu}$ is coherent, so it is continuous with respect to the topology of uniform convergence by Theorem 3.5(xiii) on p. 55. Also, for any $\epsilon > 0$ and any gambles f and g on X, if $\sup|f - g| < \epsilon$, then

$$\mathrm{C}\int f\,\mathrm{d}\nu = \inf f + \mathrm{R}\int_{\inf f}^{\sup f} \nu_*(\{x \in \mathcal{X}\colon f(x) > z\})\,\mathrm{d}z$$

and since $\{x \in \mathcal{X}\colon f(x) > z\} \subseteq \{x \in \mathcal{X}\colon g(x) + \epsilon > z\}$, and $\inf f \leq \inf g + \epsilon$,

$$\begin{aligned}
&\leq \inf g + \epsilon + \mathrm{R}\int_{\inf f}^{\sup f} \nu_*(\{x \in \mathcal{X}\colon g(x) + \epsilon > z\})\,\mathrm{d}z \\
&= \inf g + \epsilon + \mathrm{R}\int_{\inf f-\epsilon}^{\sup f-\epsilon} \nu_*(\{x \in \mathcal{X}\colon g(x) > z\})\,\mathrm{d}z
\end{aligned}$$

but, $\nu_*(\{x \in \mathcal{X}\colon g(x) > z\}) = G_{*\nu,g}(z)$, and invoking Proposition 4.65 on p. 167,

$$\begin{aligned}
&= \inf g + \epsilon + \mathrm{R}\int_{\inf g}^{\sup g} G_{*\nu,g}(z)\,\mathrm{d}z \\
&\quad + \mathrm{R}\int_{\inf f-\epsilon}^{\inf g} G_{*\nu,g}(z)\,\mathrm{d}z - \mathrm{R}\int_{\sup f-\epsilon}^{\sup g} G_{*\nu,g}(z)\,\mathrm{d}z
\end{aligned}$$

and since $0 \leq G_{*\nu,g}(z) \leq 1$, it follows that $\mathrm{R}\int_{\inf f-\epsilon}^{\inf g} G_{*\nu,g}(z)\,\mathrm{d}z \leq \inf g - \inf f + \epsilon \leq \sup|f-g| + \epsilon < 2\epsilon$, and $-\mathrm{R}\int_{\sup f-\epsilon}^{\sup g} G_{*\nu,g}(z)\,\mathrm{d}z \leq 0$, so

$$< 3\epsilon + \mathrm{C}\int g\,\mathrm{d}\nu.$$

Reversing the rôles of f and g, we find that also $\mathrm{C}\int g\,\mathrm{d}\nu < 3\epsilon + \mathrm{C}\int f\,\mathrm{d}\nu$. Concluding, if $\sup|f-g| < \epsilon$, then it must hold that

$$\left|\mathrm{C}\int f\,\mathrm{d}\nu - \mathrm{C}\int g\,\mathrm{d}\nu\right| < 3\epsilon,$$

so $\mathrm{C}\int \bullet\,\mathrm{d}\nu$ is uniformly continuous with respect to the topology of uniform convergence on $\mathcal{L}(X)$. (Note: now we have proved that $\mathrm{C}\int \bullet\,\mathrm{d}\nu$ is coherent, the bound 3ϵ in the above equation can actually be improved to ϵ; see Theorem 3.5(xiii) on p. 55.) □

We can now also prove the following quite remarkable result. The "only if" part of the proof is due to Hugo Janssen (personal communication). Recall the definition of 2-monotone lower previsions: see Definition 3.32 on p. 76.

Proposition 4.77. *A coherent lower prevision, defined on the set $\mathcal{L}(\mathcal{X})$ of all gambles on X, is 2-monotone, if and only if there is a field $\mathcal{F}$ on $\mathcal{X}$ such that the set function ν, defined by $\nu(A) := \underline{P}(I_A)$ for all A in $\mathcal{F}$, is 2-monotone, and*

$$\underline{P}(f) = \underline{\mathbf{E}}_{\underline{\mathbf{P}}_\nu}(f) = \mathrm{C}\int f\,\mathrm{d}\nu,$$

for all gambles f on X.

Proof. "if". It suffices to show that

$$\mathrm{C}\int f\vee g\,\mathrm{d}\nu + \mathrm{C}\int f\wedge g\,\mathrm{d}\nu \geq \mathrm{C}\int f\,\mathrm{d}\nu + \mathrm{C}\int g\,\mathrm{d}\nu$$

for all gambles f and g on X. This is a well-known result; see for instance Denneberg [28, p. 162, ll. 1–3, and Exercise 13.1, p. 170].

"only if". Define $\mathcal{F} := \wp(\mathcal{X})$ and $\nu(A) := \underline{P}(I_A)$ for all $A \subseteq \mathcal{X}$. The natural

extension of $\underline{\mathbf{P}}_\nu$ is given by

$$\underline{\mathbf{E}}_\nu(f) = \sup\left\{\sum_{i=1}^{n} \lambda_i \underline{P}(I_{A_i})\colon n \in \mathbb{N}, \lambda_i \geq 0, A_i \subseteq \mathcal{X}, \sum_{i=1}^{n} \lambda_i I_{A_i} \leq f\right\}.$$

It is easy to see that $\underline{\mathbf{E}}_\nu \leq \underline{P}$. Indeed, by definition, $\underline{\mathbf{E}}_\nu$ is the smallest coherent extension of $\underline{\mathbf{P}}_\nu$, and obviously, $\underline{P}$ is a coherent extension of $\underline{\mathbf{P}}_\nu$, so it should hold that $\underline{\mathbf{E}}_\nu \leq \underline{P}$.

If we can now show that, for any $\wp(\mathcal{X})$-simple gamble $f = \sum_{j=1}^{m} \mu_j I_{B_j}$, there is a particular choice of μ_j's and B_j's such that $\sum_{j=1}^{m} \mu_j \underline{P}(I_{B_j}) \geq \underline{P}(f)$, then we have proved that $\underline{\mathbf{E}}_\nu(f) \geq \underline{P}(f)$, and hence, $\underline{\mathbf{E}}_\nu(f) = \underline{P}(f)$, for all $\wp(\mathcal{X})$-simple gambles f.

Indeed, assuming that f is $\wp(\mathcal{X})$-simple, it may be written as $\mu_1 I_{B_1} + \sum_{j=2}^{m} \mu_j I_{B_j}$ with $\mu_1 \in \mathbb{R}$, $B_1 = \mathcal{X}$, $\mu_j \geq 0$ for $j > 1$, and $B_j \supsetneq B_{j+1}$ for $1 < j < m$. From the coherence and the 2-monotonicity of $\underline{P}$, it follows that

$$\begin{aligned}
\underline{P}(f) &= \mu_1 - \mu_2 + \underline{P}\left(\sum_{j=2}^{m} \mu_j B_j\right) + \underline{P}(\mu_2) \\
&\leq \mu_1 - \mu_2 + \underline{P}\left(\min\left\{\mu_2, \sum_{j=2}^{m} \mu_j B_j\right\}\right) + \underline{P}\left(\max\left\{\mu_2, \sum_{j=2}^{m} \mu_j B_j\right\}\right) \\
&= \mu_1 - \mu_2 + \underline{P}\left(\mu_2 B_2\right) + \underline{P}\left(\mu_2 + \sum_{j=3}^{m} \mu_3 B_j\right) \\
&= \mu_1 + \mu_2 \underline{P}\left(B_2\right) + \underline{P}\left(\sum_{j=3}^{m} \mu_3 B_j\right) \\
&= \mu_1 - \mu_3 + \mu_2 \underline{P}\left(B_2\right) + \underline{P}\left(\sum_{j=3}^{m} \mu_3 B_j\right) + \underline{P}(\mu_3) \\
&\leq \mu_1 - \mu_3 + \mu_2 \underline{P}\left(B_2\right) + \underline{P}\left(\min\left\{\mu_3, \sum_{j=3}^{m} \mu_j B_j\right\}\right) + \underline{P}\left(\max\left\{\mu_3, \sum_{j=3}^{m} \mu_j B_j\right\}\right) \\
&= \mu_1 - \mu_3 + \mu_2 \underline{P}\left(B_2\right) + \underline{P}\left(\mu_3 B_3\right) + \underline{P}\left(\mu_3 + \sum_{j=4}^{m} \mu_j B_j\right) \\
&= \mu_1 + \mu_2 \underline{P}\left(B_2\right) + \mu_3 \underline{P}\left(B_3\right) + \underline{P}\left(\sum_{j=4}^{m} \mu_j B_j\right)
\end{aligned}$$

$$\leq \dots$$
$$\leq \mu_1 + \sum_{j=2}^{m} \mu_j \underline{P}\big(B_j\big).$$

This ends the proof for $\wp(\mathcal{X})$-simple gambles.

Now, for any gamble f on X, there is a sequence of simple gambles f_n converging uniformly to f such that $f \geq f_n$ for all $n \in \mathbb{N}$. Hence, it holds that $\underline{\mathbf{E}}_\nu(f) \geq \underline{\mathbf{E}}_\nu(f_n) = \underline{P}(f_n)$ for every $n \in \mathbb{N}$. Since, by the coherence of $\underline{P}$, $\underline{P}(f_n)$ converges to $\underline{P}(f)$, and $\geq$ is preserved in the limit, we find that $\underline{\mathbf{E}}_\nu(f) \geq \underline{P}(f)$. □

So, any coherent lower prevision that is 2-monotone on $\mathcal{L}(X)$ is representable by a Choquet integral. Krätschmer [53, Proposition 4.3, p. 477, and Theorem 4.4(3), p. 478] has given necessary and sufficient conditions for a coherent lower prevision to be representable by a Choquet integral, with only very few restrictions on the domain of the lower prevision, based on Greco's representation theorem; he assumes that $\underline{P}$ is Stonean (whereas we assumed 2-monotonicity), which is in case of a coherent lower prevision $\underline{P}$ on $\mathcal{L}(X)$ equivalent to $\underline{P}(f \vee a) + \underline{P}(f \wedge a) = \underline{P}(f) + a$ for all gambles f (this is a weak version of co-monotone additivity).

4.3.11 Natural Extension of Nested Set Functions

Theorem 4.78. *Let ν and π be nested set functions. For any gamble f on X, it holds that*

$$\underline{\mathbf{E}}_{\underline{\mathbf{P}}_\nu}(f) = \inf f + \mathrm{R}\int_{\inf f}^{\sup f} \sup\{\nu(A)\colon A \in \operatorname{dom}\nu,\, A \subseteq \{x \in \mathcal{X}\colon f(x) > z\}\}\,\mathrm{d}z,$$
$$\underline{\mathbf{E}}_{\overline{\mathbf{P}}_\pi}(f) = \sup f - \mathrm{R}\int_{\inf f}^{\sup f} \inf\{\pi(A)\colon A \in \operatorname{dom}\pi,\, A \supseteq \{x \in \mathcal{X}\colon f(x) \leq z\}\}\,\mathrm{d}z.$$

Proof. The first equality follows from Theorem 4.36(i)&(ii) on p. 117, Theorem 3.46 on p. 85, Corollary 4.9 on p. 98, and Theorem 4.76. The second equality follows then from Proposition 4.10 on p. 99: if ν denotes the dual of

π, then

$$\begin{aligned}
&\underline{\mathbf{E}}_{\overline{\mathbf{P}}_\pi}(f) = \underline{\mathbf{E}}_{\underline{\mathbf{P}}_\nu}(f) \\
&= \inf f + \mathrm{R}\int_{\inf f}^{\sup f} \sup\{\nu(A)\colon A \in \operatorname{dom}\nu,\, A \subseteq \{x \in \mathcal{X}\colon f(x) > z\}\}\,\mathrm{d}z \\
&= \inf f + \mathrm{R}\int_{\inf f}^{\sup f} \sup\{1 - \pi(\complement A)\colon A \in \operatorname{dom}\nu,\, A \subseteq \{x \in \mathcal{X}\colon f(x) > z\}\}\,\mathrm{d}z \\
&= \sup f - \mathrm{R}\int_{\inf f}^{\sup f} \inf\{\pi(A)\colon A \in \operatorname{dom}\pi,\, A \supseteq \{x \in \mathcal{X}\colon f(x) \leq z\}\}\,\mathrm{d}z
\end{aligned}$$

□

4.3.12 Natural Extension of Belief Functions

We already constructed the natural extension of a belief function in Theorem 3.44 on p. 84; the proof of this claim is immediate from Theorem 4.76.

4.3.13 Natural Extension of Necessity and Possibility Measures

The equations below are alternative expressions for natural extension of necessity and possibility measures given by De Cooman and Aeyels [21], Eq. (1)&(2) for arbitrary gambles (*i.e.*, not necessarily $\mathcal{F}$-measurable gambles).

Theorem 4.79. *Let N be a necessity measure with necessity distribution ν, and let Π be a possibility measure with possibility distribution π. For any gamble f on X, it holds that*

$$\underline{\mathbf{E}}_{\underline{\mathbf{P}}_N}(f) = \inf f + \mathrm{R}\int_{\inf f}^{\sup f} \inf_{\substack{x\in\mathcal{X}\\ f(x)\leq z}} \nu(x)\,\mathrm{d}z,$$

$$\underline{\mathbf{E}}_{\overline{\mathbf{P}}_\Pi}(f) = \sup f - \mathrm{R}\int_{\inf f}^{\sup f} \sup_{\substack{x\in\mathcal{X}\\ f(x)\leq z}} \pi(x)\,\mathrm{d}z.$$

Proof. To prove the first equality, use Theorem 4.36(vi) on p. 117, Theorem 3.46 on p. 85, Corollary 4.9 on p. 98, and Theorem 4.76—indeed,

$$G_{*N,f}(z) = N_*(\{x \in \mathcal{X}\colon f(x) > z\}) = \inf_{\substack{x\in\mathcal{X}\\ f(x)\leq z}} \nu(x).$$

The second equality follows then from Proposition 4.10 on p. 99. □

4.4 Duality

4.4.1 Avoiding Sure Loss, Coherence, and Natural Extension: An Alternative Characterisation

There is a nice connection between lower previsions and compact convex sets of linear previsions.

Definition 4.80. We can define a map from lower previsions to sets of linear previsions. Let $\mathcal{K} \subseteq \mathcal{L}(X)$, and assume that $\mathcal{K}$ is negation invariant: $\mathcal{K} = -\mathcal{K}$. With any lower prevision $\underline{P}$ such that $\operatorname{dom} \underline{P} \subseteq \mathcal{K}$, we can associate the set of all linear behavioural extensions of $\underline{P}$ with domain $\mathcal{K}$

$$\mathbf{M}_{\underline{P}}^{\mathcal{K}} = \Big\{Q\colon Q \in \mathcal{P}^{\mathcal{K}}(X),\ (\forall f \in \operatorname{dom} \underline{P})\Big(Q(f) \geq \underline{P}(f)\Big)\Big\}. \tag{4.49}$$

The set $\mathbf{M}_{\underline{P}}^{\mathcal{L}(X)}$ is denoted by $\mathbf{M}_{\underline{P}}$.

Conversely, we can define a map from sets of linear previsions to lower previsions. With any set $\mathcal{M}$ of linear previsions whose domain includes $\mathcal{K}$, we can associate a lower prevision $\underline{\mathbf{E}}_{\mathcal{M}}^{\mathcal{K}}$ on $\mathcal{K}$, defined by

$$\underline{\mathbf{E}}_{\mathcal{M}}^{\mathcal{K}}(f) := \inf\{Q(f)\colon Q \in \mathcal{M}\}. \tag{4.50}$$

for any $f \in \mathcal{K}$.

The following theorem is a generalisation of a result by Walley [86, Sections 3.3.3&3.4.1, pp. 134–136].

Theorem 4.81. *Let $\underline{P}$ be any lower prevision, let $\mathcal{K}$ be a negation invariant subset of $\mathcal{L}(X)$, and assume that $\operatorname{dom} \underline{P} \subseteq \mathcal{K}$. Then the following propositions hold.*

(i) $\underline{P}$ avoids sure loss if and only if

$$\mathbf{M}_{\underline{P}}^{\mathcal{K}} \neq \emptyset. \tag{4.51}$$

(ii) If $\underline{P}$ avoids sure loss, then its natural extension $\underline{\mathbf{E}}_{\underline{P}}$ satisfies

$$\underline{\mathbf{E}}_{\underline{P}}(f) = \min\left\{\underline{\mathbf{E}}_Q(f)\colon Q \in \mathbf{M}_{\underline{P}}^{\mathcal{K}}\right\} \quad \textit{for any gamble } f \textit{ on } X\textit{; hence,} \tag{4.52}$$

$$= \min\left\{Q(f)\colon Q \in \mathbf{M}_{\underline{P}}^{\mathcal{K}}\right\} \quad \textit{for any } f \in \mathcal{K}. \tag{4.53}$$

(iii) $\underline{P}$ is coherent if and only if

$$\underline{P}(f) = \min\left\{Q(f)\colon Q \in \mathbf{M}_{\underline{P}}^{\mathcal{K}}\right\} \tag{4.54}$$

for any $f \in \operatorname{dom}\underline{P}$.

Proof. See Walley [86, Sections 3.3.3&3.4.1, pp. 134–136] for the case in which $\mathcal{K} = \mathcal{L}(X)$. The general case is then straightforward. Let's prove Eq. (4.52).

Let $f \in \mathcal{L}(X)$. Any $Q \in \mathbf{M}_{\underline{P}}^{\mathcal{K}}$ is a coherent (even linear) behavioural extension of $\underline{P}$. So, by Proposition 4.7 on p. 98, for any $Q \in \mathbf{M}_{\underline{P}}^{\mathcal{K}}$, $\underline{\mathbf{E}}_Q$ is a coherent behavioural extension of $\underline{\mathbf{E}}_{\underline{P}}$, and hence,

$$\underline{\mathbf{E}}_{\underline{P}}(f) \geq \inf\left\{\underline{\mathbf{E}}_Q(f)\colon Q \in \mathbf{M}_{\underline{P}}^{\mathcal{K}}\right\}.$$

If we now can show that there is a linear prevision $Q \in \mathbf{M}_{\underline{P}}^{\mathcal{K}}$ such that $\underline{\mathbf{E}}_{\underline{P}}(f) \leq \underline{\mathbf{E}}_Q(f)$, then Eq. (4.52) is established. By Walley [86, Section 3.4.1, p. 136], we know that

$$\underline{\mathbf{E}}_{\underline{P}}(f) = \min\left\{R(f)\colon R \in \mathbf{M}_{\underline{P}}\right\},$$

or equivalently, there is a linear prevision S in $\mathbf{M}_{\underline{P}}$ such that $\underline{\mathbf{E}}_{\underline{P}}(f) = S(f)$. Define Q as the restriction of S to $\mathcal{K}$: $Q := S|_{\mathcal{K}}$. Then, again by Walley [86, Section 3.4.1, p. 136],

$$\underline{\mathbf{E}}_Q(f) = \min\left\{R(f)\colon R \in \mathbf{M}_Q\right\} \leq S(f) = \underline{\mathbf{E}}_{\underline{P}}(f),$$

since $S \in \mathbf{M}_Q$.

The remainder of the proof is easy. □

The two mappings given in Definition 4.80 constitute isomorphisms between particular sets of sets of linear previsions and lower previsions. To characterise these isomorphisms, we endow the set $\mathcal{P}^{\mathcal{K}}(X)$ of linear previsions defined on a common domain $\mathcal{K} \subseteq \mathcal{L}(X)$ with the *topology of point-wise*

convergence on members of $\mathcal{K}$, or simply, the *topology of point-wise convergence*: $Q_\alpha \to Q$ if and only if $Q_\alpha(f) \to Q(f)$ for all f in $\mathcal{K}$ (with respect to the usual topology on $\mathbb{R}$).

If $\mathcal{K}$ is a linear space, and if we view $\mathcal{P}^{\mathcal{K}}(X)$ as a subset of the set $\mathcal{V}^{\mathcal{K}}(X)$ of all continuous linear real-valued maps on $\mathcal{K}$ (where $\mathcal{K}$ is endowed with the topology of uniform convergence), *i.e.*, the *topological dual of* $\mathcal{K}$, then the topology of point-wise convergence is nothing but the so-called *weak-* topology*. The weak-* topology is well-studied in the literature: it turns $\mathcal{V}^{\mathcal{K}}(X)$ into a locally convex topological vector space (see Schechter [70, 28.15(a), p. 760]) that is Hausdorff (see Schechter [70, 28.15(b), p. 760]) and paracompact (see Kelley [49, Problem N(c), p. 242]). In particular, the weak-* topology is normal (see Kelley [49, Chapter 5, Corollary 32, p. 159]): this turns out to be crucial in linking the weak-* topology on $\mathcal{V}^{\mathcal{K}}(X)$ with the topology of point-wise convergence on $\mathcal{P}^{\mathcal{K}}(X)$.

For an arbitrary negation invariant set $\mathcal{K}$ of gambles, and a gamble f in $\mathcal{K}$, the real-valued map f^* defined by

$$f^*(Q) := Q(f)$$

for any linear prevision Q in $\mathcal{P}^{\mathcal{K}}(X)$ is called *evaluation map* on $\mathcal{P}^{\mathcal{K}}(X)$ induced by f. The set of all evaluation maps on $\mathcal{P}^{\mathcal{K}}(X)$ induced by gambles in $\mathcal{K}$ is denoted by

$$\mathcal{K}^* := \{f^* : f \in \mathcal{K}\}.$$

Note that the topology of point-wise convergence is the weakest topology under which all elements of $\mathcal{K}^*$ are continuous: it is the *topology induced by* $\mathcal{K}^*$.

We first link the topology on $\mathcal{P}^{\mathcal{K}}(X)$ of point-wise convergence on members of $\mathcal{K}$, the weak-* topology on the topological dual of $\mathcal{V}^{\text{span}(\mathcal{K})}(X)$, and the subset $\mathcal{P}^{\text{span}(\mathcal{K})}(X)$ of $\mathcal{V}^{\text{span}(\mathcal{K})}(X)$, in the lemma below.

Lemma 4.82. *Let* $\mathcal{K} \subseteq \mathcal{L}(X)$*, and assume that* $\mathcal{K}$ *is negation invariant. Endow* $\mathcal{P}^{\mathcal{K}}(X)$ *with the topology of point-wise convergence, and endow* $\mathcal{V}^{\text{span}(\mathcal{K})}(X)$ *with the weak-* topology. Consider* $\mathbf{E}_\bullet^{\text{span}(\mathcal{K})}$ *as a mapping from* $\mathcal{P}^{\mathcal{K}}(X)$ *to* $\mathcal{V}^{\text{span}(\mathcal{K})}(X)$*. The following statements hold.*

(i) $\mathbf{E}_\bullet^{\text{span}(\mathcal{K})}$ *is one-to-one, and maps* $\mathcal{P}^{\mathcal{K}}(X)$ *onto* $\mathcal{P}^{\text{span}(\mathcal{K})}(X)$*.*

(ii) $\mathbf{E}_\bullet^{\text{span}(\mathcal{K})}$ *establishes an embedding of* $\mathcal{P}^{\mathcal{K}}(X)$ *into* $\mathcal{V}^{\text{span}(\mathcal{K})}(X)$*.*

(iii) $\mathcal{P}^{\text{span}(\mathcal{K})}(X)$ *is a compact subset of* $\mathcal{V}^{\text{span}(\mathcal{K})}(X)$*, and hence,* $\mathcal{P}^{\mathcal{K}}(X)$ *is a compact space.*

Proof. To simplify our notation, we shall denote $\mathbf{E}_P^{\text{span}(\mathcal{K})}$ by P'.

(i). By Proposition 4.13 on p. 100 and Proposition 4.18(v) on p. 102, every linear prevision P on $\mathcal{K}$ has a unique linear extension P' to span($\mathcal{K}$), which satisfies

$$P'\left(\sum_{i=1}^{n} \lambda_i f_i\right) = \sum_{i=1}^{n} \lambda_i P(f_i),$$

for all $n \in \mathbb{N}$, non-negative real numbers $\lambda_1, \ldots, \lambda_n$, and gambles $f_1, \ldots, f_n$ in $\mathcal{K}$; this equality simply follows from the self-conjugacy and the coherence of P (Theorem 3.5(v)&(vi) on p. 55). This establishes a one-to-one and onto mapping between $\mathcal{P}^{\mathcal{K}}(X)$ and $\mathcal{P}^{\text{span}(\mathcal{K})}(X)$—one-to-one because of the uniqueness of the linear extension, and onto because any linear prevision R on span($\mathcal{K}$) is uniquely determined by its values on $\mathcal{K}$, *i.e.*, $R = Q'$ for some linear prevision Q on $\mathcal{K}$.

(ii). First, note that the relative topology on $\mathcal{P}^{\text{span}(\mathcal{K})}(X)$, as a subset of $\mathcal{V}^{\text{span}(\mathcal{K})}(X)$ equipped with the weak-* topology, is exactly the topology of point-wise convergence on members of span($\mathcal{K}$). So we must prove that $\mathbf{E}_\bullet^{\text{span}(\mathcal{K})}$ establishes a homeomorphism between $\mathcal{P}^{\mathcal{K}}(X)$ equipped with the topology of point-wise convergence on members of $\mathcal{K}$, and its image $\mathcal{P}^{\text{span}(\mathcal{K})}(X)$ equipped with the topology of point-wise convergence on members of span($\mathcal{K}$): let Q_α denote a net in $\mathcal{P}^{\mathcal{K}}(X)$, let Q denote an element of $\mathcal{P}^{\mathcal{K}}(X)$, and let Q'_α and Q' denote the unique linear extensions of Q_α and Q to span($\mathcal{K}$); does it hold that $Q_\alpha(f) \to Q(f)$ for all $f \in \mathcal{K}$ if and only if $Q'_\alpha(g) \to Q'(g)$ for all $g \in \text{span}(\mathcal{K})$?

Clearly, if $Q'_\alpha(g) \to Q'(g)$ for all gambles $g \in \text{span}(\mathcal{K})$, then also $Q_\alpha(f) = Q'_\alpha(f) \to Q'(f) = Q(f)$ for all $f \in \mathcal{K}$.

Conversely, if $Q_\alpha(f) \to Q(f)$ for all $f \in \mathcal{K}$, then

$$Q'_\alpha\left(\sum_{i=1}^{n} \lambda_i f_i\right) = \sum_{i=1}^{n} \lambda_i Q_\alpha(f_i) \to \sum_{i=1}^{n} \lambda_i Q(f_i) = Q'\left(\sum_{i=1}^{n} \lambda_i f_i\right)$$

for all $n \in \mathbb{N}$, non-negative real numbers $\lambda_1, \ldots, \lambda_n$, and gambles $f_1, \ldots, f_n$ in $\mathcal{K}$. So, $Q'_\alpha(g) \to Q'(g)$ for all $g \in \text{span}(\mathcal{K})$.

(iii). (Method of proof due to Walley [86, Section 3.6, pp. 145–146]) Clearly,

by Theorem 3.7 on p. 56, it follows that

$$\mathcal{P}^{\text{span}(\mathcal{K})}(X) = \bigcap_{f \in \text{span}(\mathcal{K})} [f^*]^{-1}([\inf f, +\infty)),$$

where f^* is the evaluation map on $\mathcal{V}^{\text{span}(\mathcal{K})}(X)$ induced by f. So, $\mathcal{P}^{\text{span}(\mathcal{K})}(X)$ is a closed subset of $\mathcal{V}^{\text{span}(\mathcal{K})}(X)$. Also, for any $\epsilon > 0$, if $\sup|f - g| < \epsilon$ for two gambles f and g in $\text{span}(\mathcal{K})$, then for any P in $\mathcal{P}^{\text{span}(\mathcal{K})}(X)$ it holds that

$$\left|P(f) - P(g)\right| < \epsilon,$$

by the coherence of P (Theorem 3.5 on p. 55). So, $\mathcal{P}^{\text{span}(\mathcal{K})}(X)$ is also equicontinuous. Compactness of $\mathcal{P}^{\text{span}(\mathcal{K})}(X)$ is now immediate from the Banach-Alaoglu-Bourbaki theorem; see for instance Schechter [70, 28.29(UF26)].

Compactness of $\mathcal{P}^{\text{span}(\mathcal{K})}(X)$ now simply follows from (i) and (ii): since $\mathcal{P}^{\text{span}(\mathcal{K})}(X)$ is a compact subset of $\mathcal{V}^{\text{span}(\mathcal{K})}(X)$, it follows that the topological space $\mathcal{P}^{\text{span}(\mathcal{K})}(X)$, equipped with the relative topology, is compact; the inverse of $\mathbf{E}_\bullet^{\text{span}(\mathcal{K})}$ is continuous on this compact space, and maps $\mathcal{P}^{\text{span}(\mathcal{K})}(X)$ onto $\mathcal{P}^{\mathcal{K}}(X)$. But, the continuous image of a compact set is compact, whence $\mathcal{P}^{\mathcal{K}}(X)$ is compact too. □

Surprisingly, $\text{span}(\mathcal{K}^*)$ characterises all continuous real-valued mappings on $\mathcal{P}^{\mathcal{K}}(X)$:

Corollary 4.83. *Let $\mathcal{K} \subseteq \mathcal{L}(X)$, and assume that $\mathcal{K}$ is negation invariant. A real-valued mapping on $\mathcal{P}^{\mathcal{K}}(X)$ is continuous with respect to the topology of point-wise convergence on members of $\mathcal{K}$ if and only if it belongs to* $\text{span}(\mathcal{K}^*)$.

Proof. Let Λ be a real-valued mapping on $\mathcal{P}^{\mathcal{K}}(X)$. Obviously, if $\Lambda \in \text{span}(\mathcal{K}^*)$, then Λ is continuous, since if $Q_\alpha \to Q$, then

$$\Lambda(Q_\alpha) = \sum_{i=1}^{n} \lambda_i f_i^*(Q_\alpha) = \sum_{i=1}^{n} \lambda_i Q_\alpha(f_i) \to \sum_{i=1}^{n} \lambda_i Q(f_i) = \sum_{i=1}^{n} \lambda_i f_i^*(Q) = \Lambda(Q).$$

Conversely, assume that Λ is continuous on $\mathcal{P}^{\mathcal{K}}(X)$. Define the real-valued mapping Λ' on $\mathcal{P}^{\text{span}(\mathcal{K})}(X)$ by

$$\Lambda'(Q') := \Lambda \circ \left[\mathbf{E}_\bullet^{\text{span}(\mathcal{K})}\right]^{-1}(Q') = \Lambda(Q),$$

for every $Q \in \mathcal{P}^{\mathcal{K}}(X)$; where Q' is the unique linear extension of Q. By Lemma 4.82, $\mathcal{P}^{\text{span}(\mathcal{K})}(X)$ is a weak-* compact, and hence, a weak-* closed subset of $\mathcal{V}^{\text{span}(\mathcal{K})}(X)$. Moreover, the weak-* topology is paracompact (see Kelley [49, Problem N(c), p. 242]), and hence, normal (see Kelley [49, Chapter 5, Corollary 32, p. 159]). Therefore, by Tietze's extension theorem (see Kelley [49, Problem O(a), p. 242] or Willard [91, p. 103]), we can continuously extend Λ' to $\mathcal{V}^{\text{span}(\mathcal{K})}(X)$.

It is well-known that Λ' is continuous with respect to the weak-* topology if and only if it is an evaluation map on $\mathcal{V}^{\text{span}(\mathcal{K})}(X)$; see for instance Schechter [70, 28.15(c)]. Therefore, there is a gamble $f \in \text{span}(\mathcal{K})$ such that $\Lambda'(Q') = Q'(f)$ for all Q' in $\mathcal{P}^{\text{span}(\mathcal{K})}(X)$, since $\mathcal{P}^{\text{span}(\mathcal{K})}(X) \subseteq \mathcal{V}^{\text{span}(\mathcal{K})}(X)$; see for instance Kelley [49, Problem W(c), p. 108], or Schechter [70, 28.15(c)]. Consequently, there are $n \in \mathbb{N}$, real numbers $\lambda_1, \ldots, \lambda_n$, and gambles $f_1, \ldots, f_n$ in $\mathcal{K}$, such that for every Q in $\mathcal{P}^{\mathcal{K}}(X)$ (recall that $\mathcal{P}^{\mathcal{K}}(X)$ is homeomorphic to $\mathcal{P}^{\text{span}(\mathcal{K})}(X)$):

$$\Lambda(Q) = \Lambda'(Q') = Q'\left(\sum_{i=1}^{n} \lambda_i f_i\right)$$

where the right hand side can also be written as

$$= \sum_{i=1}^{n} \lambda_i Q(f_i) = \sum_{i=1}^{n} \lambda_i f_i^*(Q).$$

So, $\Lambda \in \text{span}(\mathcal{K}^*)$. □

We are now ready to prove our isomorphism. It is convenient to first introduce the following definition.

Definition 4.84. Let $\mathcal{K} \subseteq \mathcal{L}(X)$, and assume that $\mathcal{K}$ is negation invariant. A compact (with respect to the topology of point-wise convergence) convex set $\mathcal{M}$ of linear previsions on $\mathcal{K}$ is called *decomposable* if, for every $\lambda_1, \ldots, \lambda_n$ in $\mathbb{R}$, $f_1, \ldots, f_n$ in $\mathcal{K}$, and $\epsilon > 0$, there are μ^ϵ in $\mathbb{R}$, n^ϵ in $\mathbb{N}$, non-negative $\lambda_1^\epsilon, \ldots, \lambda_{n^\epsilon}^\epsilon$ in $\mathbb{R}$, and gambles $f_1^\epsilon, \ldots, f_{n^\epsilon}^\epsilon$ in $\mathcal{K}$, such that $\mu^\epsilon + \sum_{i=1}^{n^\epsilon} \lambda_i^\epsilon f_i^\epsilon \leq \sum_{i=1}^{n} \lambda_i f_i$ and

$$\min_{Q \in \mathcal{M}} \sum_{i=1}^{n} \lambda_i Q(f_i) \leq \epsilon + \mu^\epsilon + \sum_{i=1}^{n^\epsilon} \lambda_i^\epsilon \min_{Q_i \in \mathcal{M}} Q_i(f_i^\epsilon)$$

If $\mathcal{K}$ is a linear space, then every compact convex set $\mathcal{M}$ of linear previsions on $\mathcal{K}$ is decomposable: we can always choose $\mu^\epsilon = 0$, $n^\epsilon = 1$, $\lambda_1^\epsilon = 1$, and

$f_1^\epsilon = \sum_{i=1}^n \lambda_i f_i$. It then follows from the linearity of Q in $\mathcal{M}$ that $\sum_{i=1}^n \lambda_i Q(f_i) = Q(f_1^\epsilon)$.

In general, if $\mathcal{K}$ is not a linear space, then not every compact convex set of linear previsions on $\mathcal{K}$ is decomposable. For instance, let $\mathcal{X} = \{1, 2, 3\}$, $f = I_{\{1\}}$, and $g = I_{\{2\}}$, and define the linear previsions Q and R on $\{f, -f, g, -g\}$ as

$$\begin{aligned} Q(f) &= -Q(-f) = 0.1, & R(f) &= -R(-f) = 0.2, \\ Q(g) &= -Q(-g) = 0.1, & R(g) &= -R(-g) = 0.2, \end{aligned}$$

and define the compact and convex set

$$\mathcal{M} := \{\lambda Q + (1-\lambda)R \colon \lambda \in [0,1]\},$$

Then $\mathcal{M}$ is not decomposable: indeed, it's not hard to see that $\mathbf{M}^{\mathcal{K}}_{\underline{\mathbf{E}}^{\mathcal{K}}_{\mathcal{M}}} \neq \mathcal{M}$:

$$\begin{aligned} \mathbf{M}^{\mathcal{K}}_{\underline{\mathbf{E}}^{\mathcal{K}}_{\mathcal{M}}} = \Big\{ S \in \mathcal{P}^{\mathcal{K}}(X) \colon\ & Q(f) \wedge R(f) \leq S(f) \leq Q(f) \vee R(f) \text{ and} \\ & Q(g) \wedge R(g) \leq S(g) \leq Q(g) \vee R(g) \Big\}, \end{aligned}$$

for instance, $S(f) = -S(-f) = Q(f) \wedge R(f) = 0.1$ and $S(g) = -S(-g) = Q(g) \vee R(g) = 0.2$ is a linear prevision that belongs to $\mathbf{M}^{\mathcal{K}}_{\underline{\mathbf{E}}^{\mathcal{K}}_{\mathcal{M}}}$ but not to $\mathcal{M}$. As the theorem below tells us, it follows that $\mathcal{M}$ is not decomposable.

Theorem 4.85. *Let $\mathcal{K} \subseteq \mathcal{L}(X)$, and assume that $\mathcal{K}$ is negation invariant. Then $\underline{\mathbf{E}}^{\mathcal{K}}_{\bullet}$ and $\mathbf{M}^{\mathcal{K}}_{\bullet}$ establish onto and one-to-one maps between non-empty compact (with respect to the topology of point-wise convergence) convex decomposable sets of linear previsions on $\mathcal{K}$ and coherent lower previsions on $\mathcal{K}$.*

Proof. Walley [86, Section 3.6.1, pp. 145–146] proved the case in which $\mathcal{K} = \mathcal{L}(X)$. The general case is straightforward, once the condition of decomposability is recognised; let's complete the details.

Let $\mathcal{M}$ be a non-empty compact convex decomposable set of linear previsions on $\mathcal{K}$, and let $\underline{P}$ be a coherent lower prevision on $\mathcal{K}$: it suffices to show that $\underline{\mathbf{E}}^{\mathcal{K}}_{\mathcal{M}}$ is a coherent lower prevision on $\mathcal{K}$, that $\mathbf{M}^{\mathcal{K}}_{\underline{P}}$ is a compact convex decomposable set of linear previsions on $\mathcal{K}$, and that

$$\mathbf{M}^{\mathcal{K}}_{\underline{\mathbf{E}}^{\mathcal{K}}_{\mathcal{M}}} = \mathcal{M} \quad \text{and} \quad \underline{\mathbf{E}}^{\mathcal{K}}_{\mathbf{M}^{\mathcal{K}}_{\underline{P}}} = \underline{P}.$$

Indeed, $\underline{\mathbf{E}}^{\mathcal{K}}_{\mathcal{M}}(\bullet) = \inf\{Q(\bullet)\colon Q \in \mathcal{M}\}$ is a coherent lower prevision on $\mathcal{K}$ by Lemma 3.11 on p. 57: it is the lower envelope of coherent previsions.

Clearly, $\mathbf{M}^{\mathcal{K}}_{\underline{P}}$ is convex: for any Q and R in $\mathbf{M}^{\mathcal{K}}_{\underline{P}}$ and any $\lambda \in [0,1]$, it holds that $S := \lambda Q + (1-\lambda)R$ is a linear prevision on $\mathcal{K}$ (Lemma 3.10(iii)), and since $\lambda Q + (1-\lambda)R \geq \lambda \underline{P} + (1-\lambda)\underline{P} = \underline{P}$, it follows that S is a linear behavioural extension of $\underline{P}$ to $\mathcal{K}$, so S belongs to $\mathbf{M}^{\mathcal{K}}_{\underline{P}}$.

To show that $\mathbf{M}^{\mathcal{K}}_{\underline{P}}$ is compact with respect to the topology of point-wise convergence, it suffices by Lemma 4.82 to prove that $\mathbf{M}^{\operatorname{span}(\mathcal{K})}_{\underline{P}}$ is weak-* compact. Equivalently, we must show that every real-valued continuous map on $\mathbf{M}^{\operatorname{span}(\mathcal{K})}_{\underline{P}}$ achieves a minimum and a maximum; see Schechter [70, 17.26(B)&17.30(c)] (and recall that $\mathbf{M}^{\operatorname{span}(\mathcal{K})}_{\underline{P}}$ is paracompact). Let Λ be a real-valued continuous map on $\mathbf{M}^{\operatorname{span}(\mathcal{K})}_{\underline{P}}$. By Corollary 4.83, there is a gamble $f \in \operatorname{span}(\mathcal{K})$ such that that $\Lambda = f^*$. But by Theorem 4.81(iii), f^* achieves a minimum and a maximum, namely $\underline{\mathbf{E}}^{\operatorname{span}(\mathcal{K})}_{\underline{P}}(f)$ and $\overline{\mathbf{E}}^{\operatorname{span}(\mathcal{K})}_{\underline{P}}(f)$, on $\mathbf{M}^{\operatorname{span}(\mathcal{K})}_{\underline{P}}$. Hence, $\mathbf{M}^{\operatorname{span}(\mathcal{K})}_{\underline{P}}$ is weak-* compact, or equivalently, $\mathbf{M}^{\mathcal{K}}_{\underline{P}}$ is compact with respect to the topology of point-wise convergence.

Also, $\mathbf{M}^{\mathcal{K}}_{\underline{P}}$ is decomposable: let $\lambda_1, \ldots, \lambda_n$ in $\mathbb{R}$, and $f_1, \ldots, f_n$ in $\mathcal{K}$. Apply Theorem 4.81(ii) and Lemma 4.82(i) to see that

$$\begin{aligned}\underline{\mathbf{E}}_{\underline{P}}\left(\sum_{i=1}^{n}\lambda_i f_i\right) &= \min\left\{Q'\left(\sum_{i=1}^{n}\lambda_i f_i\right)\colon Q' \in \mathbf{M}^{\operatorname{span}(\mathcal{K})}_{\underline{P}}\right\}\\ &= \min\left\{\sum_{i=1}^{n}\lambda_i Q(f_i)\colon Q \in \mathbf{M}^{\mathcal{K}}_{\underline{P}}\right\}.\end{aligned}$$

Now apply the expression for natural extension, Eq. (4.1) on p. 96, and use the fact that $\underline{P}(f) = \underline{\mathbf{E}}_{\underline{P}}(f)$ for any $f \in \operatorname{dom}\underline{P}$ (which holds since $\underline{P}$ is coherent): for every $\epsilon > 0$, there are μ^ϵ in $\mathbb{R}$, n^ϵ in $\mathbb{N}$, non-negative $\lambda^\epsilon_1, \ldots, \lambda^\epsilon_{n^\epsilon}$ in $\mathbb{R}$, and gambles $f^\epsilon_1, \ldots, f^\epsilon_{n^\epsilon}$ in $\mathcal{K}$, such that $\mu^\epsilon + \sum_{i=1}^{n^\epsilon}\lambda^\epsilon_i f^\epsilon_i \leq \sum_{i=1}^{n}\lambda_i f_i$ and

$$\underline{\mathbf{E}}_{\underline{P}}\left(\sum_{i=1}^{n}\lambda_i f_i\right) \leq \epsilon + \mu^\epsilon + \sum_{i=1}^{n^\epsilon}\lambda^\epsilon_i \underline{\mathbf{E}}_{\underline{P}}(f^\epsilon_i).$$

This establishes the decomposability of $\mathbf{M}^{\mathcal{K}}_{\underline{P}}$.

To see that $\mathbf{M}^{\mathcal{K}}_{\underline{\mathbf{E}}^{\mathcal{K}}_{\mathcal{M}}} \supseteq \mathcal{M}$, observe that any linear prevision in $\mathcal{M}$ is a behavioural extension of $\underline{\mathbf{E}}^{\mathcal{K}}_{\mathcal{M}}$.

Conversely, to establish that $\mathbf{M}^{\mathcal{K}}_{\underline{\mathbf{E}}^{\mathcal{K}}_{\mathcal{M}}} \subseteq \mathcal{M}$, we show that if $Q \in \mathcal{P}^{\mathcal{K}}(X)$ but $Q \notin \mathcal{M}$, then Q is not a behavioural extension of $\underline{\mathbf{E}}^{\mathcal{K}}_{\mathcal{M}}$. Assume that $Q \notin \mathcal{M}$.

Denote by Q' the linear extension of Q to span($\mathcal{K}$), and by $\mathcal{M}'$ the set of linear extensions of elements of $\mathcal{M}$ to span($\mathcal{K}$): clearly, $\mathcal{M}'$ is weak-* compact, and $Q' \notin \mathcal{M}'$, by Lemma 4.82. So, $\{Q'\}$ and $\mathcal{M}'$ are disjoint non-empty weak-* compact convex subsets of the set $\mathcal{V}^{\text{span}(\mathcal{K})}(X)$ of continuous real-valued linear maps on span($\mathcal{K}$), if we endow span($\mathcal{K}$) with the topology of uniform convergence; note that $\mathcal{P}^{\text{span}(\mathcal{K})}(X)$ is a weak-* compact subset of $\mathcal{V}^{\text{span}(\mathcal{K})}(X)$. Hence, by the Hahn-Banach theorem (see Schechter [70, 28.4(HB19)]) there is a weak-* continuous linear functional Λ defined on $\mathcal{V}^{\text{span}(\mathcal{K})}(X)$ such that $\Lambda(Q') < \min\{\Lambda(R') \colon R' \in \mathcal{M}'\}$. Since Λ is weak-* continuous, it must be an evaluation map on $\mathcal{V}^{\text{span}(\mathcal{K})}(X)$ (see Schechter [70, 28.15(c)]), *i.e.*, there are $n \in \mathbb{N}$, real numbers $\lambda_1, \dots, \lambda_n$, and gambles $f_1, \dots, f_n$ in $\mathcal{K}$, such that

$$\begin{aligned}\sum_{i=1}^{n} \lambda_i Q(f_i) &= Q'\left(\sum_{i=1}^{n} \lambda_i f_i\right) \\ &< \min\left\{R'\left(\sum_{i=1}^{n} \lambda_i f_i\right) \colon R' \in \mathcal{M}'\right\} = \min\left\{\sum_{i=1}^{n} \lambda_i R(f_i) \colon R \in \mathcal{M}\right\}.\end{aligned}$$

By the decomposability of $\mathcal{M}$, for any $\epsilon > 0$, there are μ^ϵ_0 in $\mathbb{R}$, m^ϵ in $\mathbb{N}$, non-negative $\mu^\epsilon_1, \dots, \mu^\epsilon_{m^\epsilon}$ in $\mathbb{R}$, and $g^\epsilon_1, \dots, g^\epsilon_{m^\epsilon}$ in $\mathcal{K}$ such that

$$\begin{aligned}\mu^\epsilon_0 + \sum_{j=1}^{m^\epsilon} \mu^\epsilon_j Q(g^\epsilon_j) &\leq \sum_{i=1}^{n} \lambda_i Q(f_i) \\ &< \min\left\{\sum_{i=1}^{n} \lambda_i R(f_i) \colon R \in \mathcal{M}\right\} \leq \epsilon + \mu^\epsilon_0 + \sum_{j=1}^{m^\epsilon} \mu^\epsilon_j \min\left\{R(g^\epsilon_j) \colon R \in \mathcal{M}\right\}.\end{aligned}$$

Choosing for instance $\epsilon := \frac{1}{2}\left[\min\{\sum_{i=1}^{n} \lambda_i R(f_i) \colon R \in \mathcal{M}\} - \sum_{i=1}^{n} \lambda_i Q(f_i)\right] > 0$,[2]

$$\mu^\epsilon_0 + \sum_{j=1}^{m^\epsilon} \mu^\epsilon_j Q(g^\epsilon_j) < \mu^\epsilon_0 + \sum_{j=1}^{m^\epsilon} \mu^\epsilon_j \min\left\{R(g^\epsilon_j) \colon R \in \mathcal{M}\right\}.$$

[2]Hint: if $a \leq b < c \leq d$, then $0 < \frac{1}{2}[c - b] < c - b \leq d - a \dots$ so $a < d - \frac{1}{2}[c - b] \dots$

This strict inequality can only hold if there is some $j \in \{1, \ldots, m^\epsilon\}$ such that

$$Q(g_j^\epsilon) < \min\left\{R(g_j^\epsilon) \colon R \in \mathcal{M}\right\} = \underline{\mathbf{E}}_{\mathcal{M}}^{\mathcal{K}}(g_j^\epsilon).$$

Thus, Q is not a behavioural extension of $\underline{\mathbf{E}}_{\mathcal{M}}^{\mathcal{K}}$. □

In other words, the set of non-empty compact convex decomposable sets of linear previsions on $\mathcal{K}$ equipped with $\underline{\mathbf{E}}_{\bullet}^{\mathcal{K}}$ is isomorphic to the set of coherent lower previsions on $\mathcal{K}$ equipped with the identity map.

4.4.2 Consequences of Duality

If the domain $\mathcal{K}$ consists of a set of measurable gambles with respect to a field, we can invoke the isomorphism established in Theorem 4.30 on p. 114. Note that we endow the set of probability charges on $\mathcal{F}$ with the topology of point-wise convergence: it is immediate that this space is homeomorphic to the set of linear previsions on $I_{\mathcal{F}} \cup -I_{\mathcal{F}}$ endowed with the topology of point-wise convergence; the homeomorphism is simply $\mathbf{P}_\bullet$.

In the corollary below, we view $\mathbf{E}_{\bullet}^{\mathcal{L}_{\mathcal{F}}(X)}$ as a mapping from sets of probability charges on a field $\mathcal{F}$ to sets of linear previsions on $\mathcal{L}_{\mathcal{F}}(X)$:

$$\mathbf{E}_m^{\mathcal{L}_{\mathcal{F}}(X)} := \left\{\mathbf{E}_\mu^{\mathcal{L}_{\mathcal{F}}(X)} \colon \mu \in m\right\},$$

where m is an arbitrary set of probability charges on $\mathcal{F}$. Similarly, we view $\boldsymbol{\mu}_\bullet^{\mathcal{F}}$ as a mapping from sets of linear previsions on $\mathcal{L}_{\mathcal{F}}(X)$ to sets of probability charges on $\mathcal{F}$:

$$\boldsymbol{\mu}_{\mathcal{M}}^{\mathcal{F}} := \left\{\boldsymbol{\mu}_Q^{\mathcal{F}} \colon Q \in \mathcal{M}\right\},$$

where $\mathcal{M}$ is an arbitrary set of linear previsions on $\mathcal{L}_{\mathcal{F}}(X)$.

Corollary 4.86. *Let $\mathcal{F}$ be a field on $\mathcal{X}$. Then $\underline{\mathbf{E}}_{\bullet}^{\mathcal{L}_{\mathcal{F}}(X)} \circ \mathbf{E}_{\bullet}^{\mathcal{L}_{\mathcal{F}}(X)}$ and $\boldsymbol{\mu}_\bullet^{\mathcal{F}} \circ \mathbf{M}_\bullet^{\mathcal{L}_{\mathcal{F}}(X)}$ establish one-to-one and onto maps between non-empty compact convex sets of probability charges on $\mathcal{F}$ and coherent lower previsions on $\mathcal{L}_{\mathcal{F}}(X)$.*

Proof. Immediate from Theorem 4.30 on p. 114 and Theorem 4.85 on p. 190. Note that, because $\mathcal{L}_{\mathcal{F}}(X)$ is a linear space, decomposability of $\mathbf{E}_m^{\mathcal{L}_{\mathcal{F}}(X)}$, where m is an arbitrary non-empty compact convex set of probability charges on $\mathcal{F}$, is immediate. □

Corollary 4.86 induces another characterisation of natural extension, for instance through the lower S-integral. If $\mathcal{F}$ is a field such that $\operatorname{dom} \underline{P}$ consists of $\mathcal{F}$-measurable gambles only, let's denote by $\mathbf{m}_{\underline{P}}^{\mathcal{F}}$ the set of charges on $\mathcal{F}$, whose linear extensions to $\mathcal{L}_{\mathcal{F}}(X)$ are behavioural extensions of $\underline{P}$:

$$\mathbf{m}_{\underline{P}}^{\mathcal{F}} := \boldsymbol{\mu}_{\bullet}^{\mathcal{F}} \circ \mathbf{M}_{\underline{P}}^{\mathcal{L}_{\mathcal{F}}(X)} = \boldsymbol{\mu}_{\mathbf{M}_{\underline{P}}^{\mathcal{L}_{\mathcal{F}}(X)}}^{\mathcal{F}}, \tag{4.55}$$

The following proposition provides equivalent expressions for $\mathbf{m}_{\underline{P}}^{\mathcal{F}}$.

Proposition 4.87. *Let $\underline{P}$ be any lower prevision, let $\mathcal{F}$ be a field on $\mathcal{X}$ and assume that* $\operatorname{dom} \underline{P} \subseteq \mathcal{L}_{\mathcal{F}}(\mathrm{X})$. *Then*

$$\mathbf{m}_{\underline{P}}^{\mathcal{F}} = \left\{ \mu \in \mathcal{P}(\mathcal{F}) \colon \forall f \in \operatorname{dom} \underline{P},\ \underline{P}(f) \leq \mathrm{S} \int f \,\mathrm{d}\mu \right\}$$

and if additionally $\underline{P}$ avoids sure loss and $\underline{\mathbf{E}}_{\underline{P}}$ is 2-monotone,

$$= \left\{ \mu \in \mathcal{P}(\mathcal{F}) \colon \forall A \in \mathcal{F},\ \underline{\mathbf{E}}_{\underline{P}}(I_A) \leq \mu(A) \right\}.$$

Proof. The first equality is a consequence of Eq. (4.55) and Theorem 4.43, by which the natural extension $\underline{\mathbf{E}}_{\mu}$ of μ coincides with the lower S-integral with respect to μ. To prove the second equality, we show that $\underline{P}(f) \leq \mathrm{S} \int f \,\mathrm{d}\mu$ for all f in $\operatorname{dom} \underline{P}$ if and only if $\underline{\mathbf{E}}_{\underline{P}}(I_A) \leq \mu(A)$ for all A in $\mathcal{F}$.

Indeed, if $\underline{P}(f) \leq \mathrm{S} \int f \,\mathrm{d}\mu$ for all f, then $\mathrm{S} \int \bullet \,\mathrm{d}\mu$ is a behavioural extension of $\underline{P}$, and so by Proposition 4.7 on p. 98, it follows that also $\underline{\mathbf{E}}_{\underline{P}}$ must be a behavioural extension of the natural extension of $\mathrm{S} \int \bullet \,\mathrm{d}\mu$, which is equal to $\mathrm{S} \int \bullet \,\mathrm{d}\mu$. So $\underline{\mathbf{E}}_{\underline{P}}(g) \leq \mathrm{S} \int g \,\mathrm{d}\mu$ for all gambles g, and in particular, for all indicator gambles I_A, $A \in \mathcal{F}$.

Conversely, suppose that $\underline{P}$ avoids sure loss, $\underline{\mathbf{E}}_{\underline{P}}$ is 2-monotone, and $\underline{\mathbf{E}}_{\underline{P}}(I_A) \leq \mu(A)$ for all A in $\mathcal{F}$, and let $f \in \operatorname{dom} \underline{P}$. Since f is $\mathcal{F}$-measurable, there is a sequence f_n of $\mathcal{F}$-simple gambles converging uniformly to f (see Definition 4.25 on p. 109). Without loss of generality, we may write the f_n as $b_{0,n} + \sum_{j=1}^{m_n} b_{j,n} I_{A_{j,n}}$, with $b_{0,n}, b_{1,n}, \ldots, b_{m_n,n}$ in $\mathbb{R}$, $b_{1,n}, \ldots, b_{m_n,n}$ non-negative, and $A_{1,n}, \ldots, A_{m_n,n}$ in $\mathcal{F}$. By the coherence and 2-monotonicity of $\underline{\mathbf{E}}_{\underline{P}}$ it follows that

$$\underline{\mathbf{E}}_{\underline{P}}(f_n) = \underline{\mathbf{E}}_{\underline{P}}\left(b_{0,n} + \sum_{j=1}^{m_n} b_{j,n} I_{A_{j,n}}\right)$$
$$= b_{0,n} + \sum_{j=1}^{m_n} b_{j,n} \underline{\mathbf{E}}_{\underline{P}}\left(I_{A_{j,n}}\right) \leq b_{0,n} + \sum_{j=1}^{m_n} b_{j,n} \mu\left(I_{A_{j,n}}\right) = \mathrm{S}\int f_n \,\mathrm{d}\mu,$$

using Proposition 4.77 on p. 180 and Eq. (3.12) on p. 75. Since f_n converges uniformly to f, and $\underline{\mathbf{E}}_{\underline{P}}$ and $\mathrm{S}\int \bullet \,\mathrm{d}\mu$ are continuous with respect to the topology of uniform convergence, it follows that also

$$\underline{P}(f) \leq \underline{\mathbf{E}}_{\underline{P}}(f) = \lim_{n\to+\infty} \underline{\mathbf{E}}_{\underline{P}}(f_n) \leq \lim_{n\to+\infty} \mathrm{S}\int f_n \,\mathrm{d}\mu = \mathrm{S}\int f \,\mathrm{d}\mu,$$

which establishes the desired inequality. □

Corollary 4.88. *Let $\underline{P}$ be any lower prevision, let $\mathcal{F}$ be a field on $\mathcal{X}$ and assume that* $\operatorname{dom}\underline{P} \subseteq \mathcal{L}_{\mathcal{F}}(X)$. *Then the following propositions hold.*

(i) $\underline{P}$ avoids sure loss if and only if there is a probability charge μ on $\mathcal{F}$ such that

$$\underline{P}(f) \leq \mathrm{S}\int f \,\mathrm{d}\mu \qquad \text{for all } f \in \operatorname{dom}\underline{P}.$$

(ii) If $\underline{P}$ avoids sure loss, then

$$\underline{\mathbf{E}}_{\underline{P}}(f) = \min\left\{\mathrm{S}\underline{\int} f \,\mathrm{d}\mu \colon \mu \in \mathbf{m}_{\underline{P}}^{\mathcal{F}}\right\} \quad \text{for any gamble } f \text{ on } X\text{; hence,}$$
$$= \min\left\{\mathrm{S}\int f \,\mathrm{d}\mu \colon \mu \in \mathbf{m}_{\underline{P}}^{\mathcal{F}}\right\} \quad \text{for any } \mathcal{F}\text{-measurable gamble } f.$$

(iii) $\underline{P}$ is coherent if and only if

$$\underline{P}(f) = \min\left\{\mathrm{S}\int f \,\mathrm{d}\mu \colon \mu \in \mathbf{m}_{\underline{P}}^{\mathcal{F}}\right\} \qquad \text{for all } f \in \operatorname{dom}\underline{P}.$$

For ample fields, we actually do not have to use the lower S-integral, the S-integral is sufficient. This generalises a result by Aeyels and De Cooman [21, Section 3, p. 182] in connection with the natural extension of possibility measures.

Corollary 4.89. *Let $\underline{P}$ be any lower prevision that avoids sure loss, let $\mathcal{F}$ be an ample field on $\mathcal{X}$, and assume that all gambles in* $\operatorname{dom}\underline{P}$ *are constant on the atoms of $\mathcal{F}$. If $\underline{P}$ avoids sure loss, then*

$$\underline{\mathbf{E}}_{\underline{P}}(f) = \min\left\{ S\int [f]^{\downarrow}_{\mathcal{F}}\,\mathrm{d}\mu \colon \mu \in \mathbf{m}^{\mathcal{F}}_{\underline{P}} \right\}$$

for any gamble f on X, where the gamble $[f]^{\downarrow}_{\mathcal{F}}$ is defined as

$$[f]^{\downarrow}_{\mathcal{F}}(x) := \underline{P}_{[x]_{\mathcal{F}}}(f) = \inf_{y\in[x]_{\mathcal{F}}} f(y)$$

for any x in $\mathcal{X}$.

Proof. By Proposition 4.27, $\operatorname{dom}\underline{P} \subseteq \mathcal{L}_{\mathcal{F}}(X)$, and hence, Corollary 4.88 applies. Now, use Theorem 4.44. □

4.4.3 A Dual Characterisation of 2-Monotonicity

The following is a straightforward generalisation of a result by Walley [85].

Corollary 4.90. *A set function ν defined on a field $\mathcal{F}$ is 2-monotone if and only if for any two sets A and B in $\mathcal{F}$ such that $A \subseteq B$, there is a probability charge μ on $\mathcal{F}$ such that $\mu(A) = \nu(A)$, $\mu(B) = \nu(B)$ and $\mu(C) \geq \nu(C)$ for all $C \in \mathcal{F}$.*

Proof. See Walley [85, Corollary 6.4] in case $\mathcal{F} = \wp(\mathcal{X})$. For the sake of completeness, let's give the proof for the general case.

"if". Let D and E be sets in $\mathcal{F}$. Then, identifying $D \cap E$ with A and $D \cup E$ with B, we find that there is a probability charge μ such that

$$\nu(D \cup E) + \nu(D \cap E) = \mu(D \cup E) + \mu(D \cap E) = \mu(D) + \mu(E) \geq \nu(D) + \nu(E).$$

So ν is 2-monotone.

"only if". Let A and B be sets in $\mathcal{F}$, and assume that $A \subseteq B$. By Theorem 3.31 on p. 76 we know that $\mathbf{P}_\nu$ is coherent. By Corollary 4.88, there is a probability charge on $\mathcal{F}$ such that $\mu(C) \geq \nu(C)$ for all $C \in \mathcal{F}$, and $\mathbf{E}_\mu(I_A + I_B) = \underline{\mathbf{E}}_\nu(I_A + I_B)$. But, by Theorem 4.76 and Eq. (3.12), it follows that $\underline{\mathbf{E}}_\nu(I_A + I_B) = \nu(A) + \nu(B)$. Also, $\mathbf{E}_\mu(I_A + I_B) = \mu(A) + \mu(B)$. Since $\mu(A) \geq \nu(A)$ and $\mu(B) \geq \nu(B)$, the equality $\mu(A) + \mu(B) = \nu(A) + \nu(B)$ can only hold if $\mu(A) = \nu(A)$ and $\mu(B) = \nu(B)$. □

Chapter 5

Cauchy Extension of Lower Previsions

As we discussed in Chapter 4, any lower prevision on X that avoids sure loss can be extended to a coherent lower prevision on the set of all gambles on X, *i.e.*, to the set of all bounded real-valued random variables that are a function of X. In this chapter, we further extend a lower prevision, to a set containing also unbounded real-valued random variables. We draw inspiration from the Dunford integral (see Section 4.3.8 on p. 161), which is defined not only for gambles, but also for unbounded real-valued random variables satisfying the conditions of Definition 4.60.

5.1 Random Quantities and Extended Lower Previsions

A *random quantity* f on a random variable X is a real-valued gain, expressed in a fixed linear utility scale, that is a function of X; it is an $\mathcal{X}$–$\mathbb{R}$-mapping, interpreted as an uncertain gain: if x turns out to be the realisation of X, then we receive an amount $f(x)$ of utility. Considering random quantities on different random variables, we may write $f(X)$ in order to emphasise that f is a random quantity on X.

The set of all random quantities on X is denoted by $\mathcal{R}(X)$. The set $\mathcal{L}(X)$ of gambles on X is a subset of $\mathcal{R}(X)$, and also $\mathcal{R}(X)$ is a linear lattice—an ordered

linear space such that every two vectors have a supremum and an infimum—with respect to the point-wise addition, the point-wise scalar multiplication and the point-wise ordering; the operations, $f + g$, λf, $f \leq g$, $f \vee g$, and $f \wedge g$, for random quantities f and g and real numbers λ, can be generalised from $\mathcal{L}(X)$—defined in Section 3.3.1—to $\mathcal{R}(X)$ by verbatim translation. The definitions of $\sup f$ and $\inf f$ generalise as follows:

$$\sup f := \min\{a \in \mathbb{R}^* : a \geq f\}, \text{ and}$$
$$\inf f := \max\{a \in \mathbb{R}^* : a \leq f\},$$

where $\mathbb{R}^*$ is the set of *extended real numbers*, that is, $\mathbb{R} \cup \{-\infty, +\infty\}$.

The *extended lower prevision* $\underline{P}(f)$ of a random quantity f is defined as the supremum buying price for f; $\underline{P}(f)$ is the highest extended real number $s \in \mathbb{R}^*$ such that for any real price $t \in \mathbb{R}$ that is strictly lower than s, we are willing to pay t prior to observation of X, if we are guaranteed to receive $f(x)$ when observing $X = x$. Mathematically, we define an extended lower prevision on X as a real-valued mapping defined on some subset $\operatorname{dom} \underline{P}$, the domain of $\underline{P}$, of the set $\mathcal{R}(X)$ of random quantities. Indeed, we do not require an extended lower prevision to be defined on the set of all random quantities. Troffaes and De Cooman [79, Proposition 3(iii)] have shown that, by generalising the notions of avoiding sure loss and coherence in a straightforward way, any extended lower prevision that avoids sure loss has a least committal coherent behavioural extension to the set of all random quantities; we shall briefly summarise these results below in Section 5.2.

Extended lower previsions differ from lower previsions, defined in Section 3.3.1, in two ways: they are defined on a larger set—the set of random quantities—and they take values in a larger set—the set of extended real numbers. If $\underline{P}(f) = -\infty$, then this means that we are not willing to buy f at any price $t \in \mathbb{R}$; this can be reasonable if f is unbounded from below. If $\underline{P}(f) = +\infty$, then this means that we are willing to buy f at any price; whether this is reasonable or not is arguable, but, for instance, the Saint Petersburg paradox, introduced by Jakob Bernoulli [7], provides an instance of a random quantity, unbounded from above, that is possibly worth to be bought at any price—although this does not go without any controversy.

Similar to gambles, we can also interpret random quantities f as an uncertain loss: if x turns out to be the true value of X, we lose an amount $f(x)$

of linear utility. The *extended upper prevision* $\overline{P}(f)$ of the gamble f is then the infimum selling price for f; it is the lowest extended real number s, such that for any real price $t \in \mathbb{R}$ strictly larger than s, we are willing to receive t prior to observation of X, if we are guaranteed to lose $f(x)$ when observing $X = x$. Since a gain r is equivalent to a loss $-r$ it should hold that $\overline{P}(f) = -\underline{P}(-f)$: from any extended lower prevision $\underline{P}$ we can infer a so-called *conjugate extended upper prevision* $\overline{P}$ on $\operatorname{dom}\overline{P} = -\operatorname{dom}\underline{P}$ which represents the same behavioural dispositions. We can therefore restrict our attention to the study of extended lower previsions only, without loss of generality. Also, if we use the notation $\underline{P}$ for an extended lower prevision, $\overline{P}$ always denotes its conjugate.

It may happen that $\underline{P}$ is *self-conjugate*, that is, $\operatorname{dom}\underline{P} = \operatorname{dom}\overline{P}$ and $\underline{P}(f) = \overline{P}(f)$ for all random quantities $f \in \operatorname{dom}\underline{P}$. In such a case, we may simply write P instead of $\underline{P}$ or $\overline{P}$ whenever it is clear from the context whether we are considering either buying or selling prices (or both). We call a self-conjugate extended lower prevision P simply an *extended prevision*, and $P(f)$ represents a so-called *fair price* for the random quantity f: we are willing to buy f for any price $t < P(f)$, and we are willing to sell f for any price $t > P(f)$. Extended previsions, interpreted as fair prices, were considered by Crisma, Gigante and Millossovich [13, 12], as an extension of the work of De Finetti [26].

5.2 Inference Revisited

For the sake of completeness, let's briefly summarise the main results of Troffaes and De Cooman [79] on avoiding sure loss, coherence, and natural extension of extended lower previsions, discuss the main problem that arises, and propose a solution, which is also due to Troffaes and De Cooman [78]. This section consists mainly of an approximately verbatim translation of results by Walley [86], and observing that, despite all technical difficulties, many—but not all—of the results for lower previsions carry over to extended lower previsions. Perhaps, the reader may wish to skip to Section 5.3 and refer back to Section 5.2 only when needed; we only rely on it in motivating the need for a behavioural extension that is different from the natural extension described in Section 5.2.3, and in proving that the Dunford-type extension described further on, really is a *coherent* behavioural extension—in the following we characterise coherence for extended lower previsions: it turns out that, when extending a lower prevision to a real-valued extended

lower prevision defined on a linear space, the conditions for coherence are a verbatim generalisation of Theorem 3.6 on p. 56, however, unfortunately, the proof of this, at first sight, simple result, only follows after a rather technical analysis.

We shall accept the following axioms of rationality governing dispositions towards buying and selling random quantities; this is a literal generalisation of the rules described in Section 3.3.3.

Axiom 5.1 (Axioms of Rationality for Random Quantities)**.** For arbitrary random quantities f and g on X and arbitrary real numbers s and t the following should hold.

(i) We are disposed to buy f for any price strictly less than $\inf f$ (accepting a sure gain).

(ii) We are disposed not to buy f for any price strictly larger than $\sup f$ (avoiding a sure loss).

(iii) If we are disposed to buy f for s then we should be disposed to buy λf for λs, for any strictly positive $\lambda \in \mathbb{R}$ (scale independence).

(iv) If we are disposed to buy f for s and g for t then we should be disposed to buy $f + g$ for $s + t$ (accepting combined transactions).

(v) If we are disposed to buy f for s and $g \geq f$ then we should be disposed to buy g for s (monotonicity).

5.2.1 Avoiding Sure Loss Revisited

Definition 5.2. An extended lower prevision $\underline{P}$ on X is said to *avoid sure loss* if for every $n \in \mathbb{N}$, non-negative $\lambda_1, \ldots, \lambda_n$ in $\mathbb{R}$, and random quantities $f_1, \ldots, f_n$ in $\operatorname{dom} \underline{P}$ such that $\sum_{i=1}^{n} \lambda_i \underline{P}(f_i)$ is well-defined, we have that

$$\sup \left[\sum_{i=1}^{n} \lambda_i f_i \right] \geq \sum_{i=1}^{n} \lambda_i \underline{P}(f_i). \tag{5.1}$$

Explanation. Suppose that Eq. (5.1) fails for some $n \in \mathbb{N}$, non-negative $\lambda_1, \ldots, \lambda_n$ in $\mathbb{R}$, and random quantities $f_1, \ldots, f_n$ in $\operatorname{dom} \underline{P}$ such that $\sum_{i=1}^{n} \lambda_i \underline{P}(f_i)$

is well-defined. Then,

$$\alpha := \sum_{i=1}^{n} \lambda_i \underline{P}(f_i) > \sup\left[\sum_{i=1}^{n} \lambda_i f_i\right] =: \gamma.$$

This can only hold if $\alpha \neq -\infty$ and $\gamma \in \mathbb{R}$. It implies that we may choose a $\beta \in \mathbb{R}$ such that $\alpha > \beta > \gamma \geq \sum_{i=1}^{n} \lambda_i f_i$, which means that if we buy $\sum_{i=1}^{n} \lambda_i f_i$ at a price β, we incur a sure loss of at least $\beta - \gamma$. But, by Axiom 5.1, we are disposed to buy $\sum_{i=1}^{n} \lambda_i f_i$ at this price β, since $\alpha = \sum_{i=1}^{n} \lambda_i \underline{P}(f_i) > \beta$ (recall that $\alpha = -\infty$ does not occur). □

Compare with Definition 3.2 on p. 50: contrary to lower previsions, we cannot restrict the coëfficients $\lambda_1, \ldots, \lambda_n$ to integer values in order to characterise avoiding sure loss of extended lower previsions; roughly said, Definition 3.2(A)&(B) are not equivalent when generalised to extended lower previsions, so we need to take the strongest condition. Of course, Definition 5.2 generalises Definition 3.2: a lower prevision avoids sure loss according to Definition 3.2 if and only if it avoids sure loss according to Definition 5.2.

Why does it not suffice to consider only integer combinations? Let $\mathcal{X} = \mathbb{R}$, define the random quantity f by $f(x) := x$ for all $x \in \mathcal{X}$, and consider the extended lower prevision $\underline{P}$, with domain $\{f, -\sqrt{2}f\}$, defined by $\underline{P}(f) = 1$ and $\underline{P}(-\sqrt{2}f) = 2$. Since $n - m\sqrt{2} \neq 0$ for every n and m in $\mathbb{N}$ not both zero, we find that $\sup\left[nf - m\sqrt{2}f\right] = +\infty$ for every n and m in $\mathbb{N}$ not both zero. Consequently, the inequality

$$\sup\left[nf - m\sqrt{2}f\right] \geq n + 2m,$$

holds for every n and m in $\mathbb{N}$ not both zero, and if n and m are both zero, then the inequality also holds: we constructed an extended lower prevision $\underline{P}$ such that

$$\sup\left[\sum_{i=1}^{n} f_i\right] \geq \sum_{i=1}^{n} \underline{P}(f_i)$$

for every $n \in \mathbb{N}$, and random quantities $f_1, \ldots, f_n$ in $\operatorname{dom} \underline{P}$ such that $\sum_{i=1}^{n} \underline{P}(f_i)$ is well-defined such that $\sum_{i=1}^{n} \underline{P}(f_i)$ is well-defined; but $\underline{P}$ does not avoid sure loss: for $\lambda_1 = \sqrt{2}$, $f_1 = f$, $\lambda_2 = 1$, and $f_2 = -\sqrt{2}f$, it holds that

$$\sup\left[\lambda_1 f_1 + \lambda_2 f_2\right] = 0 < \lambda_1 + 2\lambda_2 = \lambda_1 \underline{P}(f_1) + \lambda_2 \underline{P}(f_2).$$

5.2.2 Coherence Revisited

Definition 5.3. An extended lower prevision $\underline{P}$ is called *coherent* if for every $n \in \mathbb{N}$, non-negative $\lambda_0, \lambda_1, \ldots, \lambda_n$ in $\mathbb{R}$, and random quantities $f_0, f_1, \ldots, f_n$ in $\operatorname{dom} \underline{P}$ such that $\sum_{i=1}^{n} \lambda_i \underline{P}(f_i) - \lambda_0 \underline{P}(f_0)$ is well-defined, we have that

$$\sup\left[\sum_{i=1}^{n} \lambda_i f_i - \lambda_0 f_0\right] \geq \sum_{i=1}^{n} \lambda_i \underline{P}(f_i) - \lambda_0 \underline{P}(f_0). \tag{5.2}$$

Explanation. Suppose that Eq. (5.2) fails for some $n \in \mathbb{N}$, non-negative λ_0, $\lambda_1, \ldots, \lambda_n$ in $\mathbb{R}$, and random quantities $f_0, f_1, \ldots, f_n$ in $\operatorname{dom} \underline{P}$ such that $\sum_{i=1}^{n} \lambda_i \underline{P}(f_i) - \lambda_0 \underline{P}(f_0)$ is well-defined. Then,

$$\sum_{i=1}^{n} \lambda_i \underline{P}(f_i) - \lambda_0 \underline{P}(f_0) > \sup\left[\sum_{i=1}^{n} \lambda_i f_i - \lambda_0 f_0\right].$$

The case $\lambda_0 = 0$ was explained in Section 5.2.1; assume that $\lambda_0 \neq 0$. Then,

$$\alpha := \sum_{i=1}^{n} \frac{\lambda_i}{\lambda_0} \underline{P}(f_i) - \underline{P}(f_0) > \sup\left[\sum_{i=1}^{n} \frac{\lambda_i}{\lambda_0} f_i - f_0\right] =: \beta.$$

This can only hold if $\beta \in \mathbb{R}$ and $\alpha \neq -\infty$, and therefore $\underline{P}(f_0) < +\infty$, and $\alpha_1 := \sum_{i=1}^{n} \frac{\lambda_i}{\lambda_0} \underline{P}(f_i)$ is well defined. So, $\alpha_1 - \beta > \underline{P}(f_0)$ and $\alpha_1 > -\infty$. Observe that, by the definition of β,

$$\sum_{i=1}^{n} \frac{\lambda_i}{\lambda_0} f_i - \beta \leq f_0. \tag{5.3}$$

Since $\alpha = \alpha_1 - \underline{P}(f_0)$ is well defined and $\alpha \neq -\infty$, it suffices to consider the following cases, since we know that $\underline{P}(f_0) < +\infty$:

(a) $\underline{P}(f_0) \in \mathbb{R}$. We are disposed to buy the right hand side of Eq. (5.3) for any price strictly less than $\underline{P}(f_0)$. But, we may also infer from the other assessments $\underline{P}(f_1), \ldots, \underline{P}(f_n)$ and Axiom 5.1, that we are disposed to buy the left hand side of Eq. (5.3) for any price strictly less than $\alpha_1 - \beta$. Consequently, by Axiom 5.1(v), we are disposed to buy the right hand side of Eq. (5.3), f_0, for any price strictly less than $\alpha_1 - \beta$. But this price $\alpha_1 - \beta$ is strictly larger than the supremum price $\underline{P}(f_0)$ for f_0, which points to an inconsistency in the assessments.

(b) $\underline{P}(f_0) = -\infty$. Since there is no price we are disposed to buy the right hand side of Eq. (5.3) for, there is also no price we are disposed to buy the left hand side of Eq. (5.3) for, by Axiom 5.1(v). But, since $\alpha_1 > -\infty$, we are disposed to buy $\sum_{i=1}^{n} \frac{\lambda_i}{\lambda_0} f_i$ for any price strictly smaller than α_1, which again points to a contradiction in the assessments.

□

Compare to Definition 3.3 on p. 52: again, contrary to lower previsions, we cannot restrict the coëfficients $\lambda_0, \lambda_1, \ldots, \lambda_n$ to integer values in order to characterise coherence of extended lower previsions: Definition 3.3(A)&(B) are not equivalent when generalised to extended lower previsions, so we take the strongest condition. Of course, Definition 5.3 generalises Definition 3.3: a lower prevision is coherent according to Definition 3.3 if and only if it is coherent according to Definition 5.3. We even have a slightly stronger result:

Corollary 5.4. *An extended lower prevision $\underline{P}$, defined on a subset of gambles on X, is coherent, if and only if $\underline{P}$ is a coherent lower prevision.*

Proof. Immediate, if we can show that $\underline{P}$ is real-valued. This follows from Theorem 5.5(i) below. □

By the way, Corollary 5.4 has no equivalent for avoiding sure loss: for instance, the extended lower prevision $\underline{P}$ defined on $\{0\}$ by $\underline{P}(0) := -\infty$ clearly avoids sure loss, but $\underline{P}$ is not a lower prevision.

The following theorem summarises the most important properties of coherence; it is a straightforward generalisation of Theorem 2.6.1 of Walley [86]. In the proof, we make extensive use of Appendix A.

Theorem 5.5. *Let $\underline{P}$ be a coherent extended lower prevision on X. Let f and g be random quantities on X, let f_α be a net of random quantities on X, let a be a constant random quantity on X, and let λ be a non-negative real number. Then the following statements hold whenever every term and every operation is well defined.*

(i) $\inf[f] \leq \underline{P}(f) \leq \overline{P}(f) \leq \sup[f]$

(ii) $\underline{P}(a) = \overline{P}(a) = a$

(iii) $\underline{P}(f + a) = \underline{P}(f) + a, \qquad \overline{P}(f + a) = \overline{P}(f) + a$

(iv) $f \leq g + a \implies \underline{P}(f) \leq \underline{P}(g) + a$ *and* $\overline{P}(f) \leq \overline{P}(g) + a$

(v) $\underline{P}(f) + \underline{P}(g) \leq \underline{P}(f+g) \leq \underline{P}(f) + \overline{P}(g) \leq \overline{P}(f+g) \leq \overline{P}(f) + \overline{P}(g)$

(vi) $\underline{P}(\lambda f) = \lambda \underline{P}(f), \qquad \overline{P}(\lambda f) = \lambda \overline{P}(f)$

(vii) $\underline{P}(|f|) \geq \underline{P}(f), \qquad \overline{P}(|f|) \geq \overline{P}(f)$

(viii) $\left|\underline{P}(f) - \underline{P}(g)\right| \leq \overline{P}(|f-g|), \qquad \left|\overline{P}(f) - \overline{P}(g)\right| \leq \overline{P}(|f-g|)$

(ix) $\underline{P}(|f+g|) \leq \underline{P}(|f|) + \overline{P}(|g|), \qquad \overline{P}(|f+g|) \leq \overline{P}(|f|) + \overline{P}(|g|)$

(x) $\underline{P}(f \vee g) + \underline{P}(f \wedge g) \leq \underline{P}(f) + \overline{P}(g) \leq \overline{P}(f \vee g) + \overline{P}(f \wedge g)$,
$\underline{P}(f) + \underline{P}(g) \leq \underline{P}(f \vee g) + \overline{P}(f \wedge g) \leq \overline{P}(f) + \overline{P}(g)$ *and*
$\underline{P}(f) + \underline{P}(g) \leq \overline{P}(f \vee g) + \underline{P}(f \wedge g) \leq \overline{P}(f) + \overline{P}(g)$.

(xi) $\overline{P}(|f_\alpha - f|) \to 0 \implies \underline{P}(f_\alpha) \to \underline{P}(f)$ *and* $\overline{P}(f_\alpha) \to \overline{P}(f)$

Proof. (i). Take $n = 0$, $\lambda_0 = 1$ and $f_0 = f$ in Eq. (5.2). We find that $-\inf[f] = \sup[-f] \geq -\underline{P}(f)$. Take $n = 0$, $\lambda_0 = 1$ and $f_0 = -f$ in Eq. (5.2). We find that $\sup[f] \geq -\underline{P}(-f) = \overline{P}(f)$. Take $n = 2$, $\lambda_1 = \lambda_2 = 1$, $\lambda_0 = 0$ and $f_1 = -f_2 = f$ in Eq. (5.2). We find that $0 \geq \underline{P}(f) + \underline{P}(-f) = \underline{P}(f) - \overline{P}(f)$ whenever the right hand side is well defined. By Lemma A.9(iv) we find that $\underline{P}(f) \leq \overline{P}(f)$.

(ii). This follows from (i).

(iii). Take $n = 1$, $\lambda_1 = \lambda_0 = 1$, $f_1 = f$ and $f_0 = f + \mu$ in Eq. (5.2). We find that $-\mu \geq \underline{P}(f) - \underline{P}(f+\mu)$ whenever the right hand side is well defined. Take $n = 1$, $\lambda_1 = \lambda_0 = 1$, $f_1 = f + \mu$ and $f_0 = f$ in Eq. (5.2). We find that $\mu \geq \underline{P}(f+\mu) - \underline{P}(f)$ whenever the right hand side is well defined. Therefore $\mu = \underline{P}(f+\mu) - \underline{P}(f)$ whenever the right hand side is well defined, whence $\underline{P}(f+\mu) = \underline{P}(f) + \mu$ by Lemma A.9(iv).

(iv). By Eq. (5.2) we find that $\mu \geq \sup[f-g] \geq \underline{P}(f) - \underline{P}(g)$ whenever the right hand side is well defined. Using Lemma A.9(iv) we find that $\underline{P}(f) \leq \underline{P}(g) + \mu$.

(v). We prove the first inequality. Take $n = 2$, $\lambda_1 = \lambda_2 = \lambda_0 = 1$, $f_1 = f$, $f_2 = g$ and $f_0 = f + g$. By Eq. (5.2) we find that $0 \geq \underline{P}(f) + \underline{P}(g) - \underline{P}(f+g)$ whenever the right hand side is well defined, and consequently, by Lemma A.9(iv), $\underline{P}(f) + \underline{P}(g) \leq \underline{P}(f+g)$ whenever the left hand side is well defined.

Next, we prove the second inequality. Take $n = 2$, $\lambda_1 = \lambda_2 = \lambda_0 = 1$, $f_1 = f + g$, $f_2 = -g$ and $f_0 = f$. By Eq. (5.2) we find that $0 \geq \underline{P}(f+g) + \underline{P}(-g) - \underline{P}(f) = \underline{P}(f+g) - \overline{P}(g) - \underline{P}(f)$ whenever the right hand side is well defined. Now use Lemma A.9(iv).

Next, we prove the third inequality. Take $n = 2$, $\lambda_1 = \lambda_2 = \lambda_0 = 1$, $f_1 = -f - g$, $f_2 = f$ and $f_0 = -g$. By Eq. (5.2) we find that $0 \geq \underline{P}(-f - g) + \underline{P}(f) - \underline{P}(-g) = -\overline{P}(f + g) + \underline{P}(f) + \overline{P}(g)$ whenever the right hand side is well defined. Apply Lemma A.9(iv).

Finally, we prove the fourth inequality. Take $n = 2$, $\lambda_1 = \lambda_2 = \lambda_0 = 1$, $f_1 = -f$, $f_2 = -g$ and $f_0 = -f - g$. By Eq. (5.2) we find that $0 \geq \underline{P}(-f) + \underline{P}(-g) - \underline{P}(-f - g) = -\overline{P}(f) - \overline{P}(g) + \overline{P}(f + g)$ whenever the right hand side is well defined. Invoke Lemma A.9(iv).

(vi). Take $n = 1$, $f_0 = f$, $f_1 = \lambda f$, $\lambda_0 = \lambda$ and $\lambda_1 = 1$ in Eq. (5.2). We find that $0 \geq \underline{P}(\lambda f) - \lambda \underline{P}(f)$ whenever the right hand side is well defined. Next take $n = 1$, $f_0 = \lambda f$, $f_1 = f$, $\lambda_0 = 1$ and $\lambda_1 = \lambda$ in Eq. (5.2). We find that $0 \geq \lambda \underline{P}(f) - \underline{P}(\lambda f)$ whenever the right hand side is well defined. By Lemma A.9(iv), $\underline{P}(\lambda f) = \lambda \underline{P}(f)$.

(vii). This follows from $f \leq |f|$ and (iv).

(viii). From (v) and (iv) it follows that whenever $\underline{P}(f) - \underline{P}(g)$ is well defined, we have that

$$\underline{P}(f) - \underline{P}(g) = \underline{P}(f) + \overline{P}(-g) \leq \overline{P}(f - g) \leq \overline{P}(|f - g|), \text{ and}$$
$$\underline{P}(g) - \underline{P}(f) = \underline{P}(g) + \overline{P}(-f) \leq \overline{P}(g - f) \leq \overline{P}(|f - g|).$$

We conclude that $\left|\underline{P}(f) - \underline{P}(g)\right| \leq \overline{P}(|f - g|)$ whenever $\underline{P}(f) - \underline{P}(g)$ is well defined. To prove the second inequality, notice that

$$\left|\overline{P}(f) - \overline{P}(g)\right| = \left|\underline{P}(-f) - \underline{P}(-g)\right| \leq \overline{P}(|f - g|),$$

whenever the left hand side is well defined.

(ix). This follows from $\left|f + g\right| \leq \left|f\right| + \left|g\right|$, (iv) and (v).

(x). This follows from $f \vee g + f \wedge g = f + g$ and (v).

(xi). If $\underline{P}(f)$ is real, then $\left|\underline{P}(f_\alpha) - \underline{P}(f)\right|$ is well-defined and $\underline{P}(f_\alpha) \to \underline{P}(f)$ follows from (viii).

If $\underline{P}(f) = -\infty$, then, by (viii) and Lemma A.9(iv),

$$\underline{P}(f_\alpha) \leq \underline{P}(f) + \overline{P}(|f_\alpha - f|),$$

whenever the right hand side is well-defined. But, eventually, this is the case, since, eventually, $\overline{P}(|f_\alpha - f|)$ is real: it converges to zero. It follows that, eventually, $\underline{P}(f_\alpha) = -\infty$.

If $\underline{P}(f) = +\infty$, then, by (viii) and Lemma A.9(iv),

$$\underline{P}(f) \leq \underline{P}(f_\alpha) + \overline{P}(|f_\alpha - f|),$$

whenever the right hand side is well-defined. But, eventually, this is the case, since, eventually, $\overline{P}(|f_\alpha - f|)$ is real: it converges to zero. So, eventually, $\underline{P}(f_\alpha) = +\infty$.

For $\overline{P}$ use $\left|f - f_\alpha\right| = \left|(-f) - (-f_\alpha)\right|$. □

If $\operatorname{dom}\underline{P}$ is a linear space, then we have the following simple necessary and sufficient condition for coherence; compare to Walley [86, Theorem 2.5.5, p. 75]. For extended previsions, this result was proved by Crisma, Gigante and Millossovich [12, Theorem 3.6].

Theorem 5.6. *Let $\underline{P}$ be an extended lower prevision and assume that* $\operatorname{dom}\underline{P}$ *is a linear space. Then $\underline{P}$ is coherent if and only if the following statements hold for any random quantities f and g in* $\operatorname{dom}\underline{P}$ *and any non-negative real number λ:*

(1) $\underline{P}(f) \geq \inf[f]$,

(2) $\underline{P}(\lambda f) = \lambda \underline{P}(f)$, *and*

(3) $\underline{P}(f + g) \geq \underline{P}(f) + \underline{P}(g)$, *whenever the right hand side is well defined.*

Proof. If $\underline{P}$ is coherent, then (1), (2), and (3) follow from Theorem 5.5

Conversely, suppose that $\underline{P}$ satisfies (1), (2), and (3). We prove that Eq. (5.2) holds. Take $n \in \mathbb{N}$, non-negative $\lambda_0, \lambda_1, \ldots, \lambda_n$ in $\mathbb{R}$, and random quantities $f_0, f_1, \ldots, f_n$ in $\operatorname{dom}\underline{P}$ such that $\sum_{i=1}^n \lambda_i \underline{P}(f_i) - \lambda_0 \underline{P}(f_0)$ is well-defined, and define $f = \lambda_0 f_0$, $g = \sum_{i=1}^n \lambda_i f_i$, and $h = f - g$; since $\operatorname{dom}\underline{P}$ is a linear space, f, g, and h are in $\operatorname{dom}\underline{P}$. By (2), $\underline{P}(f) = \lambda_0 \underline{P}(f_0)$, and by (2) and (3),

$$\underline{P}(g) \geq \sum_{i=1}^n \underline{P}(\lambda_i f_i) = \sum_{i=1}^n \lambda_i \underline{P}(f_i), \tag{5.4}$$

whenever the right hand side is well defined. Moreover, by (3), $\underline{P}(f) \geq \underline{P}(g) + \underline{P}(h)$ whenever the right hand side is well defined, and therefore, again by Lemma A.9(iv), also $\underline{P}(f) - \underline{P}(g) \geq \underline{P}(h)$ whenever the left hand side is well

defined. Using (1), we have that $\underline{P}(h) \geq \inf[h]$. Summarising,

$$\sup\left[\sum_{i=1}^{n} \lambda_i f_i - \lambda_0 f_0\right] = \sup[g - f] = \sup[-h] = -\inf[h]$$
$$\geq -\underline{P}(h) \geq \underline{P}(g) - \underline{P}(f)$$

whenever the right hand side is well defined, and thus by Eq. (5.4) and Lemma A.9(v),

$$\geq \sum_{i=1}^{n} \lambda_i \underline{P}(f_i) - \lambda_0 \underline{P}(f_0),$$

whenever the right hand side is well defined—which is the case by assumption. This proves that $\underline{P}$ is a coherent extended lower prevision. □

There is a subtle difference between Theorem 5.6 and its counterpart for lower previsions on gambles, Theorem 3.6 on p. 56: for Theorem 5.6 to hold, condition (2) must also hold for $\lambda = 0$. To see why, consider the extended lower prevision defined by $\underline{P}(f) = +\infty$ for all random quantities f on X: $\underline{P}$ satisfies Theorem 5.6(1)&(2)&(3) for any random quantities f and g on X and any non-negative real number λ, except for $\lambda = 0$, but, $\underline{P}$ is clearly not coherent; it does not even avoid sure loss.

Let's not forget to mention the following important result.

Lemma 5.7. *The following statements hold.*

(i) The restriction of an extended lower prevision avoiding sure loss also avoids sure loss.

(ii) The restriction of an extended coherent lower prevision is also coherent.

(iii) The restriction of an extended linear prevision to a prevision is also linear.

Proof. Immediately from Definition 5.2 and Definition 5.3. □

5.2.3 Natural Extension Revisited

Let $\underline{P}$ be any extended lower prevision, and let $\mathcal{K}$ be a set of random quantities that includes dom $\underline{P}$. As in Section 4.1 on p. 95, let's, again, carefully sum up the properties which the natural extension $\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{K}}$ of $\underline{P}$ to $\mathcal{K}$ should satisfy. First,

any behavioural disposition expressed by $\underline{P}$ should also be expressed by $\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{K}}$; it should be a behavioural extension of $\underline{P}$:

Definition 5.8. An extended lower prevision $\underline{Q}$ is called a *behavioural extension* of an extended lower prevision $\underline{P}$ if $\operatorname{dom}\underline{P} \subseteq \operatorname{dom}\underline{Q}$ and $\underline{P}(f) \leq \underline{Q}(f)$ for any random quantity $f \in \operatorname{dom}\underline{P}$.

Thus, $\operatorname{dom}\underline{P} \subseteq \mathcal{K}$ and $\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{K}}(f) \geq \underline{P}(f)$ for all f in $\operatorname{dom}\underline{P}$. Secondly, $\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{K}}$ must be coherent. And, last but not least, we want the buying prices $\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{K}}$ to be as low as possible: any coherent behavioural extension of $\underline{P}$ to $\mathcal{K}$ must also be a behavioural extension of $\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{K}}$. This can only be the case when $\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{K}}$ is the point-wise smallest coherent behavioural extension of $\underline{P}$ to $\mathcal{K}$. Summarising, defining natural extension as the least committal coherent behavioural extension, we find that natural extension for extended lower previsions is nothing but the following verbatim translation of Definition 4.2:

Definition 5.9. Let $\underline{P}$ be an extended lower prevision, and let $\operatorname{dom}\underline{P} \subseteq \mathcal{K} \subseteq \mathcal{R}(X)$. The point-wise smallest coherent behavioural extension of $\underline{P}$ to $\mathcal{K}$, if it exists, is called the *natural extension of* $\underline{P}$ *to* $\mathcal{K}$, and is denoted by $\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{K}}$.

Note that the notation $\underline{\mathbf{E}}_{\underline{P}}$ is reserved for the natural extension of a lower prevision $\underline{P}$ to the set $\mathcal{L}(X)$ of all gambles on X; see Definition 4.2. The natural extension of an extended lower prevision to the set of all random quantities will be explicitly denoted by $\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{R}(X)}$.

For now, it is still not yet clear how compatible Definition 5.9 is with Definition 4.2: is the natural extension of lower previsions to random quantities, as in Definition 5.9, an extension of the natural extension of lower previsions to gambles, as in Definition 4.2? This will follow from the following theorem, which also tells us that avoiding sure loss of $\underline{P}$ is necessary and sufficient for the existence of its natural extension $\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{K}}$. It also gives an explicit expression for $\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{K}}$, and number of criteria to check avoiding sure loss—but, note that there are some subtle but important differences from Theorem 4.3.

Theorem 5.10. *Let* $\underline{P}$ *be an extended lower prevision, and let* $\operatorname{dom}\underline{P} \subseteq \mathcal{K} \subseteq \mathcal{R}(X)$. *Define the extended lower prevision* $\underline{E}$ *on* $\mathcal{R}(X)$ *by*

$$\underline{E}(f) := \sup\left\{\alpha + \sum_{i=1}^{n} \lambda_i \underline{P}(f_i) \text{ w.d.} : \alpha \in \mathbb{R}, n \in \mathbb{N},\right.$$

$$\lambda_1, \dots, \lambda_n \geq 0,\ f_1, \dots, f_n \in \operatorname{dom} \underline{P},\ \alpha + \sum_{i=1}^{n} \lambda_i f_i \leq f \Bigg\} \quad (5.5)$$

for any random quantity $f \in \mathcal{R}(X)$. The following conditions are equivalent.

(i) $-\infty < \underline{E}(f) < +\infty$ *for some random quantity f on X.*

(ii) $-\infty < \underline{E}(f) < +\infty$ *for any gamble f on X.*

(iii) $\underline{E}$ *is a coherent extended lower prevision on $\mathcal{R}(X)$.*

(iv) The natural extension of $\underline{P}$ to $\mathcal{K}$ exists and is equal to $\underline{E}$ restricted to $\mathcal{K}$.

(v) $\underline{P}$ *has at least one coherent behavioural extension.*

(vi) $\underline{P}$ *has at least one behavioural extension that avoids sure loss.*

(vii) $\underline{P}$ *avoids sure loss.*

Proof. It suffices to establish (vii) $\Longrightarrow$ (iv) $\Longrightarrow$ (v) $\Longrightarrow$ (vi) $\Longrightarrow$ (vii) and (vii) $\Longrightarrow$ (iii) $\Longrightarrow$ (ii) $\Longrightarrow$ (i) $\Longrightarrow$ (vii).

(vii) $\Longrightarrow$ (iv). Suppose that $\underline{P}$ avoids sure loss. Before going into detail, let's sketch the proof. (a) First, we prove that $\underline{E}$ is a coherent extended lower prevision on $\mathcal{R}(X)$; it immediately follows that also its restriction to $\mathcal{K}$ is a coherent extended lower prevision. (b) We then prove that $\underline{E}(f) \geq \underline{P}(f)$ for every random quantity $f \in \operatorname{dom} \underline{P}$. This establishes that $\underline{E}$, restricted to $\mathcal{K}$, is a coherent behavioural extension of $\underline{P}$. (c) Next, we prove that, for any coherent extended lower prevision $\underline{Q}$ on $\mathcal{K}$, if $\underline{Q}(f) \geq \underline{P}(f)$ for all $f \in \operatorname{dom} \underline{P}$, then $\underline{Q}(f) \geq \underline{E}(f)$ for all $f \in \mathcal{K}$. This establishes that $\underline{E}$ is the point-wise smallest coherent behavioural extension of $\underline{P}$ to $\mathcal{K}$. Let's fill in the details now.

(a) We check that $\underline{E}$, which is an extended lower prevision defined on the linear space of all random quantities on X, satisfies the conditions of Theorem 5.6:

(1). Let f be any random quantity on X. If $\inf f$ is real, then $\underline{E}(f) \geq \inf f$ is immediate: consider $\alpha = \inf f$ and $n = 0$ in Eq. (5.5). If $\inf f = -\infty$, then the inequality $\underline{E}(f) \geq \inf f$ is immediate. The case $\inf f = +\infty$ never occurs.

(2). Let f be any random quantity on X, and let λ be any *strictly positive* real. Then the equality $\underline{E}(\lambda f) = \lambda \underline{E}(f)$ follows from Lemma A.8

(distributivity in $\mathbb{R}^*$), and the observation that, there are $\alpha \in \mathbb{R}$, $n \in \mathbb{N}$, non-negative $\lambda_1, \ldots, \lambda_n$ in $\mathbb{R}$, and random quantities $f_1, \ldots, f_n$ in $\operatorname{dom} \underline{P}$ such that $\alpha + \sum_{i=1}^n \lambda_i \underline{P}(f_i)$ is well defined and $\alpha + \sum_{i=1}^n \lambda_i f_i \leq f$, if and only if, there are $\beta \in \mathbb{R}$, $n \in \mathbb{N}$, non-negative $\kappa_1, \ldots, \kappa_n$ in $\mathbb{R}$, and random quantities $f_1, \ldots, f_n$ in $\operatorname{dom} \underline{P}$ such that $\beta + \sum_{i=1}^n \kappa_i \underline{P}(f_i)$ is well defined and $\beta + \sum_{i=1}^n \kappa_i f_i \leq \lambda f$: identify β with $\lambda\alpha$ and κ_i with $\lambda\lambda_i$.

The case $\lambda = 0$ follows if we can show that $\underline{E}(0) \leq 0$; the converse inequality already follows from (1). By Eq. (5.5), we must show that $\alpha + \sum_{i=1}^n \lambda_i \underline{P}(f_i) \leq 0$ for any $\alpha \in \mathbb{R}$, $n \in \mathbb{N}$, non-negative $\lambda_1, \ldots, \lambda_n$ in $\mathbb{R}$, and random quantities $f_1, \ldots, f_n$ in $\operatorname{dom} \underline{P}$ such that $\alpha + \sum_{i=1}^n \lambda_i \underline{P}(f_i)$ is well defined and $\alpha + \sum_{i=1}^n \lambda_i f_i \leq 0$. Indeed, since α is real, also $\sum_{i=1}^n \lambda_i \underline{P}(f_i)$ is well defined, so, since $\underline{P}$ avoids sure loss, it follows from Eq. (5.1) that

$$\sum_{i=1}^n \lambda_i \underline{P}(f_i) \leq \sup\left[\sum_{i=1}^n \lambda_i f_i\right].$$

But, since $\alpha + \sum_{i=1}^n \lambda_i f_i \leq 0$, we also find that $\sup\left[\sum_{i=1}^n \lambda_i f_i\right] \leq -\alpha$. Together with the above inequality, this implies that $\sum_{i=1}^n \lambda_i \underline{P}(f_i) \leq -\alpha$, and hence, by LemmaA.9(iv), also $\alpha + \sum_{i=1}^n \lambda_i \underline{P}(f_i) \leq 0$. This shows that $\underline{E}(0) \leq 0$, and hence, the case $\lambda = 0$ is established too.

(3). Let f and g be two random quantities on X, and suppose that $\underline{E}(f) + \underline{E}(g)$ is well defined. Then,

$$\begin{aligned}\underline{E}(f) + \underline{E}(g) = \sup\Bigg\{&\alpha + \sum_{i=1}^n \lambda_i \underline{P}(f_i) \text{ w.d.: } \alpha \in \mathbb{R},\ n \in \mathbb{N},\\ &\lambda_1, \ldots, \lambda_n \geq 0,\ f_1, \ldots, f_n \in \operatorname{dom} \underline{P},\ \alpha + \sum_{i=1}^n \lambda_i f_i \leq f\Bigg\}\\ + \sup\Bigg\{&\beta + \sum_{i=1}^m \kappa_i \underline{P}(g_i) \text{ w.d.: } \beta \in \mathbb{R},\ m \in \mathbb{N},\\ &\kappa_1, \ldots, \kappa_m \geq 0,\ g_1, \ldots, g_m \in \operatorname{dom} \underline{P},\ \beta + \sum_{i=1}^m \lambda_i g_i \leq g\Bigg\}\end{aligned}$$

and by Lemma A.9(viii),

$$= \sup\Bigg\{\alpha + \beta + \sum_{i=1}^{n} \lambda_i \underline{P}(f_i) + \sum_{i=1}^{m} \kappa_i \underline{P}(g_i) \text{ w.d.}:$$
$$\alpha, \beta \in \mathbb{R},\ n, m \in \mathbb{N},$$
$$\lambda_1, \ldots, \lambda_n, \kappa_1, \ldots, \kappa_m \geq 0,\ f_1, \ldots, f_n, g_1, \ldots, g_m \in \operatorname{dom} \underline{P},$$
$$\alpha + \sum_{i=1}^{n} \lambda_i f_i \leq f,\ \beta + \sum_{i=1}^{m} \lambda_i g_i \leq g\Bigg\}$$

and since $\alpha + \sum_{i=1}^{n} \lambda_i f_i \leq f$ and $\beta + \sum_{i=1}^{m} \lambda_i g_i \leq g$ imply that $\alpha + \beta + \sum_{i=1}^{n} \lambda_i f_i + \sum_{i=1}^{m} \lambda_i g_i \leq f + g$,

$$\leq \underline{E}(f + g).$$

We find that $\underline{E}$ is a coherent extended lower prevision on $\mathcal{R}(X)$. By Lemma 5.7, it follows that also the restriction of $\underline{E}$ to $\mathcal{K}$ is a coherent extended lower prevision.

(b) We show that $\underline{E}(f) \geq \underline{P}(f)$ for all random quantities $f \in \operatorname{dom} \underline{P}$: indeed, consider $\alpha = 0$, $n = 1$, $\lambda_1 = 1$, and $f_1 = f$ in Eq. (5.5).

(c) Let $\underline{Q}$ be a coherent extended lower prevision on $\mathcal{K}$, and assume that $\underline{Q}(f) \geq \underline{P}(f)$ for all $f \in \operatorname{dom} \underline{P}$. We must show that $\underline{Q}(f) \geq \underline{E}(f)$ for all $f \in \mathcal{K}$. Let $f \in \mathcal{K}$, then

$$\underline{E}(f) = \sup\Bigg\{\alpha + \sum_{i=1}^{n} \lambda_i \underline{P}(f_i) \text{ w.d.}: \alpha \in \mathbb{R},\ n \in \mathbb{N},$$
$$\lambda_1, \ldots, \lambda_n \geq 0,\ f_1, \ldots, f_n \in \operatorname{dom} \underline{P},\ \alpha + \sum_{i=1}^{n} \lambda_i f_i \leq f\Bigg\}$$

and applying Lemma A.9(vii),

$$= \sup\Bigg\{\alpha + \sum_{i=1}^{n} \lambda_i \underline{P}(f_i) \text{ w.d. and } > -\infty: \alpha \in \mathbb{R},\ n \in \mathbb{N},$$
$$\lambda_1, \ldots, \lambda_n \geq 0,\ f_1, \ldots, f_n \in \operatorname{dom} \underline{P},\ \alpha + \sum_{i=1}^{n} \lambda_i f_i \leq f\Bigg\}$$

and observe that, since $\underline{P}(f_i) \leq \underline{Q}(f_i)$ for all $i \in \{1, \ldots, n\}$, if $\alpha + \sum_{i=1}^{n} \lambda_i \underline{P}(f_i)$ is well defined and strictly larger than $-\infty$, then also $\alpha + \sum_{i=1}^{n} \lambda_i \underline{Q}(f_i)$ is well defined and $\alpha + \sum_{i=1}^{n} \lambda_i \underline{P}(f_i) \leq \alpha + \sum_{i=1}^{n} \lambda_i \underline{Q}(f_i)$. Hence,

$$\leq \sup \left\{ \alpha + \sum_{i=1}^{n} \lambda_i \underline{Q}(f_i) \text{ w.d.: } \alpha \in \mathbb{R},\ n \in \mathbb{N}, \right.$$
$$\left. \lambda_1, \ldots, \lambda_n \geq 0,\ f_1, \ldots, f_n \in \operatorname{dom} \underline{P},\ \alpha + \sum_{i=1}^{n} \lambda_i f_i \leq f \right\}$$

but, since $\underline{Q}$ is coherent, it follows by Theorem 5.5 that $\alpha + \sum_{i=1}^{n} \lambda_i \underline{Q}(f_i) \leq \underline{Q}(\alpha + \sum_{i=1}^{n} \lambda_i f_i) \leq \underline{Q}(f)$ whenever the left hand side is well defined. Hence,

$$\leq \underline{Q}(f).$$

(iv) $\Longrightarrow$ (v) $\Longrightarrow$ (vi). Immediate.

(vi) $\Longrightarrow$ (vii). Let $\underline{Q}$ be any behavioural extension of $\underline{P}$ that avoids sure loss. We need to show that $\underline{P}$ avoids sure loss: take any $n \in \mathbb{N}$, non-negative $\lambda_1, \ldots, \lambda_n$ in $\mathbb{R}$, and random quantities $f_1, \ldots, f_n$ in $\operatorname{dom} \underline{P}$ such that $\sum_{i=1}^{n} \lambda_i \underline{P}(f_i)$ is well-defined. If $\sum_{i=1}^{n} \lambda_i \underline{P}(f_i) = -\infty$, then Eq. (5.1) is immediate. Otherwise, since $\underline{P}(f_i) \leq \underline{Q}(f_i)$ for all $i \in \{1, \ldots, n\}$, also $\sum_{i=1}^{n} \lambda_i \underline{Q}(f_i)$ is well defined, and

$$\sum_{i=1}^{n} \lambda_i \underline{P}(f_i) \leq \sum_{i=1}^{n} \lambda_i \underline{Q}(f_i)$$

and since $\underline{Q}$ avoids sure loss,

$$\leq \sup \left[\sum_{i=1}^{n} \lambda_i f_i \right].$$

Hence, $\underline{P}$ avoids sure loss.

(vii) $\Longrightarrow$ (iii). Suppose $\underline{P}$ avoids sure loss. We already established (vii) $\Longrightarrow$ (iv), and hence, (iv) holds for $\mathcal{K} = \mathcal{R}(X)$: in particular, $\underline{E}$ is a coherent extended lower prevision on $\mathcal{R}(X)$.

(iii) $\Longrightarrow$ (ii). By Theorem 5.5(i).

(ii) $\Longrightarrow$ (i). Immediate.

(i) $\Longrightarrow$ (vii). We show that, if $\underline{P}$ does not avoid sure loss, then $\underline{E}(f) > -\infty$ implies that $\underline{E}(f) = +\infty$, for any random quantity f. So, if $\underline{P}$ does not avoid sure loss, then $\underline{E}(f) = \pm\infty$ for all random quantities f. Equivalently, if $\underline{P}$ avoids sure loss, then $-\infty < \underline{E}(f) < +\infty$ for at least one random quantity f.

Suppose that $\underline{P}$ does not avoid sure loss, and assume that $\underline{E}(f) > -\infty$. Then, there are $n \in \mathbb{N}$, non-negative $\lambda_1, \ldots, \lambda_n$ in $\mathbb{R}$, and random quantities $f_1, \ldots, f_n$ in $\operatorname{dom} \underline{P}$ such that $\sum_{i=1}^n \lambda_i \underline{P}(f_i)$ is well-defined and $\sup\left[\sum_{i=1}^n \lambda_i f_i\right] < \sum_{i=1}^n \lambda_i \underline{P}(f_i)$. Hence, there is an $\alpha \in \mathbb{R}$ such that $\sum_{i=1}^n \lambda_i f_i \leq -\alpha$ and $-\alpha < \sum_{i=1}^n \lambda_i \underline{P}(f_i)$, that is,

$$\alpha + \sum_{i=1}^{n} \lambda_i f_i \leq 0 < \alpha + \sum \lambda_i \underline{P}(f_i).$$

So, from the expression for natural extension,

$$\begin{aligned}\underline{E}(f) = \sup\Bigg\{\beta + \sum_{i=1}^{m} \kappa_i \underline{P}(g_i) \text{ w.d.: } \beta \in \mathbb{R},\, m \in \mathbb{N},\\ \kappa_1, \ldots, \kappa_m \geq 0,\, g_1, \ldots, g_m \in \operatorname{dom} \underline{P},\, \beta + \sum_{i=1}^{m} \kappa_i g_i \leq f\Bigg\}\end{aligned}$$

we get by Lemma A.9(vii)

$$\begin{aligned}= \sup\Bigg\{\beta + \sum_{i=1}^{m} \kappa_i \underline{P}(g_i) \text{ w.d. and } > -\infty\text{: } \beta \in \mathbb{R},\, m \in \mathbb{N},\\ \kappa_1, \ldots, \kappa_m \geq 0,\, g_1, \ldots, g_m \in \operatorname{dom} \underline{P},\, \beta + \sum_{i=1}^{m} \kappa_i g_i \leq f\Bigg\}\end{aligned}$$

which is obviously also equal to

$$\begin{aligned}= \sup\Bigg\{\beta + \sum_{i=1}^{m} \kappa_i \underline{P}(g_i) + \zeta\left(\alpha + \sum \lambda_i \underline{P}(f_i)\right) \text{ w.d. and } > -\infty\text{:}\\ \beta \in \mathbb{R},\, m \in \mathbb{N},\, \zeta \geq 0,\\ \kappa_1, \ldots, \kappa_m \geq 0,\, g_1, \ldots, g_m \in \operatorname{dom} \underline{P},\\ \beta + \sum_{i=1}^{m} \kappa_i g_i + \zeta\left(\alpha + \sum_{i=1}^{n} \lambda_i f_i\right) \leq f\Bigg\}\end{aligned}$$

and—this is rather subtle—since $\underline{E}(f) > -\infty$, there are $\beta \in \mathbb{R}$, $m \in \mathbb{N}$, non-negative $\kappa_1, \ldots, \kappa_m$ in $\mathbb{R}$, and random quantities $g_1, \ldots, g_m$ in $\operatorname{dom} \underline{P}$ such that $\sum_{i=1}^m \kappa_i \underline{P}(g_i)$ is well-defined, $\beta + \sum_{i=1}^m \kappa_i \underline{P}(g_i) > -\infty$, and $\beta + \sum_{i=1}^m \kappa_i g_i \leq f$. Fixing these β, m, $\kappa_1, \ldots, \kappa_m$, and $g_1, \ldots, g_m$ in the above supremum, and using the fact that $\alpha + \sum_{i=1}^n \lambda_i f_i \leq 0$, we see that

$$\geq \sup \left\{ \beta + \sum_{i=1}^m \kappa_i \underline{P}(g_i) + \zeta \left(\alpha + \sum \lambda_i \underline{P}(f_i) \right) : \zeta \geq 0 \right\}.$$

since, by construction, $\beta + \sum_{i=1}^m \kappa_i g_i + \zeta \left(\alpha + \sum_{i=1}^n \lambda_i f_i \right) \leq f$ is satisfied for all $\zeta \geq 0$. Now, since $\beta + \sum_{i=1}^m \kappa_i \underline{P}(g_i) > -\infty$ by construction, and $\alpha + \sum \lambda_i \underline{P}(f_i) > 0$, the argument tends to $+\infty$ as ζ increases, so

$$= +\infty;$$

which establishes the proof. □

Again, there are some subtle differences between Theorem 4.3 on p. 96 and Theorem 5.10. Firstly, even if there is a random quantity f on X such that $\underline{E}(f) < +\infty$, $\underline{P}$ may still incur sure loss. Indeed, let $\mathcal{X} = \mathbb{R}$, define the random quantity f on X by $f(x) := x$ for all $x \in \mathcal{X}$, and define the extended lower prevision $\underline{P}$ on $\{f, 0\}$ by $\underline{P}(f) = -\infty$, and $\underline{P}(0) = 1$. Clearly, $\underline{P}$ does not avoid sure loss: $\sup[0] < \underline{P}(0)$. Nevertheless, $\underline{E}(f) = -\infty < +\infty$; however, note that $\underline{E}(g) = \pm\infty$ for every random quantity g on X: $\underline{E}(g) = -\infty$ if g is unbounded from below, and $\underline{E}(g) = +\infty$ otherwise. In connection with Theorem 5.6, observe that $\underline{E}$ only satisfies condition (1) for random quantities that are unbounded from below, condition (2) for strictly positive λ, and condition (3): even though $\underline{E}$ is not coherent, it does have a coherent restriction. This phenomenon did not occur for lower previsions, by Theorem 4.3(i).

Secondly, it may happen that $\underline{E}(f) = \pm\infty$ for some random quantity f, even if $\underline{P}$ is a coherent lower prevision, *i.e.*, real-valued, defined on gambles only, and coherent. Indeed, for any lower prevision $\underline{P}$, $\underline{E}(f)$ is $-\infty$ whenever f is unbounded from below. As an example of a lower prevision $\underline{P}$ and a random quantity f such that $\underline{E}(f) = +\infty$, let $\mathcal{X} = \{x \in \mathbb{R} : x \geq 0\}$, let f be the identity map on $\mathcal{X}$, and define $\underline{P}$ on the set $\{f \wedge n : n \in \mathbb{N}\}$ of gambles by $\underline{P}(f \wedge n) := n$. Then, $\underline{P}$ is a coherent lower prevision; it is a restriction of the coherent lower prevision $\underline{Q}(g) := \lim_{n \to \infty} \inf_{x \geq n} g(x)$ defined for all gambles

g on X. Nevertheless, since $\underline{E}(f) \geq \underline{P}(f \wedge n) = n$ for all $n \in \mathbb{N}$, it holds that $\underline{E}(f) = +\infty$. Observe that, in this case, $\underline{E}$ is coherent.

$\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{R}(X)}$ uniquely determines the natural extension $\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{K}}$ of $\underline{P}$ to any domain $\mathcal{K}$ that includes $\operatorname{dom} \underline{P}$:

Corollary 5.11. *Let $\underline{P}$ be an extended lower prevision, and let $\operatorname{dom} \underline{P} \subseteq \mathcal{K} \subseteq \mathcal{R}(X)$. Then $\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{K}}$ exists if and only if $\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{R}(X)}$ exists, and in such a case*

$$\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{K}}(f) = \underline{\mathbf{E}}_{\underline{P}}^{\mathcal{R}(X)}(f) \text{ for all } f \in \mathcal{K}.$$

So, as before, from now on, in proofs, we can focus our attention on the natural extension $\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{R}(X)}$ of $\underline{P}$ to the set of all random quantities on X. Corollary 5.11 also tells us that the natural extension of a lower prevision $\underline{P}$ to a larger set $\mathcal{K}$ of gambles, as defined in Section 4.1, is given by the restriction of $\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{R}(X)}$ to $\mathcal{K}$: $\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{R}(X)}$ is an extension of $\underline{\mathbf{E}}_{\underline{P}}$, in the mathematical sense.

Again, an alternative, and simpler expression for natural extension is obtained when $\underline{P}$ is defined on a linear space and is coherent; also see Theorem 4.5.

Theorem 5.12. *Let $\underline{P}$ be any coherent extended lower prevision defined on a linear space, and let $\operatorname{dom} \underline{P} \subseteq \mathcal{K} \subseteq \mathcal{L}(X)$. Then the natural extension of $\underline{P}$ to $\mathcal{K}$ exists, and for any random quantity $f \in \mathcal{K}$,*

$$\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{K}}(f) := \sup\{a + \underline{P}(g) \colon a \in \mathbb{R},\ g \in \operatorname{dom} \underline{P},\ a + g \leq f\}. \tag{5.6}$$

Proof. Look at Eq. (5.5) and note that, by Theorem 5.5,

$$\sum_{i=1}^{n} \lambda_i \underline{P}(f_i) \leq \underline{P}\left(\sum_{i=1}^{n} \lambda_i f_i\right), \tag{5.7}$$

whenever the left hand side is well defined. Since we are looking for the supremum, we can replace $\sum_{i=1}^{n} \lambda_i \underline{P}(f_i)$ by $\underline{P}(g)$ with $g = \sum_{i=1}^{n} \lambda_i f_i$. □

Again, $\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{K}}$ coincides with $\underline{P}$ on $\operatorname{dom} \underline{P}$ if $\underline{P}$ is coherent; this follows directly from the definition of natural extension in case $\operatorname{dom} \underline{P} = \mathcal{K}$.

Proposition 5.13. *Let $\underline{P}$ be an extended lower prevision, and let $\operatorname{dom} \underline{P} \subseteq \mathcal{K} \subseteq \mathcal{R}(X)$. If $\underline{P}$ is coherent then $\underline{P}$ and $\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{K}}$ coincide on $\operatorname{dom} \underline{P}$.*

Proof. By the definition of natural extension, $\underline{P}$ coincides with $\underline{\mathbf{E}}_{\underline{P}}^{\operatorname{dom}\underline{P}}$ on $\operatorname{dom}\underline{P}$. But, by Theorem 5.10(iv), $\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{K}}$ is an extension of $\underline{\mathbf{E}}_{\underline{P}}^{\operatorname{dom}\underline{P}}$, and hence, also $\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{K}}$ coincides with $\underline{P}$ on $\operatorname{dom}\underline{P}$. □

Also Proposition 4.7, Proposition 4.8, and Corollary 4.9 generalise to extended lower previsions.

Proposition 5.14. *Let $\underline{P}$ and $\underline{Q}$ be extended lower previsions on X that avoid sure loss. If $\underline{Q}$ is a behavioural extension of $\underline{P}$, then $\underline{\mathbf{E}}_{\underline{Q}}^{\mathcal{R}(X)}$ is a behavioural extension of $\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{R}(X)}$ too: $\underline{\mathbf{E}}_{\underline{Q}}^{\mathcal{R}(X)}(f) \geq \underline{\mathbf{E}}_{\underline{P}}^{\mathcal{R}(X)}(f)$ for every random quantity f on X.*

Proof. If $\underline{Q}$ is a behavioural extension of $\underline{P}$, then any coherent behavioural extension of $\underline{Q}$ is also a coherent behavioural extension of $\underline{P}$; now apply the definition of natural extension to $\mathcal{R}(X)$: it is the point-wise smallest coherent behavioural extension to the set of all random quantities on X. □

Proposition 5.15. *Let $\underline{P}$ be an extended lower prevision that avoids sure loss. Let $\underline{Q}$ be any coherent behavioural extension of $\underline{P}$. Then $\underline{P}$ is equivalent to $\underline{Q}$ (that is, $\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{R}(X)} = \underline{\mathbf{E}}_{\underline{Q}}^{\mathcal{R}(X)}$) if and only if $\underline{Q}$ and $\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{R}(X)}$ coincide on* $\operatorname{dom}\underline{Q}$.

Proof. "if". Since $\underline{Q}$ is a behavioural extension of $\underline{P}$, any behavioural extension of $\underline{Q}$ is also a behavioural extension of $\underline{P}$. Hence, $\underline{\mathbf{E}}_{\underline{Q}}^{\mathcal{R}(X)} \geq \underline{\mathbf{E}}_{\underline{P}}^{\mathcal{R}(X)}$. To prove the converse inequality, let $\underline{R}$ be any coherent behavioural extension of $\underline{P}$ to the set of random quantities on X. The claim is established if we can show that $\underline{R}$ is also a behavioural extension of $\underline{Q}$. Indeed, $\underline{R} \geq \underline{\mathbf{E}}_{\underline{P}}^{\mathcal{R}(X)}$ by definition of natural extension. Since $\underline{Q} = \underline{\mathbf{E}}_{\underline{P}}^{\mathcal{R}(X)}$ on $\operatorname{dom}\underline{Q}$ it follows that also $\underline{R} \geq \underline{Q}$ on $\operatorname{dom}\underline{Q}$, which means that $\underline{R}$ is a behavioural extension of $\underline{Q}$.

"only if". Suppose $\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{R}(X)} = \underline{\mathbf{E}}_{\underline{Q}}^{\mathcal{R}(X)}$. Since $\underline{Q}$ is coherent, it follows from Proposition 4.6 that $\underline{Q}$ and $\underline{\mathbf{E}}_{\underline{Q}}^{\mathcal{R}(X)}$ coincide on $\operatorname{dom}\underline{Q}$, and hence, $\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{R}(X)}$ and $\underline{Q}$ coincide on $\operatorname{dom}\underline{Q}$. □

Corollary 5.16. *Let $\underline{P}$ be an extended lower prevision that avoids sure loss, and let* $\operatorname{dom}\underline{P} \subseteq \mathcal{J} \subseteq \mathcal{K} \subseteq \mathcal{R}(X)$. *Then*

$$\underline{\mathbf{E}}_{\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{J}}}^{\mathcal{K}}(f) = \underline{\mathbf{E}}_{\underline{P}}^{\mathcal{K}}(f), \text{ for all } f \in \mathcal{K}, \text{ and } \underline{\mathbf{E}}_{\underline{P}}^{\mathcal{J}}(f) = \underline{\mathbf{E}}_{\underline{P}}^{\mathcal{K}}(f), \text{ for all } f \in \mathcal{J}.$$

5.2.4 Duality Revisited

In order to prove duality results, we must first try to endow $\mathcal{R}(X)$ with a sufficiently strong topological structure, required to apply versions of the Hahn-Banach theorem.

Definition 5.17. For any extended lower prevision $\underline{P}$ that avoids sure loss, the extended real number $\left\|f\right\|_{\underline{P}} := \overline{\mathbf{E}}_{\underline{P}}^{\mathcal{R}(X)}(|f|)$, defined for any random quantity f on X, is called the *$\underline{P}$-norm* of f.

Definition 5.18. For any extended lower prevision $\underline{P}$ that avoids sure loss, the set $\mathcal{K}_{\underline{P}} := \left\{f \in \mathcal{R}(X)\colon \overline{\mathbf{E}}_{\underline{P}}^{\mathcal{R}(X)}(|f|) < +\infty\right\} = \left\{f \in \mathcal{R}(X)\colon \left\|f\right\|_{\underline{P}} < +\infty\right\}$ is called the *$\underline{P}$-space*.

Proposition 5.19. *Suppose $\underline{P}$ avoids sure loss. The following propositions hold.*

(i) $\mathcal{K}_{\underline{P}}$ contains any random quantity $f \in \operatorname{dom}\overline{P}$ that is bounded from below and satisfies $\overline{P}(f) < +\infty$.

(ii) If $f \in \mathcal{K}_{\underline{P}}$ and $|g| \leq |f|$, then $g \in \mathcal{K}_{\underline{P}}$.

(iii) $\mathcal{L}(X) \subseteq \mathcal{K}_{\underline{P}}$, i.e., any gamble on X belongs to the $\underline{P}$-space.

(iv) $\mathcal{K}_{\underline{P}}$, equipped with $\|\bullet\|_{\underline{P}}$, is a semi-normed linear lattice.

Proof. (i) Indeed, for any $x \in \mathcal{X}$, it holds that $|f(x)| \leq f(x) + 0 \vee (-\inf f)$: if $f(x) \geq 0$, then $|f(x)| = f(x)$, and if $f(x) < 0$, then $|f(x)| = -f(x) \leq -\inf f$; in both cases, the desired inequality follows. Hence, $|f| \leq f + 0 \vee (-\inf f)$, and therefore, by the coherence of $\overline{\mathbf{E}}_{\underline{P}}^{\mathcal{R}(X)}$ and Theorem 5.5(iv),

$$\overline{\mathbf{E}}_{\underline{P}}^{\mathcal{R}(X)}(|f|) \leq \overline{\mathbf{E}}_{\underline{P}}^{\mathcal{R}(X)}(f) + 0 \vee (-\inf f)$$

and, by definition, $\overline{\mathbf{E}}_{\underline{P}}^{\mathcal{R}(X)}$ is a behavioural extension of $\underline{P}$, so $\overline{\mathbf{E}}_{\underline{P}}^{\mathcal{R}(X)}(f) \leq \overline{P}(f)$:

$$\leq \overline{P}(f) + 0 \vee (-\inf f)$$

and since, by assumption, $\overline{P}(f) < +\infty$ and $\inf f > -\infty$,

$$< +\infty.$$

(ii). Immediate.

(iii). Immediate by the coherence of $\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{R}(X)}$ and Theorem 5.5(i): $\overline{\mathbf{E}}_{\underline{P}}^{\mathcal{R}(X)}(|f|) \leq \sup|f| < +\infty$ for any gamble f on X.

(iv). Immediate the coherence of $\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{R}(X)}$ and Theorem 5.5: $\overline{\mathbf{E}}_{\underline{P}}^{\mathcal{R}(X)}(|f|) \geq 0$ for any random quantity f, $\overline{\mathbf{E}}_{\underline{P}}^{\mathcal{R}(X)}(|\lambda f|) = |\lambda|\overline{\mathbf{E}}_{\underline{P}}^{\mathcal{R}(X)}(|f|)$ for any $\lambda \in \mathbb{R}$ and any random quantity f, and $\overline{\mathbf{E}}_{\underline{P}}^{\mathcal{R}(X)}(|f+g|) \leq \overline{\mathbf{E}}_{\underline{P}}^{\mathcal{R}(X)}(|f|) + \overline{\mathbf{E}}_{\underline{P}}^{\mathcal{R}(X)}(|g|)$, for any random quantities f and g. □

Lemma 5.20. *Let $\underline{P}$ be a real-valued coherent extended lower prevision defined on a linear lattice of random quantities, that contains all constant gambles. Let $\mathcal{D} \subseteq \operatorname{dom}\underline{P}$. Then the following conditions are equivalent.*

(A) $\overline{P}\left(\sum_{i=1}^{n} f_i\right) \geq 0$ for every $n \in \mathbb{N}$ and $f_1, \ldots, f_n$ in $\mathcal{D}$.

(B) $\overline{P}\left(\sum_{i=1}^{n} \lambda_j f_j\right) \geq 0$ for every $n \in \mathbb{N}$, non-negative $\lambda_1, \ldots, \lambda_n$ in $\mathbb{R}$, and $f_1, \ldots, f_n$ in $\mathcal{D}$.

(C) There is a linear behavioural extension Q of $\underline{P}$ such that $Q(f) \geq 0$ for every f in $\mathcal{D}$.

Proof. (A) $\Longrightarrow$ (B). Suppose that (A) holds, and assume *ex absurdo* that $\overline{P}\left(\sum_{j=1}^{n} \lambda_j f_j\right) = -\delta < 0$ for a particular choice of $n \in \mathbb{N}$, $\lambda_1, \ldots, \lambda_n \geq 0$ and $f_1, \ldots, f_n$ in $\mathcal{D}$. Let $\alpha = \overline{P}\left(\sum_{j=1}^{n} \left|f_j\right|\right)$ and let $\epsilon = \frac{\delta}{2\alpha+1}$. Since $\mathbb{Q}$ is dense in $\mathbb{R}$, there are non-negative rational numbers $\rho_j \in \mathbb{Q}$ such that $\lambda_j \leq \rho_j \leq \lambda_j + \epsilon$ for every $j \in \{1, \ldots, n\}$. By Lemma 3.4 we find that $\rho_j f_j \leq \lambda_j f_j + \epsilon\left|f_j\right|$ for every $j \in \{1, \ldots, n\}$. Let $k \in \mathbb{N}$ be a common denominator of $\rho_1, \ldots, \rho_n$ and let $m_j = k\rho_j \in \mathbb{N}$. We find that

$$\begin{aligned}\overline{P}\left(\sum_{j=1}^{n} \rho_j f_j\right) &\leq \overline{P}\left(\sum_{j=1}^{n} \lambda_j f_j + \epsilon\left|f_j\right|\right) \\ &\leq \overline{P}\left(\sum_{j=1}^{n} \lambda_j f_j\right) + \epsilon\overline{P}\left(\sum_{j=1}^{n} \left|f_j\right|\right) = -\delta + \epsilon\alpha < -\delta/2.\end{aligned}$$

We conclude that $\overline{P}\left(\sum_{j=1}^{n} m_j f_j\right) < -k\delta/2 < 0$. This contradicts (A).

(B) $\Longrightarrow$ (C). Define

$$\begin{aligned}\mathcal{E} &= \{g \in \operatorname{dom}\underline{P}; \underline{P}(g) \geq 0\}, \\ \mathcal{V} &= \{g \in \operatorname{dom}\underline{P}; g \geq \textstyle\sum_{j=1}^{n} \lambda_j f_j \text{ for some } n \in \mathbb{N}, f_j \in \mathcal{D} \cup \mathcal{E}, \lambda_j \geq 0\}.\end{aligned}$$

It is straightforward to show that $\mathcal{V}$ is a convex subset of $\operatorname{dom}\underline{P}$. By Proposition 5.19(iv), it follows that $\operatorname{dom}\underline{P}$, equipped with $\|\bullet\|_{\underline{P}}$, is a topological (semi-normed) linear space. First, we give some properties of $\mathcal{V}$.

(a) First we prove that $g \in \mathcal{V} \implies \overline{P}(g) \geq 0$. If $g \in \mathcal{V}$ then we can write that $g \geq \sum_{j=1}^{k} \lambda_j f_j + \sum_{j=1}^{l} \mu_j h_j$ with $k, l \in \mathbb{N}$, $\lambda_j, \mu_j \geq 0$, $f_j \in \mathcal{D}$ and $h_j \in \mathcal{E}$. By (B) and the coherence of $\underline{P}$ it follows that indeed

$$\overline{P}(g) \geq \overline{P}\left(\sum_{j=1}^{k} \lambda_j f_j\right) + \underline{P}\left(\sum_{j=1}^{l} \mu_j h_j\right) \geq 0 + 0.$$

(b) Next we prove that $g \in \operatorname{int}\mathcal{V} \implies \overline{P}(g) > 0$, where $\operatorname{int}\mathcal{V}$ denotes the topological interior of $\mathcal{V}$. If $g \in \operatorname{int}\mathcal{V}$ then there is an $\epsilon > 0$ such that

$$\left\{h \in \operatorname{dom}\underline{P}; \left\|h - g\right\|_{\underline{P}} < \epsilon\right\} \subseteq \mathcal{V}.$$

The inequality $\left\|h - g\right\|_{\underline{P}} < \epsilon$ is satisfied for $h = g - \epsilon/2$, which implies that $g - \epsilon/2 \in \mathcal{V}$. From (a) it follows that $\overline{P}(g - \epsilon/2) \geq 0$ and we find that indeed $\overline{P}(g) \geq \epsilon/2 > 0$.

(c) Next we prove that $\underline{P}(g) > 0 \implies g \in \operatorname{int}\mathcal{V}$. Suppose that $\underline{P}(g) > 0$. It suffices to prove that the neighbourhood

$$\left\{h \in \operatorname{dom}\underline{P}; \left\|h - g\right\|_{\underline{P}} < \underline{P}(g)/2\right\}$$

of g is a subset of $\mathcal{V}$. Assume that $\left\|h - g\right\|_{\underline{P}} < \underline{P}(g)/2$. We find that

$$\underline{P}(g) + \overline{P}(-h) \leq \overline{P}(g - h) \leq \overline{P}(|g - h|) < \underline{P}(g)/2.$$

This implies that $\overline{P}(-h) < -\underline{P}(g)/2$, or equivalently $\underline{P}(h) > \underline{P}(g)/2 > 0$. We find that $h \in \mathcal{E} \subseteq \mathcal{V}$. We conclude that $g \in \operatorname{int}\mathcal{V}$.

(d) Finally, we prove that $0 \notin \operatorname{int}\mathcal{V}$. Take the contraposition of (b) and use the fact that $\overline{P}(0) = 0 \not> 0$.

Now we are ready to apply a version of the Hahn-Banach theorem. By assumption, $1 \in \operatorname{dom}\underline{P}$, and $\underline{P}(1) = 1 > 0$. By (c) this implies that $1 \in \operatorname{int}\mathcal{V} \neq \emptyset$. Define the non-empty sets $\mathcal{A} = \mathcal{V}$ and $\mathcal{B} = \{0\}$. $\mathcal{A}$ is convex, $\operatorname{int}\mathcal{A} \neq \emptyset$,

$\mathcal{B}$ is convex and $\operatorname{int}\mathcal{A} \cap \mathcal{B} = \emptyset$. By a version of the Hahn-Banach theorem (see for instance Holmes [44, 11E, p. 63]), there is a continuous real-valued linear mapping Λ on $\operatorname{dom}\underline{P}$ such that for every $f \in \mathcal{A}$ and $g \in \mathcal{B}$ we have that $\Lambda(f) \geq \Lambda(g)$. Since $\mathcal{B}$ only contains the zero gamble, it follows that $\Lambda(f) \geq 0$ for every $f \in \mathcal{V}$.

Define the extended prevision $Q = \Lambda/\Lambda(1)$. We have that $Q(f) \geq 0$ for every $f \in \mathcal{V}$; clearly, Q is self-conjugate. So, to prove that Q is a linear extended lower prevision, it suffices to check the conditions of Theorem 5.6.

(1). We shall prove that $Q(f) = \Lambda(f)/\Lambda(1) \geq \underline{P}(f)$ for every $f \in \operatorname{dom}\underline{P}$: then, clearly, also $Q(f) \geq \inf f$. Let $f \in \operatorname{dom}\underline{P}$. The inequality is satisfied if $\Lambda(f) \geq \Lambda(1)\underline{P}(f)$. By the linearity of Λ this is equivalent to $\Lambda\left(f - \underline{P}(f)\right) \geq 0$; recall that $\underline{P}$ is assumed to be real-valued. This holds since $\underline{P}\left(f - \underline{P}(f)\right) \geq 0$ which implies that $f - \underline{P}(f) \in \mathcal{E} \subseteq \mathcal{V}$, and Λ is non-negative on $\mathcal{V}$ by construction.

(2)&(3). Immediate, since Q is a real-valued linear mapping.

It remains to prove that Q is a behavioural extension of $\underline{P}$—but this was shown above in (1)—and that is non-negative on $\mathcal{D}$: this follows at once from the observation that $\mathcal{D} \subseteq \mathcal{V}$, and that Λ is non-negative on $\mathcal{V}$ by construction.

(C) $\Longrightarrow$ (A) Suppose that Q is a linear extended lower prevision such that $\operatorname{dom}\underline{P} \subseteq \operatorname{dom} Q$, $Q(f) \geq \underline{P}(f)$ for all $f \in \operatorname{dom}\underline{P}$, and $Q(f) \geq 0$ for every $f \in \mathcal{D}$. Let $n \in \mathbb{N}$ and $f_1, \ldots, f_n \in \mathcal{D}$. It follows that $\overline{P}\left(\sum_{j=1}^n f_j\right) \geq Q\left(\sum_{j=1}^n f_j\right)$, and from (C) it follows that $Q\left(\sum_{j=1}^n f_j\right) \geq 0$, We conclude that indeed $\overline{P}\left(\sum_{j=1}^n f_j\right) \geq 0$ for every $n \in \mathbb{N}$ and every $f_1, \ldots, f_n \in \mathcal{D}$. □

Let $\mathbf{M}_{\underline{P}}^{\mathcal{K}}$ denote the set of linear extended previsions that are behavioural extensions of $\underline{P}$ to a negation invariant set of random quantities $\mathcal{K}$ that includes $\operatorname{dom}\underline{P}$.

Theorem 5.21. *Let $\underline{P}$ be any extended lower prevision, and let $\operatorname{dom}\underline{P} \subseteq \mathcal{K} \subseteq \mathcal{R}(X)$. The following statements hold.*

(i) If $\mathbf{M}_{\underline{P}}^{\mathcal{K}} \neq \emptyset$, then $\underline{P}$ avoids sure loss. Conversely, if $\underline{P}$ avoids sure loss and $\operatorname{dom}\underline{P} \subseteq \mathcal{K} \subseteq \mathcal{K}_{\underline{P}}$, then $\mathbf{M}_{\underline{P}}^{\mathcal{K}} \neq \emptyset$.

(ii) If $\underline{P}$ avoids sure loss and $\operatorname{dom}\underline{P} \subseteq \mathcal{K} \subseteq \mathcal{K}_{\underline{P}}$, then its natural extension $\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{R}(X)}$

satisfies

$$\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{R}(X)}(f) = \inf_{Q \in \mathbf{M}_{\underline{P}}^{\mathcal{K}}} \underline{\mathbf{E}}_{Q}^{\mathcal{R}(X)}(f) \qquad \text{for any } f \text{ in } \mathcal{R}(X), \tag{5.8}$$

$$\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{K}_{\underline{P}}}(f) = \min_{Q \in \mathbf{M}_{\underline{P}}^{\mathcal{K}}} \underline{\mathbf{E}}_{Q}^{\mathcal{K}_{\underline{P}}}(f) \qquad \text{for any } f \text{ in } \mathcal{K}_{\underline{P}}, \text{ and hence,} \tag{5.9}$$

$$\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{K}}(f) = \min_{Q \in \mathbf{M}_{\underline{P}}^{\mathcal{K}}} Q(f) \qquad \text{for any } f \text{ in } \mathcal{K}. \tag{5.10}$$

(iii) If $\underline{P}$ avoids sure loss and $\operatorname{dom} \underline{P} \subseteq \mathcal{K} \subseteq \mathcal{K}_{\underline{P}}$*, then $\underline{P}$ is coherent if and only if*

$$\underline{P}(f) = \min_{Q \in \mathbf{M}_{\underline{P}}^{\mathcal{K}}} Q(f) \qquad \text{for any } f \in \operatorname{dom} \underline{P}. \tag{5.11}$$

Proof. (i). If $\mathbf{M}_{\underline{P}}^{\mathcal{K}} \neq \emptyset$, choose any $Q \in \mathbf{M}_{\underline{P}}^{\mathcal{K}}$. Then for every $n \in \mathbb{N}$, non-negative $\lambda_1, \ldots, \lambda_n$ in $\mathbb{R}$, and random quantities $f_1, \ldots, f_n$ in $\operatorname{dom} \underline{P}$ such that $\sum_{i=1}^{n} \lambda_i \underline{P}(f_i)$ is well-defined, we have that

$$\sup\left[\sum_{i=1}^{n} \lambda_i f_i\right] \geq \sum_{i=1}^{n} \lambda_i Q(f_i) \geq \sum_{i=1}^{n} \lambda_i \underline{P}(f_i),$$

since Q avoids sure loss, so $\underline{P}$ avoids sure loss as well.

Conversely, if $\underline{P}$ avoids sure loss, define $\mathcal{D} := \emptyset$. Note that since $\operatorname{dom} \underline{P} \subseteq \mathcal{K} \subseteq \mathcal{K}_{\underline{P}}$, it holds that $\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{K}_{\underline{P}}}$ is real-valued. Also, since $\mathcal{D} = \emptyset$, Lemma 5.20(B) trivially applies, and by Lemma 5.20(C), there is a linear behavioural extension Q of $\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{K}_{\underline{P}}}$, and hence, of $\underline{P}$.

(ii). Let f be any gamble on X. Clearly, $\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{R}(X)}(f) \leq \inf\left\{\underline{\mathbf{E}}_{Q}^{\mathcal{R}(X)}(f) \colon Q \in \mathbf{M}_{\underline{P}}^{\mathcal{K}}\right\}$, since the natural extension of $\underline{P}$ is the point-wise smallest coherent behavioural extension of $\underline{P}$, and the natural extensions of linear extensions of $\underline{P}$ must also be behavioural extensions of $\underline{P}$.

We first show that, for any $f \in \mathcal{K}_{\underline{P}}$, there is a Q in $\mathbf{M}_{\underline{P}}^{\mathcal{K}}$ such that $\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{K}_{\underline{P}}}(f) \geq \underline{\mathbf{E}}_{Q}^{\mathcal{K}_{\underline{P}}}(f)$, establishing Eq. (5.9). Indeed, let $f \in \mathcal{K}_{\underline{P}}$, and choose $\mathcal{D} := \{\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{K}_{\underline{P}}}(f) - f\}$; note that $\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{K}_{\underline{P}}}$ is coherent, real-valued, and is defined on a linear lattice that contains all constant gambles, and moreover $\overline{\mathbf{E}}_{\underline{P}}^{\mathcal{K}_{\underline{P}}}\left(\lambda \underline{\mathbf{E}}_{\underline{P}}^{\mathcal{K}_{\underline{P}}}(f) - \lambda f\right) \geq 0$ for any $\lambda \geq 0$: so Lemma 5.20(B) applies. By Lemma 5.20(C), there is a linear behavioural extension R of $\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{K}_{\underline{P}}}$, and hence, of $\underline{P}$, such that $R\left(\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{K}_{\underline{P}}}(f) - f\right) \geq 0$,

i.e., $\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{K}_{\underline{P}}}(f) \geq R(f)$. The restriction Q of R to $\mathcal{K}$ still is a behavioural extension of $\underline{P}$: Q belongs to $\mathbf{M}_{\underline{P}}^{\mathcal{K}}$. Clearly, $R(f) \geq \underline{\mathbf{E}}_{Q}^{\mathcal{K}_{\underline{P}}}(f)$, since R is a behavioural extension of Q, by definition of Q. We conclude that $\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{K}_{\underline{P}}}(f) \geq \underline{\mathbf{E}}_{Q}^{\mathcal{K}_{\underline{P}}}(f)$.

Eq. (5.10) simply follows from the fact that $Q = \underline{\mathbf{E}}_{Q}^{\mathcal{K}}$ for any coherent prevision Q on $\mathcal{K}$; see Proposition 5.13 on p. 215.

Finally, we prove that, for any random quantity f, there is a Q in $\mathbf{M}_{\underline{P}}^{\mathcal{K}}$ such that

$$\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{R}(X)}(f) \geq \inf_{Q \in \mathbf{M}_{\underline{P}}^{\mathcal{K}}} \underline{\mathbf{E}}_{Q}^{\mathcal{R}(X)}(f),$$

establishing Eq. (5.8). This is trivially satisfied in case $\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{R}(X)}(f) = +\infty$. Note that, by the transitivity of natural extension (Corollary 5.16 on p. 216), the expression of natural extension from a linear space (Theorem 5.12 on p. 215), and Eq. (5.9), it follows that

$$\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{R}(X)}(f) = \sup_{g \in \mathcal{K}_{\underline{P}},\, g \leq f} \underline{\mathbf{E}}_{\underline{P}}^{\mathcal{K}_{\underline{P}}}(g) = \sup_{g \in \mathcal{K}_{\underline{P}},\, g \leq f} \min_{Q \in \mathbf{M}_{\underline{P}}^{\mathcal{K}}} \underline{\mathbf{E}}_{Q}^{\mathcal{K}_{\underline{P}}}(g).$$

If $\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{R}(X)}(f) = -\infty$, then, since $\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{K}_{\underline{P}}}(g)$ belongs to $\mathbb{R}$ for every $g \in \mathcal{K}_{\underline{P}}$, it must be that there is no $g \in \mathcal{K}_{\underline{P}}$ such that $g \leq f$. Consequently, also

$$\underline{\mathbf{E}}_{Q}^{\mathcal{R}(X)}(f) = \sup_{g \in \mathcal{K}_{\underline{P}},\, g \leq f} \underline{\mathbf{E}}_{Q}^{\mathcal{K}_{\underline{P}}}(g) = -\infty,$$

for any Q in $\mathbf{M}_{\underline{P}}^{\mathcal{K}}$: so in this case,

$$\inf_{Q \in \mathbf{M}_{\underline{P}}^{\mathcal{K}}} \underline{\mathbf{E}}_{Q}^{\mathcal{R}(X)}(f) = -\infty = \underline{\mathbf{E}}_{\underline{P}}^{\mathcal{R}(X)}(f),$$

and the desired inequality holds too. Finally, if $\underline{\mathbf{E}}_{Q}^{\mathcal{R}(X)}(f) \in \mathbb{R}$, then, for every $\epsilon > 0$, there is a $g_\epsilon \in \mathcal{K}_{\mathcal{K}}$ such that $g_\epsilon \leq f$ and

$$\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{R}(X)}(f) \leq \min_{Q \in \mathbf{M}_{\underline{P}}^{\mathcal{K}}} \underline{\mathbf{E}}_{Q}^{\mathcal{K}_{\underline{P}}}(g_\epsilon) + \epsilon.$$

Since $g_\epsilon \leq f$, it follows that $\underline{\mathbf{E}}_{Q}^{\mathcal{R}(X)}(g_\epsilon) \leq \underline{\mathbf{E}}_{Q}^{\mathcal{R}(X)}(f)$ for all Q in $\mathbf{M}_{\underline{P}}^{\mathcal{K}}$, and hence,

$$\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{R}(X)}(f) \leq \inf_{Q \in \mathbf{M}_{\underline{P}}^{\mathcal{K}}} \underline{\mathbf{E}}_{Q}^{\mathcal{R}(X)}(f) + \epsilon,$$

for all $\epsilon > 0$ (note that the minimum is now an infimum: we cannot guarantee that the minimum is still achieved), and therefore, also for $\epsilon = 0$: again, we recover the desired inequality.

(iii). Immediate from (ii), and the fact that, whenever $\underline{P}$ is coherent, it must holds that $\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{K}}(f) = \underline{P}(f)$ for all $f \in \operatorname{dom} \underline{P}$. □

It's nice to know when exactly $\operatorname{dom} \underline{P} \subseteq \mathcal{K}_{\underline{P}}$. The following corollary may serve as a guidance to construct extended lower previsions that satisfy exactly this requirement.

Corollary 5.22. *Suppose $\underline{P}$ avoids sure loss. Then $\operatorname{dom} \underline{P} \subseteq \mathcal{K}_{\underline{P}}$ whenever there are random quantities $f_1, \dots, f_n$ in $\operatorname{dom} \overline{P}$ that are bounded from below, such that*

(i) $\overline{P}(f_1) < +\infty, \dots, \overline{P}(f_n) < +\infty$, and

(ii) for all g in $\operatorname{dom} \underline{P}$ there are non-negative real numbers $a_0, a_1, \dots, a_n$ in $\mathbb{R}$ such that $|g| \leq a_0 + \sum_{i=1}^{n} a_i |f_i|$.

Proof. Immediate from Proposition 5.19. □

For instance, if $\underline{P}$ is a lower prevision (*i.e.*, real-valued and defined on gambles only) that avoids sure loss, then the conditions of Corollary 5.22 are satisfied, and hence, $\operatorname{dom} \underline{P} \subseteq \mathcal{K}_{\underline{P}}$.

5.2.5 Duality and Lower Integrals for Random Quantities?

The lower S-integral of a random quantity f with respect to a probability charge μ on a field $\mathcal{F}$ can be defined as follows; this is similar to Kolmogoroff's [50, Zweites Kapitel, §2, p. 663, Nr. 12] definition of the S-integral, and also similar to (but slightly more general than) Bhaskara Rao and Bhaskara Rao's [9, Definition 9.1.1, p. 231] definition of what they call the refinement integral:

$$\mathrm{S}\underline{\int} f \,\mathrm{d}\mu := \lim_{\mathcal{B} \in \mathbb{P}(\mathcal{F})} \sum_{B \in \mathcal{B}} \underline{P}_B(f)\mu(B);$$

the sum in the right-hand side is always well defined, since always $\underline{P}_B(f) < +\infty$. Similarly, we define the upper S-integral $\mathrm{S}\overline{\int} f \,\mathrm{d}\mu$, and the S-integral $\mathrm{S}\int f \,\mathrm{d}\mu$ if the lower and upper S-integral coincide. This generalises Definition 4.40 on p. 129 to random quantities. However, with the above definition, we don't have anymore that $\underline{\mathbf{E}}_{\mu}^{\mathcal{R}(X)}(\bullet)$ coincides with $\mathrm{S}\underline{\int} \bullet \,\mathrm{d}\mu$; the proof of

Theorem 4.42 on p. 130 does not generalise to random quantities, because $\underline{P}_B(f)$ may be $-\infty$, and $\mu(B) = 0$ at the same time, for some random quantity f on X and some B in $\mathcal{F}$.

For instance, the lower S-integral of the random quantity $f(\frac{1}{n}) := -n$ for all $n \in \mathbb{N}$, $n \geq 1$, and $f(x) := 0$ for all other $x \in [0,1]$, with respect to the Lebesgue measure λ on $[0,1]$, is equal to zero, but $\underline{\mathbf{E}}_\lambda^{\mathcal{R}(X)}(f) = -\infty$: indeed, $\underline{\mathbf{E}}_{\underline{P}}^{\mathcal{R}(X)}(f) = -\infty$ for any lower prevision $\underline{P}$ that avoids sure loss, and any random quantity f that is unbounded from below, simply because there are no gambles g such that $g \leq f$, so the expression for natural extension, Eq. (5.5) on p. 209, results in a supremum over the empty set, which is $-\infty$.

A similar consideration holds for the Dunford integral, which is defined in Section 4.3.8 not only for gambles, but also for random quantities, as long as they satisfy the conditions of Dunford integrability: with the same gamble f, again it holds that $\mathrm{D}\int f \,\mathrm{d}\lambda = 0$—the sequence f_n of gambles defined by $f_n := 0$ is a determining sequence for f—which is different from $\underline{\mathbf{E}}_\lambda^{\mathcal{R}(X)}(f)$.

Another consequence of these observations, is that we have no equivalent of Corollary 4.88 on p. 195 for extended lower previsions: there are coherent extended lower previsions $\underline{R}$ and random quantities f in $\operatorname{dom}\underline{R}$ such that, for every field $\mathcal{F}$ and every set m of probability charges on $\mathcal{F}$, still

$$\underline{R}(f) \notin \left\{ \inf_{\mu\in m} \underline{\mathbf{E}}_\mu^{\mathcal{R}(X)}(f), \inf_{\mu\in m} \mathrm{S}\underline{\int} f \,\mathrm{d}\mu, \inf_{\mu\in m} \mathrm{D}\int f \,\mathrm{d}\mu \right\};$$

we have to assume that f is Dunford integrable with respect to all μ in m for the last expression to have any meaning. Concluding, there is no *general* equivalence between sets of probability charges and coherent extended lower previsions on $\mathcal{R}(X)$ similar to Corollary 4.88, neither through natural extension, lower S-integrals, or Dunford integrals.

We now turn to the following question: are there coherent extended lower previsions, not necessarily defined only on gambles, that are representable by natural extension, lower S-integrals, or Dunford integrals? As the Dunford integral already guarantees the existence of an extension of a probability charge to a real-valued coherent extended prevision, which may be defined also on unbounded random quantities, we try to construct extended lower previsions from lower previsions through Cauchy sequences, similar to the construction of the Dunford integral, and we prove that these

are representable through a lower envelope of Dunford integrals.

This yields two results at once: (i) a representation of real-valued coherent extended lower previsions through Dunford integrals, and (ii) a way to extend coherent lower previsions to coherent extended lower previsions that are still real-valued, and that are not overly conservative—recall that $\underline{E}_{\underline{P}}^{\mathcal{R}(X)}(f) = -\infty$ for any lower prevision $\underline{P}$ that avoids sure loss and any random quantity f unbounded from below; simply taking the natural extension of a lower prevision, we are not disposed to buy any random quantity that is unbounded from below: that's rather conservative.

5.3 Extension of Lower Previsions to Essentially Bounded Random Quantities

5.3.1 Null Sets and Null Random Quantities

From now on, unless stated otherwise, we shall assume that $\underline{P}$ is a coherent lower prevision defined on $\mathcal{L}(X)$, and $\overline{P}$ is its conjugate upper prevision. If we start out with a lower prevision $\underline{Q}$ that avoids sure loss, then we can always end up with such a $\underline{P}$ by natural extension of $\underline{Q}$ as explained in Chapter 4. We shall only rely on the coherence of $\underline{P}$, and on nothing else.

A first step towards an extension of lower previsions to possibly unbounded random quantities, is initiated by the observation that a coherent lower prevision may be invariant under a change of the values of gambles in some states $x \in \mathcal{X}$. For instance, suppose we throw a pebble and when it lands, the distance between its centre and a reference point near us is measured. Consider the measured distance X (in metres); X is a random variable, and it takes values in the set of non-negative reals. Prior to the throw, we are guaranteed to receive a reward $f(x) := x \wedge 100$ in units of some linear utility: we receive the gamble f. Alternatively, consider the unbounded random quantity $g(x) := x$ for every $x \leq 100$ and $g(x) := -x$ otherwise. If we are very confident that we cannot throw a pebble further than 100m, then f is equivalent with g: changing this gamble f at any state $x \geq 100$ does not change our expected reward.

This shows that, depending on your beliefs, the values of a gamble can sometimes be modified in some states without changing our expected re-

ward. Already observe that there is a similar phenomenon in the theory of integration: for example, changing the value of the integrand in a countable number of points of the real line does not change the value of the Lebesgue integral. Therefore, it does not matter whether the integrand is bounded on these points or not.

Definition 5.23. A subset A of $\mathcal{X}$ is called *$\underline{P}$-null* if $\overline{P}(I_A) = 0$. The set of all $\underline{P}$-null sets is denoted by $\mathcal{N}_{\underline{P}}$.

The following theorem states that $\underline{P}$-null sets are just those sets for which we are absolutely certain—prepared to bet at all odds on the fact—that they do not contain the outcome of X.

Theorem 5.24. *A subset A of $\mathcal{X}$ is $\underline{P}$-null if and only if $\underline{P}(-KI_A + \epsilon I_{\complement A}) > 0$ for all $K \geq 0$ and $\epsilon > 0$.*

Proof. If A could contain the outcome of X, accepting $-KI_A + \epsilon I_{\complement A}$ could result in an arbitrary large loss by choosing K sufficiently large and ϵ sufficiently small. Hence, intuitively, it is clear that the condition yields a sufficient condition for A to be a $\underline{P}$-null set. Mathematically, this follows from the coherence of $\underline{P}$:

$$0 < \underline{P}(-KI_A + \epsilon I_{\complement A}) \leq K\underline{P}(-I_A) + \epsilon\overline{P}(I_{\complement A}),$$

and hence,

$$K\overline{P}(I_A) < \epsilon\overline{P}(I_{\complement A}),$$

for all $K \geq 0$ and $\epsilon > 0$. This can only be satisfied if $\overline{P}(I_A) = 0$. Indeed, if $\overline{P}(I_A) \neq 0$ then any $K > \epsilon\frac{\overline{P}(I_{\complement A})}{\overline{P}(I_A)}$ will violate the inequality.

Conversely, assume that I_A is $\underline{P}$-null. Then $\underline{P}(-I_A) = -\overline{P}(I_A) = 0$ and $\underline{P}(I_{\complement A}) = \underline{P}(1 - I_A) = 1 - \underline{P}(-I_A) = 1$. Using these equalities, we find that

$$\underline{P}(-KI_A + \epsilon I_{\complement A}) \geq K\underline{P}(-I_A) + \epsilon\underline{P}(I_{\complement A}) = \epsilon > 0.$$

for all $K \geq 0$ and $\epsilon > 0$. □

Proposition 5.25. *The following statements hold.*

(i) $\emptyset$ is a $\underline{P}$-null set. $\mathcal{X}$ is not a $\underline{P}$-null set.

(ii) If A is a $\underline{P}$-null set and $B \subseteq A$, then B is a $\underline{P}$-null set.

(iii) If $A_1, \ldots, A_n$ are $\underline{P}$-null sets, then $\bigcup_{i=1}^{n} A_i$ is a $\underline{P}$-null set.

Proof. (i) follows from [86, Theorem 2.7.4(b)]. (ii) follows from [86, Theorem 2.7.4(a)&(c)]. (iii) follows from [86, Theorem 2.7.4(a)&(j)]. □

Corollary 5.26. *The following statements hold.*

(i) $\mathcal{N}_{\underline{P}}$ is a proper ideal of subsets of $\mathcal{X}$.

(ii) $(\mathcal{N}_{\underline{P}}, \subseteq)$ is a directed set.

Proof. (i). Simply observe that Proposition 5.25(i)&(ii)&(iii) are the defining properties of a proper ideal.

(ii). Recall that a directed set is a partially pre-ordered set in which any two elements are dominated by another element, *i.e.*, which satisfies the composition property. This follows from Proposition 5.25(iii). Indeed, if A and $B \in \mathcal{N}_{\underline{P}}$ then $A \subseteq C$ and $B \subseteq C$ for $C = A \cup B \in \mathcal{N}_{\underline{P}}$. □

Definition 5.27. A random quantity f is called *$\underline{P}$-null* if for every $\epsilon > 0$,

$$\overline{P}(\{x \in \mathcal{X} \colon |f(x)| > \epsilon\}) = 0.$$

The set of all $\underline{P}$-null random quantities is denoted by $\mathcal{R}^0_{\underline{P}}$.

A random quantity is $\underline{P}$-null if, for all $\epsilon > 0$, we are absolutely certain—prepared to bet at all odds, in the sense of Theorem 5.24, on the fact—that its absolute value will not exceed ϵ.

Coherence does not imply downward continuity, but still, why not $\overline{P}(\{x \in \mathcal{X} \colon |f(x)| \neq 0\}) = 0$ as a definition? Clearly, it implies our definition, but it is too restrictive as demonstrated by the following example, adapted from Bhaskara Rao and Bhaskara Rao [9, Proposition 4.2.7(ii) and Example 2.3.5(1)]. Let $\mathcal{X} = \mathbb{N}$. Let $\overline{P}$ be an upper prevision defined by $\overline{P}(I_A) := 0$ if A is finite and $\overline{P}(I_A) := 1$ otherwise; $\overline{P}$ is coherent and can therefore be extended to a coherent upper prevision on all gambles through natural extension. Let $f(n) := 1/(n+1)$. Then, as intuitively expected, f is a $\underline{P}$-null gamble, but $\overline{P}(\{n \in \mathbb{N} \colon |f(n)| \neq 0\}) = 1 \neq 0$.

Definition 5.28. Two random quantities f and g are *equal almost everywhere with respect to* $\underline{P}$ if $f - g$ is $\underline{P}$-null. In this case, we use the notation $f = g$ a.e. $\underline{P}$.

Definition 5.29. A random quantity f is *dominated almost everywhere with respect to* $\underline{P}$ by a random quantity g if there is a $\underline{P}$-null random quantity N such that $f \leq g + N$. In this case, we use the notation $f \leq g$ a.e. $\underline{P}$. We also define the expressions $f \leq +\infty$ a.e. $\underline{P}$ and $-\infty \leq f$ a.e. $\underline{P}$ to be true for every random quantity f.

The following proposition gives some properties of $\underline{P}$-null random quantities.

Proposition 5.30. *Let $A \subseteq \mathcal{X}$. Let f, g, h, f_1, f_2, g_1, and g_2 be random quantities. Let a and b be real numbers. Then the following statements hold.*

(i) A is a $\underline{P}$-null set if and only if I_A is a $\underline{P}$-null gamble.

(ii) If A is $\underline{P}$-null, then $I_A f$ is $\underline{P}$-null.

(iii) If f and g are $\underline{P}$-null, then $\left|f\right|$, $af + bg$, $f \vee g$ and $f \wedge g$ are $\underline{P}$-null.

(iv) If $\left|f\right| \leq \left|g\right|$ a.e. $\underline{P}$ and g is $\underline{P}$-null, then f is $\underline{P}$-null.

(v) If $f \leq g$ a.e. $\underline{P}$ and $g \leq h$ a.e. $\underline{P}$, then $f \leq h$ a.e. $\underline{P}$.

(vi) $f \leq g$ a.e. $\underline{P}$ and $g \leq f$ a.e. $\underline{P}$ if and only if $f = g$ a.e. $\underline{P}$.

(vii) Assume that a and b are non-negative. If $f_1 \leq f_2$ a.e. $\underline{P}$ and $g_1 \leq g_2$ a.e. $\underline{P}$, then $af_1 + bg_1 \leq af_2 + bg_2$ a.e. $\underline{P}$.

(viii) If $f = g$ a.e. $\underline{P}$ and $g = h$ a.e. $\underline{P}$, then $f = h$ a.e. $\underline{P}$.

(ix) If $f_1 = f_2$ a.e. $\underline{P}$ and $g_1 = g_2$ a.e. $\underline{P}$, then $\left|f_1\right| = \left|f_2\right|$ a.e. $\underline{P}$, $af_1 + bg_1 = af_2 + bg_2$ a.e. $\underline{P}$, $f_1 \vee g_1 = f_2 \vee g_2$ a.e. $\underline{P}$ and $f_1 \wedge g_1 = f_2 \wedge g_2$ a.e. $\underline{P}$.

Proof. (i). This follows from Definition 5.23 and Definition 5.27.

(ii). For any $\epsilon > 0$, we have that $\{x \in \mathcal{X} \colon |I_A(x)f(x)| > \epsilon\} \subseteq A$. Using the monotonicity of $\overline{P}$, we find that $I_A f$ is a $\underline{P}$-null random quantity.

(iii). Treating sum and scalar multiplication separately and assuming that $a \neq 0$ (the case $a = 0$ is immediate), this follows from

$$\{x \in \mathcal{X} \colon ||f|(x)| > \epsilon\} = \{x \in \mathcal{X} \colon |f(x)| > \epsilon\},$$
$$\{x \in \mathcal{X} \colon |af(x)| > \epsilon\} = \{x \in \mathcal{X} \colon |f(x)| > \epsilon/|a|\},$$
$$\{x \in \mathcal{X} \colon |f(x) + g(x)| > \epsilon\} \subseteq \{x \in \mathcal{X} \colon |f(x)| > \epsilon/2\} \cup \{x \in \mathcal{X} \colon |g(x)| > \epsilon/2\},$$

and the monotonicity and sub-additivity of $\overline{P}$. The maximum and the minimum of f and g can be written as linear combinations of f, g and $|f-g|$,

$$f \vee g = (f+g+|f-g|)/2, \quad f \wedge g = (f+g-|f-g|)/2,$$

so these cases follow from the previous ones.

(iv). By definition of $|f| \leq |g|$ a.e. $\underline{P}$ there is a $\underline{P}$-null random quantity N such that $|f| \leq |g| + N$. We find that

$$\{x \in \mathcal{X} \colon |f(x)| > \epsilon\} \subseteq \{x \in \mathcal{X} \colon |g(x)| > \epsilon/2\} \cup \{x \in \mathcal{X} \colon |N(x)| > \epsilon/2\}.$$

Now use the monotonicity and sub-additivity of $\overline{P}$.

(v). By definition there are two $\underline{P}$-null random quantities N and M such that $f \leq g + N$ and $g \leq h + M$. It follows that $f \leq h + (N+M)$. By (iii) $N+M$ is a $\underline{P}$-null random quantity. We find that $f \leq h$ a.e. $\underline{P}$.

(vi). "if". Obvious. "only if". By definition there are two $\underline{P}$-null random quantities N and M such that $f \leq g + N$ and $g \leq f + M$. It follows that $-|M| \leq f - g \leq |N|$, which implies that $\left|f-g\right| \leq |N| \vee |M|$. By (iii) $|N| \vee |M|$ is a $\underline{P}$-null random quantity. It follows from (iv) that $f-g$ is also a $\underline{P}$-null random quantity. We find that $f = g$ a.e. $\underline{P}$.

(vii). By definition there are two $\underline{P}$-null random quantities N and M such that $f_1 \leq f_2 + N$ and $g_1 \leq g_2 + M$. It follows that $af_1 + bg_1 \leq af_2 + bg_2 + (aN + bM)$. By (iii) $aN + bM$ is a $\underline{P}$-null random quantity. We find that $af_1 + bg_1 \leq af_2 + bg_2$ a.e. $\underline{P}$.

(viii). By definition there are two $\underline{P}$-null random quantities N and M such that $f = g + N$ and $g = h + M$. It follows that $f = h + (N+M)$. By (iii) $N+M$ is a $\underline{P}$-null random quantity. We find that $f = h$ a.e. $\underline{P}$.

(ix) By definition there are two $\underline{P}$-null random quantities N and M such that $f_1 = f_2 + N$ and $g_1 = g_2 + M$. It follows that $\left|f_2\right| - |N| \leq \left|f_1\right| \leq \left|f_2\right| + |N|$. By (iii) $-|N|$ and $|N|$ are $\underline{P}$-null random quantities. From (vi) we find that $\left|f_1\right| = \left|f_2\right|$ a.e. $\underline{P}$. It also follows that $af_1 + bg_1 = af_2 + bg_2 + (aN + bM)$. By (iii) $aN + bM$ is a $\underline{P}$-null random quantity. We find that $af_1 + bg_1 = af_2 + bg_2$ a.e. $\underline{P}$. The maximum and the minimum of f_1 and g_1 can be written as linear combinations of f_1, g_1 and $\left|f_1 - g_1\right|$, and a similar statement holds for the maximum and the minimum of f_2 and g_2, so these cases follow from the previous ones. □

Corollary 5.31. *The set of $\underline{P}$-null random quantities constitutes a linear lattice with respect to the point-wise ordering of random quantities.*

Corollary 5.32. *The binary relation $\leq$ a.e. $\underline{P}$ is a partial pre-order on the set of random quantities with associated equivalence relation $=$ a.e. $\underline{P}$.*

5.3.2 Null Gambles

Lemma 5.33. *Let f be a gamble. Let a be a real number. If $\overline{P}(|f|) = 0$ then $\underline{P}(af) = \overline{P}(af) = 0$.*

Proof. Assume that $\overline{P}(|f|) = 0$. Then we also have $\overline{P}(|af|) = |a|\overline{P}(|f|) = 0$. It follows from the coherence of $\underline{P}$ that

$$0 = -\overline{P}(|af|) = \underline{P}(-|af|) \leq \underline{P}(af) \leq \overline{P}(af) \leq \overline{P}(|af|) = 0.$$

We may therefore conclude that $\underline{P}(af) = \overline{P}(af) = 0$. □

Proposition 5.34. *Let f and g be gambles . Let a and b be real numbers. Then the following statements hold.*

(i) f is a $\underline{P}$-null gamble if and only if $\overline{P}(|f|) = 0$.

(ii) If f and g are $\underline{P}$-null gambles then $\underline{P}(af + bg) = \overline{P}(af + bg) = 0$.

Proof. (i). Immediate from Lemma 4.61 on p. 162, with a constant sequence $f_n := f$.

(ii). By (i) we have that $\overline{P}(|f|) = 0$ and $\overline{P}(|g|) = 0$. From Lemma 5.33 it follows that $\overline{P}(af) = \underline{P}(af) = 0$ and $\overline{P}(bg) = \underline{P}(bg) = 0$. Using the super-linearity of the coherent $\underline{P}$, the sub-linearity of $\overline{P}$ and $\underline{P}(h) \leq \overline{P}(h)$ for every gamble h, we find that

$$0 = \underline{P}(af) + \underline{P}(bg) \leq \underline{P}(af + bg) \leq \overline{P}(af + bg) \leq \overline{P}(af) + \overline{P}(bg) = 0,$$

so indeed $\underline{P}(af + bg) = \overline{P}(af + bg) = 0$. □

5.3.3 Essentially Bounded Random Quantities

We now explain how lower and upper previsions can be extended easily to random quantities that are bounded on the complement of a $\underline{P}$-null set.

Definition 5.35. Let f be a random quantity. Then the following conditions are equivalent; if any (hence all) of them are satisfied, we say that f is $\underline{P}$-*essentially bounded above*.

(A) There is a $\underline{P}$-null set $A \subset \mathcal{X}$ such that f is bounded above on $\complement A$.

(B) There is a gamble g such that $f \leq g$ a.e. $\underline{P}$.

(C) There is a gamble g such that $\overline{P}(\{x \in \mathcal{X} \colon f(x) > g(x)\}) = 0$.

Proof of equivalence. (A) $\implies$ (B). Let $A \subset \mathcal{X}$ be a $\underline{P}$-null set such that f is bounded above on $\complement A$. Define the *gamble* $g := (f\,\complement A) \vee 0$ and define the random quantity $N := I_A f$, then $f \leq g + N$. If we can prove that N is a $\underline{P}$-null random quantity, then (B) is established. Let $\epsilon > 0$. Since $|N(x)| > \epsilon$ implies that $x \in A$, we find that

$$\{x \in \mathcal{X} \colon |N(x)| > \epsilon\} \subseteq A.$$

Now use the monotonicity of $\overline{P}$ (which again follows from the coherence of $\underline{P}$).

(B) $\implies$ (C). Let g be a gamble such that $f \leq g$ a.e. $\underline{P}$. Then there is a $\underline{P}$-null random quantity N such that $f \leq g + N$. We also have that $f \leq g + |N|$. It follows that if $f(x) > g(x) + 1$ then $|N(x)| > 1$. Define the gamble $h := g + 1$. Then we find that

$$\{x \in \mathcal{X} \colon f(x) > h(x)\} \subseteq \{x \in \mathcal{X} \colon |N(x)| > 1\}.$$

Now use the monotonicity of the coherent $\overline{P}$ and the fact that N is a $\underline{P}$-null random quantity.

(C) $\implies$ (A). Let g be a gamble such that $\overline{P}(\{x \in \mathcal{X} \colon f(x) > g(x)\}) = 0$. Define $A := \{x \in \mathcal{X} \colon f(x) > g(x)\}$ then A is a $\underline{P}$-null set. Since $I_{\complement A} f \leq I_{\complement A} g$ it follows that f is bounded above on $\complement A$. This establishes (A). □

Definition 5.36. A random quantity f is called $\underline{P}$-*essentially bounded below* if $-f$ is $\underline{P}$-essentially bounded above.

Definition 5.37. Let f be a random quantity. Then the following conditions are equivalent; if any (hence all) of them are satisfied, we say that f is $\underline{P}$-*essentially bounded*; the set of all $\underline{P}$-essentially bounded random quantities is denoted by $\mathcal{K}_{\underline{P}}^{\sharp}(X)$.

(A) f is $\underline{P}$-essentially bounded above and below.

(B) $|f|$ is $\underline{P}$-essentially bounded above.

(C) There is a gamble $f_\mathfrak{b}$ such that $f = f_\mathfrak{b}$ a.e. $\underline{P}$.

Proof of equivalence. (A) $\Longrightarrow$ (B). Assume that there are gambles g_1 and g_2 such that $g_1 \leq f$ a.e. $\underline{P}$ and $f \leq g_2$ a.e. $\underline{P}$. This implies that there are $\underline{P}$-null random quantities N_1 and N_2 such that $g_1 \leq f + N_1$ and $f \leq g_2 + N_2$ It follows that $|f| \leq |g_1| + |g_2| + |N_1| + |N_2|$. We find that $|f| \leq |g_1| + |g_2|$ a.e. $\underline{P}$.

(B) $\Longrightarrow$ (C). Assume that there is a gamble g such that $\overline{P}(\{x \in \mathcal{X} \colon |f(x)| > g(x)\}) = 0$. Define $A := \{x \in \mathcal{X} \colon |f(x)| > g(x)\}$ and $f_\mathfrak{b} := f \mathbb{C} A$. If we can show that $f - f_\mathfrak{b}$ is a $\underline{P}$-null random quantity then $f = f_\mathfrak{b}$ a.e. $\underline{P}$. Let $\epsilon > 0$. Then

$$\{x \in \mathcal{X} \colon |f(x) - f_\mathfrak{b}(x)| > \epsilon\} \subseteq A.$$

Now use the monotonicity of the coherent $\overline{P}$.

(C) $\Longrightarrow$ (A). Notice that $f = f_\mathfrak{b}$ a.e. $\underline{P}$ implies that $f \leq f_\mathfrak{b}$ a.e. $\underline{P}$ and $f_\mathfrak{b} \leq f$ a.e. $\underline{P}$. □

Corollary 5.38. *$\mathcal{K}^\sharp_{\underline{P}}(X)$ is a linear lattice.*

Proof. By Definition 5.37(C), we can write $\mathcal{K}^\sharp_{\underline{P}}(X)$ as $\{f+N \colon f \in \mathcal{L}(X),\ N \in \mathcal{R}^0_{\underline{P}}\}$. This is a linear space by Corollary 5.31 on p. 229.

Is it also a lattice? By Definition 5.37(B), it is immediate that a random quantity f on X is essentially bounded if and only if $|f|$ is essentially bounded, and hence $f \in \mathcal{K}^\sharp_{\underline{P}}(X)$ if and only if $|f| \in \mathcal{K}^\sharp_{\underline{P}}(X)$. Since we already proved that $\mathcal{K}^\sharp_{\underline{P}}(X)$ is a linear space, it follows that $\mathcal{K}^\sharp_{\underline{P}}(X)$ is a lattice: $f \vee g$ and $f \wedge g$ can be written as linear combinations of $|f - g|$ and $|f + g|$. □

5.3.4 Extension to Essentially Bounded Random Quantities

Since for every $\underline{P}$-essentially bounded random quantity f, there is a gamble $f_\mathfrak{b}$ such that $f = f_\mathfrak{b}$ a.e. $\underline{P}$, we can define $\underline{P}(f_\mathfrak{b})$ and $\overline{P}(f_\mathfrak{b})$ to be the lower and upper prevision of f. This extends the domain of $\underline{P}$ and $\overline{P}$ from the set of all gambles to the set of all $\underline{P}$-essentially bounded random quantities. However, we still have to check whether the lower prevision does not depend on the particular choice of $f_\mathfrak{b}$. This is the subject of the following proposition and corollary.

Proposition 5.39. *Let f and g be gambles. If $f \leq g$ a.e. $\underline{P}$ then $\underline{P}(f) \leq \underline{P}(g)$ and $\overline{P}(f) \leq \overline{P}(g)$.*

Proof. Assume that $f \leq g$ a.e. $\underline{P}$. By Definition 5.29, there is a $\underline{P}$-null random quantity N' such that $f \leq g + N'$. But, this means that there is also a $\underline{P}$-null gamble N such that $f \leq g + N$. Indeed, take $N = |N'| \wedge |f - g|$. Then N is $\underline{P}$-null because $N \leq |N'|$, and N is a gamble because $\sup|N| \leq \sup|f - g| < +\infty$.

It follows from the coherence of $\underline{P}$ that

$$\begin{aligned}\underline{P}(f) &\leq \underline{P}(g + N) \leq \underline{P}(g) + \overline{P}(N),\\ \overline{P}(f) &\leq \overline{P}(g + N) \leq \overline{P}(g) + \overline{P}(N).\end{aligned}$$

By Proposition 5.34(i) we know that $\overline{P}(N) = 0$. We conclude that $\underline{P}(f) \leq \underline{P}(g)$ and $\overline{P}(f) \leq \overline{P}(g)$. □

Corollary 5.40. *Let f and g be gambles. If $f = g$ a.e. $\underline{P}$ then $\underline{P}(f) = \underline{P}(g)$ and $\overline{P}(f) = \overline{P}(g)$.*

Definition 5.41. For any $\underline{P}$-essentially bounded random quantity f, we define the extended lower and upper previsions by $\underline{P}^\sharp(f) := \underline{P}(f_\flat)$ and $\overline{P}^\sharp(f) := \overline{P}(f_\flat)$, where $f_\flat$ is any gamble on X such that $f = f_\flat$ a.e. $\underline{P}$. The gamble $f_\flat$ is called a *determining gamble for f with respect to $\underline{P}$*, or simply a *determining gamble for f*.

Note that, if f is $\underline{P}$-essentially bounded, then it follows from Definition 5.37(C) that $f_\flat$ exists, and it follows from Corollary 5.40 that the lower and upper prevision of f are independent of the choice of the determining gamble $f_\flat$.

Theorem 5.42. *$\underline{P}^\sharp$ and $\overline{P}^\sharp$ are extensions of $\underline{P}$ and $\overline{P}$, respectively.*

Proof. Let f be a gamble on $\mathcal{X}$. Then f is a determining gamble for itself. It follows from Definition 5.41 that $\overline{P}^\sharp(f) = \overline{P}(f)$ and $\underline{P}^\sharp(f) = \underline{P}(f)$. □

Theorem 5.43. *$\underline{P}^\sharp$ is a coherent extended lower prevision on $\mathcal{K}_{\underline{P}}^\sharp(X)$, and $\overline{P}^\sharp$ is its conjugate.*

Proof. Obviously, for any $\underline{P}$-essentially bounded random quantity f, it holds that $\underline{P}^\sharp(f) = \underline{P}(f_\flat) = -\overline{P}(-f_\flat) = -\overline{P}^\sharp(-f)$, so, $\overline{P}^\sharp$ is indeed the conjugate of $\underline{P}^\sharp$.

Since, by Corollary 5.38 on p. 232, $\mathcal{K}_{\underline{P}}^{\sharp}(X)$ is a linear space, it suffices to check that $\underline{P}^{\sharp}$ satisfies the conditions of Theorem 5.6 on p. 206. Let f and g be random quantities in $\mathcal{K}_{\underline{P}}^{\sharp}(X)$, and let λ be a non-negative real number.

(1). We must show that $\underline{P}^{\sharp}(f) \geq \inf f$. If $\inf f = -\infty$, then the claim is immediate. Otherwise, if $\inf f > -\infty$, let $f_{\flat}$ be any determining gamble for f: $f_{\flat} = f$ a.e. $\underline{P}$. Then, also $f_{\flat} \vee \inf f = f \vee \inf f$ a.e. $\underline{P}$, so, $f_{\flat} \vee \inf f$ is a determining gamble for $f \vee \inf f = f$. Hence,

$$\underline{P}^{\sharp}(f) = \underline{P}(f_{\flat} \vee \inf f) \geq \inf f,$$

since $f_{\flat} \vee \inf f \geq \inf f$.

(2). We must show that $\underline{P}^{\sharp}(\lambda f) = \lambda \underline{P}^{\sharp}(f)$. If $\lambda = 0$, then the claim is evident. If $\lambda > 0$, then $f_{\flat}$ is a determining gamble for f if and only if $\lambda f_{\flat}$ is a determining gamble for λf, and hence,

$$\underline{P}^{\sharp}(\lambda f) = \underline{P}(\lambda f_{\flat}) = \lambda \underline{P}(f_{\flat}) = \lambda \underline{P}^{\sharp}(f).$$

(3). We are left to show that $\underline{P}^{\sharp}(f+g) \geq \underline{P}^{\sharp}(f) + \underline{P}^{\sharp}(g)$. If $f_{\flat}$ is a determining gamble for f, and $g_{\flat}$ is a determining gamble for g, then $f_{\flat}+g_{\flat}$ is a determining gamble for $f+g$, whence:

$$\underline{P}^{\sharp}(f+g) = \underline{P}(f_{\flat}+g_{\flat}) \geq \underline{P}(f_{\flat}) + \underline{P}(g_{\flat}) = \underline{P}^{\sharp}(f) + \underline{P}^{\sharp}(g).$$

So, $\underline{P}^{\sharp}$ is coherent. □

5.3.5 Examples

Let $\underline{P} := \underline{P}_{\mathcal{X}}$ be the vacuous lower prevision on $\mathcal{L}(X)$. Then the empty set is the only $\underline{P}$-null set. The only $\underline{P}$-null random quantity is 0. We have that $\mathcal{L}(X) = \mathcal{K}_{\underline{P}}^{\sharp}(X)$, $\underline{P} = \underline{P}^{\sharp}$ and $\overline{P} = \overline{P}^{\sharp}$.

More generally, let $\underline{P} := \underline{P}_A$ be the vacuous lower prevision on $\mathcal{L}(X)$ with respect to a non-empty subset A of $\mathcal{X}$: $\underline{P}(f) = \inf_{x \in A} f(x)$. Then, $\mathcal{N}_{\underline{P}} = \{B \subseteq \mathcal{X} \colon B \cap A = \emptyset\}$, $\mathcal{R}_{\underline{P}}^{0} = \{f \in \mathcal{R}(\mathcal{X}) \colon I_A f = 0\}$, $\mathcal{K}_{\underline{P}}^{\sharp}(X) = \{f \in \mathcal{R}(\mathcal{X}) \colon I_A f \in \mathcal{L}(X)\}$, and

$$\underline{P}^{\sharp}(f) = \inf_{x \in A} f(x), \tag{5.12}$$

for any f in $\mathcal{K}_{\underline{P}}^{\sharp}(X)$, *i.e.*, any f that is bounded on A.

For another example, let $\mathcal{F}$ be a field, let μ be a probability charge on $\mathcal{F}$, and let $\underline{P} := \underline{\mathbf{E}}_\mu$. A set A is a $\underline{P}$-null set if and only if it is a μ-null set, as defined for instance by Bhaskara Rao and Bhaskara Rao [9, Definition 4.2.2], *i.e.*, if and only if $\mu^*(A) = 0$; this is immediate from Theorem 4.36(v) on p. 117. Consequently, a random quantity is a $\underline{P}$-null random quantity if and only if it is a μ-null function, as for instance defined by Bhaskara Rao and Bhaskara Rao [9, Definition 4.2.4], *i.e.*, if and only if $\{x \in \mathcal{X} \colon |f(x)| > \epsilon\}$ is a μ-null set for all $\epsilon > 0$. A random variable is $\underline{P}$-essentially bounded if and only if it is μ-essentially bounded, again as defined for instance by Bhaskara Rao and Bhaskara Rao [9, Definition 4.2.8], *i.e.*, if and only if it is bounded on the complement of a μ-null set. In fact, if f is a $\underline{P}$-essentially bounded random quantity, and f has an $\mathcal{F}$-simple determining gamble $f_\flat$, then f is Dunford integrable with respect to μ, and

$$\underline{P}^\sharp(f) = \mathrm{D}\int f \,\mathrm{d}\mu;$$

indeed, simply observe that any determining gamble $f_\flat$ for f determines a determining sequence $f_n := f_\flat$ for f, in the sense of Definition 4.60.

We shall prove further on that, for an arbitrary lower prevision $\underline{Q}$ that avoids sure loss, and whose domain is a subset of $\mathscr{L}_{\mathcal{F}}(X)$, it holds that

$$\underline{\mathbf{E}}^\sharp_{\underline{Q}}(f) = \min_{\mu \in \mathbf{m}^{\mathcal{F}}_{\underline{Q}}} \mathrm{D}\int f \,\mathrm{d}\mu,$$

for any $\underline{\mathbf{E}}_{\underline{Q}}$-essentially bounded random quantity f that has an $\mathcal{F}$-simple determining gamble $f_\flat$: on this set of random quantities, the extended lower prevision $\underline{\mathbf{E}}^\sharp_{\underline{Q}}$ is representable by a lower envelope of Dunford integrals. Of course, having an $\mathcal{F}$-simple determining gamble is quite a strong condition in general, even if $\mathcal{F}$ is the power set on $\mathcal{X}$. Therefore, further on, we shall prove a more general result; see Theorem 5.76 on p. 265.

5.4 Extension of Lower Previsions by Hazily Convergent Cauchy Sequences

5.4.1 Introduction

Instead of approximating a random quantity f by a gamble $f_\flat$ that is equal to f outside a null set, we now shall try approximating f by a sequence of gambles f_n, converging *hazily* to f, *i.e.*, converging to f outside a null set (see Definition 5.44 on p. 236): in this way, we may extend a lower prevision $\underline{P}$ to a larger set of random quantities, *i.e.*, larger than $\mathcal{K}_{\underline{P}}^{\sharp}(X)$. However, it is easy to construct different sequences of gambles, that converge hazily to the same random quantity f, but for which the lower prevision doesn't converge to the same number. In Proposition 5.48 on p. 239, we shall give sufficient conditions for such candidate sequences to converge to the same lower prevision, which are very similar to the conditions of Dunford integrability; compare to Definition 4.60 on p. 162. In essence, this comes down to requiring f_n to be a Cauchy sequence with respect to the $\underline{P}$-norm. The random quantities for which this Cauchy construction is possible are called *previsible*, and we shall prove that the resulting extension yields coherent extended lower and upper previsions.

A very reasonable way of approximating an unbounded random quantity is by a sequence of cuts: we shall prove that every previsible random variable can be approximated by a sequence of cuts. Moreover, the converse also holds: if a random variable can be approximated by cuts, then it is previsible. We shall prove an analogon of the Lebesgue dominated convergence theorem, in order to establish this.

Finally, we shall show that, for linear previsions, our extension coincides with the Dunford integral, and, for arbitrary lower previsions, our extension is given by a lower envelope of Dunford integrals.

5.4.2 Hazy Convergence

Definition 5.44. Let f be a random quantity, and let f_α be a net of random quantities. Then the following conditions are equivalent; if any (hence all) of them are satisfied, we say that *the net f_α converges $\underline{P}$-hazily to f* and we use the notation $f_\alpha \xrightarrow{\underline{P}} f$.

(A) For every $\epsilon > 0$, $\lim_{\alpha} \overline{P}(\{x \in \mathcal{X} \colon |f(x) - f_\alpha(x)| > \epsilon\}) = 0$.

(B) For every $\epsilon > 0$, eventually $\overline{P}(\{x \in \mathcal{X} \colon |f(x) - f_\alpha(x)| > \epsilon\}) < \epsilon$.

Proof of equivalence. We first prove that (A) implies (B). Let $\epsilon > 0$ and $\delta > 0$. Then (A) implies that there is a $N_{\epsilon,\delta}$ such that for every $\alpha \geq N_{\epsilon,\delta}$

$$\overline{P}(\{x \in \mathcal{X} \colon |f(x) - f_\alpha(x)| > \epsilon\}) < \delta.$$

Take $\delta := \epsilon$.

Next we prove that (B) implies (A). Let $\epsilon > 0$ and $\delta > 0$. Then (B) implies that there is an N_ϵ such that for every $\alpha \geq N_\epsilon$

$$\overline{P}(\{x \in \mathcal{X} \colon |f(x) - f_\alpha(x)| > \epsilon\}) < \epsilon.$$

Let $M_{\epsilon,\delta} := N_{\min\{\epsilon,\delta\}}$. It follows that for every $\alpha \geq M_{\epsilon,\delta}$ we have that

$$\overline{P}(\{x \in \mathcal{X} \colon |f(x) - f_\alpha(x)| > \min\{\epsilon, \delta\}\}) < \min\{\epsilon, \delta\} < \delta.$$

Since

$$\{x \in \mathcal{X} \colon |f(x) - f_\alpha(x)| > \epsilon\}) \subseteq \{x \in \mathcal{X} \colon |f(x) - f_\alpha(x)| > \min\{\epsilon, \delta\}\}).$$

we find that indeed

$$\overline{P}(\{x \in \mathcal{X} \colon |f(x) - f_\alpha(x)| > \epsilon\}) < \delta.$$

for every $\alpha \geq M_{\epsilon,\delta}$. □

Proposition 5.45. *Let f and g be random quantities. Let f_α be a net of random quantities. Then the following statements hold.*

(i) If $f_\alpha \xrightarrow{P} f$ and $f = g$ a.e. $\underline{P}$ then $f_\alpha \xrightarrow{P} g$.

(ii) Conversely, if $f_\alpha \xrightarrow{P} f$ and $f_\alpha \xrightarrow{P} g$ then $f = g$ a.e. $\underline{P}$.

Proof. The proof follows from

$$\begin{aligned}\{x \in \mathcal{X} \colon |f_\alpha(x) - g(x)| > \epsilon\} \\ \subseteq \{x \in \mathcal{X} \colon |f_\alpha(x) - f(x)| > \epsilon/2\} \cup \{x \in \mathcal{X} \colon |f(x) - g(x)| > \epsilon/2\},\end{aligned}$$

$$\{x \in \mathcal{X} \colon |f(x) - g(x)| > \epsilon\} \subseteq \{x \in \mathcal{X} \colon |f(x) - f_\alpha(x)| > \epsilon/2\} \cup \{x \in \mathcal{X} \colon |f_\alpha(x) - g(x)| > \epsilon/2\},$$

and the monotonicity and sub-additivity of the coherent $\overline{P}$. □

Proposition 5.46. *Let f and g be two random quantities. Let f_α and g_α be two nets of random quantities. Assume that $f_\alpha \xrightarrow{\underline{P}} f$ and $g_\alpha \xrightarrow{\underline{P}} g$. Let a and b be real numbers. Then $af_\alpha + bg_\alpha$, $|f_\alpha|$, $f_\alpha \vee g_\alpha$ and $f_\alpha \wedge g_\alpha$ converge $\underline{P}$-hazily to $af + bg$, $|f|$, $f \vee g$ and $f \wedge g$, respectively.*

Proof. For the linear combination, we treat addition and scalar multiplication separately. For the addition we have that

$$\{x \in \mathcal{X} \colon |f(x) + g(x) - f_\alpha(x) - g_\alpha(x)| > \epsilon\} \subseteq \{x \in \mathcal{X} \colon |f(x) - f_\alpha(x)| > \epsilon/2\} \cup \{x \in \mathcal{X} \colon |g(x) - g_\alpha(x)| > \epsilon/2\}.$$

Using the monotonicity and sub-additivity of the coherent $\overline{P}$, it follows that

$$\begin{aligned}\lim_{n\to\infty} \overline{P}(\{x \in \mathcal{X} \colon |f(x) + g(x) - f_\alpha(x) - g_\alpha(x)| > \epsilon\}) &\leq \lim_{n\to\infty} \overline{P}(\{x \in \mathcal{X} \colon |f(x) - f_\alpha(x)| > \epsilon/2\}) \\ &\quad + \lim_{n\to\infty} \overline{P}(\{x \in \mathcal{X} \colon |g(x) - g_\alpha(x)| > \epsilon/2\}) \\ &= 0 + 0,\end{aligned}$$

whence indeed $f_\alpha + g_\alpha \xrightarrow{\underline{P}} f + g$.

For scalar multiplication, we may assume that the scalar a is non-zero (the case $a = 0$ is immediate). Then

$$\lim_{n\to\infty} \overline{P}(\{x \in \mathcal{X} \colon |af(x) - af_\alpha(x)| > \epsilon\}) = \lim_{n\to\infty} \overline{P}(\{x \in \mathcal{X} \colon |f(x) - f_\alpha(x)| > \epsilon/|a|\}) = 0,$$

whence $af_\alpha \xrightarrow{\underline{P}} af$.

For the absolute value we have

$$\{x \in \mathcal{X} \colon \big||f(x)| - |f_\alpha(x)|\big| > \epsilon\} \subseteq \{x \in \mathcal{X} \colon |f(x) - f_\alpha(x)| > \epsilon\}.$$

Again use the monotonicity of the coherent $\overline{P}$.

The maximum and the minimum of f and g can be written as linear combinations of f, g and $|f-g|$,

$$f \vee g = (f+g+|f-g|)/2, \quad f \wedge g = (f+g-|f-g|)/2,$$

so this case follows from the previous cases. □

5.4.3 Previsibility

Lemma 5.47. *Let f be a non-negative gamble. Then, for every $\epsilon > 0$ there is a $\delta_\epsilon > 0$ such that for every $A \subseteq \mathcal{X}$, if $\overline{P}(A) < \delta_\epsilon$ then $\overline{P}(fA) < \epsilon$.*

Proof. Let $\epsilon > 0$. If $\sup f = 0$ then $f \equiv 0$ and $\overline{P}(fA) = 0$ independently of $A \subseteq \mathcal{X}$, whence $\overline{P}(fA) < \epsilon$. Hence, we may assume that $\sup f > 0$. Define $\delta_\epsilon := \epsilon / \sup f$. If $\overline{P}(A) < \delta_\epsilon$, then we find that $\overline{P}(fA) \leq \sup f \overline{P}(A) < \sup f \delta_\epsilon = \epsilon$. □

Proposition 5.48. *Let f_n and g_n be two sequences of gambles converging $\underline{P}$-hazily to a random quantity h. Suppose that*

$$\lim_{n,m\to\infty} \overline{P}(|f_n - f_m|) = 0, \quad \text{and} \quad \lim_{n,m\to\infty} \overline{P}(|g_n - g_m|) = 0,$$

i.e., that f_n and g_n are Cauchy with respect to the $\underline{P}$-norm. Then the limits $\lim_{n\to\infty} \overline{P}(f_n)$ and $\lim_{n\to\infty} \overline{P}(g_n)$ exist, are finite and coincide. Also the limits $\lim_{n\to\infty} \underline{P}(f_n)$ and $\lim_{n\to\infty} \underline{P}(g_n)$ exist, are finite and coincide.

Proof. First we prove that the limits exist and are finite. This follows from the inequalities (see the properties of coherence listed in Theorem 3.5 on p. 55)

$$|\overline{P}(f_n) - \overline{P}(f_m)| \leq \overline{P}(|f_n - f_m|), \qquad |\underline{P}(f_n) - \underline{P}(f_m)| \leq \overline{P}(|f_n - f_m|),$$
$$|\overline{P}(g_n) - \overline{P}(g_m)| \leq \overline{P}(|g_n - g_m|), \qquad |\underline{P}(g_n) - \underline{P}(g_m)| \leq \overline{P}(|g_n - g_m|).$$

Since the right hand sides converge to zero, the left hand sides must converge to zero too. This means that $\overline{P}(f_n)$, $\underline{P}(f_n)$, $\overline{P}(g_n)$ and $\underline{P}(g_n)$ are Cauchy sequences. By the completeness of $\mathbb{R}$, their limits exist and are finite.

We prove $\lim_{n\to\infty} \overline{P}(f_n) = \lim_{n\to\infty} \overline{P}(g_n)$ and $\lim_{n\to\infty} \underline{P}(f_n) = \lim_{n\to\infty} \underline{P}(g_n)$. Define the gamble $N_n := |f_n - g_n|$. Again by the coherence of $\underline{P}$ (Theorem 3.5

on p. 55), we have that

$$|\underline{P}(f_n) - \underline{P}(g_n)| \leq \overline{P}(N_n), \qquad |\overline{P}(f_n) - \overline{P}(g_n)| \leq \overline{P}(N_n).$$

The proof is complete if we can show that $\overline{P}(N_n)$ converges to zero. This is what we now set out to do.

For every $n \in \mathbb{N}$ and every $A \subseteq \mathcal{X}$, define $a_n(A) := \overline{P}(N_n A)$. We must prove that $a_n(\mathcal{X})$ converges to zero.

Every a_n is an element of the function space $\mathbb{R}^{\wp(\mathcal{X})}$. Equip this function space with the topology of uniform convergence on $\wp(\mathcal{X})$. From the completeness of $\mathbb{R}$, it follows that $\mathbb{R}^{\wp(\mathcal{X})}$ is complete with respect to the topology of uniform convergence on $\wp(\mathcal{X})$; see for instance Schechter [70, Section 19.12].

We first claim that a_n converges with respect to the topology of uniform convergence on $\wp(\mathcal{X})$. Indeed, consider $A \subseteq \mathcal{X}$, then, using the coherence of $\underline{P}$ (in particular, using that $|\overline{P}(h) - \overline{P}(h')| \leq \overline{P}(|h - h'|)$, $\big||h| - |h'|\big| \leq |h - h'| \leq |h| + |h'|$, $|h|A \leq |h|$ for all gambles h and h', and the monotonicity and sub-linearity of $\overline{P}$), we find that

$$\begin{aligned}
|a_n(A) - a_m(A)| &= \Big|\overline{P}\big(|f_n - g_n|A\big) - \overline{P}\big(|f_m - g_m|A\big)\Big| \\
&\leq \overline{P}\Big(\big||f_n - g_n| - |f_m - g_m|\big|A\Big) \\
&\leq \overline{P}\Big(\big|\,(f_n - g_n) - (f_m - g_m)\,\big|A\Big) \\
&\leq \overline{P}\Big(\big|\,(f_n - f_m) - (g_n - g_m)\,\big|\Big) \\
&\leq \overline{P}\big(|f_n - f_m|\big) + \overline{P}\big(|g_n - g_m|\big).
\end{aligned}$$

Since the right hand side converges to zero independently of A, it follows that a_n is Cauchy with respect to the topology of uniform convergence on $\wp(\mathcal{X})$. By the completeness of $\mathbb{R}^{\wp(\mathcal{X})}$ with respect to this topology, we find that a_n indeed converges uniformly on $\wp(\mathcal{X})$.

Uniform convergence implies point-wise convergence, so for every $A \subseteq \mathcal{X}$ we can define $a(A) := \lim_{n\to\infty} a_n(A)$. We must prove that $a(\mathcal{X}) = 0$.

Let $\epsilon > 0$. By the convergence of a_n with respect to the topology of uniform convergence on $\wp(\mathcal{X})$, there is an $M_\epsilon \in \mathbb{N}$ such that for every $A \subseteq \mathcal{X}$ and every $n \geq M_\epsilon$

$$|a_n(A) - a(A)| < \epsilon. \tag{5.13}$$

By Lemma 5.47, there is a $\delta_\epsilon > 0$ such that for every $A \subseteq \mathcal{X}$ we have that if $\overline{P}(A) < \delta_\epsilon$ then $\overline{P}(N_{M_\epsilon}A) = a_{M_\epsilon}(A) < \epsilon$. Since $a(A) \leq |a(A) - a_{M_\epsilon}(A)| + a_{M_\epsilon}(A)$, it follows from Eq. (5.13) that for all $A \subseteq \mathcal{X}$

$$\overline{P}(A) < \delta_\epsilon \implies a(A) < 2\epsilon. \tag{5.14}$$

Define $B := \{x \in \mathcal{X} : N_{M_\epsilon}(x) \neq 0\}$, then $N_{M_\epsilon} \complement B = 0$. This implies that $\overline{P}(N_{M_\epsilon} \complement B) = a_{M_\epsilon}(\complement B) = 0$. From Eq. (5.13) it follows that $a(\complement B) < \epsilon$.

Next we prove that $a(\mathcal{X}) < 5\epsilon$.

1. Consider the case $\overline{P}(B) = 0$. Then $a(B) = \lim_{n\to\infty} \overline{P}(N_n B) = 0$ since $0 \leq \overline{P}(N_n B) \leq \sup N_n \overline{P}(B) = 0$ for every $n \in \mathbb{N}$. By the sub-additivity of $\overline{P}$ it follows that $a(\mathcal{X}) \leq a(B) + a(\complement B) < 0 + \epsilon < 5\epsilon$.

2. Now consider the other case $\overline{P}(B) > 0$. Since f_n and g_n converge $\underline{P}$-hazily to h, it follows from Proposition 5.46 that $N_n = |f_n - g_n|$ converges $\underline{P}$-hazily to 0. By the definition of $\underline{P}$-hazy convergence, this implies that for the ϵ, M_ϵ and δ_ϵ constructed above, there is a $K_\epsilon \geq M_\epsilon$, such that for all $n \geq K_\epsilon$
$$\overline{P}(\{x \in \mathcal{X} : N_n(x) > \epsilon/\overline{P}(B)\}) < \delta_\epsilon. \tag{5.15}$$

   Define $C := \{x \in \mathcal{X} : N_{K_\epsilon}(x) \leq \epsilon/\overline{P}(B)\}$. By the sub-additivity of the coherent $\overline{P}$ we have that $a(\mathcal{X}) \leq a(B \cap C) + a(B \cap \complement C) + a(\complement B)$. We now investigate each term of this sum.

   (i) By Eq. (5.13) we get $a(B \cap C) < a_{K_\epsilon}(B \cap C) + \epsilon$, since $K_\epsilon \geq M_\epsilon$. Since $N_{K_\epsilon}(x) \leq \epsilon/\overline{P}(B)$ for all $x \in C$ and $\overline{P}(B \cap C) \leq \overline{P}(B)$, we find that $a_{K_\epsilon}(B \cap C) = \overline{P}(N_{K_\epsilon}B \cap C) \leq (\epsilon/\overline{P}(B))\overline{P}(B \cap C) \leq \epsilon$. Therefore $a(B \cap C) < 2\epsilon$.

   (ii) We claim that $a(B \cap \complement C) < 2\epsilon$. By Eq. (5.15) it follows that $\overline{P}(\complement C) < \delta_\epsilon$. The claim is established using $\overline{P}(B \cap \complement C) \leq \overline{P}(\complement C)$ and Eq. (5.14).

   (iii) We have already proved that $a(\complement B) < \epsilon$.

In both cases it follows that $a(\mathcal{X}) < 5\epsilon$. Since this holds for any $\epsilon > 0$, we may conclude that indeed $a(\mathcal{X}) = 0$. □

Definition 5.49. A random quantity f is called *$\underline{P}$-previsible* if there is a sequence f_n of gambles such that

(i) f_n converges $\underline{P}$-hazily to f, and

(ii) f_n is a Cauchy sequence with respect to the $\underline{P}$-norm.

The sequence f_n is called a *determining sequence of gambles for f with respect to $\underline{P}$*, or simply a *determining sequence for f*. The set of all $\underline{P}$-previsible random quantities is denoted by $\mathcal{K}_{\underline{P}}^{\mathsf{x}}(X)$. For any $\underline{P}$-previsible random quantity f, we define the extended lower and upper previsions $\underline{P}^{\mathsf{x}}(f) := \lim_{n\to\infty} \underline{P}(f_n)$ and $\overline{P}^{\mathsf{x}}(f) := \lim_{n\to\infty} \overline{P}(f_n)$, where f_n is any determining sequence for f.

Recall the definition of the $\underline{P}$-norm, Definition 5.17 on p. 217. So, since $\underline{P}$ is a coherent lower prevision defined on $\mathcal{L}(X)$, condition (ii) simply means that

$$\lim_{n,m\to\infty} \left\|f_n - f_m\right\|_{\underline{P}} = \lim_{n,m\to\infty} \overline{\mathbf{E}}_{\underline{P}}(|f_n - f_m|) = \lim_{n,m\to\infty} \overline{P}(|f_n - f_m|) = 0,$$

whence, if a random quantity f is $\underline{P}$-previsible, then, by Proposition 5.48, $\underline{P}^{\mathsf{x}}(f)$ and $\overline{P}^{\mathsf{x}}(f)$ exist, are finite and are independent of the choice of the determining sequence f_n for f.

Obviously, every gamble f is $\underline{P}$-previsible, since the constant sequence $f_n := f$ is a determining sequence for f: $\mathcal{L}(X) \subseteq \mathcal{K}_{\underline{P}}^{\mathsf{x}}(X)$. Moreover, every $\underline{P}$-essentially bounded random quantity f is $\underline{P}$-previsible, since the constant sequence $f_n := f_\flat$ is a determining sequence for f, if $f_\flat$ is a determining gamble for f: $\mathcal{L}(X) \subseteq \mathcal{K}_{\underline{P}}^{\sharp}(X) \subseteq \mathcal{K}_{\underline{P}}^{\mathsf{x}}(X)$. We even have a stronger result:

Theorem 5.50. *$\underline{P}^{\mathsf{x}}$ and $\overline{P}^{\mathsf{x}}$ are extensions of $\underline{P}^{\sharp}$ and $\overline{P}^{\sharp}$, respectively.*

Proof. Let $f \in \mathcal{K}_{\underline{P}}^{\sharp}(X)$. Let $f_\flat$ be a determining gamble for f. Define $f_n := f_\flat$ for every $n \in \mathbb{N}$. Then f_n is a determining sequence for f. It is immediate from Definition 5.49 that $\overline{P}^{\mathsf{x}}(f) = \overline{P}(f_\flat) = \overline{P}^{\sharp}(f)$ and $\underline{P}^{\mathsf{x}}(f) = \underline{P}(f_\flat) = \underline{P}^{\sharp}(f)$. □

Theorem 5.51. *Let f and g be $\underline{P}$-previsible random quantities. Let f_n be a determining sequence for f and let g_n be a determining sequence for g. Let a and b be real numbers. Then $af_n + bg_n$, $|f_n|$, $f_n \vee g_n$ and $f_n \wedge g_n$ are determining sequences for $af + bg$, $|f|$, $f \vee g$ and $f \wedge g$, respectively. Hence, $af + bg$, $|f|$, $f \vee g$ and $f \wedge g$ are $\underline{P}$-previsible.*

Proof. $\underline{P}$-hazy convergence follows from Proposition 5.46 on p. 238. To check

the Cauchy condition, use the inequalities

$$\overline{P}\Big(\big|(af_n + bg_n) - (af_m + bg_m)\big|\Big) \leq |a|\overline{P}\big(|f_n - f_m|\big) + |b|\overline{P}\big(|g_n - g_m|\big),$$
$$\overline{P}\Big(\big||f_n| - |f_m|\big|\Big) \leq \overline{P}\big(|f_n - f_m|\big)$$

and the fact that the maximum and the minimum of f and g can be written as linear combinations of f, g and $|f - g|$. □

Corollary 5.52. *$\mathcal{K}^{\mathbf{x}}_{\underline{P}}(\mathcal{X})$ is a linear lattice.*

Proposition 5.53. *$\underline{P}^{\mathbf{x}}$ is a coherent extended lower prevision on $\mathcal{K}^{\mathbf{x}}_{\underline{P}}(\mathcal{X})$, and $\overline{P}^{\mathbf{x}}$ is its conjugate.*

Proof. Obviously, for any $\underline{P}$-previsible random quantity f, it holds that $\underline{P}^{\mathbf{x}}(f) = \lim_{n\to\infty} \underline{P}(f_n) = \lim_{n\to\infty}(-\overline{P}(-f_n)) = -\lim_{n\to\infty} \overline{P}(-f_n) = -\overline{P}^{\mathbf{x}}(-f)$, so, $\overline{P}^{\mathbf{x}}$ is indeed the conjugate of $\underline{P}^{\mathbf{x}}$.

Since, by Corollary 5.52, $\mathcal{K}^{\mathbf{x}}_{\underline{P}}(X)$ is a linear space, it suffices to check that $\underline{P}^{\mathbf{x}}$ satisfies the conditions of Theorem 5.6 on p. 206. Indeed, let f and g be random quantities in $\mathcal{K}^{\mathbf{x}}_{\underline{P}}(X)$, and let λ be a non-negative real number.

(1). We must show that $\underline{P}^{\mathbf{x}}(f) \geq \inf f$. If $\inf f = -\infty$, then the claim is immediate. Otherwise, if $\inf f > -\infty$, note that, if f_n is a determining sequence for f, then $f_n \vee \inf f$ is a determining sequence for $f \vee \inf f = f$, by Theorem 5.51. Hence,

$$\underline{P}^{\mathbf{x}}(f) = \lim_{n\to\infty} \underline{P}(f_n \vee \inf f) \geq \inf f,$$

since $f_n \vee \inf f \geq \inf f$, and since $\underline{P}$ is coherent.

(2). We must show that $\underline{P}^{\mathbf{x}}(\lambda f) = \lambda \underline{P}^{\mathbf{x}}(f)$. If $\lambda = 0$, then the claim is evident. If $\lambda > 0$, then, by Theorem 5.51, f_n is a determining sequence for f if and only if λf_n is a determining sequence for λf, and hence, using the coherence of $\underline{P}$,

$$\underline{P}^{\mathbf{x}}(\lambda f) = \lim_{n\to\infty} \underline{P}(\lambda f_n) = \lim_{n\to\infty} \lambda \underline{P}(f_n) = \lambda \lim_{n\to\infty} \underline{P}(f_n) = \lambda \underline{P}^{\mathbf{x}}(f).$$

(3). We are left to show that $\underline{P}^{\mathbf{x}}(f+g) \geq \underline{P}^{\mathbf{x}}(f) + \underline{P}^{\mathbf{x}}(g)$. If f_n is a determining sequence for f, and g_n is a determining sequence for g, then, by Theorem 5.51, $f_n + g_n$ is a determining sequence for $f+g$, whence, again using the coherence

of $\underline{P}$:

$$\underline{P}^{\mathbf{x}}(f+g) = \lim_{n\to\infty} \underline{P}(f_n+g_n) \geq \lim_{n\to\infty} \left(\underline{P}(f_n)+\underline{P}(g_n)\right)$$
$$= \lim_{n\to\infty} \underline{P}(f_n) + \lim_{n\to\infty} \underline{P}(g_n) = \underline{P}^{\mathbf{x}}(f) + \underline{P}^{\mathbf{x}}(g).$$

So, $\underline{P}^{\mathbf{x}}$ is coherent. □

As a result of the coherence of $\underline{P}^{\mathbf{x}}$, it follows that $\underline{P}^{\mathbf{x}}$ satisfies all of the properties listed in Theorem 5.5 on p. 203. There are a few additional properties, which are worth mentioning here:

Proposition 5.54. *Let f and g be $\underline{P}$-previsible random quantities. If $f \leq g$ a.e. $\underline{P}$ then $\underline{P}^{\mathbf{x}}(f) \leq \underline{P}^{\mathbf{x}}(g)$ and $\overline{P}^{\mathbf{x}}(f) \leq \overline{P}^{\mathbf{x}}(g)$.*

Proof. If $f \leq g$ a.e. $\underline{P}$, then, by definition, there is a $\underline{P}$-null random quantity N such that $f \leq g + N$. Let h_n be a determining sequence for $g - f$. Since N is $\underline{P}$-null, it holds that $g - f = g - f + N$ a.e. $\underline{P}$, and so, by Proposition 5.45, h_n also converges $\underline{P}$-hazily to $g - f + N$: consequently, h_n is also a determining sequence for $g - f + N$. By Theorem 5.51, we find that $(h_n)^+ := h_n \vee 0$ is also a determining sequence for $(g - f + N)^+ = g - f + N$ (recall that $f \leq g + N$). It follows that

$$\underline{P}^{\mathbf{x}}(g) - \underline{P}^{\mathbf{x}}(f) \geq \underline{P}^{\mathbf{x}}(g-f) = \lim_{n\to\infty} \underline{P}(h_n) = \underline{P}^{\mathbf{x}}(g-f+N) = \lim_{n\to\infty} \underline{P}((h_n)^+) \geq 0,$$
$$\overline{P}^{\mathbf{x}}(g) - \overline{P}^{\mathbf{x}}(f) \geq \underline{P}^{\mathbf{x}}(g-f) = \lim_{n\to\infty} \underline{P}(h_n) = \underline{P}^{\mathbf{x}}(g-f+N) = \lim_{n\to\infty} \underline{P}((h_n)^+) \geq 0.$$

This completes the proof. □

Definition 5.55. For any $\underline{P}$-previsible random quantity f, the *$\underline{P}$-previsible norm of f* is defined by $\left\|f\right\|^{\mathbf{x}}_{\underline{P}} := \overline{P}^{\mathbf{x}}(|f|)$.

Note that the $\underline{P}$-previsible norm coincides with the restriction of the $\underline{P}^{\mathbf{x}}$-norm to the set of $\underline{P}$-previsible random quantities. Consequently, it holds that $\mathcal{K}^{\mathbf{x}}_{\underline{P}}(X)$, equipped with $\|\bullet\|^{\mathbf{x}}_{\underline{P}}$, is a semi-normed linear lattice (see Proposition 5.19(iv)).

Theorem 5.56. *Let f be a $\underline{P}$-previsible random quantity. Then f is a $\underline{P}$-null random quantity if and only if $\left\|f\right\|^{\mathbf{x}}_{\underline{P}} = 0$.*

Proof. Define $A_\epsilon := \{x \in \mathcal{X} \colon |f(x)| > \epsilon\}$.

First we show that the condition is sufficient. Suppose $\left\|f\right\|_{\underline{P}}^{\mathbf{x}} = 0$. Let $\epsilon > 0$. Since $|f| \geq \epsilon A_\epsilon \geq 0$ it follows from the monotonicity of $\overline{P}^{\mathbf{x}}$ (Theorem 5.5(iv)) that $0 = \left\|f\right\|_{\underline{P}}^{\mathbf{x}} \geq \epsilon \overline{P}(A_\epsilon) \geq 0$. We find that $\overline{P}(A_\epsilon) = 0$ for every $\epsilon > 0$, *i.e.*, f is $\underline{P}$-null.

Next we show that the condition is necessary. Suppose that f is a $\underline{P}$-null random quantity, *i.e.*, $\overline{P}(A_\epsilon) = 0$ for every $\epsilon > 0$. We prove that the constant sequence $f_n := 0$ is a determining sequence for $|f|$. For $\underline{P}$-hazy convergence, we need that $\overline{P}(\{x \in \mathcal{X} \colon \left||f(x)| - f_n(x)\right| > \epsilon\})$ converges to zero for every $\epsilon > 0$. Indeed, since $\{x \in \mathcal{X} \colon \left||f(x)| - f_n(x)\right| > \epsilon\} = A_\epsilon$ and $\overline{P}(A_\epsilon) = 0$ for every $\epsilon > 0$ we have $\underline{P}$-hazy convergence. The Cauchy condition is satisfied since we have a constant sequence. We find that $f_n := 0$ is a determining sequence for $|f|$. It follows that $\overline{P}^{\mathbf{x}}(|f|) = \overline{P}(0) = 0$. □

Finally, we show how sequences of previsible random quantities may also be used as determining sequences.

Lemma 5.57. *Let f be a $\underline{P}$-previsible random quantity. Let f_n be a determining sequence for f. Then f_n converges to f with respect to the $\underline{P}$-previsible norm: $\overline{P}^{\mathbf{x}}(|f - f_n|) \to 0$.*

Proof. By Theorem 5.51, $|f_m - f_n|$ is a determining sequence for $|f - f_n|$. Using Definition 5.49, it follows that

$$\lim_{n\to\infty} \overline{P}^{\mathbf{x}}(|f - f_n|) = \lim_{n\to\infty} \left(\lim_{m\to\infty} \overline{P}(|f_m - f_n|) \right) = \lim_{n,m\to\infty} \overline{P}(|f_m - f_n|) = 0.$$

This ends the proof. □

Theorem 5.58. *Let f_n be a sequence of $\underline{P}$-previsible random quantities. Suppose that*

(i) f_n converges $\underline{P}$-hazily to f,

(ii) $\lim_{n,m\to\infty} \overline{P}^{\mathbf{x}}(|f_n - f_m|) = 0$.

Then f is $\underline{P}$-previsible, $\underline{P}^{\mathbf{x}}(f) = \lim_{n\to\infty} \underline{P}^{\mathbf{x}}(f_n)$ and $\overline{P}^{\mathbf{x}}(f) = \lim_{n\to\infty} \overline{P}^{\mathbf{x}}(f_n)$.

Proof. Fix n in $\mathbb{N}$. Since f_n is $\underline{P}$-previsible and by Lemma 5.57, there is a

gamble g_n such that

$$\begin{aligned} \overline{P}^{\mathbf{x}}(|f_n - g_n|) &< 1/n, \\ \overline{P}(\{x \in \mathcal{X} \colon |f_n(x) - g_n(x)| > 1/n\}) &< 1/n. \end{aligned} \tag{5.16}$$

We prove that the g_n constitute a determining sequence for f.

First, we show that the sequence g_n converges $\underline{P}$-hazily to f. Let $\epsilon > 0$. Since f_n converges $\underline{P}$-hazily to f, there is an $N_\epsilon \in \mathbb{N}$ such that

$$\overline{P}(\{x \in \mathcal{X} \colon |f_n(x) - f(x)| > \epsilon/2\}) < \epsilon/3 \tag{5.17}$$

for every $n \geq N_\epsilon$. Define $M_\epsilon := \max\{N_\epsilon, \lceil 2/\epsilon \rceil\}$. Then for every $n \geq M_\epsilon$ we have that

$$\begin{aligned} &\{x \in \mathcal{X} \colon |g_n(x) - f(x)| > \epsilon\} \\ &\qquad \subseteq \{x \in \mathcal{X} \colon |g_n(x) - f_n(x)| > 1/M_\epsilon\} \cup \{x \in \mathcal{X} \colon |f_n(x) - f(x)| > \epsilon/2\}. \end{aligned}$$

It follows from Eq. (5.16), Eq. (5.17) and the sub-additivity of the coherent $\overline{P}$ that

$$\overline{P}(\{x \in \mathcal{X} \colon |g_n(x) - f(x)| > \epsilon\}) \leq 1/M_\epsilon + \epsilon/3 < \epsilon/2 + \epsilon/2 = \epsilon,$$

for every $n \geq M_\epsilon$. We find that g_n converges $\underline{P}$-hazily to f.

Next, we have that

$$\left\|g_n - g_m\right\|_{\underline{P}} = \left\|g_n - g_m\right\|_{\underline{P}}^{\mathbf{x}} \leq \left\|g_n - f_n\right\|_{\underline{P}}^{\mathbf{x}} + \left\|f_n - f_m\right\|_{\underline{P}}^{\mathbf{x}} + \left\|f_m - g_m\right\|_{\underline{P}}^{\mathbf{x}}.$$

Since all terms of the sum on the right side of the inequality converge to zero, it follows that $\left\|g_n - g_m\right\|_{\underline{P}}$ converges to zero, *i.e.*, g_n is Cauchy in $\underline{P}$-norm. We conclude that g_n is a determining sequence for f, and it follows that f is $\underline{P}$-previsible.

Finally, we have that

$$\left\|f_n - f\right\|_{\underline{P}}^{\mathbf{x}} \leq \left\|f_n - g_n\right\|_{\underline{P}}^{\mathbf{x}} + \left\|g_n - f\right\|_{\underline{P}}^{\mathbf{x}}.$$

Since both terms of the sum on the right side of the inequality converge to zero (see (5.16) and Lemma 5.57), it follows that $\left\|f_n - f\right\|_{\underline{P}}^{\mathbf{x}}$ converges to zero. Now use Theorem 5.5(xi) on p. 203. □

5.5 Approximation by Cuts

In this section, we establish a *constructive* way in order to determine whether a random quantity is previsible, and to find a determining sequence for an arbitrary previsible random quantity. In constructing determining sequences, it seems sensible to look at so-called cuts:

Definition 5.59. Let f be a random quantity. For a and b non-negative real numbers the *(a,b)-cut of* f is defined by

$$f_{a,b}(x) := \begin{cases} -b & \text{if } f(x) < -b, \\ a & \text{if } f(x) > a, \\ f(x) & \text{otherwise.} \end{cases}$$

If a_n and b_n are sequences of non-negative real numbers converging to infinity, we shall now establish that f_{a_n,b_n} is a determining sequence for f with respect to $\underline{P}$ if and only if f is $\underline{P}$-previsible, in which case $\underline{P}^{\mathrm{x}}(f) = \lim_{n\to\infty} \underline{P}(f_{a_n,b_n})$. The proof of this fact is simple, if we first introduce some new mathematical machinery.

5.5.1 A New Type of Measurability

Definition 5.60. Let f be a random quantity. Then the following conditions are equivalent; if any (hence all) of them are satisfied, we say that f is $\underline{P}$-*measurable*.

(A) There is a sequence of simple gambles f_n converging $\underline{P}$-hazily to f.

(B) For every $\epsilon > 0$ there is a partition $\{F_0, F_1, \ldots, F_n\}$ of $\mathcal{X}$ such that $\overline{P}(F_0) < \epsilon$ and $|f(x) - f(x')| < \epsilon$ for every $x, x' \in F_i$, $i = 1, \ldots, n$.

(C) There are two sequences a_n and b_n of non-negative real numbers converging to infinity such that f_{a_n,b_n} converges $\underline{P}$-hazily to f.

(D) For every two sequences a_n and b_n of non-negative real numbers converging to infinity we have that f_{a_n,b_n} converges $\underline{P}$-hazily to f.

Proof of equivalence of Definition 5.60(A)*&*(B). First we prove that (A) implies (B). Let $\epsilon > 0$. Since f_n converges $\underline{P}$-hazily to f, there is an m_ϵ in $\mathbb{N}$ such that

$$\overline{P}(\{x \in \mathcal{X} \colon |f(x) - f_{m_\epsilon}(x)| > \epsilon/3\}) < \epsilon.$$

Let $g := f_{m_\epsilon} = \sum_{i=1}^{n} c_i E_i$. Let $F_0 := \{x \in \mathcal{X} : |f(x) - g(x)| > \epsilon/3\}$. We have that $\overline{P}(F_0) < \epsilon$. Let $F_i := E_i \cap \complement F_0$, $i = 1, \ldots, n$. Since $\{E_1, \ldots, E_n\}$ is a partition of $\mathcal{X}$, it follows that $\{F_0, F_1, \ldots, F_n\}$ is also a partition of $\mathcal{X}$. For $i = 1, \ldots, n$ we have for every $x, x' \in F_i$ that $g(x) = g(x') = c_i$, $|f(x) - g(x)| \leq \epsilon/3$ and $|f(x') - g(x')| \leq \epsilon/3$. We find that

$$|f(x) - f(x')| = |f(x) - g(x) + g(x') - f(x')| \leq \epsilon/3 + \epsilon/3 < \epsilon.$$

This establishes (A) $\Longrightarrow$ (B).

Next we prove that (B) $\Longrightarrow$ (A). Let $n \in \mathbb{N} \setminus \{0\}$, and let $\{F_{n,0}, F_{n,1}, \ldots, F_{n,k_n}\}$ be a partition of $\mathcal{X}$ such that $\overline{P}(F_{n,0}) < 1/n$ and $|f(x) - f(x')| < 1/n$ for every $x, x' \in F_{n,i}$, $i = 1, \ldots, k_n$. Fix some $x_{n,i}$ in $F_{n,i}$, $i = 1, \ldots, k_n$. Define the simple gamble

$$f_n := \sum_{i=1}^{k_n} f(x_{n,i}) F_{n,i}.$$

We claim that f_n converges $\underline{P}$-hazily to f. Let $\epsilon > 0$. Define $M_\epsilon := \lceil 1/\epsilon \rceil$. By construction of f_n we have for every $n \geq M_\epsilon$ that

$$\{x \in \mathcal{X} : |f(x) - f_n(x)| > \epsilon\} \subseteq F_{n,0}.$$

It follows that

$$\overline{P}(\{x \in \mathcal{X} : |f(x) - f_n(x)| > \epsilon\}) \leq \overline{P}(F_{n,0}) < 1/n.$$

This establishes (B) $\Longrightarrow$ (A). □

To prove equivalence of Definition 5.60(B)&(C)&(D) we first need to prove some lemma's. All of these use Definition 5.60(A) or (B) for $\underline{P}$-measurability.

Lemma 5.61. *Every gamble is $\underline{P}$-measurable.*

Proof. Let f be a gamble. Let $\epsilon > 0$. Since f is bounded there is a sequence f_n of simple gambles converging uniformly to f, *i.e.*, there is an $N_\epsilon \in \mathbb{N}$ such that for every $n \geq N_\epsilon$ and every $x \in \mathcal{X}$ we have that $|f(x) - f_n(x)| < \epsilon$. This implies that

$$\overline{P}(\{x \in \mathcal{X} : |f(x) - f_n(x)| > \epsilon\}) = \overline{P}(\emptyset) = 0$$

for all $n \geq N_\epsilon$. We find that f_n converges $\underline{P}$-hazily to f. We conclude that f is $\underline{P}$-measurable. □

Definition 5.62. A random quantity f is called *$\underline{P}$-smooth* if for every $\epsilon > 0$ there is a $k_\epsilon > 0$ such that $\overline{P}(\{x \in \mathcal{X} \colon |f(x)| > k_\epsilon\}) < \epsilon$.

Lemma 5.63. *Every $\underline{P}$-measurable random quantity is $\underline{P}$-smooth.*

Proof. Let f be a $\underline{P}$-measurable random quantity. Let $\epsilon > 0$. By Definition 5.60(B) there is a partition $\{F_0, F_1, \ldots, F_n\}$ of $\mathcal{X}$ such that $\overline{P}(F_0) < \epsilon$ and $|f(x) - f(x')| < \epsilon$ for every $x, x' \in F_i$, $i = 1, \ldots, n$.

Fix $i \in \{1, \ldots, n\}$. Fix x' in F_i. Then $|f(x)| \leq |f(x')| + \epsilon$ for every $x \in F_i$. It follows that f is bounded on F_i.

Next, define

$$k := \max_{i=1}^{n} \sup\{|f(x)| \colon x \in F_i\},$$

then $|f(x)| > k$ only if $x \in F_0$. By the monotonicity of the coherent $\overline{P}$ we find that

$$\overline{P}(\{x \in \mathcal{X} \colon |f(x)| > k\}) \leq \overline{P}(F_0) < \epsilon.$$

This establishes the claim. □

Lemma 5.64. *Let f be a random quantity. Let f_α be a net of $\underline{P}$-measurable random quantities. If f_α converges $\underline{P}$-hazily to f then f is $\underline{P}$-measurable.*

Proof. We verify Definition 5.60(B). Let $\epsilon > 0$. Since f_α converges $\underline{P}$-hazily to f there is an α_0 such that

$$\overline{P}(\{x \in \mathcal{X} \colon |f(x) - f_{\alpha_0}(x)| > \epsilon/4\}) < \epsilon/2.$$

Define $A := \{x \in \mathcal{X} \colon |f(x) - f_{\alpha_0}(x)| > \epsilon/4\}$. Then $\overline{P}(A) < \epsilon/2$ and $|f(x) - f_{\alpha_0}(x)| \leq \epsilon/4$ for all $x \in \complement A$.

Since f_{α_0} is $\underline{P}$-measurable there is a partition $\{F_0, F_1, \ldots, F_n\}$ of $\mathcal{X}$ that satisfies $\overline{P}(F_0) < \epsilon/2$ and $|f_{\alpha_0}(x) - f_{\alpha_0}(x')| < \epsilon/2$ for all $x, x' \in F_i$, $i = 1, \ldots, n$.

Define $E_0 := A \cup F_0$ and $E_i := (\complement A) \cap F_i$, $i = 1, \ldots, n$. Since $\{F_0, F_1, \ldots, F_n\}$ is a partition of $\mathcal{X}$, we have that $\{E_0, E_1, \ldots, E_n\}$ is also a partition of $\mathcal{X}$. It follows that

$$\overline{P}(E_0) = \overline{P}(A \cup F_0) \leq \overline{P}(A) + \overline{P}(F_0) < \epsilon/2 + \epsilon/2 = \epsilon$$

and

$$\begin{aligned}|f(x) - f(x')| &= |f(x) - f_{\alpha_0}(x) + f_{\alpha_0}(x) - f_{\alpha_0}(x') + f_{\alpha_0}(x') - f(x')| \\ &\leq |f(x) - f_{\alpha_0}(x)| + |f_{\alpha_0}(x) - f_{\alpha_0}(x')| + |f_{\alpha_0}(x') - f(x')| \\ &< \epsilon/4 + \epsilon/2 + \epsilon/4 = \epsilon.\end{aligned}$$

for all $x, x' \in (\complement A) \cap F_i = E_i$. □

Proof of the equivalence of Definition 5.60(B)*&*(C)*&*(D). First, we prove that (B) implies (D). Since f is $\underline{P}$-measurable, f is also $\underline{P}$-smooth (Lemma 5.63). This means that for every $\delta > 0$ there is a $k_\delta > 0$ such that

$$\overline{P}(\{x \in \mathcal{X} \colon |f(x)| > k_\delta\}) < \delta. \tag{5.18}$$

Let a_n and b_n be two sequences of non-negative real numbers converging to infinity. For these sequences there is an N_δ in $\mathbb{N}$ such that for every $n \geq N_\delta$ we have that $a_n \geq k_\delta$ and $b_n \geq k_\delta$.

Let $\epsilon > 0$. For every $n \geq N_\delta$ we find that

$$\begin{aligned}\{x \in \mathcal{X} \colon |f_{a_n,b_n}(x) - f(x)| > \epsilon\} &\subseteq \{x \in \mathcal{X} \colon |f(x)| > \epsilon + \min\{a_n, b_n\}\} \\ &\subseteq \{x \in \mathcal{X} \colon |f(x)| > \epsilon + k_\delta\} \\ &\subseteq \{x \in \mathcal{X} \colon |f(x)| > k_\delta\}\end{aligned}$$

Now use the monotonicity of the coherent $\overline{P}$ and inequality Eq. (5.18) to conclude that f_{a_n,b_n} converges $\underline{P}$-hazily to f.

It is clear that (D) implies (C).

Finally we prove that (C) implies (B). Assume there are two sequences a_n and b_n of non-negative real numbers converging to infinity such that f_{a_n,b_n} converges $\underline{P}$-hazily to f. Since f_{a_n,b_n} is bounded, it follows from Lemma 5.61 that f_{a_n,b_n} is $\underline{P}$-measurable for every $n \in \mathbb{N}$. Now use Lemma 5.64. □

5.5.2 A Dominated Convergence Theorem

Lemma 5.65. *Let f be a random quantity. Let f_n be a sequence of gambles converging $\underline{P}$-hazily to f. Then there is a sequence g_n of gambles converging $\underline{P}$-hazily to f such that $|g_n| \leq 2|f|$ for every $n \in \mathbb{N}$.*

Proof. Since f_n converges $\underline{P}$-hazily to f there is a subsequence f_{n_k} such that for all k in $\mathbb{N}$

$$\overline{P}(\{x \in \mathcal{X}\colon |f(x) - f_{n_k}(x)| > 1/k\}) < 1/k.$$

Define $A_k := \{x \in \mathcal{X}\colon |f(x) - f_{n_k}(x)| > 1/k\}$ then we have for every $k \in \mathbb{N}$ that

$$\overline{P}(A_k) < 1/k \tag{5.19}$$

and for every $x \in \complement A_k$ that

$$|f(x) - f_{n_k}(x)| \leq 1/k. \tag{5.20}$$

Define

$$g_k(x) := \begin{cases} f_{n_k}(x) & \text{if } x \in \complement A_k \text{ and } |f_{n_k}(x)| > 2/k, \\ 0 & \text{otherwise.} \end{cases}$$

Since f_{n_k} is bounded, we have that g_{n_k} is bounded as well.

We now show that $|g_k| \leq 2|f|$ for all $k \in \mathbb{N}$. Consider x in $\mathcal{X}$.

1. If $x \in A_k$ or $|f_{n_k}(x)| \leq 2/k$ then $|g_k(x)| = 0$ and the inequality $|g_k(x)| \leq 2|f(x)|$ is satisfied.

2. Assume that $x \in \complement A_k$ and $|f_{n_k}(x)| > 2/k$. By Eq. (5.20) we have that

$$|g_k(x)| = |f_{n_k}(x)| \leq |f_{n_k}(x) - f(x)| + |f(x)| \leq 1/k + |f(x)|. \tag{5.21}$$

From Eq. (5.20) and $|f_{n_k}(x)| > 2/k$ we find that also

$$1/k \geq |f_{n_k}(x)| - |f(x)| > 2/k - |f(x)|$$

which implies that

$$|f(x)| > 1/k. \tag{5.22}$$

From Eq. (5.21) and Eq. (5.22) we find that indeed $|g_k(x)| \leq 2|f(x)|$.

In both cases we have that $|g_k(x)| \leq 2|f(x)|$. We conclude that $|g_k| \leq 2|f|$.

Finally, we prove that g_k converges $\underline{P}$-hazily to f. Let $\epsilon > 0$. Then

$$\begin{aligned} &\{x \in \mathcal{X}\colon |f(x) - g_k(x)| > \epsilon\} \\ &\qquad = \{x \in A_k\colon |f(x) - g_k(x)| > \epsilon\} \cup \{x \in \complement A_k\colon |f(x) - g_k(x)| > \epsilon\}. \end{aligned}$$

If $k > 1/\epsilon$ then it follows from Eq. (5.20) that $\{x \in \bigcap A_k \colon |f(x) - g_k(x)| > \epsilon\} = \emptyset$. Using Eq. (5.19) and the monotonicity and sub-linearity of $\overline{P}$, we find that for every $k > 1/\epsilon$

$$\overline{P}(\{x \in \mathcal{X} \colon |f(x) - g_k(x)| > \epsilon\}) \leq \overline{P}(A_k) + \overline{P}(\emptyset) < 1/k + 0.$$

This ends the proof. □

Lemma 5.66. *Let f be a positive $\underline{P}$-previsible random quantity. Then, for every $\epsilon > 0$ there is a $\delta_\epsilon > 0$ such that for every $A \subseteq \mathcal{X}$, if $\overline{P}(A) < \delta_\epsilon$ then $\overline{P}^{\mathrm{x}}(fA) < \epsilon$.*

Proof. Let $\epsilon > 0$. By Lemma 5.57 there is a gamble g such that $\overline{P}^{\mathrm{x}}(|f - g|) < \epsilon/2$. Define $\delta_\epsilon := \epsilon/(2\sup|g|)$. Let $A \subseteq \mathcal{X}$ and assume that $\overline{P}^{\mathrm{x}}(A) = \overline{P}(A) < \delta_\epsilon$. Then, taking into account the coherence of $\underline{P}^{\mathrm{x}}$ (Theorem 5.5 on p. 203),

$$\begin{aligned}\overline{P}^{\mathrm{x}}(fA) &\leq \overline{P}^{\mathrm{x}}(|f - g|A) + \overline{P}^{\mathrm{x}}(|g|A) = \overline{P}^{\mathrm{x}}(|f - g|A) + \overline{P}(|g|A) \\ &\leq \overline{P}^{\mathrm{x}}(|f - g|) + \sup|g|\overline{P}(A) < \epsilon/2 + \epsilon/2 = \epsilon.\end{aligned}$$

This ends the proof. □

Theorem 5.67. *Let f be a $\underline{P}$-previsible random quantity. Let g be a random quantity. Suppose that $|g| \leq |f|$ a.e. $\underline{P}$. Then g is $\underline{P}$-previsible if and only if g is $\underline{P}$-measurable.*

Proof. Assume $\underline{P}$-previsibility of g. $\underline{P}$-measurability follows from Definition 5.49, Lemma 5.61 and Lemma 5.64.

Conversely, assume that g is $\underline{P}$-measurable. By Lemma 5.65 there is a sequence g_n of gambles converging $\underline{P}$-hazily to g such that $g_n \leq 2|g|$ for every $n \in \mathbb{N}$.

Let $A \subseteq \mathcal{X}$. Since $|g| \leq |f|$ a.e. $\underline{P}$ we have that $|g_n| \leq 2|f|$ a.e. $\underline{P}$ which implies that $|g_n - g_m|A \leq 4|f|A$ a.e. $\underline{P}$ for every $n, m \in \mathbb{N}$. Applying Proposition 5.54 on p. 244, we find that

$$\overline{P}(|g_n - g_m|A) \leq \overline{P}^{\mathrm{x}}(4|f|A) \tag{5.23}$$

for every $A \subseteq \mathcal{X}$ and every $n, m \in \mathbb{N}$.

Now we show that $\left\|g_n - g_m\right\|_{\underline{P}}$ converges to zero. Let $\epsilon > 0$. Then, by Lemma 5.66 there is a $\delta_\epsilon > 0$ such that for every $A \subseteq \mathcal{X}$

$$\overline{P}^{\mathrm{x}}(A) < \delta_\epsilon \implies \overline{P}^{\mathrm{x}}(4|f|A) < \epsilon. \tag{5.24}$$

Since g_n converges $\underline{P}$-hazily to g, there is an N_ϵ in $\mathbb{N}$ such that for all $n, m \geq N_\epsilon$ we have that

$$\begin{aligned}\overline{P}(\{x \in \mathcal{X} \colon |g_n(x) - g_m(x)| > \epsilon\}) \\ \leq \overline{P}(\{x \in \mathcal{X} \colon |g_n(x) - g(x)| > \epsilon/2\}) + \overline{P}(\{x \in \mathcal{X} \colon |g(x) - g_m(x)| > \epsilon/2\}) \\ < \delta_\epsilon/2 + \delta_\epsilon/2 = \delta_\epsilon.\end{aligned}$$

Define $B_{nm} := \{x \in \mathcal{X} \colon |g_n(x) - g_m(x)| > \epsilon\})$ then for all $n, m \geq N_\epsilon$ we have that $\overline{P}^{\mathbf{x}}(B_{nm}) < \delta_\epsilon$. From Eq. (5.24) it follows that

$$\overline{P}^{\mathbf{x}}(4|f|B_{nm}) < \epsilon$$

for all $n, m \geq N_\epsilon$. We also have that

$$|g_n(x) - g_m(x)| \leq \epsilon.$$

for all $x \in \complement B_{nm}$. Using Eq. (5.23) it follows that

$$\begin{aligned}\overline{P}(|g_n - g_m|) &\leq \overline{P}(|g_n - g_m|B_{nm}) + \overline{P}(|g_n - g_m|\complement B_{nm}) \\ &\leq \overline{P}^{\mathbf{x}}(4|f|B_{nm}) + \epsilon\overline{P}^{\mathbf{x}}(\complement B_{nm}) < \epsilon + \epsilon = 2\epsilon.\end{aligned}$$

for all $n, m \geq N_\epsilon$.

We conclude that g_n is a determining sequence for g, so g is $\underline{P}$-previsible. □

The next theorem is a dominated convergence theorem for our extension $\underline{P}^{\mathbf{x}}$.

Theorem 5.68. *Let g be a $\underline{P}$-previsible random quantity. Let f_n be a sequence of $\underline{P}$-measurable random quantities such that $|f_n| \leq |g|$ a.e. $\underline{P}$ for every $n \in \mathbb{N}$. Let f be a random quantity. Then the following statements are equivalent.*

(i) f_n converges $\underline{P}$-hazily to f.

(ii) f is $\underline{P}$-previsible and $\left\|f - f_n\right\|_{\underline{P}}^{\mathbf{x}}$ converges to zero and hence, $\underline{P}^{\mathbf{x}}(f_n) \to \underline{P}^{\mathbf{x}}(f)$ and $\overline{P}^{\mathbf{x}}(f_n) \to \overline{P}^{\mathbf{x}}(f)$.

Proof. By Theorem 5.67 f_n is $\underline{P}$-previsible for every $n \in \mathbb{N}$.

We first prove that (i) implies (ii). Assume that f_n converges $\underline{P}$-hazily to f. Then (ii) follows from Theorem 5.58, if we can show that f_n is a Cauchy sequence, that is, $\lim_{n,m\to\infty} \left\|f_n - f_m\right\|_{\underline{P}}^{\mathbf{x}} = 0$.

Let $\epsilon > 0$. By Lemma 5.66 there is a $\delta_\epsilon > 0$ such that for every $A \subseteq \mathcal{X}$, if $\overline{P}^{\mathbf{x}}(A) < \delta_\epsilon$ then $\overline{P}^{\mathbf{x}}(|g|A) < \epsilon$. Since $|f_n| \leq |g|$ a.e. $\underline{P}$ for every $n \in \mathbb{N}$, it follows from Proposition 5.54 on p. 244 that also

$$\overline{P}^{\mathbf{x}}(A) < \delta_\epsilon \implies \overline{P}^{\mathbf{x}}(|f_n|A) < \epsilon \tag{5.25}$$

for every $n \in \mathbb{N}$.

Since f_n converges $\underline{P}$-hazily to f there is a N_ϵ in $\mathbb{N}$ such that for every $n \geq N_\epsilon$

$$\overline{P}(\{x \in \mathcal{X} \colon |f(x) - f_n(x)| > \epsilon/2\}) < \delta_\epsilon/2.$$

This implies that for every $n, m \geq N_\epsilon$

$$\overline{P}(\{x \in \mathcal{X} \colon |f_n(x) - f_m(x)| > \epsilon\}) < \delta_\epsilon.$$

Define $E_{nm} := \{x \in \mathcal{X} \colon |f_n(x) - f_m(x)| > \epsilon\}$, then

$$\begin{aligned} \overline{P}^{\mathbf{x}}(E_{nm}) &< \delta_\epsilon, \\ |f_n - f_m| \rfloor E_{nm} &\leq \epsilon, \end{aligned}$$

for every $n, m \geq N_\epsilon$. By Eq. (5.25) we also have that $\overline{P}^{\mathbf{x}}(|f_n|E_{nm}) < \epsilon$ and $\overline{P}^{\mathbf{x}}(|f_m|E_{nm}) < \epsilon$. for every $n, m \geq N_\epsilon$. Using the coherence of $\overline{P}^{\mathbf{x}}$, and Theorem 5.5 on p. 203 in particular, we find that

$$\begin{aligned} \overline{P}^{\mathbf{x}}(|f_n - f_m|) &\leq \overline{P}^{\mathbf{x}}(|f_n - f_m|E_{nm}) + \overline{P}^{\mathbf{x}}(|f_n - f_m| \rfloor E_{nm}) \\ &\leq \overline{P}^{\mathbf{x}}(|f_n|E_{nm}) + \overline{P}^{\mathbf{x}}(|f_m|E_{nm}) + \epsilon \\ &< \epsilon + \epsilon + \epsilon = 3\epsilon, \end{aligned}$$

for every $n, m \geq N_\epsilon$. This establishes that (i) implies (ii).

Next we prove that (ii) implies (i). Assume that f is $\underline{P}$-previsible and $\left\|f - f_n\right\|_{\underline{P}}^{\mathbf{x}}$ converges to zero. We need to prove that f_n converges $\underline{P}$-hazily to f. This means that for every $\epsilon_1, \epsilon_2 > 0$ there is a $N_{\epsilon_1,\epsilon_2}$ such that for every

$n \geq N_{\epsilon_1,\epsilon_2}$ we have that

$$\overline{P}(\{x \in \mathcal{X} \colon |f(x) - f_n(x)| > \epsilon_1\}) < \epsilon_2.$$

This will be established if, for every $\epsilon_1, \epsilon_2 > 0$, we can prove the existence of a sequence of sets E_n (which may depend on ϵ_1 and ϵ_2) such that $\overline{P}(E_n) < \epsilon_2$ and $|f - f_n| \complement E_n \leq \epsilon_1$ hold for every $n \geq N_{\epsilon_1,\epsilon_2}$. Indeed, suppose that the second inequality holds, then

$$\complement E_n \subseteq \{x \in \mathcal{X} \colon |f(x) - f_n(x)| \leq \epsilon_1\},$$

which implies that

$$E_n \supseteq \{x \in \mathcal{X} \colon |f(x) - f_n(x)| > \epsilon_1\},$$

and now use the first inequality and the monotonicity of the coherent $\overline{P}$.

Let $\epsilon_1, \epsilon_2 > 0$. Define $r := \epsilon_1/3$ and $\epsilon := \epsilon_1\epsilon_2/10$.

Define $g_n := |f - f_n|$. Since $\left\|f - f_n\right\|_{\underline{P}}^{\mathbf{x}}$ converges to zero there is an $N_\epsilon \in \mathbb{N}$ such that for every $n \geq N_\epsilon$

$$\overline{P}^{\mathbf{x}}(g_n) < \epsilon. \tag{5.26}$$

g_n is non-negative and $\underline{P}$-previsible. So it follows from Definition 5.49, Lemma 5.57, and the coherence of $\underline{P}^{\mathbf{x}}$ (Theorem 5.5(xi) on p. 203) that there are numbers M_ϵ, $K_{r,\epsilon/r} \in \mathbb{N}$, and a sequence of non-negative (use Theorem 5.51 on p. 242) gambles h_n, such that whenever $n \geq \max\{M_\epsilon, K_{r,\epsilon/r}\}$ it holds that

$$|\overline{P}^{\mathbf{x}}(g_n) - \overline{P}(h_n)| < \epsilon, \tag{5.27}$$

$$\overline{P}(\{x \in \mathcal{X} \colon |g_n(x) - h_n(x)| > r\}) < \epsilon/r. \tag{5.28}$$

Let $L_{r,\epsilon} = \max\{N_\epsilon, M_\epsilon, K_{r,\epsilon/r}\}$. Define $A_n := \{x \in \mathcal{X} \colon |g_n(x) - h_n(x)| > r\}$, then for all $n \geq L_{r,\epsilon}$ it holds that

$$\overline{P}(A_n) < \epsilon/r, \tag{5.29}$$

$$|g_n - h_n| \complement A_n \leq r. \tag{5.30}$$

Define $B_n := \{x \in \mathcal{X} \colon h_n(x) > r\}$, then for all n it holds that (recall that Z_n is

non-negative, so $B_n \leq h_n/r$)

$$\overline{P}(B_n) \leq \overline{P}(h_n)/r, \tag{5.31}$$

$$|h_n(x)| \mathbb{C} B_n \leq r. \tag{5.32}$$

Finally, define $E_n := A_n \cup B_n$ then we have for every $n \geq L_{r,\epsilon}$ that

$$\begin{aligned}\overline{P}(E_n) &\leq \overline{P}(A_n) + \overline{P}(B_n) < \epsilon/r + \overline{P}(h_n)/r < \epsilon/r + \left(\overline{P}^{\mathbf{x}}(g_n) + \epsilon\right)/r \\ &< 3\epsilon/r = \tfrac{9}{10}\epsilon_2 < \epsilon_2,\end{aligned}$$

where we subsequently used the inequalities Eq. (5.29), Eq. (5.31), Eq. (5.27) and Eq. (5.26). We also have that

$$|f - f_n| \mathbb{C} E_n = g_n \mathbb{C} E_n \leq (h_n + r) \mathbb{C} E_n \leq r + r = \tfrac{2}{3}\epsilon_1 < \epsilon_1,$$

where we subsequently used the inequalities Eq. (5.30) and Eq. (5.32). This establishes that (ii) implies (i). □

5.5.3 Previsibility by Cuts

Corollary 5.69. *Let a_n and b_n be two sequences of non-negative real numbers converging to infinity. A random quantity f is $\underline{P}$-previsible if and only if f_{a_n,b_n} is a determining sequence for f.*

Proof. If f_{a_n,b_n} is a determining sequence for f then f is $\underline{P}$-previsible by Definition 5.49. This proves sufficiency.

To prove necessity, suppose that f is $\underline{P}$-previsible, then by Theorem 5.67 f is $\underline{P}$-measurable. It follows that the sequence f_{a_n,b_n} converges $\underline{P}$-hazily to f. Also the gamble f_{a_n,b_n} is $\underline{P}$-measurable by Lemma 5.61, and $|f_{a_n,b_n}| \leq |f|$. From Theorem 5.68 it follows that

$$\begin{aligned}\lim_{n,m\to\infty} \left\|f_{a_n,b_n} - f_{a_m,b_m}\right\|_{\underline{P}}^{\mathbf{x}} &\leq \lim_{n\to\infty} \left\|f_{a_n,b_n} - f\right\|_{\underline{P}}^{\mathbf{x}} + \lim_{m\to\infty} \left\|f - f_{a_m,b_m}\right\|_{\underline{P}}^{\mathbf{x}} \\ &= 0 + 0 = 0.\end{aligned}$$

We find that f_{a_n,b_n} is a determining sequence for f. □

5.5.4 The 2-Monotone Case: A Choquet Integral for Previsible Random Variables

In this section, we generalise the representation theorem for 2-monotone lower previsions $\underline{P}$ (Proposition 4.77 on p. 180) to extended lower previsions $\underline{P}^{\mathrm{x}}$.

In Section 4.3.10, we defined the Choquet integral of an arbitrary gamble, with respect to an arbitrary 2-monotone set function ν. Let's generalise the Choquet integral to random quantities:

Definition 5.70. Let $\mathcal{F}$ be a field on $\mathcal{X}$ and let ν be a 2-monotone set function on $\mathcal{F}$. Let f be any random quantity on X. Let $G_{*\nu,f}$ be the *lower decreasing distribution function* of f with respect to ν, that is,

$$G_{*\nu,f}(z) := \nu_*(\{x \in \mathcal{X} \colon f(x) > z\}),$$

for any $z \in \mathbb{R}$.The *Choquet integral* of f with respect to ν is defined as

$$\begin{aligned} \mathrm{C}\int f \,\mathrm{d}\nu &:= \mathrm{R}\int_{-\infty}^{0} [G_{*\nu,f}(z) - 1]\,\mathrm{d}z + \mathrm{R}\int_{0}^{+\infty} G_{*\nu,f}(z)\,\mathrm{d}z \\ &= \lim_{a\to+\infty} \mathrm{R}\int_{-a}^{0} [G_{*\nu,f}(z) - 1]\,\mathrm{d}z + \lim_{b\to+\infty} \mathrm{R}\int_{0}^{b} G_{*\nu,f}(z)\,\mathrm{d}z, \end{aligned}$$

whenever the sum in the right hand side is well defined, and in such a case, we say that f is *Choquet integrable* with respect to ν.

Proof of integrability and existence of the limits. Observe that since $G_{*\nu,f}(z) - 1$ is non-increasing in z, it's Riemann integrable over $[-a, 0]$ by Proposition 4.68 on p. 167. Also, since $G_{*\nu,f}(z) - 1 \leq 0$ for all z in $\mathbb{R}$, it follows that $\mathrm{R}\int_{-a}^{0}[G_{*\nu,f}(z) - 1]\,\mathrm{d}z$ is non-increasing in a: thus, the limit of $\mathrm{R}\int_{-a}^{0}[G_{*\nu,f}(z) - 1]\,\mathrm{d}z$ for $a \to +\infty$ must exist (and coincides with the infimum over a).

Riemann integrability of $G_{*\nu,f}(z)$ over $[0, b]$ and existence of the limit for $b \to +\infty$ is proved in a similar way. □

Lemma 5.71. *Let $f \colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}$. Then $\lim_{a,b\to\infty} f(a, b) = f^*$ if and only if $\lim_{n\to\infty} f(a_n, b_n) = f^*$ for every two sequences a_n and b_n of non-negative real numbers converging to infinity.*

Proof. "if". Assume that $f(a, b)$ does not converge to f^* as $a, b \to \infty$. This means that there is an $\epsilon > 0$ such that for every $R > 0$ there are $a_R, b_R \geq R$ such

that $|f(a_R, b_R) - f^*| \geq \epsilon$. This holds in particular for every $R = n$, $n \in \mathbb{N}$. Hence, there are two sequences a_n and b_n such that $a_n, b_n \geq n$ and $|f(a_n, b_n) - f^*| \geq \epsilon$. Consequently, there are two sequences a_n and b_n of non-negative real numbers converging to infinity for which $f(a_n, b_n)$ does not converge to f^* as $n \to \infty$.

"only if". Let a_n and b_n be any two sequences of non-negative real numbers converging to infinity, that is, for every $R > 0$ there is an $N_R \in \mathbb{N}$ such that $n \geq N_R$ implies that $a_n, b_n \geq R$. Assuming that $\lim_{a,b\to\infty} f(a, b) = f^*$, for every $\epsilon > 0$ there is an $R_\epsilon > 0$ such that $a, b \geq R_\epsilon$ implies that $|f(a, b) - f^*| < \epsilon$. Hence, it holds in particular that $|f(a_n, b_n) - f^*| < \epsilon$ for every $n \geq N_{R_\epsilon}$, or equivalently, $\lim_{n\to\infty} f(a_n, b_n) = f^*$. □

Theorem 5.72. *Let $\underline{P}$ be a 2-monotone coherent lower prevision on $\mathcal{L}(X)$. Let ν be any 2-monotone set function, defined on a field $\mathcal{F}$, such that $\underline{P} = \underline{\mathbf{E}}_\nu$. If a random quantity f is $\underline{P}$-previsible, then f is Choquet integrable, and*

$$\underline{P}^{\mathbf{x}}(f) = \mathrm{C}\int f \,\mathrm{d}\nu.$$

Proof. By Proposition 4.77 on p. 180, there always is a 2-monotone set function ν defined on a field $\mathcal{F}$, such that $\underline{P} = \underline{\mathbf{E}}_\nu$.

Assume that f is $\underline{P}$-previsible. By Corollary 5.69, for every two sequences of non-negative real numbers converging to infinity, f_{a_n,b_n} is a determining sequence for f. It follows that

$$\underline{P}^{\mathbf{x}}(f) = \lim_{n\to\infty} \underline{P}(f_{a_n,b_n})$$

since f_{a_n,b_n} is a determining sequence for f. But, f_{a_n,b_n} is a gamble, so we already know that $\underline{P}(f_{a_n,b_n}) = \mathrm{C}\int f_{a_n,b_n} \,\mathrm{d}\nu$:

$$= \lim_{n\to\infty} \left(\mathrm{R}\int_{0\wedge\inf f_{a_n,b_n}}^{0} [G_{*\nu,f_{a_n,b_n}}(z) - 1] \,\mathrm{d}z + \mathrm{R}\int_{0}^{0\vee\sup f_{a_n,b_n}} G_{*\nu,f_{a_n,b_n}}(z) \,\mathrm{d}z \right)$$

and, since $-a_n \leq 0 \wedge \inf f_{a_n,b_n}$ and $0 \vee \sup f_{a_n,b_n} \leq b_n$, and moreover since $\mathrm{R}\int_{-a_n}^{0\wedge\inf f_{a_n,b_n}} [G_{*\nu,f_{a_n,b_n}}(z) - 1] \,\mathrm{d}z = 0$ and $\mathrm{R}\int_{0\vee\sup f_{a_n,b_n}}^{b_n} G_{*\nu,f_{a_n,b_n}}(z) \,\mathrm{d}z = 0$, we can also write, applying Proposition 4.65 on p. 167,

$$= \lim_{n\to\infty} \left(\mathrm{R}\int_{-a_n}^{0} [G_{*\nu,f_{a_n,b_n}}(z) - 1] \,\mathrm{d}z + \mathrm{R}\int_{0}^{b_n} G_{*\nu,f_{a_n,b_n}}(z) \,\mathrm{d}z \right)$$

but now, since $G_{*\nu,f_{a_n,b_n}}(z) = G_{*\nu,f}(z)$ for every z in $[-a_n, b_n)$:

$$= \lim_{n\to\infty} \left(\mathrm{R} \int_{-a_n}^{0} [G_{*\nu,f}(z) - 1]\,\mathrm{d}z + \mathrm{R} \int_{0}^{b_n} G_{*\nu,f}(z)\,\mathrm{d}z \right)$$

and, this holds for every two sequences a_n and b_n of non-negative real numbers converging to infinity. Hence, applying Lemma 5.71:

$$\begin{aligned} &= \lim_{a,b\to\infty} \left(\mathrm{R} \int_{-a}^{0} [G_{*\nu,f}(z) - 1]\,\mathrm{d}z + \mathrm{R} \int_{0}^{b} G_{*\nu,f}(z)\,\mathrm{d}z \right) \\ &= \lim_{a\to\infty} \mathrm{R} \int_{-a}^{0} [G_{*\nu,f}(z) - 1]\,\mathrm{d}z + \lim_{b\to\infty} \mathrm{R} \int_{0}^{b} G_{*\nu,f}(z)\,\mathrm{d}z \end{aligned}$$

and it also follows in particular that this sum is well defined. So, f is Choquet integrable, and

$$= \mathrm{C} \int f\,\mathrm{d}\nu.$$

The theorem is established. □

Finally, note that Choquet integrability with respect to ν, does not necessarily imply $\underline{\mathbf{E}}_\nu$-previsibility: for instance, it may happen that $\mathrm{C}\int f\,\mathrm{d}\nu$ is not finite even if f is Choquet integrable. However, *for any Choquet integrable random quantity*, we do have that

$$\mathrm{C} \int f\,\mathrm{d}\nu = \lim_{n\to\infty} \underline{\mathbf{E}}_\nu(f_{a_n,b_n}),$$

where a_n and b_n are arbitrary sequences of non-negative real numbers, and the limit on the right hand side is independent of the choice of these sequences; this is immediate from the proof of Theorem 5.72.

5.6 Lower Envelopes of Dunford Integrals

We now show how the extension $\underline{P}^{\mathbf{x}}$ of a lower prevision $\underline{P}$ is representable as the lower envelope of a set of Dunford integrals with respect to probability charges on a field $\mathcal{F}$. First, we need to prove some lemmas.

Lemma 5.73. *Let $\underline{P}$ and $\underline{Q}$ be two coherent lower prevision defined on the set $\mathscr{L}(X)$ of all gambles on X. If $\underline{Q}$ is a behavioural extension of $\underline{P}$, then the following statements hold.*

(i) Any determining sequence for a random quantity f with respect to $\underline{P}$ is also a determining sequence for f with respect to $\underline{Q}$.

(ii) If a random quantity is $\underline{P}$-previsible, then it is also $\underline{Q}$-previsible.

(iii) $\underline{Q}^{\mathrm{x}}$ is a behavioural extension of $\underline{P}^{\mathrm{x}}$.

Proof. Let f_n be a determining sequence for f with respect to $\underline{P}$. We prove that f_n is also a determining sequence for f with respect to $\underline{Q}$.

Since $\underline{Q}$ dominates $\underline{P}$ we have that $\overline{Q}(h) \leq \overline{P}(h)$ for every gamble h. It follows that for every $\epsilon > 0$ and every $n, m \in \mathbb{N}$

$$0 \leq \overline{Q}(\{x \in \mathcal{X} \colon |f - f_n| > \epsilon\}) \leq \overline{P}(\{x \in \mathcal{X} \colon |f - f_n| > \epsilon\})$$
$$0 \leq \overline{Q}(|f_n - f_m|) \leq \overline{P}(|f_n - f_m|)$$

From Definition 5.49 it follows that the right hand sides converge to zero. It follows that the left hand sides converge to zero too. We find that f_n is a determining sequence for f with respect to $\underline{Q}$.

Clearly, this implies that f is $\underline{Q}$-previsible whenever f is $\underline{P}$-previsible.

Since $\underline{Q}$ dominates $\underline{P}$, $\underline{Q}^{\mathrm{x}}$ also dominates $\underline{P}^{\mathrm{x}}$ for every $\underline{P}$-previsible random quantity f. Indeed, let f_n be a determining sequence for f with respect to $\underline{P}$. We have already proved that f_n is also a determining sequence for f with respect to $\underline{Q}$. Since $\underline{Q}(f_n) \geq \underline{P}(f_n)$ for every $n \in \mathbb{N}$ we find that $\lim_{n\to\infty} \underline{Q}(f_n) \geq \lim_{n\to\infty} \underline{P}(f_n)$, since the limits on both sides exist. It follows that $\underline{Q}^{\mathrm{x}}(f) \geq \underline{P}^{\mathrm{x}}(f)$ for every $\underline{P}$-previsible random quantity f. □

Recall Definition 4.11 on p. 99 and Theorem 4.62 on p. 163: a gamble f is μ-integrable if and only if $\underline{\mathbf{E}}_\mu(f) = \overline{\mathbf{E}}_\mu(f)$, if and only if f is Dunford integrable with respect to μ, if and only if f is S-integrable.

Lemma 5.74. *Let $\mathcal{F}$ be a field on $\mathcal{X}$, let μ be a probability charge on $\mathcal{F}$, and let f be any random quantity on X. Then the following statements are equivalent.*

(i) f is Dunford integrable with respect to μ.

(ii) f is $\underline{\mathbf{E}}_\mu$-previsible and has a determining sequence f_n with respect to $\underline{\mathbf{E}}_\mu$ consisting of μ-integrable gambles only.

(iii) *f is $\underline{\mathbf{E}}_\mu$-previsible and has a determining sequence f_n with respect to $\underline{\mathbf{E}}_\mu$ consisting of $\mathcal{F}$-measurable gambles only.*

(iv) *f is $\underline{\mathbf{E}}_\mu$-previsible and has a determining sequence f_n with respect to $\underline{\mathbf{E}}_\mu$ consisting of $\mathcal{F}$-simple gambles only.*

If any (and hence, all) of these conditions is satisfied, then

$$\underline{\mathbf{E}}_\mu^{\mathrm{x}}(f) = \mathrm{D}\int f\,\mathrm{d}\mu = \lim_{n\to\infty}\mathbf{E}_\mu(f_n) = \lim_{n\to\infty}\mathrm{D}\int f_n\,\mathrm{d}\mu.$$

Proof. (iv) $\Longrightarrow$ (iii) $\Longrightarrow$ (ii). Immediate from Proposition 4.28 on p. 112.

(ii) $\Longrightarrow$ (i). Since f_n is a determining sequence for f with respect to $\underline{P}$, it holds that $\overline{\mathbf{E}}_\mu(|f_n - f_m|) \to 0$, and, for all $\epsilon > 0$, $\overline{\mathbf{E}}_\mu(\{x \in \mathcal{X}\colon |f_n - f| > \epsilon\}) \to 0$.

The first condition can be written as $\mathrm{D}\int|f_n - f_m|\,\mathrm{d}\mu \to 0$; indeed, by Theorem 4.42 on p. 130 we have that $\underline{\mathbf{E}}_\mu = \mathrm{S}\underline{\int}\bullet\,\mathrm{d}\mu$, by Theorem 4.41 on p. 129 it follows that $\mathrm{S}\underline{\int}\bullet\,\mathrm{d}\mu$ is a 2-monotone coherent lower prevision, and so by Proposition 4.18(vi), it follows that the domain of $\mathbf{E}_\mu = \mathrm{S}\int\bullet\,\mathrm{d}\mu$ is a linear lattice: thus, for any n and m in $\mathbb{N}$, we find that $|f_n - f_m|$ belongs to $\operatorname{dom}\mathbf{E}_\mu$. Now, by Theorem 4.62 on p. 163, the S-integral and the D-integral coincide on $\operatorname{dom}\mathbf{E}_\mu$, so, indeed $\overline{\mathbf{E}}_\mu(|f_n - f_m|) = \mathrm{D}\int|f_n - f_m|\,\mathrm{d}\mu \to 0$.

By Theorem 4.36(v) on p. 117, the second condition, $\overline{\mathbf{E}}_\mu(\{x \in \mathcal{X}\colon |f_n - f| > \epsilon\}) \to 0$, can be written as $\mu^*(\{x \in \mathcal{X}\colon |f_n - f| > \epsilon\}) \to 0$.

Now use the fact that the f_n are μ-integrable, or equivalently, Dunford integrable with respect to μ, and apply Theorem 4.64 on p. 166 to see that f is Dunford integrable with respect to μ, with Dunford integral

$$\mathrm{D}\int f\,\mathrm{d}\mu = \lim_{n\to\infty}\mathrm{D}\int f_n\,\mathrm{d}\mu = \lim_{n\to\infty}\mathbf{E}_\mu(f_n) = \lim_{n\to\infty}\underline{\mathbf{E}}_\mu(f_n) = \underline{\mathbf{E}}_\mu^{\mathrm{x}}(f),$$

where the last equality follows from the fact that f is $\underline{\mathbf{E}}_\mu$-previsible, with determining sequence f_n.

(i) $\Longrightarrow$ (iv). If f is Dunford integrable, then there is a sequence of $\mathcal{F}$-simple gambles f_n such that $\mathrm{D}\int|f_n - f_m|\,\mathrm{d}\mu \to 0$, and, for all $\epsilon > 0$, $\mu^*(\{x \in \mathcal{X}\colon |f(x) - f_n(x)| > \epsilon\}) \to 0$. These conditions can be written as $\overline{\mathbf{E}}_\mu(|f_n - f_m|) \to 0$, and, for all $\epsilon > 0$, $\overline{\mathbf{E}}_\mu(\{x \in \mathcal{X}\colon |f(x) - f_n(x)| > \epsilon\}) \to 0$, so, f_n is a determining sequence for f with respect to $\underline{\mathbf{E}}_\mu$. □

Lemma 5.75. *Let $\mathcal{F}$ be a field on $\mathcal{X}$, and let $\underline{Q}$ be a lower prevision that avoids sure loss, and whose domain consists of $\mathcal{F}$-measurable gambles only, i.e.,* $\operatorname{dom}\underline{Q} \subseteq \mathcal{L}_{\mathcal{F}}(X)$. *Let f be an $\underline{\mathbf{E}}_{\underline{Q}}$-previsible random quantity that has a determining sequence f_n with respect to $\underline{\mathbf{E}}_{\underline{Q}}$ consisting of $\mathcal{F}$-simple gambles only. Then the following statements hold.*

(i) f_n is a determining sequence for f with respect to every probability charge μ in $\mathbf{m}^{\mathcal{F}}_{\underline{Q}}$, and hence, f is Dunford integrable with respect to every μ in $\mathbf{m}^{\mathcal{F}}_{\underline{Q}}$.

(ii) For every linear behavioural extension R of $\underline{\mathbf{E}}^{\mathbf{x}}_{\underline{Q}}$, there is a probability charge μ in $\mathbf{m}^{\mathcal{F}}_{\underline{Q}}$ such that $R(f) = \mathrm{D}\int f\,\mathrm{d}\mu$.

(iii) Conversely, for every probability charge μ in $\mathbf{m}^{\mathcal{F}}_{\underline{Q}}$, there is a linear behavioural extension R of $\underline{\mathbf{E}}^{\mathbf{x}}_{\underline{Q}}$ such that $\mathrm{D}\int f\,\mathrm{d}\mu = R(f)$.

Proof. (i). Let $\mu \in \mathbf{m}^{\mathcal{F}}_{\underline{Q}}$. By definition (see Eq. (4.55) on p. 194), $\underline{\mathbf{E}}_{\mu}$ is a behavioural extension of $\underline{Q}$, and therefore also of $\underline{\mathbf{E}}_{\underline{Q}}$, so the claim follows from Lemma 5.73(i), Lemma 5.74, and the fact that each f_n is $\mathcal{F}$-simple.

(ii). Let R be any linear behavioural extension of $\underline{\mathbf{E}}^{\mathbf{x}}_{\underline{Q}}$. Define the probability charge μ on $\mathcal{F}$ by $\mu(A) := R(I_A)$ for all A in $\mathcal{F}$. First, we show that μ belongs to $\mathbf{m}^{\mathcal{F}}_{\underline{Q}}$ (defined in Eq. (4.55) on p. 194): $R|_{\mathcal{L}_{\mathcal{F}}(X)}$ is a coherent behavioural extension of the linear prevision $\mathbf{P}_{\mu}$, and so, by Proposition 4.14 on p. 100, since $\mathcal{L}_{\mathcal{F}}(X) \subseteq \mathbf{E}_{\mu}$, for any $\mathcal{F}$-measurable gamble g it follows that

$$\underline{\mathbf{E}}_{\mu}(g) = \underline{\mathbf{E}}_{R|_{\mathcal{L}_{\mathcal{F}}(X)}}(g)$$

but $R|_{\mathcal{L}_{\mathcal{F}}(X)}$ is coherent, and therefore coincides with its natural extension on its domain,

$$= R(g)$$

but, by assumption, R is a behavioural extension of $\underline{\mathbf{E}}^{\mathbf{x}}_{\underline{Q}}$, so

$$\geq \underline{\mathbf{E}}^{\mathbf{x}}_{\underline{Q}}(g)$$

and so, if g is in $\operatorname{dom}\underline{Q} \subseteq \mathcal{L}_{\mathcal{F}}(X)$, we may applying Theorem 5.42 on p. 233 and Theorem 5.50 on p. 242 to find that

$$= \underline{\mathbf{E}}_{\underline{Q}}(g).$$

So, $\underline{\mathbf{E}}_{\mu}$ is a behavioural extension of $\underline{\mathbf{E}}_{\underline{Q}}\Big|_{\mathcal{L}_{\mathcal{F}}(X)}$. And since $\operatorname{dom}\underline{Q}$ consists of $\mathcal{F}$-measurable gambles only, it follows that $\underline{\mathbf{E}}_{\mu}$ is a behavioural extension of $\underline{Q}$, and therefore, μ belongs to $\mathbf{m}_{\underline{Q}}^{\mathcal{F}}$.

By (i), it follows that f is Dunford integrable with respect to μ, with determining sequence f_n. It remains to prove that $R(f) = \mathrm{D}\int f\,\mathrm{d}\mu$.

Indeed, since each f_n is $\mathcal{F}$-simple, we may write $f_n = \sum_{j=1}^{m_n} a_{j,n} I_{A_{j,n}}$, with $a_{1,n}, \ldots, a_{m_n,n}$ in $\mathbb{R}$ and $A_{1,n}, \ldots, A_{m_n,n}$ in $\mathcal{F}$, and it easily follows that $R(f_n) = \mathrm{D}\int f_n\,\mathrm{d}\mu$ using the linearity of R and Definition 3.19 on p. 65. Since, by (i), f_n is a determining sequence for μ, it follows that $\mathrm{D}\int f_n\,\mathrm{d}\mu \to \mathrm{D}\int f\,\mathrm{d}\mu$. Can we also show that $R(f_n) \to R(f)$? Of course: by Lemma 5.57 on p. 245, it follows that $\overline{\mathbf{E}}_{\underline{Q}}^{\mathbf{x}}(|f - f_n|) \to 0$; but, $\overline{\mathbf{E}}_{\underline{Q}}^{\mathbf{x}}(|f - f_n|) \geq R(|f - f_n|)$ for all $n \in \mathbb{N}$, so, it must hold that $R(|f - f_n|) \to 0$ as well. By the coherence of R (see Theorem 5.5(xi) on p. 203), we find that $R(f) = \lim_{n\to\infty} R(f_n)$.

(iii). Conversely, let μ be any probability charge in $\mathbf{m}_{\underline{Q}}^{\mathcal{F}}$. Define the extended lower prevision

$$\underline{S}(g) := \sup\left\{\underline{\mathbf{E}}_{\underline{Q}}^{\mathbf{x}}(g-h) + \mathrm{D}\int h\,\mathrm{d}\mu\colon \right.$$
$$\left. h \in \operatorname{dom}\underline{\mathbf{E}}_{\underline{Q}}^{\mathbf{x}},\ h \text{ Dunford integrable with respect to } \mu\right\}$$

for all random quantities g in $\operatorname{dom}\underline{\mathbf{E}}_{\underline{Q}}^{\mathbf{x}}$. It is easy to show that $\underline{S}$ is a coherent behavioural extension of $\underline{\mathbf{E}}_{\underline{Q}}^{\mathbf{x}}$. Clearly, it is a behavioural extension of $\underline{\mathbf{E}}_{\underline{Q}}^{\mathbf{x}}$: choose $h = 0$. We are left to show that $\underline{S}$ is coherent. Clearly, $\underline{S}(g_1 + g_2) \geq \underline{S}(g_1) + \underline{S}(g_2)$, and $\underline{S}(\lambda g) = \lambda\underline{S}(g)$ whenever $\lambda > 0$, follow easily from the definition of $\underline{S}$, simply by using the coherence of $\underline{\mathbf{E}}_{\underline{Q}}^{\mathbf{x}}$. It remains to prove that $\underline{S}(0) = 0$. Using Lemma 5.74, it follows that $\mathrm{D}\int h\,\mathrm{d}\mu = \underline{\mathbf{E}}_{\mu}^{\mathbf{x}}(h)$ for every h that is Dunford integrable with respect to μ. Also, by Lemma 5.73, and the fact that $\underline{\mathbf{E}}_{\mu}$ is a behavioural extension of $\underline{\mathbf{E}}_{\underline{Q}}$ (since μ belongs to $\mathbf{m}_{\underline{Q}}^{\mathcal{F}}$), it also

follows that $\overline{\mathbf{E}}^{\mathbf{x}}_{\mu}(h) \leq \overline{\mathbf{E}}^{\mathbf{x}}_{\underline{Q}}(h)$ for any such h. So,

$$\begin{aligned}
\underline{S}(0) &= \sup\left\{\underline{\mathbf{E}}^{\mathbf{x}}_{\underline{Q}}(0-h) + \mathrm{D}\int h\,\mathrm{d}\mu\colon \right. \\
&\qquad \left. h \in \operatorname{dom}\underline{\mathbf{E}}^{\mathbf{x}}_{\underline{Q}},\ h \text{ Dunford integrable with respect to } \mu\right\} \\
&\leq \sup\left\{\underline{\mathbf{E}}^{\mathbf{x}}_{\underline{Q}}(0-h) + \overline{\mathbf{E}}^{\mathbf{x}}_{\underline{Q}}(h)\colon \right. \\
&\qquad \left. h \in \operatorname{dom}\underline{\mathbf{E}}^{\mathbf{x}}_{\underline{Q}},\ h \text{ Dunford integrable with respect to } \mu\right\} \\
&= 0,
\end{aligned}$$

and since $\underline{S}$ is a behavioural extension of $\underline{\mathbf{E}}^{\mathbf{x}}_{\underline{Q}}$, it follows also that $\underline{S}(0) \geq 0$, so, $\underline{S}(0) = 0$. We conclude that $\underline{S}$ is a coherent behavioural extension of $\underline{\mathbf{E}}^{\mathbf{x}}_{\underline{Q}}$.

Now, since f is Dunford integrable with respect to μ (by (i)), it also follows that $\underline{S}(f) \geq \mathrm{D}\int f\,\mathrm{d}\mu$ (choose $h = f$), and

$$\begin{aligned}
\overline{S}(f) = -\underline{S}(-f) &= -\sup\left\{\underline{\mathbf{E}}^{\mathbf{x}}_{\underline{Q}}(-f-h) + \mathrm{D}\int h\,\mathrm{d}\mu\colon \right. \\
&\qquad \left. h \in \operatorname{dom}\underline{\mathbf{E}}^{\mathbf{x}}_{\underline{Q}},\ h \text{ Dunford integrable with respect to } \mu\right\} \\
&= \inf\left\{\overline{\mathbf{E}}^{\mathbf{x}}_{\underline{Q}}(f+h) - \mathrm{D}\int h\,\mathrm{d}\mu\colon \right. \\
&\qquad \left. h \in \operatorname{dom}\underline{\mathbf{E}}^{\mathbf{x}}_{\underline{Q}},\ h \text{ Dunford integrable with respect to } \mu\right\} \\
&\leq \mathrm{D}\int f\,\mathrm{d}\mu
\end{aligned}$$

(choose $h = -f$). So, $\underline{S}(f) = \overline{S}(f) = \mathrm{D}\int f\,\mathrm{d}\mu$.

Concluding, by Theorem 5.21(i) on p. 220, it follows that $\mathbf{M}_S^{\mathcal{K}_S} \neq \emptyset$; take any R in $\mathbf{M}_S^{\mathcal{K}_S}$. But, since $\underline{S}$ is a behavioural extension of $\underline{\mathbf{E}}^{\mathbf{x}}_{\underline{Q}}$, it follows that $\mathcal{K}_{\underline{\mathbf{E}}^{\mathbf{x}}_{\underline{Q}}} \subseteq \mathcal{K}_{\underline{S}}$, and hence, we find that R is indeed a linear behavioural extension of $\underline{\mathbf{E}}^{\mathbf{x}}_{\underline{Q}}$. Finally, since $\underline{S}(f) \leq R(f) \leq \overline{S}(f)$ by construction of R, we find that $R(f) = \mathrm{D}\int f\,\mathrm{d}\mu$. □

Here's the icing on the cake.

Theorem 5.76. *Let $\mathcal{F}$ be a field on $\mathcal{X}$, and let $\underline{Q}$ be a lower prevision that avoids sure loss, and whose domain consists of $\mathcal{F}$-measurable gambles only, i.e.,* $\operatorname{dom} \underline{Q} \subseteq \mathcal{L}_{\mathcal{F}}(X)$. *Then, for every $\underline{\mathbf{E}}_{\underline{Q}}$-previsible random quantity f that has a determining sequence with respect to $\underline{\mathbf{E}}_{\underline{Q}}$ consisting of $\mathcal{F}$-measurable gambles only, it holds that*

$$\underline{\mathbf{E}}^{\mathbf{x}}_{\underline{Q}}(f) = \min_{\mu \in \mathbf{m}^{\mathcal{F}}_{\underline{Q}}} \mathrm{D} \int f \,\mathrm{d}\mu.$$

Proof. Note that f is Dunford integrable with respect to all μ in $\mathbf{m}^{\mathcal{F}}_{\underline{Q}}$, by Lemma 5.74. Also note that $\operatorname{dom} \underline{\mathbf{E}}^{\mathbf{x}}_{\underline{Q}} \subseteq \mathcal{K}_{\underline{\mathbf{E}}^{\mathbf{x}}_{\underline{Q}}}$, since $\underline{\mathbf{E}}^{\mathbf{x}}_{\underline{Q}}$ is a real-valued coherent lower prevision. Applying Theorem 5.21(iii) on p. 220, it follows that

$$\underline{\mathbf{E}}^{\mathbf{x}}_{\underline{Q}}(f) = \min_{R \in \mathbf{M}_{\underline{\mathbf{E}}^{\mathbf{x}}_{\underline{Q}}}} \min_{\operatorname{dom} \underline{\mathbf{E}}^{\mathbf{x}}_{\underline{Q}}} R(f)$$

and by by Lemma 5.74, the sequence f_n may assumed to be $\mathcal{F}$-simple instead of $\mathcal{F}$-measurable. Applying Lemma 5.75,

$$= \min_{\mu \in \mathbf{m}^{\mathcal{F}}_{\underline{Q}}} \mathrm{D} \int f \,\mathrm{d}\mu.$$

□

Chapter 6

Optimality under Uncertainty

Let us consider a system, to which we may apply an action a, freely chosen from a set A of available actions. We want to find the *optimal actions* from A, *i.e.*, those actions that perform best according to some criterion. For instance, it is often assumed that each action induces a real-valued gain J_a: in that case, an action a^* is considered optimal in A if it induces the highest gain among all actions in A. More generally, we may wish to find the set $\operatorname{opt}(A)$ of all optimal actions in A, *i.e.*, the set of all actions that induce the highest gain.

If there is no uncertainty regarding the gains J_a, $a \in A$, then the solution to this problem is simply given by

$$\operatorname{opt}(A) = \arg\max_{a \in A} J_a.$$

Note that $\operatorname{opt}(A)$ may be empty; however, if we assume that the set $\{J_a : a \in A\}$ is a compact subset of $\mathbb{R}$—this holds if A is a finite set—then $\operatorname{opt}(A)$ contains at least one element. Secondly, note that even if $\operatorname{opt}(A)$ contains more than one action, all actions a in $\operatorname{opt}(A)$ induce the same gain J_a; so, if, in the end, the gain is all that matters, it suffices to identify only one action a^* in $\operatorname{opt}(A)$—often, this greatly simplifies the analysis.

In many situations, the gain J_a induced by an action a is influenced by variables which may not be well-known. Assuming that these variables can be modelled through a random variable X, it is customary to consider the

gain J_a as a random quantity on X; random quantities were introduced in Section 5.1. We view J_a as a real-valued gain that is a function of X, and that is expressed in a fixed utility scale; so, J_a is an $\mathcal{X}$–$\mathbb{R}$-mapping, interpreted as an uncertain gain: if x turns out to be the realisation of X, and we apply action a, then we receive an amount of utility $J_a(x)$. Which action should we choose?

6.1 A Classical Approach: Maximising Expected Utility

Let's consider a common case, where our knowledge about X is modelled through a probability charge μ on a field $\mathcal{F}$ on $\mathcal{X}$, as we discussed in Section 3.5.2 on p. 61 ff. and Section 4.3.2 on p. 109 ff. Usually, assuming that the gains J_a are bounded random quantities, the field $\mathcal{F}$ is chosen such that all gains J_a, $a \in A$, are $\mathcal{F}$-measurable; see Definition 4.25 on p. 109, and Propositions 4.26&4.27. Then, a common way to arrive at a set of optimal actions, goes through their *expected utility*:

$$P(J_a) := \mathrm{D}\int J_a \,\mathrm{d}\mu = \mathrm{S}\int J_a \,\mathrm{d}\mu = \mathbf{E}_\mu(J_a),$$

where the right hand side is given by the Dunford integral, S-integral, or $\mathbf{P}_\mu$-integral respectively; the equality of all these integrals was established in Theorem 4.62 on p. 163, and integrability of the random quantities J_a follows from their $\mathcal{F}$-measurability, *i.e.*, Proposition 4.28 on p. 112. For simplicity, we shall assume that μ is independent of a: the action a does not influence our beliefs about X. This is called *act-state independence*.

If we interpret μ as a prevision $\mathbf{P}_\mu$, then $P(J_a)$ is nothing but the prevision of J_a obtained through natural extension of $\mathbf{P}_\mu$: in this interpretation, the expected utility of J_a corresponds to the prevision of J_a, *i.e.*, the fair price for J_a, as explained on p. 48. So, as far as it makes sense to *maximise expected utility*:

$$\mathrm{opt}\,(A) := \arg\max_{a\in A} P(J_a). \tag{6.1}$$

Again, $\mathrm{opt}\,(A)$ may be empty, but if we assume for instance that the set $\{J_a : a \in A\}$ is compact with respect to the topology of uniform convergence, which holds for instance if A is finite, then $\mathrm{opt}\,(A)$ contains at least one

element. Moreover, even if opt (A) contains more than one action, all actions a in opt (A) induce the same expected gain $P(J_a)$; so, if, in the end, the expected gain is all that matters, it again suffices to identify only one action a^* in opt (A), which may simplify the analysis.

6.2 Generalising Maximal Expected Utility: Why and How?

6.2.1 The Tossing Machine

To see why maximising expected utility, *i.e.*, Eq. (6.1), is not always a desirable criterion for selecting all optimal actions from a set A of actions, consider a tossing machine, to which we may apply either one of the following actions:

- a_0: The machine tosses the coin, and whatever the outcome, we receive nothing, *i.e.*, there is neither gain nor loss.

- a_1: The machine tosses the coin, and if the outcome is heads, we receive two units of utility. Otherwise, we lose one unit of utility.

- a_2: The machine tosses the coin, and if the outcome is heads, we lose one unit of utility. Otherwise, we gain two units of utility.

In the manual of the machine, it is only stated that when taking an action in $A = \{a_0, a_1, a_2\}$, the coin will turn up either heads or tails. The gains J_{a_1} and J_{a_2} depend on this outcome, say X. Therefore, we shall consider them as random quantities on X; obviously, we can also consider J_{a_0} as a constant random quantity on X. The random variable X assumes values in $\mathcal{X} = \{H, T\}$, with H for heads and T for tails, so

$$\begin{aligned} J_{a_0}(H) &= 0, & J_{a_1}(H) &= 2, & J_{a_2}(H) &= -1, \\ J_{a_0}(T) &= 0, & J_{a_1}(T) &= -1, & J_{a_2}(T) &= 2. \end{aligned}$$

Without any further information regarding the mechanism of the tossing machine, only three solutions for opt (A) are reasonable, once we recognise

that the gains and the given information share a symmetry:

$$\begin{aligned}\mathrm{opt}_1(A) &= \{J_{a_0}\},\\ \mathrm{opt}_2(A) &= \{J_{a_1}, J_{a_2}\}, \text{ and}\\ \mathrm{opt}_3(A) &= \{J_{a_0}, J_{a_1}, J_{a_2}\}.\end{aligned}$$

Indeed, these are the only sets which reflect the symmetry of the problem: our beliefs regarding the machine's coin do not change when switching heads with tails, so J_{a_1} can be optimal if and only if J_{a_2} is optimal, because J_{a_1} transforms into J_{a_2} and *vice versa* when switching heads with tails.

Which one is the most reasonable? Should we select a_0? Should we not select a_0? Or, should we arbitrary apply any action in A? For sure, selecting a_0 is the safest choice: we are guaranteed not to lose. On the other hand, we may gain 2 units of utility applying a_1 or a_2, whereas applying a_0 we shall not gain anything at all: actions a_1 and a_2 have more potential. However, arguably, it may be criticised that we have insufficient knowledge about the machine in order to weigh the risk against the potential: therefore, we might as well consider all actions in A as optimal. These are arguments in favour of each of the three choices $\mathrm{opt}_1(A)$, $\mathrm{opt}_2(A)$, and $\mathrm{opt}_3(A)$.

On the other hand, $\mathrm{opt}_1(A)$ excludes the potential of gaining two units of utility. Moreover, as $\mathrm{opt}_1(A)$ contains one action only, it excludes the possibility of learning: we don't have much information to start from, so, since we don't have act-state independence *a priori* (it's not stated in the manual), it could be worth trying also actions that are not risk averse, in order to gain information about how we should choose our actions in the future. The set $\mathrm{opt}_2(A)$ is larger, but the actions a_1 and a_2 may be too risky: perhaps, we are not willing to lose one unit of utility, even though we possibly gain two units. Apparently, this leaves us with $\mathrm{opt}_3(A)$, the largest set of the three, as the most reasonable set of optimal actions.

Let's try and investigate how maximising expected utility solves this problem. First of all, note that $\mathcal{F} = \wp(\mathcal{X}) = \{\emptyset, \{H\}, \{T\}, \{H, T\}\}$ is the unique field such that all random quantities J_{a_0} ,J_{a_1}, and J_{a_2} are $\mathcal{F}$-measurable, and note that any probability charge μ on $\mathcal{F}$ is uniquely determined by its value,

say, on $\{H\}$: if we denote $\mu(\{H\})$ by p, then μ is given by

$$\begin{aligned} \mu(\emptyset) &= 0, & \mu(\{H,T\}) &= 1, \\ \mu(\{H\}) &= p, & \mu(\{T\}) &= 1-p. \end{aligned}$$

The machine's manual does not give us a clue about the value of p, so, in principle, we are not able to maximise expected utility; unless we hypothetically assume that there is a value $p \in [0,1]$ governing the probabilities of the machine's coin: then, applying Definition 3.19 on p. 65,

$$\begin{aligned} P(J_{a_0}) &= \mathrm{D}\int J_{a_0}\,\mathrm{d}\mu = 0, \\ P(J_{a_1}) &= \mathrm{D}\int J_{a_1}\,\mathrm{d}\mu = 3p-1, \text{ and} \\ P(J_{a_2}) &= \mathrm{D}\int J_{a_2}\,\mathrm{d}\mu = 2-3p. \end{aligned}$$

As a function of the hypothetical value $p \in [0,1]$, the set of optimal actions in A, in the sense of Eq. (6.1), maximising expected utility, is

$$\operatorname{opt}(A) = \begin{cases} \{a_2\}, & \text{if } 0 \le p < \frac{1}{2}, \\ \{a_1, a_2\}, & \text{if } p = \frac{1}{2}, \\ \{a_1\}, & \text{if } \frac{1}{2} < p \le 1. \end{cases}$$

What do we observe?

- For almost any value of p, maximising expected utility yields a unique optimal action.
- For almost any value of p, maximising expected utility yields neither $\operatorname{opt}_1(A)$, nor $\operatorname{opt}_2(A)$, nor $\operatorname{opt}_3(A)$, and hence, does not reflect the symmetry of the problem regarding heads and tails; only in the marginal case $p = \frac{1}{2}$, it yields one of the suggested solutions, namely, $\operatorname{opt}_2(A)$.
- The risk-averse action a_0 *never* belongs to the set of optimal actions: apparently, maximising expected utility tends against risk aversion.

Concluding, maximising expected utility may not be desirable, because:

- Independently of the probability charge μ, it yields almost always a

unique optimal action, which is rather surprising in case we have only little information: it cannot seriously model indecision.

- If it is not *a priori* clear how to identify μ, an (necessarily arbitrary) choice of μ may have an unidentifiable effect on what we call optimal. In particular, maximising expected utility does not incorporate robustness.

- It may not reflect the symmetry of the problem, unless we impose this symmetry on μ itself. However, often there simply is no probability charge μ reflecting the symmetry of the problem, which at the same time also reflects our beliefs regarding X. For instance, in the above example, symmetry dictates $p = \frac{1}{2}$, but nevertheless, this value of p does not correspond to the given information: we have no clue about the value of p.

- Apparently, maximising expected utility may tend against risk aversion.

6.2.2 Assumptions and Notation

Convinced of the desirability of going beyond maximising expected utility, we now discuss a number of ways to derive criteria of optimality based on coherent extended lower previsions, instead of probability charges. Of course, we wish to retain maximising expected utility as a special case, *i.e.*, when we can represent our beliefs through a probability charge, these criteria should coincide with maximising expected utility.

Therefore, we must be able to link coherent extended lower previsions with probability charges, or, more generally, with expected utility. This requires us to impose a number of technical limitations on the extended lower previsions modelling our beliefs.

Recall how, in the previous chapters, we modelled our beliefs about a random variable X through the assessment of supremum buying prices $\underline{Q}(f)$ for random quantities f in some subset $\operatorname{dom}\underline{Q}$ of the set $\mathcal{R}(X)$ of all random quantities on X; we called the mapping $\underline{Q}$ an extended lower prevision (see Section 3.2 on p. 42 ff., and Section 5.1 on p. 197 ff.). If this $\underline{Q}$ was not vulnerable to so-called Dutch book arguments, then we said that $\underline{Q}$ avoided sure loss (see Definition 5.2 on p. 200), and in this case, we explained how to derive from

$\underline{Q}$, through natural extension (see Section 5.2.3 on p. 207 ff., Definition 5.9 in particular), and convergence methods similar to Dunford integration (see Section 5.4.3 on p. 239 ff., Definition 5.49 on p. 241 in particular), a coherent extended lower prevision $\underline{P}$ defined on a much larger set $\operatorname{dom}\underline{P}$ of random quantities. Moreover, for a number of special cases, we explained how this extension $\underline{P}$ could be viewed as the lower envelope of integrals with respect to probability charges, or, more generally, as a lower envelope of linear extended previsions: the key to this result was essentially that $\underline{Q}$ had a real-valued extension $\underline{P}$ defined on a linear lattice that contained all constant random quantities; see Lemma 5.20 on p. 218.

For instance, if we start with an extended lower prevision $\underline{Q}$ that avoids sure loss, we end up with such a coherent extended lower prevision by taking

$$\underline{P} := \underline{\mathbf{E}}_{\underline{Q}}^{\mathcal{R}(X)}\Big|_{\mathcal{K}_{\underline{Q}}},$$

where $\mathcal{K}_{\underline{Q}}$ is the linear lattice of random quantities f for which $\overline{\mathbf{E}}_{\underline{Q}}^{\mathcal{R}(X)}(|f|) < +\infty$; note that $\mathcal{K}_{\underline{Q}}$ contains also all bounded random quantities, and hence, all constant random quantities. Alternatively, if we start with a lower prevision $\underline{Q}$ (*i.e.*, $\underline{Q}$ real-valued and defined on bounded random quantities only) that avoids sure loss, then we also end up with such a coherent extended lower prevision by taking

$$\underline{P} := \underline{\mathbf{E}}_{\underline{Q}}^{\mathbf{x}},$$

which is defined on the linear lattice of all $\underline{\mathbf{E}}_{\underline{Q}}$-previsible random quantities; again, all bounded random quantities belong to this linear lattice. So, ***from now on, we shall assume that $\underline{P}$ is a real-valued coherent extended lower prevision, defined on linear lattice of random quantities on X that contains at least the set $\mathbb{R}(X)$ of all constant random quantities on X, and all random quantities J_a induced by actions a in A.*** Of course, there are extended lower previsions that do not satisfy this requirement, but nevertheless, we shall impose these restrictions, as they lead to a number of very practical consequences, and as they cover all cases in which we start with an extended lower prevision $\underline{Q}$ such that $\operatorname{dom}\underline{Q} \subseteq \mathcal{K}_{\underline{Q}}$, which can be guaranteed to hold by restricting $\underline{Q}$ to be real-valued, and careful choice of the domain of $\underline{Q}$, as explained in Corollary 5.22 on p. 223; for instance, it always holds in case $\underline{Q}$ is a lower prevision.

Let's briefly summarise the consequences of this assumption.

(i) The set $\operatorname{dom}\underline{P}$ equipped with the $\underline{P}$-norm, which is defined by $\left\|f\right\|_{\underline{P}} := \overline{P}(|f|)$ for all f in $\operatorname{dom}\underline{P}$, is a topological vector space—note that $|f|$ belongs to $\operatorname{dom}\underline{P}$ whenever f does, since $\operatorname{dom}\underline{P}$ is a linear lattice. However, it turns out that the $\underline{P}$-norm topology is slightly too weak for the purpose of this work: we need to strengthen it with point-wise convergence on members of $\mathcal{X}$ (also called the weak convergence). Concretely, we shall say that a net f_α in $\operatorname{dom}\underline{P}$ converges to f in $\operatorname{dom}\underline{P}$ if

$$\begin{cases} \overline{P}(|f_\alpha - f|) \to 0, \text{ and} \\ f_\alpha(x) \to f(x), \text{ for all } x \in \mathcal{X}; \end{cases} \tag{6.2}$$

unless explicitly stated otherwise, $\operatorname{dom}\underline{P}$ ***is assumed to be endowed with the topology induced by this convergence.*** It turns $\operatorname{dom}\underline{P}$ into a locally convex topological vector space, which also happens to be Hausdorff. A topological basis at 0 consists for instance of the convex sets

$$\{f \in \operatorname{dom}\underline{P} \colon \overline{P}(|f|) < \epsilon \text{ and } f(x) < \delta(x)\},$$

for $\epsilon > 0$, and $\delta(x) > 0$ for all $x \in \mathcal{X}$. It has more open sets and more closed sets than the $\underline{P}$-norm topology and the weak topology, but it has less compact sets than the $\underline{P}$-norm topology and the weak topology. In any case, this topology is weaker than the topology of uniform convergence. Note that in case $\mathcal{X}$ is finite, it reduces to the weak topology, which is in that case also equivalent to the topology of uniform convergence.

Interestingly, if $\underline{P} = \underline{\mathbf{E}}^{\mathrm{x}}_{Q}$, then the set $\mathcal{L}(X)$ of bounded random quantities is dense in $\operatorname{dom}\underline{P}$ with respect to this topology; this is an immediate consequence of Corollary 5.69 on p. 256 (previsibility by cuts) and Lemma 5.57 on p. 245 (convergence in $\underline{P}$-norm of determining sequences).

(ii) For ease of notation, ***we shall denote set*** $\mathbf{M}_{\underline{P}}^{\operatorname{dom}\underline{P}}$ ***of linear extended previsions on*** $\operatorname{dom}\underline{P}$ ***that are behavioural extensions of*** $\underline{P}$ ***by*** $\mathcal{M}$***, or by*** $\mathcal{M}_{\underline{P}}$ ***in case we study more than one extended lower prevision at the same time.*** This set $\mathcal{M}$ is non-empty, convex, and compact with respect to the topology of point-wise convergence on members of $\operatorname{dom}\underline{P}$, and

it uniquely determines $\underline{P}$:

$$\underline{P}(f) = \min_{Q \in \mathcal{M}} Q(f),$$

for any f in $\operatorname{dom} \underline{P}$. If $\underline{P} = \underline{\mathbf{E}}^{\mathbf{x}}_{\underline{Q}}$ and all random quantities in $\operatorname{dom} \underline{P}$ have $\mathcal{F}$-measurable determining sequences with respect to $\underline{\mathbf{E}}_{\underline{Q}}$, then we even have that

$$\underline{P}(f) = \min_{\mu \in \mathbf{m}^{\mathcal{F}}_{\underline{Q}}} \mathrm{D} \int f \, \mathrm{d}\mu,$$

i.e., $\mathcal{M}$ is fully characterised by a non-empty convex compact set of probability charges $\mathbf{m}^{\mathcal{F}}_{\underline{Q}}$; see Theorem 5.76 on p. 265.

6.2.3 A Technical Lemma About Preorders

A relation that is reflexive and transitive is called a *preorder*, and such preorders are used to model preference. In this section, we prove a technical but very useful lemma about the existence of maximal elements with respect to preorders; it's an abstraction of a result proved by De Cooman and Troffaes [23].

Let $\mathcal{V}$ be any set, and let $\trianglerighteq$ be any preorder on $\mathcal{V}$. An element v of a subset $\mathcal{S}$ of $\mathcal{V}$ is called $\trianglerighteq$ *-maximal* in $\mathcal{S}$ if, for all w in $\mathcal{S}$, $w \trianglerighteq v$ implies $v \trianglerighteq w$. The set of $\trianglerighteq$ -maximal elements is denoted by

$$\max_{\trianglerighteq}(\mathcal{S}) := \left\{ v \in \mathcal{S} \colon (\forall w \in \mathcal{S})(w \trianglerighteq v \implies v \trianglerighteq w) \right\}. \tag{6.3}$$

For any v in $\mathcal{S}$, we also define the *up-set* of v relative to $\mathcal{S}$ as

$$\uparrow^{\mathcal{S}}_{\trianglerighteq} v := \{ w \in \mathcal{S} \colon w \trianglerighteq v \}.$$

Lemma 6.1. *Let $\mathcal{V}$ be a Hausdorff topological space. Let $\trianglerighteq$ be any preorder on $\mathcal{V}$ such that for any v in $\mathcal{V}$, the set $\uparrow^{\mathcal{V}}_{\trianglerighteq} v$ is closed. Then, for any non-empty compact subset $\mathcal{S}$ of $\mathcal{V}$, the following statements hold.*

(i) For every v in $\mathcal{S}$, the set $\uparrow^{\mathcal{S}}_{\trianglerighteq} v$ is non-empty and compact.

(ii) The set $\max_{\trianglerighteq}(\mathcal{S})$ of $\trianglerighteq$ -maximal elements of $\mathcal{S}$ is non-empty.

(iii) For every v in $\mathcal{S}$, there is a $\trianglerighteq$ -maximal element w of $\mathcal{S}$ such that $w \trianglerighteq v$.

Proof. (i). Since $\trianglerighteq$ is reflexive, it follows that $v \trianglerighteq v$, so $\uparrow^{\mathcal{S}}_{\trianglerighteq} v$ is non-empty. Is it compact? Clearly, $\uparrow^{\mathcal{S}}_{\trianglerighteq} v = \mathcal{S} \cap \uparrow^{\mathcal{V}}_{\trianglerighteq} v$, so $\uparrow^{\mathcal{S}}_{\trianglerighteq} v$ is the intersection of a compact set and a closed set, and therefore $\uparrow^{\mathcal{S}}_{\trianglerighteq} v$ must be compact too.

(ii). Let $\mathcal{S}'$ be any subset of the non-empty compact set $\mathcal{S}$ that is linearly ordered with respect to $\trianglerighteq$. If we can show that $\mathcal{S}'$ has an upper bound in $\mathcal{S}$ with respect to $\trianglerighteq$, then we can infer from a version of Zorn's lemma [70, (AC7), p. 144] (which also holds for preorders) that $\mathcal{S}$ has a $\trianglerighteq$ -maximal element. Let then $\{v_1, v_2, \ldots, v_n\}$ be an arbitrary finite subset of $\mathcal{S}'$. We can assume without loss of generality that $v_1 \trianglerighteq v_2 \trianglerighteq \cdots \trianglerighteq v_n$, and consequently $\uparrow^{\mathcal{S}}_{\trianglerighteq} v_1 \subseteq \uparrow^{\mathcal{S}}_{\trianglerighteq} v_2 \subseteq \cdots \subseteq \uparrow^{\mathcal{S}}_{\trianglerighteq} v_n$. This implies that the intersection $\bigcap_{k=1}^{n} \uparrow^{\mathcal{S}}_{\trianglerighteq} v_k = \uparrow^{\mathcal{S}}_{\trianglerighteq} v_1$ of these up-sets is non-empty: the collection $\left\{\uparrow^{\mathcal{S}}_{\trianglerighteq} v \colon v \in \mathcal{S}'\right\}$ of compact and hence closed ($\mathcal{V}$ is Hausdorff) subsets of $\mathcal{S}$ has the finite intersection property. Consequently, since $\mathcal{S}$ is compact, the intersection $\bigcap_{v \in \mathcal{S}'} \uparrow^{\mathcal{S}}_{\trianglerighteq} v$ is non-empty as well, and this is the set of upper bounds of $\mathcal{S}'$ in $\mathcal{S}$ with respect to $\trianglerighteq$. So, by Zorn's lemma, $\mathcal{S}$ has a $\trianglerighteq$ -maximal element: $\max_{\trianglerighteq}(\mathcal{S})$ is non-empty.

(iii). Combine (i) and (ii) to show that the non-empty compact set $\uparrow^{\mathcal{S}}_{\trianglerighteq} v$ has a maximal element w with respect to $\trianglerighteq$. It is then a trivial step to prove that w is also $\trianglerighteq$ -maximal in $\mathcal{S}$: we must show that for any u in $\mathcal{S}$, if $u \trianglerighteq w$, then $w \trianglerighteq u$. But, if $u \trianglerighteq w$, then also $u \trianglerighteq v$ since $w \trianglerighteq v$ by construction. Hence, $u \in \uparrow^{\mathcal{S}}_{\trianglerighteq} v$, and since w is $\trianglerighteq$ -maximal in $\uparrow^{\mathcal{S}}_{\trianglerighteq} v$, it follows that $w \trianglerighteq u$. □

6.3 $\underline{P}$-Maximality

6.3.1 Pair-Wise Choice

Maximality is a criterion of optimality based on pair-wise choice. Let's briefly explain the basics of pair-wise choice, and how it leads to optimality.

We model pair-wise choice by a relation R on A: for a pair of actions a and b in A, we say that aRb, if we are disposed to choose a whenever we have the choice between only a and b. Clearly, R is a relation on A. What properties must R satisfy? If we are presented with only a single action a in A, then we have no choice but to choose a. So, for any a in A, it always holds that aRa: R is reflexive. If we are presented with only two actions a and b in A, we have no choice but to choose a, or to choose b. So, for any actions a and b in A, it always holds that aRb or bRa: R is complete.

Definition 6.2. A complete and reflexive relation R on A is called a *choice*

relation.

So, a choice relation determines optimality on pairs: it tells us which actions we may choose from a given pair of actions. But, how does a choice relation lead to optimality when the set of actions contains more than two elements?

Looking at optimality through pair-wise choice can be traced back to at least Condorcet [16]—his method is perhaps most clearly explained in his voting examples; see for instance Condorcet [16, pp. lvj–lxix, 4.^{e} Exemple]. More recently, Sen [73] has studied optimality based on pair-wise choice under very general assumptions. In agreement with Condorcet's method, given a complete and reflexive relation R on A—a choice relation—we should select (see Sen [73, p. 55, Eq. (1)])

$$\operatorname{opt}_R(A) := \{a \in A \colon (\forall b \in A)(aRb)\}. \tag{6.4}$$

This simply means that we are disposed to choose a from A, whenever, for all b in A, we are disposed to choose a from $\{a, b\}$. In this way, pair-wise choice, *i.e.*, a notion of optimality on pairs, leads to a notion of optimality on larger sets. The mapping $\operatorname{opt}_R(\bullet)$, as a mapping from subsets of A to subsets of A, is called a social choice function, if $\operatorname{opt}_R(B)$ is non-empty for every non-empty subset B of A:

Definition 6.3. A mapping $\operatorname{opt}(\bullet) : \wp(A) \to \wp(A)$, satisfying $\operatorname{opt}(B) \subseteq B$ for all $B \subseteq A$, and $\operatorname{opt}(B) \neq \emptyset$ whenever $B \neq \emptyset$, is called a *social choice function*.

Of course, as we already noted in Section 3.2 on p. 42 ff., it is an important observation that not all reasonable social choice functions are representable by Eq. (6.4), *i.e.*, by choice relations.

Sen [73] extensively studied the interplay between general social choice functions $\operatorname{opt}(\bullet)$, choice relations R, and social choice functions $\operatorname{opt}_R(\bullet)$ induced by such relations R, as in Eq. (6.4). As lower previsions naturally lead to pair-wise preference, Walley [86] inferred an optimality criterion based on lower previsions, essentially invoking Eq. (6.4). Let's briefly explain how this works.

Suppose we are given a *strict partial order*, *i.e.*, a transitive and anti-reflexive relation, $>$ on A, which models strict preference on A: we say that $a > b$, whenever we strictly prefer a over b. In terms of pair-wise choice, $a > b$ means that we shall never choose b from $\{a, b\}$. In effect, from $\{a, b\}$, we shall

only choose a. So, any choice relation $R_>$ that is compatible with the choices implied by $>$, must satisfy that $a > b$ if and only if $aR_>b$ and not $bR_>a$. But, this property uniquely determines $R_>$.

Indeed, since $R_>$ is a choice relation, it must be complete: it always holds that $aR_>b$ or $bR_>a$. So, not $bR_>a$ implies $aR_>b$. Therefore, $aR_>b$ and not $bR_>a$, is equivalent to not $bR_>a$. Hence $a > b$ is equivalent to not $bR_>a$, or, equivalently, $a \not> b$ is equivalent to $bR_>a$: a is not strictly preferred to b, if and only if we are disposed to choose b from $\{a, b\}$. So, within Sen's framework [73, p. 55, Eq. (1)], given a notion of strict preference $>$, the only way to end up with a set of optimal actions is by choosing R equal to $\not>$:

$$\operatorname{opt}_{\not>}(A) := \{a \in A \colon (\forall b \in A)(a \not> b)\}.$$

Since $>$ is a strict partial order, it follows that $\not>$ is complete and reflexive, and hence, a choice relation. Moreover, $\not>$ is acyclic, so, as Sen [73, p. 55] remarks, it follows that $\operatorname{opt}_{\not>}(A)$ is non-empty whenever A is finite. In particular, $\operatorname{opt}_{\not>}(\bullet)$ is a social choice function on finite sets. We shall call actions in $\operatorname{opt}_{\not>}(A)$ *$>$-maximal* because they correspond to the elements of A that are maximal with respect to the strict partial order $>$, *i.e.*, they correspond to those actions in A for which there is no action in A that is strictly preferred to it:

$$\operatorname{opt}_{\not>}(A) = \max_>(A) := \{a \in A \colon (\forall b \in A)(a \not> b)\}. \tag{6.5}$$

Now, following Walley [86, Sections 3.7–3.9], we can easily derive a strict partial order from the coherent extended lower prevision $\underline{P}$, which has the interpretation of a strict preference. Recall that the extended lower prevision $\underline{P}(f)$ of a random quantity f has a behavioural interpretation as a supremum acceptable price for buying f: $\underline{P}(f)$ is the highest real number $s \in \mathbb{R}$ such that for any price $t \in \mathbb{R}$ that is strictly lower than s, we are willing to pay t prior to observing X, if we are guaranteed to receive $f(x)$ once $X = x$ has been observed. This allows us to define a strict partial order $>_{\underline{P}}$ on $\operatorname{dom}\underline{P}$ whose interpretation is that of strict preference. We also introduce a preorder $\geqslant_{\underline{P}}$, which satisfies $\max_{>_{\underline{P}}}(\bullet) = \max_{\geqslant_{\underline{P}}}(\bullet)$; recall that we defined maximality for preorders in Eq. (6.3) on p. 275, and maximality for strict partial orders in Eq. (6.5) on p. 278.

Definition 6.4. For any two random quantities f and g in $\operatorname{dom}\underline{P}$, we say that

f is strictly preferred to g with respect to $\underline{P}$, and write $f >_{\underline{P}} g$, if

$$\underline{P}(f - g) > 0 \text{ or } (f \geq g \text{ and } f \neq g).$$

We say that *f is preferred to g with respect to $\underline{P}$*, and we write $f \geqslant_{\underline{P}} g$, if

$$\underline{P}(f - g) > 0 \text{ or } f \geq g.$$

Indeed, assuming act-state independence, if $f \geq g$ and $f \neq g$ then we should strictly prefer f to g, since f can only induce a higher gain than g. On the other hand, $\underline{P}(f - g) > 0$ expresses that we are willing to pay some strictly positive price to exchange g for f, which again means that we strictly prefer f to g. The preference relation $\geqslant_{\underline{P}}$ is a preorder: it is reflexive and transitive. Moreover, $\geqslant_{\underline{P}}$ is the weakest preorder—*i.e.*, comparing the fewest pairs—that agrees with $>_{\underline{P}}$: $f \geqslant_{\underline{P}} g$ if and only if either $f >_{\underline{P}} g$ or $f = g$. Also, $\geqslant_{\underline{P}}$ is anti-symmetric: if $f \geqslant_{\underline{P}} g$ and $g \geqslant_{\underline{P}} f$ then $f = g$. A reflexive, transitive, and anti-symmetric relation is called a *partial order*, so, $\geqslant_{\underline{P}}$ is actually a partial order on $\operatorname{dom} \underline{P}$.

It is clear that we can also use the coherent extended lower prevision $\underline{P}$ to express a strict preference between any two actions a and b in A, based on their gains J_a and J_b: if $J_a >_{\underline{P}} J_b$, then the uncertain gain J_a is strictly preferred to the uncertain gain J_b, and therefore the action a should also be strictly preferred to b. In such a case, we may also write $a >_{\underline{P}} b$.

Similarly, we shall write $a \geqslant_{\underline{P}} b$ if $J_a \geqslant_{\underline{P}} J_b$, and $a \geq b$ if $J_a \geq J_b$. Note that, whereas $\geqslant_{\underline{P}}$ and $\geq$ are partial orders on $\operatorname{dom} \underline{P}$, they may only be preorders on A, since it may happen that $J_a = J_b$ for two different actions a and b in A.

The relation $>_{\underline{P}}$ is anti-reflexive and transitive: this follows from the coherence of $\underline{P}$ (in particular, $\underline{P}(f - f) = 0$ and $\underline{P}(h - f) \geq \underline{P}(h - g) + \underline{P}(g - f)$). So, it is indeed a strict partial order on $\operatorname{dom} \underline{P}$, and therefore also on A. Hence, as we explained above, this leads to the following criterion of optimality (see Walley [86, Section 3.9.2, p. 161]):

Definition 6.5. An action a in A is called *$\underline{P}$-maximal* in A if no action in A is strictly preferred to a with respect to $>_{\underline{P}}$:

$$\operatorname{opt}_{\not<_{\underline{P}}}(A) = \max_{>_{\underline{P}}}(A) = \max_{\geqslant_{\underline{P}}}(A) = \{a \in A \colon (\forall b \in A)(a \not<_{\underline{P}} b)\}.$$

Proof of equality. All equalities are immediate consequences of the definitions

of $\operatorname{opt}_{\not<\underline{P}}(A)$ and $\max_{>\underline{P}}(A)$, except for the equality $\max_{\geqslant\underline{P}}(A) = \{a \in A\colon (\forall b \in A)(a \not<\underline{P} b)\}$. Indeed, since for any a and b in A, it holds that $a \not<\underline{P} b$, if and only if $J_a \not<\underline{P} J_b$, if and only if $J_a \not\leqslant\underline{P} J_b$ or $J_a = J_b$, if and only if $J_b \geqslant\underline{P} J_a$ implies $J_a = J_b$, if and only if—using the fact that $\geqslant\underline{P}$ is a partial order on $\operatorname{dom}\underline{P}$—$J_b \geqslant\underline{P} J_a$ implies $J_a \geqslant\underline{P} J_b$, if and only if $b \geqslant\underline{P} a$ implies $a \geqslant\underline{P} b$. So, by Eq. (6.3) on p. 275, it follows that that $\max_{\geqslant\underline{P}}(A) = \{a \in A\colon (\forall b \in A)(a \not<\underline{P} b)\}$. □

It's convenient to denote by $\mathcal{J}_A$ the set of gain random quantities induced by actions in A,

$$\mathcal{J}_A := \{J_a\colon a \in A\}.$$

The $\underline{P}$-maximal actions in A are precisely those actions whose induced gain is a maximal element of $\mathcal{J}_A$ with respect to the strict partial order $>\underline{P}$, or equivalently, with respect to the partial order $\geqslant\underline{P}$, so

$$a \in \operatorname{opt}_{\not<\underline{P}}(A) \iff J_a \in \operatorname{opt}_{\not<\underline{P}}(\mathcal{J}_A)\,.$$

Let's now give a number of important properties of $\underline{P}$-maximality.

6.3.2 Monotonicity

The more determinate our beliefs, the smaller the set of $\underline{P}$-maximal actions:

Theorem 6.6. *If $\underline{Q}$ is a behavioural extension of $\underline{P}$, then* $\max_{>\underline{Q}}(A) \subseteq \max_{>\underline{P}}(A)$.

Proof. Suppose that a is $\underline{Q}$-maximal in A. Then, for all b in A, it holds that $a \not<\underline{Q} b$, *i.e.*, $\underline{Q}(J_b - J_a) \leq 0$, and $J_b \not\geqslant J_a$ or $J_b = J_a$. But, if $\underline{Q}(J_b - J_a) \leq 0$, then it must also hold that $\underline{P}(J_b - J_a) \leq 0$, since that $\underline{Q}$ is a behavioural extension of $\underline{P}$. So, for all b in A, it holds that $\underline{P}(J_b - J_a) \leq 0$, and $J_b \not\geqslant J_a$ or $J_b = J_a$, *i.e.*, $a \not<\underline{P} b$: a must be $\underline{P}$-maximal as well. □

6.3.3 $\underline{P}$-Maximality Through Point-Wise Maximality

In this section, we derive a simple way to find $\underline{P}$-maximal actions, and we show how $\underline{P}$-maximality generalises maximising expected utility. The following theorem generalises a result by Walley [86, Section 3.9.2, p. 161].

Theorem 6.7. *An action a in A is $\underline{P}$-maximal if and only if it is $\geq$-maximal and $\overline{P}(J_a - J_b) \geq 0$ for all actions b in A:*

$$\max_{\geqslant_{\underline{P}}}(A) = \max_{>_{\underline{P}}}(A) = \left\{a \in \max_{\geq}(A) : (\forall b \in A)(\overline{P}(J_a - J_b) \geq 0)\right\}$$

and if $\mathcal{J}_A$ is compact, then a in A is $\underline{P}$-maximal if and only if a is $\geq$-maximal and $\overline{P}(J_a - J_b) \geq 0$ for all $\geq$-maximal actions b in A:

$$= \left\{a \in \max_{\geq}(A) : (\forall b \in \max_{\geq}(A))(\overline{P}(J_a - J_b) \geq 0)\right\}$$

Proof. First, note that, for any random quantity f in $\operatorname{dom}\underline{P}$, the set $\uparrow_{\geq}^{\operatorname{dom}\underline{P}} f$ is closed. Indeed, if a net g_α in $\uparrow_{\geq}^{\operatorname{dom}\underline{P}} f$ converges to g in $\operatorname{dom}\underline{P}$, *i.e.*,

$$\overline{P}(|g - g_\alpha|) \to 0 \text{ and } \forall x \in \mathcal{X}\colon g_\alpha(x) \to g(x),$$

then g belongs to $\uparrow_{\geq}^{\operatorname{dom}\underline{P}} f$: since $g_\alpha(x) \geq f(x)$ for all $x \in \mathcal{X}$, it follows that also $\lim_\alpha g_\alpha(x) = g(x) \geq f(x)$ for all $x \in \mathcal{X}$, so $g \in \uparrow_{\geq}^{\operatorname{dom}\underline{P}} f$. Hence, $\uparrow_{\geq}^{\operatorname{dom}\underline{P}} f$ is closed, and therefore Lemma 6.1 on p. 275 applies on $\operatorname{dom}\underline{P}$ with preorder $\geq$.

By the definition of $\underline{P}$-maximality, it easily follows that

$$\max_{>_{\underline{P}}}(A) = \left\{a \in A\colon (\forall b \in A)\Big(\overline{P}(J_a - J_b) \geq 0 \text{ and } (J_a \not\leq J_b \text{ or } J_a = J_b)\Big)\right\},$$

but, if a does not belong to $\max_{\geq}(A)$, then there is a b in A such that $J_b \geq J_a$ and $J_b \neq J_a$, and hence, it cannot hold that $J_a \not\leq J_b$ or $J_a = J_b$. Therefore, any a in A such that $J_a \not\leq J_b$ or $J_a = J_b$ for all b in A must be $\geq$-maximal:

$$= \left\{a \in \max_{\geq}(A) : (\forall b \in A)\Big(\overline{P}(J_a - J_b) \geq 0 \text{ and } (J_a \not\leq J_b \text{ or } J_a = J_b)\Big)\right\}$$

but, if a belongs to $\max_{\geq}(A)$, then it must hold that $J_a \not\leq J_b$ or $J_a = J_b$ for all b in A. Indeed, if $J_a \leq J_b$ and $J_a \neq J_b$, then J_a cannot belong to $\max_{\geq}(A)$, a contradiction. So,

$$= \left\{a \in \max_{\geq}(A) : (\forall b \in A)\Big(\overline{P}(J_a - J_b) \geq 0\Big)\right\}$$

which establishes the first equality. Now assume that $\mathcal{J}_A$ is compact. Consider b in A. If a belongs to $\max_{\geq}(A)$, then, by Lemma 6.1(iii), there is a $\geq$-maximal c in A such that $J_c \geq J_b$. But, since $\overline{P}(J_a - J_b) \geq \overline{P}(J_a - J_c)$ whenever $J_c \geq J_b$, this implies that $\overline{P}(J_a - J_b) \geq 0$ whenever $\overline{P}(J_a - J_c) \geq 0$ for all $\geq$-maximal actions c in A. So, if $\overline{P}(J_a - J_c) \geq 0$ for all c in $\max_{\geq}(A)$, then this must also hold for all b in A; and conversely, it is obvious that if $\overline{P}(J_a - J_b) \geq 0$ for all b in A, then the same also holds for all c in $\max_{\geq}(A)$. We find

$$= \left\{a \in \max_{\geq}(A) : (\forall c \in \max_{\geq}(A))\left(\overline{P}(J_a - J_c) \geq 0\right)\right\}$$

which establishes the second equality. □

As a consequence, if all random quantities in $\mathcal{J}_A$ are $\geq$-maximal, *i.e.*, if $J_a \not\leq J_b$ or $J_a = J_b$ for all a and b in A, then the set of $\underline{P}$-maximal actions in A are exactly those actions a in A such that $\overline{P}(J_a - J_b) \geq 0$ for all actions b in A. If additionally $\underline{P}$ is self-conjugate, *i.e.*, if $\underline{P}(f) = \overline{P}(f)$ for all f in $\operatorname{dom}\underline{P}$ and using $P(f)$ as a notation for both $\underline{P}(f)$ and $\overline{P}(f)$, this condition means that P-maximal actions a maximise their prevision P.

Corollary 6.8. *Suppose that $\underline{P}$ is self-conjugate. The following statements hold.*

(i) If $\mathcal{J}_A = \max_{\geq}(\mathcal{J}_A)$, then $\max_{>_{\underline{P}}}(A) = \arg\max_{a \in A} P(J_a)$.

(ii) If $\mathcal{J}_A$ is compact, then $\max_{>_{\underline{P}}}(A) = \arg\max_{a \in \max_{\geq} A} P(J_a)$.

Proof. Apply Theorem 6.7, and observe that, for any a and b in A, $P(J_a - J_b) \geq 0$ if and only if $P(J_a) \geq P(J_b)$ by the self-conjugacy and coherence of $\underline{P}$. □

So, in case $\underline{P}$ is self-conjugate, $\underline{P}$-maximality corresponds in essence to maximising expected utility as defined in Eq. (6.1) on p. 268. Note that, in Corollary 6.8(i), we don't need to impose compactness of $\mathcal{J}_A$.

6.3.4 Existence of Dominating $\underline{P}$-Maximal Actions

$\underline{P}$-maximal actions do not always exist: not every preordered set has maximal elements. A fairly general sufficient condition for the existence of $\underline{P}$-maximal elements in A is that $\mathcal{J}_A$ should be compact with respect to the topology introduced in Eq. (6.2) on p. 274. This generalises a result mentioned by Walley [86, Section 3.9.2]. In fact, we prove a stronger result in Theorem 6.9 below,

which turns out to be very important in proving that the dynamic programming approach works for $\underline{P}$-maximality in Chapter 7, p. 299 ff. Theorem 6.9 is a slightly stronger version of an earlier result proved by De Cooman and Troffaes [23]: Theorem 6.9 does not hold only for extended lower previsions defined on bounded random quantities only, as was the case for the earlier version, and secondly, it requires compactness of $\mathcal{J}_A$ with respect to a topology that is weaker than the topology of uniform convergence (and therefore has more compact sets).

Theorem 6.9. *If $\mathcal{J}_A$ is non-empty and compact, then for every action a in A, there is a $\underline{P}$-maximal action b in A such that $b \gtrcurlyeq_{\underline{P}} a$.*

Proof. We already observed that $\max_{>_{\underline{P}}}(A) = \max_{\gtrcurlyeq_{\underline{P}}}(A)$; see Definition 6.5. So, by Lemma 6.1(iii) on p. 275, it would suffice to prove that for any random quantity f in $\operatorname{dom}\underline{P}$, the set $\uparrow^{\operatorname{dom}\underline{P}}_{\gtrcurlyeq_{\underline{P}}} f$ is closed. Unfortunately, $\uparrow^{\operatorname{dom}\underline{P}}_{\gtrcurlyeq_{\underline{P}}} f$ is not closed in general. Therefore, instead, for a given random quantity J_a in $\mathcal{J}_A$, we shall identify a random quantity J_b in $\max_{\gtrcurlyeq_{\underline{P}}}(\mathcal{J}_A)$ such that $J_b \gtrcurlyeq_{\underline{P}} J_a$, using preorders $\trianglerighteq$ for which $\uparrow^{\operatorname{dom}\underline{P}}_{\trianglerighteq} f$ *is* closed.

Indeed, consider the point-wise order $\geq$ (which is not only a preorder, but even a partial order) defined by $f \geq g$ whenever $f(x) \geq g(x)$ for all $x \in \mathcal{X}$, and the preorder $\gtrcurlyeq_{\underline{P}}$ defined by $f \gtrcurlyeq_{\underline{P}} g$ whenever $\underline{P}(f - g) \geq 0$. We first show that the sets $\uparrow^{\operatorname{dom}\underline{P}}_{\geq} f$ and $\uparrow^{\operatorname{dom}\underline{P}}_{\gtrcurlyeq_{\underline{P}}} f$ are closed, for any random quantity f in $\operatorname{dom}\underline{P}$. The closedness of $\uparrow^{\operatorname{dom}\underline{P}}_{\geq} f$ has been demonstrated in the proof of Theorem 6.7. Let's prove that $\uparrow^{\operatorname{dom}\underline{P}}_{\gtrcurlyeq_{\underline{P}}} f$ is closed.

For instance, let g_α be any net in $\uparrow^{\operatorname{dom}\underline{P}}_{\gtrcurlyeq_{\underline{P}}} f$, and suppose that g_α converges to a random quantity g in $\operatorname{dom}\underline{P}$:

$$\overline{P}(|g - g_\alpha|) \to 0 \text{ and } \forall x \in \mathcal{X}\colon g_\alpha(x) \to g(x).$$

We must show that g belongs to $\uparrow^{\operatorname{dom}\underline{P}}_{\gtrcurlyeq_{\underline{P}}} f$. Indeed, since $\overline{P}(|g - g_\alpha|) \to 0$, it follows by Theorem 3.5(xi) on p. 55 that $\overline{P}(g - g_\alpha) \to \overline{P}(g - g) = 0$. Also, for every α, it holds that $\underline{P}(g - f) \geq \underline{P}(g - g_\alpha) + \underline{P}(g_\alpha - f) \geq \underline{P}(g - g_\alpha)$, since $g_\alpha \in \uparrow^{\operatorname{dom}\underline{P}}_{\gtrcurlyeq_{\underline{P}}} f$. Therefore, $\underline{P}(g - f) \geq \lim_\alpha \underline{P}(g - g_\alpha) = 0$, so $g \in \uparrow^{\operatorname{dom}\underline{P}}_{\gtrcurlyeq_{\underline{P}}} f$.

Now, given J_a, we construct J_b as follows. By Lemma 6.1(iii), there is a $\geq$-maximal element J_c in $\mathcal{J}_A$ such that $J_c \geq J_a$. In particular, it follows that $J_c \gtrcurlyeq_{\underline{P}} J_a$, so, if J_c happens to be $\underline{P}$-maximal, then we may take $b = c$ and the theorem is established. If J_c is not $\underline{P}$-maximal, then

(a) by Theorem 6.7, there is at least one action d in $\max_{\geq}(A)$ such that $\overline{P}(J_c - J_d) < 0$,

(b) by Lemma 6.1(i) $\uparrow^{\mathcal{J}_A}_{\succcurlyeq_{\underline{P}}} J_d$ is compact, and therefore, since $\overline{P}$ is continuous (see Theorem 3.5(xi)), the mapping $\overline{P}$ achieves a maximum on $\uparrow^{\mathcal{J}_A}_{\succcurlyeq_{\underline{P}}} J_d$, say, at $J_e \in \uparrow^{\mathcal{J}_A}_{\succcurlyeq_{\underline{P}}} J_d$. For every random quantity J_s in $\uparrow^{\mathcal{J}_A}_{\succcurlyeq_{\underline{P}}} J_d$, it holds that

$$\overline{P}(J_e - J_s) \geq \overline{P}(J_e) + \underline{P}(-J_s) = \overline{P}(J_e) - \overline{P}(J_s) \geq 0, \tag{6.6}$$

(c) by Lemma 6.1(iii) there is a $\geq$-maximal element J_b in $\mathcal{J}_A$ such that $J_b \geq J_e$.

We now show that J_b is $\underline{P}$-maximal in $\mathcal{J}_A$, and that $J_b \succcurlyeq_{\underline{P}} J_a$.

But, by construction, J_b is $\geq$-maximal in $\mathcal{J}_A$, so, by Theorem 6.7 it suffices to show that $\overline{P}(J_b - J_s) \geq 0$ for all s in $\max_{\geq}(A)$. Indeed, if $J_s \in \uparrow^{\mathcal{J}_A}_{\succcurlyeq_{\underline{P}}} J_d$, then $\overline{P}(J_b - J_s) \geq \overline{P}(J_e - J_s)$ since $J_b \geq J_e$ by construction, and $\overline{P}(J_e - J_s) \geq 0$ by Eq. (6.6). On the other hand, if $J_s \notin \uparrow^{\mathcal{J}_A}_{\succcurlyeq_{\underline{P}}} J_d$, then $\underline{P}(J_s - J_d) < 0$, and hence, since $J_b \geq J_e$,

$$\overline{P}(J_b - J_s) \geq \overline{P}(J_e - J_s) \geq \underline{P}(J_e - J_d) + \overline{P}(J_d - J_s) > 0,$$

because $\underline{P}(J_e - J_d) \geq 0$ by construction of J_e, and $\overline{P}(J_d - J_s) = -\underline{P}(J_s - J_d) > 0$ since $J_s \notin \uparrow^{\mathcal{J}_A}_{\succcurlyeq_{\underline{P}}} J_d$.

So, J_b is $\underline{P}$-maximal in $\mathcal{J}_A$. It remains to show that $J_b \succcurlyeq_{\underline{P}} J_a$. Indeed, since $J_b \geq J_e$ and $J_c \geq J_a$,

$$\underline{P}(J_b - J_a) \geq \underline{P}(J_e - J_c) \geq \underline{P}(J_e - J_d) + \underline{P}(J_d - J_c) > 0,$$

because $\underline{P}(J_e - J_d) \geq 0$ by construction of J_e, and $\underline{P}(J_d - J_c) = -\overline{P}(J_c - J_d) > 0$ by construction of J_d. □

6.4 $\mathcal{M}$-Maximality

6.4.1 Robustifying Preference

In the theory of *Bayesian sensitivity analysis*, also called *Bayesian robustness*, belief is modelled through a set **m** of probability measures, which is assumed to contain the 'true', but unknown, probability measure μ; see for instance Berger [6], in case **m** corresponds to set of posterior probability distributions.

A similar model is obtained in *quasi-Bayesian theory*, where sets of probability measures represent partial preference orderings that satisfy certain systems of axioms, similar to the axioms of rationality, given in Axiom 3.1 on p. 49; see Giron and Rios [35] in case $\mathcal{X}$ is a finite set; also see Seidenfeld, Schervish and Kadane [72] for a very general representation of preferences, using sets of probability/utility pairs.

Central to these theories is that a random quantity f is to be preferred to a random quantity g whenever it is preferred under all candidate models $\mu \in \mathbf{m}$. This leads to the definition (or representation) of a 'robustified' preference ordering $\succcurlyeq_{\mathbf{m}}$. Let us denote by $P_\mu(f)$ the expected utility of f with respect to μ, *i.e.*, say that the P_μ are linear behavioural extensions of the μ in $\mathbf{m}$ defined on a common domain $\mathcal{K} \subseteq \mathcal{R}(X)$; for simplicity, assume that $\mathcal{K}$ is a linear space, and $\mathcal{J}_A \subseteq \mathcal{K}$. Then, for each μ we already have a preference ordering, namely $\succcurlyeq_{P_\mu}$, defined in Definition 6.4. These preference orderings induce the following robustified preference:

Definition 6.10. For any a and b in A, we say that $a \succcurlyeq_{\mathbf{m}} b$ whenever $a \succcurlyeq_{P_\mu} b$ for all $\mu \in \mathbf{m}$.

We have already established an onto and one-to-one correspondence between coherent extended lower previsions of the form $\underline{P} = \underline{\mathbf{E}}^{\mathbf{x}}_{\underline{Q}}$ (where $\underline{Q}$ is a lower prevision that avoids sure loss) and convex compact sets of probability charges on $\wp(\mathcal{X})$; see Theorem 5.76 on p. 265. For these lower previsions, the above definition can also be written as:

Definition 6.11. For any a and b in A, we say that $a \succcurlyeq_{\mathcal{M}} b$ whenever $a \succcurlyeq_R b$ for all $R \in \mathcal{M}$.

Recall that $\mathcal{M}$ denotes the set of extended linear previsions on $\operatorname{dom} \underline{P}$ which are behavioural extensions of $\underline{P}$.

We shall take the above definition as a general definition of robustified preference $\succcurlyeq_{\mathcal{M}}$: the set $\mathcal{M}$ is assumed to contain the 'true', but unknown, prevision (*i.e.*, expected utility operator) R.

Since $\mathcal{M}$ is assumed to be compact and convex, it is not difficult to show that the partial orders $\succcurlyeq_{\mathcal{M}}$ and $\succcurlyeq_{\underline{P}}$ on $\operatorname{dom} \underline{P}$ are one and the same, simply through observation that $\underline{P}$ is the lower envelope of $\mathcal{M}$. Indeed, since $\underline{P}(f) = \min_{R \in \mathcal{M}} R(f)$ for all f in $\operatorname{dom} \underline{P}$, for any a and b in A it holds that $J_a \geq J_b$ or $\underline{P}(J_a - J_b) > 0$, if and only if $J_a \geq J_b$ or $\min_{R \in \mathcal{M}} R(J_a - J_b) > 0$, and, since this

minimum is actually reached for some R in $\mathcal{M}$, this holds if and only if $J_a \geq J_b$ or $R(J_a - J_b) > 0$ for all R in $\mathcal{M}$. So, $\succcurlyeq_{\mathcal{M}}$ is equal to $\succcurlyeq_{\underline{P}}$.

As a result, Bayesian robustness can be modelled through $\underline{P}$-maximality, whenever (i) $\mathbf{m}$ induces a compact and convex set of previsions (expected utility operators) defined on a common linear space $\mathcal{K} \subseteq \mathcal{R}(\mathcal{X})$, and (ii) the lower envelope $\underline{P}$ of this set of previsions satisfies the conditions of Section 6.2.2 on p. 272 ff. Note that these assumptions are in particular satisfied when $\mathcal{X}$ is a finite space, $\mathbf{m}$ is a convex and compact set of probability measures on $\wp(\mathcal{X})$, and $\mathcal{K} = \mathcal{R}(\mathcal{X})$, *i.e.*, the case discussed by Giron and Rios [35].

6.4.2 Robustifying Choice Functions: E-Admissibility

But there is in the literature yet another notion of optimality that can be associated with the compact convex set of extended linear previsions $\mathcal{M}$: an action a can be considered optimal in A if it is a maximal element of A with respect to the preorder $\succcurlyeq_R$ for *some* $R \in \mathcal{M}$. This notion of optimality is a special case of what Levi [54, Section 4.8] calls 'E-admissibility' (we have a unique utility function, whereas Levi allows for a convex set of utility functions, and secondly, we only consider the static case here, whereas E-admissibility also extends to dynamic systems). This notion of optimality does not generally coincide with the ones associated with the preorders $\succcurlyeq_{\mathcal{M}}$ and $\succcurlyeq_{\underline{P}}$, unless the set $\mathcal{J}_A$ is convex; see for instance Walley [86, Section 3.9.5, pp. 162–163]. We are therefore led to consider the following notion of optimality:

Definition 6.12. An action a in A is called $\mathcal{M}$-*maximal*, if it is R-maximal for some R in $\mathcal{M}$:

$$\operatorname{opt}_{\mathcal{M}}(A) := \bigcup_{R \in \mathcal{M}} \operatorname{opt}_{\not\prec_R}(A) = \bigcup_{R \in \mathcal{M}} \max_{\succcurlyeq_R}(A)\,. \tag{6.7}$$

6.4.3 Monotonicity

The more determinate our beliefs, the smaller the set of $\mathcal{M}$-maximal actions (this is similar to Theorem 6.6 on p. 280).

Theorem 6.13. *If $\underline{Q}$ is a behavioural extension of $\underline{P}$, then* $\max_{\mathcal{M}_{\underline{Q}}}(A) \subseteq \max_{\mathcal{M}_{\underline{P}}}(A)$.

Proof. If $\underline{Q}$ is a behavioural extension of $\underline{P}$, then any linear behavioural extension S of $\underline{Q}$ is also a linear behavioural extension of $\underline{P}$, and hence, a linear behavioural extension of an element—namely, the restriction of S to $\operatorname{dom}\underline{P}$—of $\mathcal{M}_{\underline{P}}$.

Suppose that a is $\mathcal{M}_{\underline{Q}}$-maximal in A. Then, there is a linear behavioural extension S_a of $\underline{Q}$ such a is S_a-maximal in A. But, since S_a is also a linear behavioural extension of an R_a in $\mathcal{M}_{\underline{P}}$, it follows from Theorem 6.6 on p. 280 that a is R_a-maximal, and hence $\mathcal{M}_{\underline{P}}$-maximal. □

6.4.4 $\mathcal{M}$-Maximality Through Point-Wise Maximality

Theorem 6.14. *An action a in A is $\mathcal{M}$-maximal if and only if it is $\geq$-maximal in A and there is an R_a in $\mathcal{M}$ such that $R_a(J_a - J_b) \geq 0$ for every action b in A:*

$$\max_{>_{\underline{P}}}(A) = \left\{a \in \max_{\geq}(A) : (\exists R_a \in \mathcal{M})(\forall b \in A)(R_a(J_a - J_b) \geq 0)\right\},$$

and if $\mathcal{J}_A$ is compact, then a in A is $\underline{P}$-maximal if and only if a is $\geq$-maximal in A and there is an R_a in $\mathcal{M}$ such that $R_a(J_a - J_b) \geq 0$ for every $\geq$-maximal action b in A:

$$= \left\{a \in \max_{\geq}(A) : (\exists R_a \in \mathcal{M})(\forall b \in \max_{\geq}(A))(R_a(J_a - J_b) \geq 0)\right\}.$$

Proof. Immediate from Eq. (6.7) and Theorem 6.7 on p. 280. □

6.4.5 When $\underline{P}$-Maximality and $\mathcal{M}$-Maximality Coincide

In any case, $\mathcal{M}$-maximality implies $\underline{P}$-maximality (see Walley [86, Sect. 3.9.4, p. 162, ll. 26–28]):

Proposition 6.15. *If a is $\mathcal{M}$-maximal in A, then it is also $\underline{P}$-maximal in A.*

Proof. If a is $\mathcal{M}$-maximal in A, then, by Theorem 6.14, a is $\geq$-maximal and there is a linear extended prevision R_a in $\mathcal{M}$ such that $R_a(J_a - J_b) \geq 0$ for all b in A. But, by definition of $\mathcal{M}$, R_a is a behavioural extension of $\underline{P}$: $R_a(f) \geq \underline{P}(f)$ for all f in $\operatorname{dom}\underline{P}$. Hence, a is $\geq$-maximal and $\overline{P}(J_a - J_b) \geq R_a(J_a - J_b) \geq 0$ for all b in A. Applying Theorem 6.7, we find that a is $\underline{P}$-maximal in A. □

The converse does not need to hold: we have demonstrated in Section 6.2.1 on p. 269 ff. that $\operatorname{opt}_{\mathcal{M}}(A)$ does not need to coincide with $\operatorname{opt}_{\not<_{\underline{P}}}(A)$; also see Walley [86, Section 3.9.9, p. 165] for a counterexample. As already

mentioned before, the converse is however guaranteed to hold if A contains no more than two elements (Schervish, Seidenfeld, Kadane, and Levi [71]) or if $\mathcal{J}_A$ is convex (Walley [86, Section 3.9.5, "maximality theorem"]):

Theorem 6.16. *If A contains no more than two actions, then* $\operatorname{opt}_{\mathcal{M}}(A) = \operatorname{opt}_{\not<_{\underline{P}}}(A)$.

Proof. Immediate. □

Theorem 6.17. *If $\mathcal{J}_A$ is convex, then* $\operatorname{opt}_{\mathcal{M}}(A) = \operatorname{opt}_{\not<_{\underline{P}}}(A)$.

Proof. (Proof adapted from Walley [86, Section 3.9.5, "maximality theorem"]) Let $a \in A$ and define $\mathcal{K} = \{J_a - J_b : b \in A\}$. Since $\mathcal{J}_A$ is convex, so is $\mathcal{K}$. By Theorems 6.7&6.14, it suffices to prove that $\overline{P}(f) \geq 0$ for all f in $\mathcal{K}$ if and only if there is a Q in $\mathcal{M}$ such that $Q(f) \geq 0$ for all f in $\mathcal{K}$.

"if". Immediate since $\overline{P}(f) \geq Q(f)$ for all f in $\mathcal{K}$.

"only if". Assume that $\overline{P}(f) \geq 0$ for all f in $\mathcal{K}$. Since $\mathcal{K}$ is convex, condition (B) of Lemma 5.20 on p. 218 is satisfied, so, condition (C) of Lemma 5.20 must be satisfied as well: there is a linear behavioural extension R of $\underline{P}$ such that $R(f) \geq 0$ for all f in $\mathcal{K}$. Now take the restriction Q of R to $\operatorname{dom}\underline{P}$: Q belongs to $\mathcal{M}$ and satisfies the desired property. □

6.5 $\underline{P}$-Maximinity and $\underline{P}$-Maximaxity

6.5.1 Worst-Case and Best-Case Ranking

Another common generalisation of maximising expected utility ranks actions according to the lower (or upper) prevision of their gain gambles:

Definition 6.18. An action a in A is called $\underline{P}$-*maximin* in A if it is maximal in A with respect to the preorder $\sqsupseteq_{\underline{P}}$ defined by $a \sqsupseteq_{\underline{P}} b$ whenever

$$\underline{P}(J_a) > \underline{P}(J_b) \text{ or } J_a \geq J_b.$$

Similarly, an action a in A is called $\underline{P}$-*maximax* in A if it is maximal in A with respect to the preorder $\sqsupseteq_{\overline{P}}$ defined by $a \sqsupseteq_{\overline{P}} b$ whenever

$$\overline{P}(J_a) > \overline{P}(J_b) \text{ or } J_a \geq J_b.$$

For an axiomatic study of $\underline{P}$-maximin, we refer to Gilboa and Schmeidler [34]. For some reason, $\underline{P}$-maximax does not enjoy the same popularity as

$\underline{P}$-maximin, but it has been applied as a criterion of optimality in the theory of Markov decision processes, see Satia and Lave [68].

6.5.2 No Monotonicity

More determinate beliefs usually do not lead to a smaller set $\underline{P}$-maximin or $\underline{P}$-maximax actions. Indeed, consider the example of Section 6.2.1 on p. 269 ff.: without any further information about the coin, $\underline{P}$ is the vacuous lower prevision on $\{H, T\}$, and the $\underline{P}$-maximin solution is $\{a_0\}$. Suppose we receive additional information: the coin will always land heads up. This information is modelled by taking $\underline{P}$ equal to the vacuous lower prevision on $\{H\}$; the corresponding $\underline{P}$-maximin solution is $\{a_1\}$, which is clearly no subset of $\{a_0\}$.

6.5.3 Through Pointwise Maximality

Again, point-wise maximality helps (also see Walley [86, Sect. 3.9.7, pp. 163–164]):

Theorem 6.19. *An action a in A is $\underline{P}$-maximin if and only if it is $\geq$-maximal and $\underline{P}(J_a) \geq \underline{P}(J_b)$ for all actions b in A:*

$$\max_{\sqsupseteq_{\underline{P}}}(A) = \left\{a \in \max_{\geq}(A) : (\forall b \in A)(\underline{P}(J_a) \geq \underline{P}(J_b))\right\}$$

and if $\mathcal{J}_A$ is compact, then a in A is $\underline{P}$-maximin if and only if a is $\geq$-maximal and $\underline{P}(J_a) \geq \underline{P}(J_b)$ for all $\geq$-maximal actions b in A:

$$= \arg \max_{a \in \max_{\geq}(A)} \underline{P}(J_a).$$

Proof. Similar to the proof of Theorem 6.7 on p. 280. □

Theorem 6.20. *An action a in A is $\underline{P}$-maximax if and only if it is $\geq$-maximal and $\overline{P}(J_a) \geq \overline{P}(J_b)$ for all actions b in A:*

$$\max_{\sqsupseteq_{\overline{P}}}(A) = \left\{a \in \max_{\geq}(A) : (\forall b \in A)(\overline{P}(J_a) \geq \overline{P}(J_b))\right\}$$

and if $\mathcal{J}_A$ is compact, then a in A is $\underline{P}$-maximax if and only if a is $\geq$-maximal and $\overline{P}(J_a) \geq \overline{P}(J_b)$ for all $\geq$-maximal actions b in A:

$$= \arg \max_{a \in \max_{\geq}(A)} \overline{P}(J_a).$$

Proof. Similar to the proof of Theorem 6.7 on p. 280. □

Corollary 6.21. *The following statements hold.*

(i) If $\mathcal{J}_A = \max_{\geq}(\mathcal{J}_A)$, then

$$\max_{\sqsupseteq_{\underline{P}}}(A) = \arg\max_{a \in A} \underline{P}(J_a) \quad \textit{and} \quad \max_{\sqsupseteq_{\overline{P}}}(A) = \arg\max_{a \in A} \overline{P}(J_a).$$

(ii) If $\mathcal{J}_A$ is compact, then

$$\max_{\sqsupseteq_{\underline{P}}}(A) = \arg \max_{a \in \max_{\geq} A} \underline{P}(J_a) \quad \textit{and} \quad \max_{\sqsupseteq_{\overline{P}}}(A) = \arg \max_{a \in \max_{\geq} A} \overline{P}(J_a).$$

6.5.4 $\underline{P}$-Maximin and $\underline{P}$-Maximax Imply $\underline{P}$-Maximality

The following proposition connects $\underline{P}$-maximinity and $\underline{P}$-maximaxity with $\underline{P}$-maximality.

Proposition 6.22. *If an action a is $\underline{P}$-maximin or $\underline{P}$-maximax in A, then it also $\underline{P}$-maximal in A.*

Proof. By the coherence of $\underline{P}$, it follows that $\overline{P}(J_a - J_b) \geq \overline{P}(J_a) - \overline{P}(J_b)$ and $\overline{P}(J_a - J_b) \geq \underline{P}(J_a) - \underline{P}(J_b)$ for any a and b in A. So, if a is $\underline{P}$-maximin or $\underline{P}$-maximax in A, then, by these inequalities and Theorem 6.19 or Theorem 6.20, it follows that a is $\geq$-maximal in A and $\overline{P}(J_a - J_b) \geq 0$ for all b in A. But, by Theorem 6.7 on p. 280, this means that a is $\underline{P}$-maximal in A. □

6.5.5 $\underline{P}$-Maximax Implies $\mathcal{M}$-Maximality

$\underline{P}$-maximin does not necessarily imply $\mathcal{M}$-maximality: that was demonstrated by the example of Section 6.2.1 on p. 269 ff. However, $\underline{P}$-maximax does imply $\mathcal{M}$-maximality.

Proposition 6.23. *If an action a is $\underline{P}$-maximax in A, then it also $\mathcal{M}$-maximal in A.*

Proof. Suppose that a is $\underline{P}$-maximax in A, *i.e.*, by Theorem 6.20, a is $\geq$-maximal, and $\overline{P}(J_a) \geq \overline{P}(J_b)$ for all b in A. Since $\overline{P}(J_a) = \max_{Q \in \mathcal{M}} Q(J_a)$, there is an R in $\mathcal{M}$ such that $R(J_a) = \overline{P}(J_a)$. So, for all b in A,

$$R(J_a) = \overline{P}(J_a) \geq \overline{P}(J_b) = \max_{Q \in \mathcal{M}} Q(J_b) \geq R(J_b),$$

and therefore, a must be $\mathcal{M}$-maximal as well, by Theorem 6.14. □

6.5.6 Existence of Dominating $\underline{P}$-Maximin and $\underline{P}$-Maximax Actions

The existence of dominating $\underline{P}$-maximin and $\underline{P}$-maximax actions, which will turn out to be very important in dynamic programming, is again guaranteed if $\mathcal{J}_A$ is compact.

Theorem 6.24. *If $\mathcal{J}_A$ is non-empty and compact, then for every action a in A, there is a $\underline{P}$-maximin action b in A such that $b \sqsupseteq_{\underline{P}} a$. Similarly, for every action a in A, there is a $\underline{P}$-maximax action b in A such that $b \sqsupseteq_{\overline{P}} a$.*

Proof. By Lemma 6.1(iii), there is a $\geq$-maximal element J_c in $\mathcal{J}_A$ such that $J_c \geq J_a$. In particular, it follows that $J_c \sqsupseteq_{\underline{P}} J_a$, so, if J_c happens to be $\underline{P}$-maximin, then we may take $b = c$ and the theorem is established. If J_c is not $\underline{P}$-maximin, then

(a) by Theorem 6.19, there is at least one action d in $\max_{\geq}(A)$ such that $\underline{P}(J_c) < \underline{P}(J_d)$,

(b) by the continuity of $\underline{P}$ (see Theorem 3.5(xi)) and the compactness of $\mathcal{J}_A$, $\underline{P}$ achieves a maximum on $\mathcal{J}_A$, say, at J_e. For every random quantity J_s in $\mathcal{J}_A$, it holds that

$$\underline{P}(J_e) \geq \underline{P}(J_s). \tag{6.8}$$

(c) by Lemma 6.1(iii) there is a $\geq$-maximal element J_b in $\mathcal{J}_A$ such that $J_b \geq J_e$.

We now show that J_b is $\underline{P}$-maximin in $\mathcal{J}_A$, and that $J_b \sqsupseteq_{\underline{P}} J_a$ (note that this b has been constructed independently of a).

By construction, J_b is $\geq$-maximal in $\mathcal{J}_A$, so, by Theorem 6.19 it suffices to show that $\underline{P}(J_b) \geq \underline{P}(J_s)$ for all s in A. But this follows from $\underline{P}(J_b) \geq \underline{P}(J_e)$ since $J_b \geq J_e$ by construction, and $\underline{P}(J_e) \geq \underline{P}(J_s)$ by Eq. (6.8).

So, J_b is $\underline{P}$-maximin in $\mathcal{J}_A$. It remains to show that $J_b \sqsupseteq_{\underline{P}} J_a$. Indeed, since $J_b \geq J_e$ and $J_c \geq J_a$,

$$\underline{P}(J_b) \geq \underline{P}(J_e) \geq \underline{P}(J_d) > \underline{P}(J_c) \geq \underline{P}(J_a),$$

because $\underline{P}(J_e) \geq \underline{P}(J_d)$ by construction of J_e, and $\underline{P}(J_d) > \underline{P}(J_c)$ by construction of J_d.

The construction of a dominating $\underline{P}$-maximax action is similar. □

6.6 Interval Dominance and Weak $\underline{P}$-Maximality

6.6.1 Definition

Yet another criterion of optimality is obtained as follows:

Definition 6.25. An action a in A is called *weakly $\underline{P}$-maximal* in A if it is maximal in A with respect to the preorder $\sqsupseteq_{\underline{P}}$ defined by $a \sqsupseteq_{\underline{P}} b$ whenever

$$\underline{P}(J_a) > \overline{P}(J_b) \text{ or } J_a \geq J_b.$$

The inequality $\underline{P}(J_a) > \overline{P}(J_b)$ is sometimes also called *interval dominance*: it says that the interval $[\underline{P}(J_a), \overline{P}(J_a)]$ is completely on the right side of the interval $[\underline{P}(J_b), \overline{P}(J_b)]$.

6.6.2 Weak?

By the coherence of $\underline{P}$, it follows that $\underline{P}(J_a - J_b) \geq \underline{P}(J_a) - \overline{P}(J_b)$, so $\underline{P}$-maximality implies weak $\underline{P}$-maximality; whence the name weak $\underline{P}$-maximality.

Proposition 6.26. *If a is $\underline{P}$-maximal in A, then it is also weakly $\underline{P}$-maximal in A.*

6.6.3 Monotonicity

The more determinate our beliefs, the smaller the set of weakly $\underline{P}$-maximal actions (this is similar to Theorem 6.6 on p. 280 and Theorem 6.13 on p. 286).

Theorem 6.27. *If $\underline{Q}$ is a behavioural extension of $\underline{P}$, then* $\max_{\sqsupseteq_{\underline{Q}}}(A) \subseteq \max_{\sqsupseteq_{\underline{P}}}(A)$.

Proof. Suppose that a is $\trianglerighteq_{\underline{Q}}$-maximal in A. Then, for all b in A, it holds that $\overline{Q}(J_a) \geq \underline{Q}(J_b)$, and $J_b \not\geq J_a$ or $J_b = J_a$. But, if $\overline{Q}(J_a) \geq \underline{Q}(J_b)$, then it must also hold that $\overline{P}(J_a) \geq \underline{P}(J_b)$, since that $\underline{Q}$ is a behavioural extension of $\underline{P}$. So, for all b in A, it holds that $\overline{P}(J_a) \geq \underline{P}(J_b)$, and $J_b \not\geq J_a$ or $J_b = J_a$: a must be weakly $\underline{P}$-maximal as well. □

6.6.4 Weak $\underline{P}$-Maximality Through Point-Wise Maximality

Also for this type of optimality, point-wise maximality helps.

Theorem 6.28. *An action a in A is weakly $\underline{P}$-maximal if and only if it is $\geq$-maximal and $\overline{P}(J_a) \geq \underline{P}(J_b)$ for all actions b in A:*

$$\max_{\trianglerighteq_{\underline{P}}}(A) = \left\{a \in \max_{\geq}(A) : (\forall b \in A)(\overline{P}(J_a) \geq \underline{P}(J_b))\right\}$$

and if $\mathcal{J}_A$ is compact, then a in A is weakly $\underline{P}$-maximal if and only if a is $\geq$-maximal and $\overline{P}(J_a) \geq \underline{P}(J_b)$ for all $\geq$-maximal actions b in A:

$$= \left\{a \in \max_{\geq}(A) : (\forall b \in \max_{\geq}(A))(\overline{P}(J_a) \geq \underline{P}(J_b))\right\}$$

Proof. Similar to the proof of Theorem 6.7 on p. 280. □

6.6.5 Existence of Dominating Weakly $\underline{P}$-Maximal Actions

Theorem 6.29. *If $\mathcal{J}_A$ is non-empty and compact, then for every action a in A, there is a weakly $\underline{P}$-maximal action b in A such that $b \trianglerighteq_{\underline{P}} a$.*

Proof. By Lemma 6.1(iii), there is a $\geq$-maximal element J_c in $\mathcal{J}_A$ such that $J_c \geq J_a$. In particular, it follows that $J_c \trianglerighteq_{\underline{P}} J_a$, so, if J_c happens to be weakly $\underline{P}$-maximal, then we may take $b = c$ and the theorem is established. If J_c is not weakly $\underline{P}$-maximal, then

(a) by Theorem 6.28, there is at least one action d in $\max_{\geq}(A)$ such that $\overline{P}(J_c) < \underline{P}(J_d)$,

(b) by the continuity of $\underline{P}$ (see Theorem 3.5(xi)) and the compactness of $\mathcal{J}_A$, $\underline{P}$ achieves a maximum on $\mathcal{J}_A$, say, at J_e. For every random quantity J_s in $\mathcal{J}_A$, it holds that

$$\underline{P}(J_e) \geq \underline{P}(J_s). \tag{6.9}$$

(c) by Lemma 6.1(iii) there is a $\geq$-maximal element J_b in $\mathcal{J}_A$ such that $J_b \geq J_e$.

We now show that J_b is weakly $\underline{P}$-maximal in $\mathcal{J}_A$, and that $J_b \supseteq_{\underline{P}} J_a$ (note that this b has been constructed independently of a).

By construction, J_b is $\geq$-maximal in $\mathcal{J}_A$, so, by Theorem 6.28 it suffices to show that $\overline{P}(J_b) \geq \underline{P}(J_s)$ for all s in A. But this follows from $\overline{P}(J_b) \geq \overline{P}(J_e)$ since $J_b \geq J_e$ by construction, and $\overline{P}(J_e) \geq \underline{P}(J_e) \geq \underline{P}(J_s)$ by Eq. (6.9).

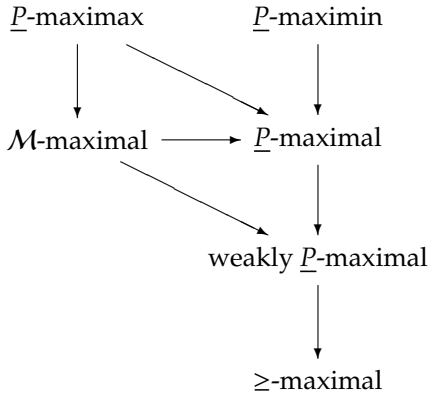
So, J_b is weakly $\underline{P}$-maximal in $\mathcal{J}_A$. It remains to show that $J_b \supseteq_{\underline{P}} J_a$. Indeed, since $J_b \geq J_e$ and $J_c \geq J_a$,

$$\underline{P}(J_b) \geq \underline{P}(J_e) \geq \underline{P}(J_d) > \overline{P}(J_c) \geq \overline{P}(J_a),$$

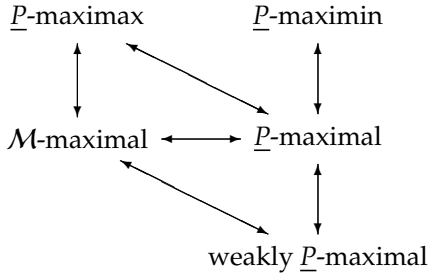
because $\underline{P}(J_e) \geq \underline{P}(J_d)$ by construction of J_e, and $\underline{P}(J_d) > \underline{P}(J_c)$ by construction of J_d. □

6.7 Summary of Optimality Criteria

- In general, the following implications hold:



- If $\underline{P}$ is self-conjugate, then $\mathcal{M} = \{P\}$, and all criteria of optimality coincide (except of course for $\geq$-maximality):



- All criteria derived directly from pair-wise preference (such as $\underline{P}$-maximality, weak $\underline{P}$-maximality, $\geq$-maximality, $\underline{P}$-maximinity, and $\underline{P}$-maximaxity) satisfy the following principle: if $\mathcal{J}_A$ is compact, then for every action a in A there is an optimal b in A such that b is preferred to a.

- All refine $\max_{\geq}$ ($\bullet$): any optimal action a in A is $\geq$-maximal in A.

- For any of these criteria, except for $\underline{P}$-maximinity and $\underline{P}$-maximaxity, if $\underline{Q}$ dominates $\underline{P}$, then optimality with respect to $\underline{Q}$, implies optimality with respect to $\underline{P}$. Roughly said, the more determinate our beliefs, the smaller the set of optimal actions.

The last property, which we have called *monotonicity*, has a nice intuitive interpretation: the more determinate our beliefs, the more decisive we are. $\underline{P}$-maximinity and $\underline{P}$-maximaxity violate this principle: they imply strong decisiveness even if our beliefs are very weak. This is demonstrated by the example of Section 6.2.1 on p. 269 ff.: $\underline{P}$-maximin yields a unique solution, namely a_0, "the safest choice".

Part II

Dynamic Programming

Chapter 7

Dynamic Programming with Uncertain Gain

In this chapter, we shall generalise the optimisation technique of dynamic programming for discrete-time systems with an uncertain gain function, assuming that uncertainty about the gain function is described by an extended lower prevision $\underline{P}$. We shall show that, in general, only for $\underline{P}$-maximality and $\mathcal{M}$-maximality an optimal feedback can be constructed by solving a Bellman-like equation. This result is due to De Cooman and Troffaes [24, 23].

7.1 Introduction

The main objective in optimal control is to find out how a system can be influenced, or controlled, in such a way that its behaviour satisfies certain requirements, while at the same time maximising a given gain function. A very efficient method for solving optimal control problems for discrete-time systems is the recursive *dynamic programming* technique, introduced by Richard Bellman [4].

7.1.1 The Principle of Optimality

In Figure 7.1 we depict a situation where a system can go from state a to state c through state b in three ways: following the paths $\alpha\beta$, $\alpha\gamma$ and $\alpha\delta$. We denote the gains associated with these paths by $J_{\alpha\beta}$, $J_{\alpha\gamma}$ and $J_{\alpha\delta}$ respectively.

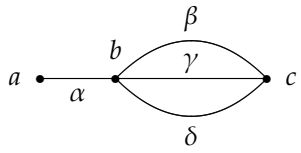


Figure 7.1: Principle of Optimality

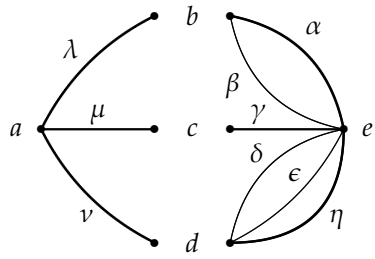


Figure 7.2: Dynamic Programming

Assume that path $\alpha\gamma$ is optimal, meaning that $J_{\alpha\gamma} > J_{\alpha\beta}$ and $J_{\alpha\gamma} > J_{\alpha\delta}$. Then it follows that path γ is the optimal way to go from b to c. To see this, observe that $J_{\alpha\nu} = J_\alpha + J_\nu$ for $\nu \in \{\beta, \gamma, \delta\}$ (we shall assume throughout that gains are additive along paths) and derive from the inequalities above that $J_\gamma > J_\beta$ and $J_\gamma > J_\delta$. This simple observation, which Bellman called the *principle of optimality*, forms the basis for the recursive technique of dynamic programming for solving an optimal control problem. To see how this is done in principle, consider the situation depicted in Figure 7.2. Suppose we want to find the optimal way to go from state a to state e. After one time step, we can reach the states b, c and d from state a, and the optimal paths from these states to the final state e are known to be α, γ and η, respectively. To find the optimal path from a to e, we only need to compare the costs $J_\lambda + J_\alpha$, $J_\mu + J_\gamma$ and $J_\nu + J_\eta$ of the respective candidate optimal paths $\lambda\alpha$, $\mu\gamma$ and $\nu\eta$, since the principle of optimality tells us that the paths $\lambda\beta$, $\nu\delta$ and $\nu\epsilon$ cannot be optimal: if they were, then so would be the paths β, δ and ϵ. This, written down in a more formal language, is what is essentially known as *Bellman's equation*. It allows us to solve an optimal control problem fairly efficiently through a recursive procedure, by calculating optimal paths backwards from the final state.

7.1.2 Uncertain Gain

In applications, it may happen that the gain function, which associates a gain with every possible control action and the resulting behaviour of the system, is not well known. This problem is most often treated by modelling the uncertainty about the gain by means of a probability measure, and by maximising the *expected gain* under this probability measure, as in Section 6.1 on p. 268 ff. Due to the linearity of the expectation operator, this approach does not change the nature of the optimisation problem in any essential way, and the usual dynamic programming method can therefore still be applied.

As an example, consider the simple linear system described by

$$x_{k+1} = ax_k + bu_k, \quad k = 0, \ldots, N-1 \tag{7.1}$$

where $x_k \in \mathbb{R}$ denote the system state and $u_k \in \mathbb{R}$ the control at time k, and where a and b are non-zero real numbers. Given an initial state x_0 and a sequence $u_\bullet$ of successive controls $u_0, u_1, \ldots, u_{N-1}$, the systems goes through the successive states $x_1, x_2, \ldots, x_N$ determined by Eq. (7.1), and we assume that with this control there is associated a gain

$$J(x_0, u_\bullet, \omega) = -\sum_{k=0}^{N-1} [x_k^2 + \omega u_k^2],$$

where ω is some strictly positive real constant. Solving the present optimal control problem consists in finding a control $u_\bullet$ that brings the system at time N in a given final state x_f, while at the same time maximising the gain $J(x_0, u_\bullet, \omega)$. The dynamic programming approach achieves this by reasoning backwards in time. First, the control u_{N-1} is determined that maximises the gain

$$-x_{N-1}^2 - \omega u_{N-1}^2 = -\Big(\frac{x_f - bu_{N-1}}{a}\Big)^2 - \omega u_{N-1}^2.$$

This control also determines a unique x_{N-1}, and the procedure is then repeated by finding a control u_{N-2} that maximises the gain $x_{N-2}^2 + \omega u_{N-2}^2 \ldots$ The principle of optimality then ensures that the $u_\bullet$ found in this recursive manner indeed solves the optimal control problem. When ω is not well known, and only its probability distribution is given, the optimal control problem is solved by maximising the expected value of the gain, *i.e.*, by maximising

expected utility, which can in this special example be done by replacing ω with its expectation.

We have argued in Chapter 6 that optimality cannot always be modelled adequately through maximising expected utility, because, roughly speaking, there may not be enough information available to identify a single probability measure. In those cases, it is more appropriate to represent the available knowledge by an *extended coherent lower prevision*, or what is in many cases mathematically equivalent, a *set of probability measures*. This approach has been applied to estimation and control by for instance Wolfe [93], Chevé and Congar [10], Utkin and Gurov [81], and Kozine and Utkin [52]. In the example above, it may for instance happen that the probability distribution for ω is only known to belong to a given set: e.g., ω is normally distributed with mean zero, but the variance is only known to belong to an interval $[\underline{\sigma}^2, \overline{\sigma}^2]$; or ω itself is only known to belong to an interval $[\underline{\omega}, \overline{\omega}]$.

Two questions now arise naturally. First of all, how should we formulate the optimal control problem: what does it mean for a control to be optimal with respect to an *uncertain gain function*, where the uncertainty is represented through an extended coherent lower prevision? In Chapter 6, under the assumption that our beliefs are not affected by our actions, we have identified five different optimality criteria, each with a different interpretation (although they coincide for precise probability models), and we have studied the relations between them. A second question, which we shall address in this chapter, is whether it is still possible to solve the corresponding optimal control problems using the ideas underlying Bellman's dynamic programming method? We shall show in Section 7.2 that this is the case for only two of the five optimality criteria we have studied: only for these a generalised principle of optimality holds, and the optimal controls are solutions of suitably generalised Bellman-like equations. In order to arrive at this conclusion, we study the properties that an abstract notion of optimality should satisfy for the Bellman approach to work. To illustrate how our ideas can be implemented, we shall present two numerical example in Sections 7.3&7.4.

Of course, other authors, such as Satia and Lave [68], White and Eldeib [90], Givan, Leach, and Dean [36], and Harmanec [41], have extended the dynamic programming algorithm to systems with uncertain gain and/or uncertain dynamics, where the uncertainty is modelled by a set of probability distributions. But none of them seem to have questioned in what sense their

generalised dynamic programming method leads to optimal paths. Here we approach the problem from the opposite, and in our opinion, more logical side: one should *first* define a notion of optimality and investigate whether the dynamic programming argument holds for it, rather than blindly "generalise" Bellman's algorithm without showing that it actually yields optimal controls.

In the remainder of this section, we introduce the basic concepts and notation used in the remainder of this chapter.

7.1.3 The System

For a and b in $\mathbb{N}$, the set of natural numbers c that satisfy $a \leq c \leq b$ is denoted by $[a, b]$. Let

$$x_{k+1} = f(x_k, u_k, k)$$

describe a discrete-time dynamical system with $k \in \mathbb{N}$, $x_k \in \mathcal{X}$ and $u_k \in \mathcal{U}$. The set $\mathcal{X}$ is the state space (e.g., $\mathbb{R}^n$, $n \in \mathbb{N} \setminus \{0\}$), and the set $\mathcal{U}$ is the control space (e.g., $\mathbb{R}^m$, $m \in \mathbb{N} \setminus \{0\}$). The map $f \colon \mathcal{X} \times \mathcal{U} \times \mathbb{N} \to \mathcal{X}$ describes the evolution of the state in time: given the state $x_k \in \mathcal{X}$ and the control $u_k \in \mathcal{U}$ at time $k \in \mathbb{N}$, it returns the next state x_{k+1} of the system. For practical reasons, we impose a final time N beyond which we are not interested in the dynamics of the system. Moreover, it may happen that not all states and controls are allowed at all times: we demand that x_k should belong to a set of *admissible states* $\mathcal{X}_k$ at every instant $k \in [0, N]$, and that u_k should belong to a set of *admissible controls* $\mathcal{U}_k$ at every instant $k \in [0, N-1]$, where $\mathcal{X}_k \subseteq \mathcal{X}$ and $\mathcal{U}_k \subseteq \mathcal{U}$ are given. The set $\mathcal{X}_N$ may be thought of as the set we want the state to end up in at time N.

7.1.4 Paths

A *path* is a triple $(x, k, u_\bullet)$, where $x \in \mathcal{X}$ is a state, $k \in [0, N]$ a time instant, and $u_\bullet \colon [k, N-1] \to \mathcal{U}$ a sequence of controls. Such a path fixes a unique state trajectory $x_\bullet \colon [k, N] \to \mathcal{X}$, which is defined recursively through $x_k = x$ and $x_{\ell+1} = f(x_\ell, u_\ell, \ell)$ for every $\ell \in [k, N-1]$. It is said to be *admissible* if $x_\ell \in \mathcal{X}_\ell$ for every $\ell \in [k, N]$ and $u_\ell \in \mathcal{U}_\ell$ for every $\ell \in [k, N-1]$. We denote the unique map from the empty set $\emptyset$ to $\mathcal{U}$ by $u_\emptyset$. If $k = N$, the control $u_\bullet$ does nothing: it is equal to $u_\emptyset$. The unique path starting and ending at time $k = N$ in $x \in \mathcal{X}$

is denoted by $(x, N, u_\emptyset)$.

The set of admissible paths starting in the state $x \in \mathcal{X}_k$ at time $k \in [0, N]$ is denoted by $\mathcal{U}(x, k)$, i.e.,

$$\mathcal{U}(x,k) = \{(x,k,u_\bullet)\colon (x,k,u_\bullet) \text{ admissible path}\}.$$

For example, $\mathcal{U}(x, N) = \{(x, N, u_\emptyset)\}$ whenever $x \in \mathcal{X}_N$ and $\mathcal{U}(x, N) = \emptyset$ otherwise.

If we consider a path with final time M different from N, then we write $(x, k, u_\bullet)_M$ (assume $k \le M \le N$). Observe that $(x, k, u_\bullet)_k$ can be identified with $(x, k, u_\emptyset)_k$; it is the unique path (of length zero) starting and ending at time k in x. Let $0 \le k \le \ell \le m$. Two paths $(x, k, u_\bullet)_\ell$ and $(y, \ell, v_\bullet)_m$ can be concatenated if $y = x_\ell$. The concatenation is denoted by $(x, k, u_\bullet, \ell, v_\bullet)_m$ or by $(x, k, u_\bullet)_\ell \oplus (y, \ell, v_\bullet)_m$. It represents the path that starts in state x at time k, and results from applying control u_i for times $i \in [k, \ell - 1]$ and control v_i for times $i \in [\ell, m - 1]$. In particular,

$$(x,k,u_\bullet)_\ell = (x,k,u_\bullet)_k \oplus (x,k,u_\bullet)_\ell = (x,k,u_\bullet)_\ell \oplus (x_\ell,\ell,u_\bullet)_\ell.$$

The set of admissible paths starting in state $x \in \mathcal{X}_k$ at time $k \in [0, N]$ and ending at time $\ell \in [k, N]$ is denoted by $\mathcal{U}(x, k)_\ell$. In particular we have that $\mathcal{U}(x, k)_k = \{(x, k, u_\emptyset)_k\}$ if $x \in \mathcal{X}_k$, and $\mathcal{U}(x, k)_k = \emptyset$ otherwise. Moreover, for any $(x, k, u_\bullet)_\ell \in \mathcal{U}(x, k)_\ell$ and any $\mathcal{V} \subseteq \mathcal{U}(x_\ell, \ell)$, we use the notation

$$(x,k,u_\bullet)_\ell \oplus \mathcal{V} = \{(x,k,u_\bullet)_\ell \oplus (x_\ell,\ell,v_\bullet)\colon (x_\ell,\ell,v_\bullet) \in \mathcal{V}\}.$$

7.1.5 The Gain Function

We assume that applying the control action $u \in \mathcal{U}$ to the system in state $x \in \mathcal{X}$ at time $k \in [0, N-1]$ yields a real-valued gain $g(x, u, k, \omega)$. Moreover, reaching the final state $x \in \mathcal{X}$ at time N also yields a gain $h(x, \omega)$. The parameter ω collects all variables that influence the gain. If we knew its value, say ω^*, we would know the gains to be $g(x, u, k, \omega^*)$ and $h(x, \omega^*)$. As it is, the value of ω is uncertain, so we shall consider ω as the uncertain outcome of a random variable Ω that takes values in a set $\mathbf{\Omega}$. So the gains are uncertain as well, and we shall consider them as random quantities on Ω. It is important to note that Ω only influences the gains; it has no effect on the system dynamics,

which are assumed to be known perfectly well. We shall also assume that our beliefs about Ω are influenced nor by the initial state of the system, nor by the sequence of controls applied to the system: essentially, this is what we have called *act-state independence* in Chapter 6—beware of the fact that the standard terminology is rather unfortunate in the context of systems theory, as the word "state" in "act-state dependence" refers to the random variable Ω, and not the state of the system.

We shall only consider the important case that the gains are *additive along paths*, i.e., with a path $(x, k, u_\bullet)$ we associate a gain $J(x, k, u_\bullet, \omega)$ given by:

$$J(x, k, u_\bullet, \omega) = \sum_{i=k}^{N-1} g(x_i, u_i, i, \omega) + h(x_N, \omega),$$

for any $\omega \in \boldsymbol{\Omega}$ (gain additivity). If $M < N$, we also use the notation

$$J(x, k, u_\bullet, \omega)_M = \sum_{i=k}^{M-1} g(x_i, u_i, i, \omega).$$

It will be convenient to associate a zero gain with an empty control action: for $k \in [0, N]$ we let $J(x, k, u_\bullet, \omega)_k = 0$.

The main objective of optimal control can now be formulated as follows: given that the system is in the initial state $x \in \mathcal{X}$ at time $k \in [0, N]$, find a control sequence $u_\bullet \colon [k, N-1] \to \mathcal{U}$ resulting in an admissible path $(x, k, u_\bullet)$ such that the corresponding gain $J(x, k, u_\bullet, \omega)$ is maximal. Moreover, we would like this control sequence $u_\bullet$ to be such that its value u_k at time k is a function of x and k only, since in that case the control can be realised through state feedback.

If ω is known, then the problem reduces to the classical problem of dynamic programming, first studied and solved by Bellman [4]. We shall assume here that the available information about the true state of the world is modelled through an *extended coherent lower prevision* $\underline{P}$ defined on a sufficiently large set $\operatorname{dom} \underline{P}$ of random quantities on Ω. A special case of this obtains when $\underline{P}$ is an extended linear prevision. Recall from Chapter 5 that extended linear previsions are precise probability models; they are *previsions* or *fair prices* in the sense of de Finetti [27], and in many cases, they can be interpreted as expectation operators associated with probability measures, or more generally, probability charges.

For a given path $(x, k, u_\bullet)$, the corresponding gain $J(x, k, u_\bullet, \omega)$ can be seen as a real-valued map on $\mathbf{\Omega}$, which is denoted by $J(x, k, u_\bullet)$ and is called the *gain random quantity* associated with $(x, k, u_\bullet)$. In the same way we define the gain random quantities $g(x_k, u_k, k)$, $h(x_N)$ and $J(x, k, u_\bullet)_M$. There is gain additivity: $J(x, k, u_\bullet, \ell, v_\bullet)_m = J(x, k, u_\bullet)_\ell + J(x_\ell, \ell, v_\bullet)_m$ for $k \leq \ell \leq m \leq N$, and $J(x, k, u_\bullet)_k = 0$. We denote by $\mathcal{J}(x, k)$ the set of gain random quantities for admissible paths from initial state $x \in \mathcal{X}_k$ at time $k \in [0, N]$:

$$\mathcal{J}(x, k) = \{J(x, k, u_\bullet) \colon (x, k, u_\bullet) \in \mathcal{U}(x, k)\}.$$

We shall assume that $\underline{P}$ ***is a real-valued coherent extended lower prevision, and that*** $\operatorname{dom} \underline{P}$ ***is a linear lattice that contains at least all constant random quantities and all gain random quantities*** $g(x_k, u_k, k)$ ***and*** $h(x_N)$***, for all*** $k \in [0, N-1]$***, all*** x_k ***in*** $\mathcal{X}_k$***, all*** u_k ***in*** $\mathcal{U}_k$***, and all*** x_N ***in*** $\mathcal{X}_N$**.** As a consequence, we can endow $\operatorname{dom} \underline{P}$ with a topology that suits our purpose very well, and we can also associate with $\underline{P}$ a compact convex set $\mathcal{M}$ of extended linear previsions defined on $\operatorname{dom} \underline{P}$ such that $\underline{P}(f) = \min_{Q \in \mathcal{M}} Q(f)$ for any f in $\operatorname{dom} \underline{P}$; this has been discussed in Section 6.2.2 on p. 272 ff.

7.2 Conditions for Dynamic Programming Under a General Notion of Optimality

In Chapter 6, we have discussed five different ways of associating optimal paths with a lower prevision $\underline{P}$, all of which occur in the literature. We now propose to find out whether, for these different types of optimality, we can use the ideas behind the dynamic programming method to solve the corresponding optimal control problems.

To this end, we take a closer look at Bellman's analysis as described in Section 7.1, and we investigate which properties a generic notion of optimality must satisfy for his method to work. Let us therefore assume that there is some property, called $*$*-optimality*, which a path in a given set of paths $\mathcal{P}$ either has or does not have. If a path in $\mathcal{P}$ has this property, we say that it is $*$-optimal in $\mathcal{P}$. We shall denote the set of the $*$-optimal elements of $\mathcal{P}$ by $\operatorname{opt}_*(\mathcal{P})$. By definition, $\operatorname{opt}_*(\mathcal{P}) \subseteq \mathcal{P}$. Further on, we shall apply our findings to the various instances of $*$-optimality described above.

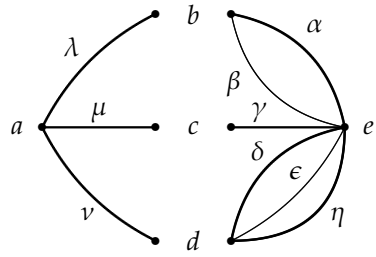


Figure 7.3: A More General Type of Dynamic Programming

We may also view opt_* as an operator on subsets of a space $\mathcal{P}$, that maps any subset T of $\mathcal{P}$ to a (possibly empty) subset $\text{opt}_*(T)$ of that set T. Social choice functions are a particular type of optimality operators, which additionally satisfy $\text{opt}_*(T) \neq \emptyset$ whenever $T \neq \emptyset$.

7.2.1 The Principle of Optimality

Consider Figure 7.3, where we want to find the $*$-optimal paths from state a to state e. Suppose that after one time step, we can reach the states b, c and d from state a. The $*$-optimal paths from these states to the final state e are known to be α, γ, and δ and η, respectively. For the dynamic programming approach to work, we need to be able to infer from this a generalised form of the Bellman equation, stating essentially that the $*$-optimal paths from a to e, *a priori* given by $\text{opt}_*(\{\lambda\alpha, \lambda\beta, \mu\gamma, \nu\delta, \nu\epsilon, \nu\eta\})$, are actually also given by $\text{opt}_*(\{\lambda\alpha, \mu\gamma, \nu\delta, \nu\eta\})$, i.e., the $*$-optimal paths in the set of concatenations of λ, μ and ν with the respective $*$-optimal paths α, γ, and δ and η. It is therefore necessary to exclude that the concatenations $\lambda\beta$ and $\nu\epsilon$ with the non-$*$-optimal paths β and ϵ can be $*$-optimal. This amounts to requiring that the operator opt_* should satisfy some appropriate generalisation of Bellman's *principle of optimality* that will allow us to conclude that $\lambda\beta$ and $\nu\epsilon$ cannot be $*$-optimal because then β and ϵ would be $*$-optimal as well. Definition 7.3 below provides a precise general formulation.

7.2.2 Insensitivity to The Omission of Non-Optimal Paths

But, perhaps surprisingly for someone familiar with the traditional form of dynamic programming, opt_* should satisfy an *additional* property: the omission of the non-$*$-optimal paths $\lambda\beta$ and $\nu\epsilon$ from the set of candidate $*$-optimal paths should not have any effect on the actual $*$-optimal paths: we need that

$$\text{opt}_*(\{\lambda\alpha, \lambda\beta, \mu\gamma, \nu\delta, \nu\epsilon, \nu\eta\}) = \text{opt}_*(\{\lambda\alpha, \mu\gamma, \nu\delta, \nu\eta\}).$$

This is obviously true for the simple type of optimality that we have looked at in Section 7.1, but it need not be true for the more abstract types that we want to consider here. Equality will be guaranteed if opt_* is insensitive to the omission of non-$*$-optimal elements from $\{\lambda\alpha, \lambda\beta, \mu\gamma, \nu\delta, \nu\epsilon, \nu\eta\}$, in the following sense.

Definition 7.1. Consider a set $S \neq \emptyset$ and an *optimality operator* opt_* defined on the set $\wp(S)$ of subsets of S such that $\text{opt}_*(T) \subseteq T$ for all $T \subseteq S$. Elements of $\text{opt}_*(T)$ are called *$*$-optimal in T*. The optimality operator opt_* is called *insensitive to the omission of non-$*$-optimal elements from S* if $\text{opt}_*(S) = \text{opt}_*(T)$ for all T such that $\text{opt}_*(S) \subseteq T \subseteq S$.

The following proposition gives an interesting sufficient condition for this insensitivity in case optimality is associated with a (family of) strict partial order(s): it suffices that every non-optimal path is strictly dominated by an optimal path.

Proposition 7.2. *Let S be a non-empty set provided with a family of strict partial orders $>_j$, $j \in J$. Define for $T \subseteq S$, $\text{opt}_{>_j}(T) = \left\{a \in T : (\forall b \in T)(b \not> _j a)\right\}$ as the set of maximal elements of T with respect to $>_j$, and let $\text{opt}_J(T) = \bigcup_{j\in J} \text{opt}_{>_j}(T)$. Then $\text{opt}_{>_j}$, $j \in J$ and opt_J are optimality operators. If for some $j \in J$,*

$$(\forall a \in S \setminus \text{opt}_{>_j}(S))(\exists b \in \text{opt}_{>_j}(S))(b >_j a), \tag{7.2}$$

then $\text{opt}_{>_j}$ is insensitive to the omission of non-$>_j$-optimal elements from S. If Eq. (7.2) holds for all $j \in J$, then opt_J is insensitive to the omission of non-J-optimal elements from S.

Proof. Consider j in J, and assume that Eq. (7.2) holds for this j. Let $\text{opt}_{>_j}(S) \subseteq T \subseteq S$, then we must prove that $\text{opt}_{>_j}(S) = \text{opt}_{>_j}(T)$. First of all, if $a \in \text{opt}_{>_j}(S)$

then $b \not>_j a$ for all b in S, and *a fortiori* for all b in T, whence $a \in \mathrm{opt}_{>_j}(T)$. Consequently, $\mathrm{opt}_{>_j}(S) \subseteq \mathrm{opt}_{>_j}(T)$. Conversely, let $a \in \mathrm{opt}_{>_j}(T)$ and assume *ex absurdo* that $a \notin \mathrm{opt}_{>_j}(S)$. It then follows from Eq. (7.2) that there is some c in $\mathrm{opt}_{>_j}(S)$ and therefore in T such that $c >_j a$, which contradicts $a \in \mathrm{opt}_{>_j}(T)$.

Next, assume that Eq. (7.2) holds for all $j \in J$. Let $\mathrm{opt}_J(S) \subseteq T \subseteq S$, then we must prove that $\mathrm{opt}_J(S) = \mathrm{opt}_J(T)$. Consider any $j \in J$, then $\mathrm{opt}_{>_j}(S) \subseteq \mathrm{opt}_J(S) \subseteq T \subseteq S$, so we may infer from the first part of the proof that $\mathrm{opt}_{>_j}(S) = \mathrm{opt}_{>_j}(T)$. By taking the union over all $j \in J$, we find that indeed $\mathrm{opt}_J(S) = \mathrm{opt}_J(T)$. □

We are now ready for a precise formulation of the dynamic programming approach for solving optimal control problems associated with general types of optimality. We assume that we have some type of optimality, called $*$-optimality, that allows us to associate with the set of admissible paths $\mathcal{U}(x,k)$ starting at time k in initial state x, an optimality operator opt_* defined on the set $\wp(\mathcal{U}(x,k))$ of subsets of $\mathcal{U}(x,k)$. For each such subset $\mathcal{V}$, $\mathrm{opt}_*(\mathcal{V})$ is then the set of admissible paths that are $*$-optimal in $\mathcal{V}$. The principle of optimality states that the optimality operators associated with the various $\mathcal{U}(x,k)$ should be related in a special way.

Definition 7.3 (Principle of Optimality)**.** $*$-optimality satisfies the *principle of optimality* if it holds for all $k \in [0,N]$, $x \in \mathcal{X}_k$, $\ell \in [k,N]$ and $(x,k,u_\bullet)$ in $\mathcal{U}(x,k)$ that if $(x,k,u_\bullet)$ is $*$-optimal in $\mathcal{U}(x,k)$, then $(x_\ell,\ell,u_\bullet)$ is $*$-optimal in $\mathcal{U}(x_\ell,\ell)$.

This may also be expressed as:

$$\mathrm{opt}_*(\mathcal{U}(x,k)) \subseteq \bigcup_{(x,k,u_\bullet)_\ell \in \mathcal{U}(x,k)_\ell} (x,k,u_\bullet)_\ell \oplus \mathrm{opt}_*(\mathcal{U}(x_\ell,\ell))\,.$$

The Bellman equation now states that applying the optimality operator to the right hand side suffices to achieve equality. (Usually this is stated with $\ell = k+1$.)

Theorem 7.4 (Bellman Equation)**.** *Let $k \in [0,N]$ and $x \in \mathcal{X}_k$. Assume that $*$-optimality satisfies the principle of optimality, and that the optimality operator* opt_* *for $\mathcal{U}(x,k)$ is insensitive to the omission of non-$*$-optimal elements from $\mathcal{U}(x,k)$. Then*

for all $\ell \in [k, N]$*:*

$$\mathrm{opt}_*\left(\mathcal{U}(x,k)\right) = \mathrm{opt}_*\left(\bigcup_{(x,k,u)_\ell \in \mathcal{U}(x,k)_\ell} (x,k,u)_\ell \oplus \mathrm{opt}_*\left(\mathcal{U}(x_\ell,\ell)\right)\right),$$

that is, a path is $*$*-optimal if and only if it is a* $*$*-optimal concatenation of an admissible path* $(x,k,u_\bullet)_\ell$ *and a* $*$*-optimal path of* $\mathcal{U}(x_\ell,\ell)$*.*

Proof. Fix k in $[0,N]$, $\ell \in [k,N]$ and $x \in \mathcal{X}_k$. Define

$$\mathcal{V}_1 = \bigcup_{(x,k,u)_\ell \in \mathcal{U}(x,k)_\ell} (x,k,u)_\ell \oplus \mathrm{opt}_*\left(\mathcal{U}(x_\ell,\ell)\right), \text{ and,}$$
$$\mathcal{V}_2 = \bigcup_{(x,k,u)_\ell \in \mathcal{U}(x,k)_\ell} (x,k,u)_\ell \oplus \Big(\mathcal{U}(x_\ell,\ell) \setminus \mathrm{opt}_*\left(\mathcal{U}(x_\ell,\ell)\right)\Big).$$

Obviously, $\mathcal{U}(x,k) = \mathcal{V}_1 \cup \mathcal{V}_2$ and $\mathcal{V}_1 \cap \mathcal{V}_2 = \emptyset$. We have to prove that $\mathrm{opt}_*\left(\mathcal{U}(x,k)\right) = \mathrm{opt}_*\left(\mathcal{V}_1\right)$. By the principle of optimality, no path in $\mathcal{V}_2$ is $*$-optimal in $\mathcal{U}(x,k)$, so $\mathcal{V}_2 \cap \mathrm{opt}_*\left(\mathcal{U}(x,k)\right) = \emptyset$. This implies that $\mathrm{opt}_*\left(\mathcal{U}(x,k)\right) \subseteq \mathcal{V}_1 \subseteq \mathcal{U}(x,k)$, and since opt_* is assumed to be insensitive to the omission of non-$*$-optimal elements from $\mathcal{U}(x,k)$, it follows that $\mathrm{opt}_*\left(\mathcal{U}(x,k)\right) = \mathrm{opt}_*\left(\mathcal{V}_1\right)$. □

Let us now apply these general results to the specific types of optimality introduced in Chapter 6. For all five optimality operators $\mathrm{opt}_{\geqslant_{\underline{P}}}$, $\mathrm{opt}_{\mathcal{M}}$, $\mathrm{opt}_{\sqsupseteq_{\underline{P}}}$, $\mathrm{opt}_{\sqsupseteq_{\overline{P}}}$, and $\mathrm{opt}_{\supseteq_{\underline{P}}}$, we shall check whether we can use a Bellman equation to solve the corresponding optimal control problem.

7.2.3 $\underline{P}$-Maximality

We first consider the optimality operator $\mathrm{opt}_{\geqslant_{\underline{P}}}$ that selects from a set of paths S those that are the maximal elements of S with respect to the preorder $\geqslant_{\underline{P}}$. The following lemma roughly states that strict preference amongst paths with respect to $>_{\underline{P}}$ is preserved under concatenation and truncation. As a result, the principle of optimality with respect to $\underline{P}$-maximality holds.

Lemma 7.5. *Let* $k \in [0,N]$ *and* $\ell \in [k,N]$*. Consider the paths* $(x,k,u_\bullet)_\ell$ *in* $\mathcal{U}(x,k)_\ell$ *and* $(x_\ell,\ell,v_\bullet)$*,* $(x_\ell,\ell,w_\bullet)$ *in* $\mathcal{U}(x_\ell,\ell)$*. Then* $(x_\ell,\ell,v_\bullet) >_{\underline{P}} (x_\ell,\ell,w_\bullet)$ *if and only if* $(x,k,u_\bullet)_\ell \oplus (x_\ell,\ell,v_\bullet) >_{\underline{P}} (x,k,u_\bullet)_\ell \oplus (x_\ell,\ell,w_\bullet)$*.*

Proof. Let f, g, and h be random quantities in $\operatorname{dom} \underline{P}$. The statement is proved if we can show that $g >_{\underline{P}} h$ if and only if $f + g >_{\underline{P}} f + h$. But $\underline{P}(g - h) > 0$ if and only if $\underline{P}((f + g) - (f + h)) = \underline{P}(g - h) > 0$, $g \geq h$ if and only if $f + g \geq f + h$, and finally, $g \neq h$ if and only if $f + g \neq f + h$. This establishes the equivalence. □

Proposition 7.6 (Principle of Optimality)**.** *Let $k \in [0, N]$, $x \in \mathcal{X}_k$ and $(x, k, u^*_\bullet) \in \mathcal{U}(x, k)$. If $(x, k, u^*_\bullet)$ is $\underline{P}$-maximal in $\mathcal{U}(x, k)$ then $(x_\ell, \ell, u^*_\bullet)$ is $\underline{P}$-maximal in $\mathcal{U}(x_\ell, \ell)$ for all $\ell \in [k, N]$.*

Proof. If $(x_\ell, \ell, u^*_\bullet)$ is not $\underline{P}$-maximal, then there is a path $(x_\ell, \ell, u_\bullet)$ such that $(x_\ell, \ell, u_\bullet) >_{\underline{P}} (x_\ell, \ell, u^*_\bullet)$. By Lemma 7.5 we find that

$$(x, k, u^*_\bullet)_\ell \oplus (x_\ell, \ell, u_\bullet) >_{\underline{P}} (x, k, u^*_\bullet)_\ell \oplus (x_\ell, \ell, u^*_\bullet) = (x, k, u^*_\bullet).$$

This means that $(x, k, u^*_\bullet)_\ell \oplus (x_\ell, \ell, u_\bullet)$ is preferred to $(x, k, u^*_\bullet)$, and therefore $(x, k, u^*_\bullet)$ cannot be $\underline{P}$-maximal, a contradiction. □

As a direct consequence of Theorem 6.9 on p. 283, and Proposition 7.2, we see that if $\mathcal{J}(x, k)$ is compact, then the optimality operator $\operatorname{opt}_{\gtrdot_{\underline{P}}}$ associated with $\mathcal{U}(x, k)$ is insensitive to the omission of non-$\underline{P}$-maximal elements. Together with Proposition 7.6 and Theorem 7.4, this allows us to infer a Bellman equation for $\underline{P}$-maximality.

Corollary 7.7. *Let $k \in [0, N]$ and $x \in \mathcal{X}_k$. If $\mathcal{J}(x, k)$ is compact, then for all $\ell \in [k, N]$*

$$\operatorname{opt}_{\gtrdot_{\underline{P}}}(\mathcal{U}(x, k)) = \operatorname{opt}_{\gtrdot_{\underline{P}}}\left(\bigcup_{(x,k,u)_\ell \in \mathcal{U}(x,k)_\ell} (x, k, u)_\ell \oplus \operatorname{opt}_{\gtrdot_{\underline{P}}}(\mathcal{U}(x_\ell, \ell)) \right), \tag{7.3}$$

that is, a path is $\underline{P}$-maximal if and only if it is a $\underline{P}$-maximal concatenation of an admissible path $(x, k, u_\bullet)_\ell$ and a $\underline{P}$-maximal path of $\mathcal{U}(x_\ell, \ell)$.

Corollary 7.7 results in a procedure to calculate all $\underline{P}$-maximal paths. Indeed, $\operatorname{opt}_{\gtrdot_{\underline{P}}}(\mathcal{U}(x, N)) = \{u_\emptyset\}$ for every $x \in \mathcal{X}_N$, and $\operatorname{opt}_{\gtrdot_{\underline{P}}}(\mathcal{U}(x, k))$ can be calculated recursively through Eq. (7.3). It also provides a method for constructing a $\underline{P}$-maximal feedback: for every $x \in \mathcal{X}_k$, choose any $(x, k, u^*_\bullet(x, k)) \in \operatorname{opt}_{\gtrdot_{\underline{P}}}(\mathcal{U}(x, k))$. Then $\phi(x, k) = u^*_k(x, k)$ realises a $\underline{P}$-maximal feedback.

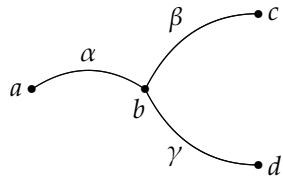


Figure 7.4: A Counterexample

7.2.4 $\mathcal{M}$-Maximality

We now turn to the optimality operator $\text{opt}_{\mathcal{M}}$, defined through Eq. (6.7) on p. 286. If we recall Proposition 7.2, we see that $\text{opt}_{\mathcal{M}}$ is insensitive to the omission of non-$\mathcal{M}$-maximal elements of $\mathcal{U}(x,k)$ whenever $\mathcal{J}(x,k)$ is compact. By Proposition 7.6, $\text{opt}_{\mathcal{M}}$ satisfies the principle of optimality (indeed, if a path is $\mathcal{M}$-maximal, then it must be P-maximal for some $P \in \mathcal{M}$, and by the proposition any truncation of it is also P-maximal, hence also $\mathcal{M}$-maximal). This means that the Bellman equation also holds for $\mathcal{M}$-maximality under similar conditions as for $\underline{P}$-maximality. As shown in Theorem 6.17 on p. 288, both types of optimality coincide if $\mathcal{J}(x,k)$ is convex.

7.2.5 $\underline{P}$-Maximinity and $\underline{P}$-Maximaxity

We come to the types of optimality associated with the strict partial orders $\sqsupset_{\underline{P}}$ and $\sqsupset_{\overline{P}}$. It follows from Theorem 6.24 on p. 291 and Proposition 7.2 that if $\mathcal{J}(x,k)$ is compact, the optimality operator $\text{opt}_{\sqsupset_{\underline{P}}}$ is insensitive to the omission of non-$\underline{P}$-maximin paths from $\mathcal{U}(x,k)$, and a similar observation holds for $\text{opt}_{\sqsupset_{\overline{P}}}$. But, as the following counterexample shows, the principle of optimality holds for neither $\underline{P}$-maximinity, nor $\underline{P}$-maximaxity, and therefore the dynamic programming approach may not work here. Essentially, this is because the preorders $\sqsupseteq_{\underline{P}}$ and $\sqsupseteq_{\overline{P}}$ are not vector orderings—they are not compatible with gain additivity: contrary to expected gains, lower and upper expected gains are not additive.

Example 7.8. Consider the dynamical system depicted in Figure 7.4. Let $\mathbf{\Omega} = \{\sharp, \flat\}$ and denote the random quantity that maps $\sharp \mapsto x$ and $\flat \mapsto y$ by $\langle x, y\rangle$. Let $\underline{P}$ be the vacuous lower prevision on Ω, defined by $\underline{P}(\langle x, y\rangle) = \min\{x, y\}$. Assume that $J(\alpha) = \langle 2, 0\rangle$, $J(\beta) = \langle 0, -1\rangle$ and $J(\gamma) = \langle -2, 0\rangle$ (there is zero gain

associated with the final state). Then $\alpha\beta \not\trianglerighteq_{\underline{P}} \alpha\gamma$: indeed, $\langle 2,-1\rangle$ does not dominate $\langle 0,0\rangle$ point-wise, and $\underline{P}(\langle 2,-1\rangle) = \min\{2,-1\} \not> \min\{0,0\} = \underline{P}(\langle 0,0\rangle)$ or equivalently $\langle 0,0\rangle$ maximises the smallest expected gain. Hence, we find that $\alpha\gamma$ is $\underline{P}$-maximin. But $\beta \sqsupset_{\underline{P}} \gamma$: indeed, $\underline{P}(\langle 0,-1\rangle) = \min\{0,-1\} > \min\{-2,0\} = \underline{P}(\langle -2,0\rangle)$, which means that γ is not $\underline{P}$-maximin. So $\underline{P}$-maximinity does not satisfy the principle of optimality.

It's not so hard to construct a similar counterexample for $\underline{P}$-maximaxity.

7.2.6 Weak $\underline{P}$-Maximality

It can be shown easily that if $\mathcal{J}(x,k)$ is compact, the optimality operator $\text{opt}_{\trianglerighteq_{\underline{P}}}$ on $\mathcal{U}(x,k)$ is insensitive to the omission of non-weakly-$\underline{P}$-maximal paths from $\mathcal{U}(x,k)$. But, as the following counterexample shows, we cannot guarantee that the principle of optimality holds for weak $\underline{P}$-maximality, and therefore the dynamic programming approach may not work here. Again, this is because the partial order $\trianglerighteq_{\underline{P}}$ is not compatible with gain additivity. It also indicates that by solving the Bellman-type equation advocated in [41], we will not necessarily get paths that are optimal in the sense described above.

Example 7.9. Consider again the dynamical system depicted in Figure 7.4. As before, let $\boldsymbol{\Omega} = \{\sharp, \flat\}$, let $\underline{P}$ be the vacuous lower prevision on Ω, and denote the random quantity that maps $\sharp \mapsto x$ and $\flat \mapsto y$ by $\langle x, y\rangle$. Assume that $J(\alpha) = \langle 2,0\rangle$, $J(\beta) = \langle 0,0\rangle$ and $J(\gamma) = \langle -1,-1\rangle$ (there is zero gain associated with the final state). Then $\alpha\beta \not\trianglerighteq_{\underline{P}} \alpha\gamma$: indeed, $\langle 2,0\rangle$ does not dominate $\langle 1,-1\rangle$ point-wise, and, $\underline{P}(\langle 2,0\rangle) = \min\{2,0\} \not> \max\{1,-1\} = \overline{P}(\langle 1,-1\rangle)$. Hence, we find that $\alpha\gamma$ is weakly $\underline{P}$-maximal. But $\beta \sqsupset_{\underline{P}} \gamma$: indeed, $\underline{P}(\langle 0,0\rangle) = \min\{0,0\} > \max\{-1,-1\} = \overline{P}(\langle -1,-1\rangle)$, which means that γ is not weakly $\underline{P}$-maximal. So, weak $\underline{P}$-maximality does not satisfy the principle of optimality.

7.3 Example: Pol&Jo versus Renthouse

Suppose we have a total amount of money x at our disposal, which we can invest into two companies, named Pol&Jo and Renthouse, and which we shall simply denote by 0 and 1. We denote our investment in company 0 by u_0, and in company 1 by u_1. Observe that x, u_0 and u_1 are non-negative real

numbers, and $u_0 + u_1 \leq x$. The total gain is

$$J(x, u_0, u_1) = \omega_0 u_0 + \omega_1 u_1 + \omega_2 (x - u_0 - u_1),$$

where $\omega_0 > 0$, $\omega_1 > 0$ are gain factors (for companies 0 and 1), and $\omega_2 > 0$ is the devaluation factor (of the money we have not invested). We wish to maximise the gain, but, we are uncertain about ω_0, ω_1 and ω_2. We know that $\omega_0 = 1 + g_0 + \epsilon$ and $\omega_1 = 1 + g_1 + \epsilon$. g_0 and g_1 model the productivities of the companies, and ϵ models economical variations that affect each company in the same way, such as the global economical state. We do not make any assumption about the dependence between g_0, g_1, ϵ and ω_2. We only know that $g_0 \in [0.0, 0.3]$, $g_1 \in [0.1, 0.2]$, $\epsilon \in [-0.1, 0.2]$ and $\omega_2 \in [0.85, 0.95]$. This leads to the following lower prevision on the set $\mathscr{L}(\Omega_0, \Omega_1, \Omega_2)$ of bounded random quantities on $(\Omega_0, \Omega_1, \Omega_2)$:

$$\underline{P}(f) = \inf \Big\{ f(1 + g_0 + \epsilon, 1 + g_1 + \epsilon, \omega_2) : \\ g_0 \in [0.0, 0.3], g_1 \in [0.1, 0.2], \epsilon \in [-0.1, 0.2], \omega_2 \in [0.85, 0.95] \Big\}.$$

This $\underline{P}$ is intuitively acceptable as a supremum buying price for any bounded random quantity f on $\Omega_0 \times \Omega_1 \times \Omega_2$, as we have explained in Section 3.5.1 on p. 60, Eq. (3.5) in particular.

We formulate this problem in terms of a dynamical system. If we define $x_0 = x$ and, recursively $x_{k+1} = x_k - u_k$, the total gain is precisely equal to $J(x, u_\bullet, 0)$, with $g(x_k, u_k, k, \omega) = \omega_k u_k$ and $h(x_2, \omega) = \omega_2 x_2$. Each state x_k represents the money we can invest in companies $\ell \geq k$, and should therefore be non-negative. Obviously, there is gain additivity.

Note that the gain random quantities $\omega_0 u_0$, $\omega_1 u_1$, and $\omega_2 x_2$, are $\underline{P}$-essentially bounded, so, for the purpose of optimal control, it suffices to consider the (real-valued) extended lower prevision $\underline{P}^\sharp$ defined on the set (linear lattice) of all $\underline{P}$-essentially bounded random quantities; see Chapter 5, Section 5.3 on p. 225 ff, Eq. (5.12) on p. 234 in particular. We now wish to find all u_0 and u_1 such that the gain $J(x, u_0, u_1)$ is $\underline{P}^\sharp$-maximal. Observe that this is a two-dimensional optimisation problem. For any initial state x, the set $\mathscr{J}(x, 0)$ of admissible gain random quantities is compact: the set $A = \{(u_0, u_1) : u_0 \geq 0, u_1 \geq 0, u_0 + u_1 \leq 1\}$ is closed and bounded, and therefore compact with respect to the usual topology on $\mathbb{R}^2$, and $J(x, \bullet, \bullet)$, as a mapping from A to

$\mathcal{K}^\sharp_{\underline{P}}(\Omega_1, \Omega_2, \Omega_3)$, is continuous: if $u_0^\alpha \to u_0$ and $u_1^\alpha \to u_1$, then

$$u_0^\alpha \omega_0 + u_1^\alpha \omega_1 + (x - u_1^\alpha - u_2^\alpha)\omega_2 \to u_0\omega_0 + u_1\omega_1 + (x - u_1 - u_2)\omega_2$$

for all $(\omega_0, \omega_1, \omega_2)$ in $\mathbf{\Omega}_0 \times \mathbf{\Omega}_1 \times \mathbf{\Omega}_2$ (point-wise convergence), and

$$\begin{aligned}\overline{P}^\sharp(&|J(x, u_0^\alpha, u_1^\alpha) - J(x, u_0, u_1)|) \\ &\leq (1 + 0.3 + 0.2 - 0.85)|u_0^\alpha - u_0| + (1 + 0.2 + 0.2 - 0.85)|u_0^\alpha - u_0| \to 0\end{aligned}$$

($\underline{P}^\sharp$-norm convergence).

So, there is gain additivity, and $\mathcal{J}(x, 0)$ is compact, so Corollary 7.7 applies: we can solve this problem using dynamic programming.

For $k = 1$, we find that the control $u_1 = x_1$ is optimal from state x_1 at time 1. Indeed, first observe that all controls are maximal with respect to the point-wise order. In that case, optimality of u_1 is equivalent to $\overline{P}^\sharp(J(x_1, u_1, 1) - J(x_1, v_1, 1)) \geq 0$ for all v_1. This holds if and only if

$$\begin{aligned}\sup\Big\{(1 + g_1 + \epsilon - \omega_2)(u_1 - v_1)\colon \\ g_1 \in [0.1, 0.2], \epsilon \in [-0.1, 0.2], \omega_2 \in [0.85, 0.95]\Big\} \geq 0,\end{aligned}$$

and thus, if and only if $u_1 \geq v_1$ for all v_1. Hence, optimal paths maximise u_1. The highest u_1 we can choose such that x_2 is still non-negative is $u_1 = x_1$.

For $k = 0$, the dynamic programming argument says that we only have to consider concatenations of $(x_0, u_0, 0)_1$ with optimal paths from state $x_1 = x_0 - u_0$, of which there is only one, $(x_1, x_1, 1)$, as we showed. Again all controls are maximal with respect to the point-wise order. But

$$\overline{P}^\sharp\Big(J((x_0, u_0, 0)_1 \oplus (x_0 - u_0, x_0 - u_0, 1)) - J((x_0, v_0, 0)_1 \oplus (x_0 - v_0, x_0 - v_0, 1))\Big) \geq 0$$

also holds for any u_0 and any v_0. Indeed, the inequality is equivalent to

$$\sup\Big\{(g_0 - g_1)(u_0 - v_0)\colon g_0 \in [0.0, 0.3], g_1 \in [0.1, 0.2], \epsilon \in [-0.1, 0.2]\Big\} \geq 0$$

which obviously holds for any choice of u_0 and v_0. Thus, all paths $(x_0, u_0, 0)_1 \oplus (x_0 - u_0, x_0 - u_0, 1)$ are optimal.

In conclusion, the information implies that we should invest all money x,

but we cannot infer how we should divide x over the two companies.

By our dynamic programming approach we have have managed to solve this two-dimensional optimisation problem by reducing it to two one-dimensional ones, which are each very easy to solve. In the more general case of uncertain investment with n companies, we initially have a n-dimensional optimisation problem, and dynamic programming reduces this to n very simple one-dimensional optimisation problems.

Let's now discuss a more sophisticated example.

7.4 Example: Robust Sequence Alignment

7.4.1 Introduction

Aligning genetic sequences is a very widely used and important technique in bioinformatics, see for instance Mount [58]. To give a few examples, through sequence alignment we can determine evolutionary relationships among species, and in particular, we can reconstruct phylogenetic trees. An alignment may also reveal functional regions in genetic sequences. Such information may for example lead to the discovery of new or improved drug treatments, or may help in deciding what treatment is best fitted for a particular patient genotype. Sequence alignment is also a handy tool in predicting structural and biochemical properties of sequences.

The alignment problem is usually formulated as an optimisation problem. Basically, positive scores are assigned to matches, and negative scores are assigned to mismatches and gaps. These scores are summarised in what is called a *score matrix*. We aim to find the alignment with the highest total score. This approach has two benefits: (i) it allows us to characterise the optimal ("best") alignment from all possible alignments in an objective way, and (ii) the highest score, corresponding to the best alignment, provides us with an objective measure of the quality of this alignment. Moreover, Needleman and Wunsch [60] have developed an efficient dynamic programming algorithm to calculate the optimal alignment of a small number of sequences (say, two or three sequences). In this article, we will focus on *pair-wise* sequence alignment, that is, the alignment of only two sequences.

Clearly, aligning genetic sequences relies heavily on the choice of the score matrix: how should we reward matches, and how should punish gaps

and mismatches? In practice, a large number of score matrices are being used, and precise choice of the score matrix relies on additional assumptions about the sequences under study. For example, when using PAM score matrices, introduced by Dayhoff, Schwartz, and Orcutt [15], and on which we will focus here, the following assumptions are made (a more detailed explanation follows in Section 7.4.2):

- the evolutionary distance of the sequences to their closest common ancestor is known,
- evolution is in an equilibrium point,
- in this equilibrium point, there is evolutionary reversibility—any point mutation is as probable as its reverse,
- point mutations at different locations in the sequence are i.i.d., and
- point mutations at different times are i.i.d.

Different evolutionary distances induce different score matrices. These matrices are denoted by PAM(T), where T denotes the evolutionary distance between the sequences under study and their closest common ancestor. Of course, the above assumptions are not meant to summarise the current state of the art—much research in molecular evolution is devoted to generalising these assumptions.

Estimating the evolutionary distance is a major issue in molecular evolution, especially when comparing short sequences. Indeed, 'estimation bias usually occurs when the sequence length is short so that stochastic effects are strong' (Gu and Li [39, p. 5899, right column, ll. 25–27]). In many cases, one can only rely on the sequences under study to estimate evolutionary distance—no additional information is available.

One approach is somehow to guess the evolutionary distance from the similarity of the two sequences. Typically, PAM250 is chosen if the sequences are 20% similar, PAM120 if they are 40% similar, PAM60 if they are 60% similar, etc. It is however not entirely clear how in general similarity percentages can be derived from two sequences, prior to alignment.

Another approach consists in solving the optimisation problem not for one, but for a set of PAM matrices, or even with different other methods, and then choose the method that returns the highest optimal score. The

```
          X         10        20
H-alpha VL-SPADKTNVKAAWGKVGAHAGEYGAEA
          | | |   |   | | ||||     | | ||
H-beta  VHLTPEEKSAVTALWGKV--NVDEVGGEA
          X        10          20
```

Figure 7.5: An extract from a possible alignment of hemoglobin alpha and beta chains, produced by 'Alion' (Nevill-Manning, Huang, and Brutlag [61]).

performance of different alignment methods has been studied, and one of the interesting results that have come out of such studies is that 'for different pairs many different methods create the best alignments', and hence, that 'if a method that could select the best alignment method for each pair existed, a significant improvement of the alignment quality could be gained' (see Elofsson [32]). However, in practice it is computationally unfeasible to try out a large numbers of methods and to tune all parameters (such as evolutionary distance, gap penalty, etc.) for each one of them.

We shall investigate whether a bias in the evolutionary distance also leads to a bias in the optimal alignment. In particular, we shall generalise the well-known Needleman-Wunsch algorithm (see Needleman and Wunsch [60]) in order to determine whether an alignment, or parts of it, are insensitive to the evolutionary distance in an interval. In order to do so, we rely on the extension of the dynamic programming formalism developed before.

7.4.2 Optimal Sequence Alignment

What is Sequence Alignment?

A sequence alignment consists of writing two (or more) sequences in rows, and writing similar characters in the same column. In doing so, one is allowed to introduce so-called *gaps*, denoted by a dash '-' in either one of the sequences. Assuming that the sequences are derived from a common ancestor sequence, matches correspond to conserved regions, mismatches correspond to mutations and gaps correspond to deletions or insertions, briefly called *indels*, in either one of the sequences. Figure 7.5 gives an example of an amino acid alignment.

It is convenient to represent alignments in a grid, as depicted in Figure 7.6. All paths from the upper left corner to the lower right corner represent pos-

sible alignments. The path drawn in Figure 7.6 corresponds to the alignment given in Figure 7.5. A diagonal move introduces no gaps, a downwards move introduces a gap in the upper sequence, a rightwards move introduces a gap in the lower sequence.

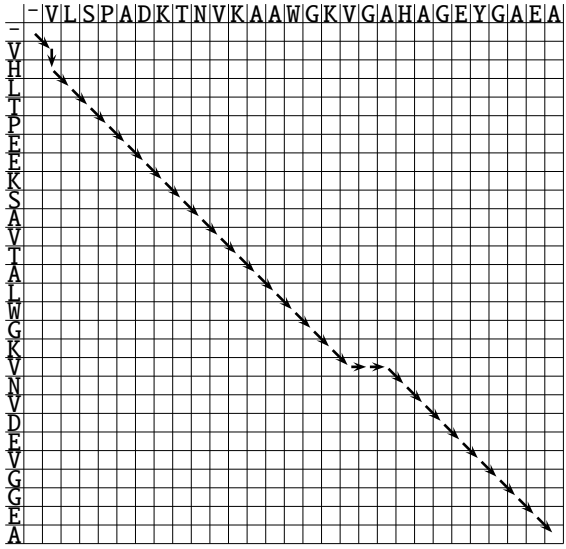


Figure 7.6: Alignments can be conveniently represented in a grid.

When trying to explain evolutionary relationships between sequences, we should identify the alignment that has the highest chance of being the result of an evolutionary process. That is, we try to explain the alignment as the result of evolution from a common ancestor.

We first show how evolutionary dynamics can be described on the level of genetic sequences. Then we show how a score matrix is obtained from these dynamics, and how the resulting optimisation problem indeed identifies the alignment that has the highest chance of being the result of evolution from a common ancestor.

Evolutionary Sequence Dynamics

The PAM ('point accepted mutation') matrices are widely accepted as the standard scoring system when looking for evolutionary relationships in protein sequences. They are related to the evolution of amino acid sequences

described by a Markov model for amino acid substitution (see Dayhoff, Schwartz, and Orcutt [15]). Indels, which introduce alignment gaps, are not modelled by PAM and are treated separately. We will only give a very brief description of the basic ideas underlying the dynamics. A more extensive discussion and improvements of this approach can be found in for instance Dayhoff, Schwartz, and Orcutt [15], Jones, Taylor, and Thornton [46], Benner, Cohen, and Gonnet [5], and Müller and Vingron [59].

Let $A_t(i)$ denote the amino acid at site i at (discrete) time t of a sequence of length N. It is first assumed that amino acids mutate independently at each site of the sequence. This implies that the probability of the sequence A_t to evolve to the sequence A_{t+s} is equal to

$$P[A_{t+s}|A_t] = \prod_{i=1}^{N} P[A_{t+s}(i)|A_t(i)]. \tag{7.4}$$

Hence, it suffices to know only the probabilities $P[A_{t+s}(i)|A_t(i)]$ at each site i of the sequence. It is also assumed that amino acids mutate independently in time,

$$P[A_{t+s}(i)|A_t(i)] = \prod_{r=t}^{t+s-1} P[A_{r+1}(i)|A_r(i)]. \tag{7.5}$$

We thus only need to know the probabilities $P[A_{r+1}(i)|A_r(i)]$ at each site i and time r.

Finally, assuming that the transition probabilities are identically distributed in time and space, $P[A_{r+1}(i)|A_r(i)]$ does not depend on the actual values of r and i, but only on the amino acids $A_r(i)$ and $A_{r+1}(i)$. Hence, if we know for any pair (a, b) of amino acids the probability $P[b|a]$ of a being substituted by b after one unit of time, then we also know the probability of any sequence A_t evolving to A_{t+s}, through Eqs. (7.4) and (7.5). Under the assumptions made so far, this establishes that we can model evolution of amino acid sequences through a Markov model.

It is convenient to assume that evolution from ancestors to descendants is modelled by the same Markov process as the evolution from descendants to ancestors, that is, that the Markov process is *time-reversible*. Assuming $P[b|a] > 0$ for all amino acid pairs (a, b), the Markov process attains a stationary distribution π after a sufficient long time. Moreover, π is independent of the

initial distribution, and is the unique solution of

$$\sum_a P[b|a]\pi[a] = \pi[b]. \tag{7.6}$$

Assuming we attained this equilibrium, the process is time-reversible if and only if (Ross [67])

$$P[b|a]\pi[a] = P[a|b]\pi[b]. \tag{7.7}$$

Consider two amino acid sequences, B and C, that have evolved from a common ancestor A in t time units. Assuming time-reversibility, and assuming that all amino acids in A are i.i.d. according to the stationary distribution π, evolution from A to B and C in t time units is equivalent to evolution from B to A in t time units, and then from A to C in t time units. But this is equivalent to evolution from B to C in $2t$ time units. Hence, we can calculate the probability of B and C having evolved from a common ancestor in t time units simply by calculating the probability of C having evolved from B in $2t$ time units.

In practice, the transition probabilities $P[b|a]$ of the Markov model are estimated using a large dataset of sequences that have already been aligned (originally, sequences from closely related species were considered, that is, sequences of at least 85% similarity). Many generalisations of this model have been developed, dropping stationarity of the transition probabilities, allowing different transition probabilities on different sites, etc.

A Log Likelihood Ratio Scoring

Using the Markov model for amino acid evolution, a scoring matrix is derived that has the interpretation of a log likelihood ratio. The entries of the matrix are roughly given by (up to a normalisation factor)

$$s_t(a,b) = \log \frac{L_{\text{evol}}[a,b](t)}{L_{\text{rand}}[a,b]}, \tag{7.8}$$

that is, the logarithm of the likelihood that a and b are aligned as a consequence of the evolutionary Markov process from a common ancestor t time units ago, divided by the likelihood that a and b are aligned 'by chance', that is, as a consequence of a multinomial process, where amino acid frequencies are obtained from the same data used to construct the Markov model. A

positive score $s_t(a, b)$ means that a and b are more likely to be aligned by evolution than by chance, a negative score means the opposite. Remark that $s_t(a, b) = s_t(b, a)$.

To obtain a score for sequences, recall that we assumed different sites on sequences to be independent. Hence, the log likelihood ratio of two aligned sequences B and C—of equal length and without gaps—is obtained by adding the log likelihood ratios at each site of the sequences:

$$S_t(B, C) = \sum_{i=1}^{N} s_t(B(i), C(i)) \tag{7.9}$$

If we interpret the score as a gain and sequences as paths, we already observe that there is 'gain additivity'; we shall rely on this property when applying dynamic programming.

Gap Scoring

More generally, let B be a sequence of length N, and let C be a sequence of length M. Consider any alignment u of B and C, and denote the characters (amino acids or gaps) at site i in the alignment by $B_u(i)$ and $C_u(i)$. The score of the alignment is given by

$$S_t(B, C)(u) = \sum_{i=1}^{K} s_t(B_u(i), C_u(i)), \tag{7.10}$$

where K is the length of the alignment. If both $B_u(i)$ and $C_u(i)$ are amino acids, the $s_t(B_u(i), C_u(i))$ is given by the log likelihood ratio (Eq. (7.8)). If either one of them, say $B_u(i)$, is a gap, then the score is given by minus the *gap opening penalty* g if $B_u(i-1)$ is not a gap, and by minus the *gap extension penalty* r if $B_u(i-1)$ is a gap (g and r are positive).

Choice of Score Matrix and Gap Penalties

As argued before, the score for a pair of amino acids is given by Eq. (7.8). This score rewards alignments that are more likely by evolution than 'by chance', and punishes alignments that are less likely by evolution than 'by chance'.

Gap openings are less likely than gap extensions, and therefore the gap opening penalty g is chosen substantially higher than the gap extension

penalty r. The gap penalties should also be chosen relative to the range of scores in the score matrix. If the gap penalty is too high, gaps will never appear in the optimal alignment. And if it is too low, too many gaps will appear in the optimal alignment.

Much research has been devoted to analysing how the score matrix and gap penalties should be chosen. The choice of the score matrix is based mainly on the evolutionary dynamical model and estimates of the evolutionary distance. Through statistical analysis, appropriate gap opening and extension penalties have been motivated for various score matrices (see for instance Pearson [64]).

One result is that a good choice for the score matrix, and consequently also a good choice for the gap penalties, can be made based on the evolutionary distance between sequences and their closest common ancestor.

Needleman-Wunsch Algorithm

Finding the optimal alignment is at first sight an extremely hard computational task. The number of possible alignments of two sequences of length N grows exponentially with N. Even for sequences of modest length, computing power is far from able to compare that many sequences in a reasonable amount of time.

Dynamic programming provides a method for exponentially reducing the number of alignments that need to be considered in order to find the optimal one (see Needleman and Wunsch [60]). We shall not discuss the original algorithm here; instead, we shall immediately discuss a generalised version of the algorithm in Section 7.4.4 further on, as the original version is then simply obtained as a special case.

7.4.3 Modelling Evolutionary Distance by a Coherent Lower Prevision

In Section 7.4.2, it was argued that a good choice of the score matrix and the gap penalties can be made based on the evolutionary distance between the sequences under study and their closest common ancestor. Unfortunately, for short sequences, estimation of evolutionary distance is subject to serious bias due to stochastic effects (see Gu and Li [39]). Instead of somehow trying to improve evolutionary distance estimates between short sequences

by reducing stochastic effects—this may well be impossible—we propose a different approach.

Instead, does a bias in the evolutionary distance also lead to a bias in the optimal alignment? Or, in other words, how sensitive is the alignment to the chosen value for evolutionary distance? It is well-known that optimal alignment is quite sensitive to the choice of the score matrix, especially for long sequences (see Elofsson [32]). But for short sequences, this does not need to be the case. To give an extreme example: if we would find that the optimal alignment is independent of the evolutionary distance, we also should not have to worry about it.

Lower previsions provide the perfect tool for performing such a sensitivity analysis. Let us briefly recall the essentials of the theory of lower previsions that we need and apply them to the alignment problem.

Let $\mathcal{T} = \{t \in \mathbb{R} \colon t \geq 0\}$ be the space of possible evolutionary distances t between two sequences B and C and their closest common ancestor. Let T denote the corresponding random variable, which takes values in $\mathcal{T}$. Assume that the only additional information we have about T is that it takes a value in the interval $[t_1, t_2]$, for some $t_1 \leq t_2$. This information is modelled by a vacuous lower prevision relative to $[t_1, t_2]$ (see Section 3.5.1 on p. 60 and Section 5.3 on p. 225 ff, Eq. (5.12) on p. 234 in particular), and leads to the following strict preference relation between alignments (see Definition 6.4 on p. 278; note that for simplicity of exposition, we omit the point-wise ordering):

Definition 7.10 (Preference)**.** Let u and v be two alignments (of B and C). Then, u is said to be *strictly preferred to* v, and we write $u >_{[t_1,t_2]} v$, if

$$\inf_{t \in [t_1,t_2]} [S_t(B,C)(u) - S_t(B,C)(v)] > 0. \tag{7.11}$$

The total scores $S_\bullet(B,C)(u)$, interpreted as random quantities on T, are assumed to be bounded on $[t_1, t_2]$; this is usually the case, and it allows us to apply the dynamic programming results of this chapter without any further complications.

If $u >_{[t_1,t_2]} v$ then there is an $\epsilon > 0$ such that $S_t(B,C)(u) > S_t(B,C)(v) + \epsilon$ for every $t \in [t_1, t_2]$. This means that, independently of the evolutionary distance in $[t_1, t_2]$, u is (uniformly) a strictly better alignment of B and C than v. In such a case, we should of course prefer u over v.

The optimisation problem can now also be restated. Usually, the partial order $>_{[t_1,t_2]}$ will not have a greatest element. Therefore, it makes more sense to look for undominated, or maximal elements, as we have argued in Section 6.3.1 on p. 276 ff.

Definition 7.11 (Maximality)**.** Say an alignment u is *maximal* with respect to $[t_1, t_2]$ if $v \not>_{[t_1,t_2]} u$, that is, if

$$\sup_{t\in[t_1,t_2]} [S_t(B, C)(u) - S_t(B, C)(v)] \geq 0, \tag{7.12}$$

for all possible alignments v of B and C.

The idea behind this definition is that, if we do not prefer any other alignment v over u, then we should consider u as a good alignment candidate. Through pair-wise comparison, the information we have does not allow us to make a better choice than u. An efficient algorithm for finding all maximal alignments will be given in Section 7.4.4. But let us first make a few important remarks.

Firstly, the notion of maximality generalises the classical notion of optimality. Indeed, if $t_1 = t_2 = t$ then any maximal alignment actually maximises the score $S_t(B, C)(v)$ over all possible alignments v (this is a very simple instance of Corollary 6.8 on p. 282).

Secondly, it is often argued that it is important to find *the* best alignment. But, when looking for maximal alignments, we do not obtain a single solution, but rather a set of solutions—perhaps even a pretty large set. At first sight, this may seem undesirable. Nevertheless, even a set of best possible alignments can be useful:

- If we obtain a large set, then this simply means that we have insufficient information in order to construct the best alignment.
- We might be lucky and find that there is only one maximal alignment. If that is the case, we actually also know that this alignment is insensitive to any assumptions made about evolutionary distance in the interval $[t_1, t_2]$.
- More generally, there may be certain constant patterns in the set of maximal alignments, i.e., it may happen that certain regions are consistently aligned over the whole set of maximal alignments. We then

do not only know that these regions are optimally aligned, but also that they are insensitive to any assumptions made about evolutionary distance in the interval $[t_1, t_2]$.

7.4.4 Finding Maximal Alignments Through Dynamic Programming

We briefly discuss how the dynamic programming algorithm is implemented to find all maximal alignments.

Let B be a sequence of length N, and let C be a sequence of length M. First, finding maximal alignments of B and C is restated in terms of finding the maximal paths of a dynamical system. This is done by interpreting alignments as paths of a dynamical system, and scores as gains associated with that path. Figure 7.6 on p. 319 illustrates how to interpret pair-wise alignment as a dynamical system. The grid represents the state space. At each point in the grid we can move either rightwards, downwards, or along the diagonal (except at the right and bottom borders). The gain associated with a move from position (i, j) if the previous move was p, is given by

$$G_t(i, j, p, \downarrow) = \begin{cases} r_t, & \text{if } p = \downarrow \\ g_t, & \text{otherwise} \end{cases}$$

$$G_t(i, j, p, \rightarrow) = \begin{cases} r_t, & \text{if } p = \rightarrow \\ g_t, & \text{otherwise} \end{cases}$$

$$G_t(i, j, p, \searrow) = S_t(B(i), C(j))$$

The gain associated with a path is simply given by the sum of the gains of each move.

The gain depends on the evolutionary distance t. Since the gain also depends on the previous move we must extend the state space with an additional state variable p at each point (i, j) in order to remember our previous move. Otherwise, we cannot apply the dynamical programming formalism.

Let $\mathcal{U}(i, j, p)$ denote the set of all paths from (i, j, p) to the right bottom corner. Observe that p denotes the previous move, $p \in \{\downarrow, \rightarrow, \searrow\}$, which is needed in order to calculate the gain (in order to tell the difference between a gap opening and a gap extension). Let $\mathcal{V}(i, j, p)$ denote the set of maximal

paths from (i, j, p) to the bottom right corner, that is,

$$\mathcal{V}(i, j, p) = \max_{>_{[t_1,t_2]}} \mathcal{U}(i, j, p) \tag{7.13}$$

It is convenient to define $\mathcal{V}(i, j, p) = \emptyset$ whenever $i > N$ or $j > M$. Observe that $\mathcal{U}(i, j, p)$ is a finite set for every state (i, j, p). Hence, the compactness condition under which the generalised Bellman equation holds is trivially satisfied [24].

Theorem 7.12 (Generalised Bellman Equation)**.** *For any state* (i, j, p) *the following equality holds:*

$$\mathcal{V}(i, j, p) = \max_{>_{[t_1,t_2]}} \Big\{ (i, j, p; \downarrow) \oplus \mathcal{V}(i + 1, j, \downarrow), \quad (i, j, p; \rightarrow) \oplus \mathcal{V}(i, j + 1, \rightarrow), \\ (i, j, p; \searrow) \oplus \mathcal{V}(i + 1, j + 1, \searrow) \Big\} \tag{7.14}$$

where $(i, j, p; \downarrow) \oplus \mathcal{V}(i + 1, j, \downarrow)$ *denotes the set of all concatenations of the downward move from state* (i, j, p)*, with a maximal path from state* $(i + 1, j, \downarrow)$*, etc.*

Eq. (7.14) yields an efficient recursive algorithm to calculate the set of all maximal paths $\mathcal{V}(0, 0, \searrow)$, and hence, all maximal alignments. It solves a global maximisation problem by solving $3MN$ smaller maximisation problems (see Figure 7.7).

7.4.5 Test Case

As a demonstration of our approach, consider two sequences of length 80:

```
VLSPADKTNVKAAWGKVGAHAGYGAEALERMFLSFPTTKTYFPHFDLSHGSAVKGHGKKVAKALSAVHLDDMPNALSALS
MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLVSRLLVVYPWTQRFFESFGDLSTPDAVMGNPKVKAHGKKVLGAFSDGL
```

The relation between these sequences is quite well known, but, for the sake of exposition, let's assume we don't know anything about the evolutionary distance between these two sequences. Let's therefore consider the collection of *all* PAM matrices ever considered in sequence alignment: PAM2, PAM3, . . . , up to PAM450: PAM2 corresponds to a very short evolutionary distance, and PAM450 corresponds to an extremely long evolutionary distance (see Dayhoff, Schwartz, and Orcutt [15]).

```
** initialisation **
for p=|,-,\
  MAX(N,M,p)={(M,N,p)}
  for i=0 to N
    MAX(i,M+1,p)={}
  next i
  for j=0 to M
    MAX(N+1,j,p)={}
  next j
next p
** dynamic programming **
for i=N to 0
  for j=M to 0
    for p=|,-,\
      if (i<N) or (j<M)
        ** Bellman **
        MAX(i,j,p)=max{
          (i,j,p;|)+MAX(i+1,j,|),
          (i,j,p;-)+MAX(i,j+1,-),
          (i,j,p;\)+MAX(i+1,j+1,\)
        }
    next p
  next j
next i
```

Figure 7.7: A rough sketch of the algorithm for calculating maximal alignments.

Running the above algorithm, we find a very modest number of optimal alignments: the set of all alignments has only 18 maximal elements (see Table 7.1).

Moreover, within this set, there are surprisingly long robust segments: in all of the 18 maximal alignments, the pattern

```
V--LSPADKTNVKAAWGKVGAHAGYGAEALE??--??FLSFPTTKTYFPHF-DLSHGSA????????????????????????????????
MVHLTPEEKSAVTALWGKVNVDEV-GGEALG??VS??LVVYPWTQRFFESFGDLSTPDA????????????????????????????????
```

is present. Consequently, the choice of the PAM matrix is irrelevant to the alignment of these subsequences. The simulation, implemented through a very simple, hardly optimised C++ program, takes 5 minutes, 37 seconds, and 660 milliseconds, on a 1.8GHz Mobile Intel® Pentium with 512 megabytes of memory. As a comparison, on the same machine, this document was LaTeX'ed from raw LaTeX source in 27 seconds and 491 milliseconds.

This example demonstrates one possible way of how the theory of lower previsions can be applied in bioinformatics, allowing us to substantially weaken assumptions we have to make about data, for instance about the evolutionary distance. In this example, we did that by means of an interval rather than using a point estimate: it turns out that a good alignment (or set of alignments) still can be found in an efficient way, through a generalisation of the well-known Needleman-Wunsch algorithm.

1	V--LSPADKTNVKAAWGKVGAHAGYGAEALE----RMFLSFPTTKTYFPHF-DLSHGSA------VKGHGKKVAKALSAVHLDDMPNALSALS MVHLTPEEKSAVTALWGKVNVDEV-GGEALGRLVSRLLVVYPWTQRFFESFGDLSTPDAVMGNPKVKAHGKKVLGAFS--------DGL----
2	V--LSPADKTNVKAAWGKVGAHAGYGAEALE----RMFLSFPTTKTYFPHF-DLSHGSA------VKGHGKKVAKALSAVHLDDMPNALSALS MVHLTPEEKSAVTALWGKVNVDEV-GGEALGRLVSRLLVVYPWTQRFFESFGDLSTPDAVMGNPKVKAHGKKV--------LGAFSDGL----
3	V--LSPADKTNVKAAWGKVGAHAGYGAEALE----RMFLSFPTTKTYFPHF-DLSHGSAV------KGHGKKVAKALSAVHLDDMPNALSALS MVHLTPEEKSAVTALWGKVNVDEV-GGEALGRLVSRLLVVYPWTQRFFESFGDLSTPDAVMGNPKVKAHGKKVLGAFS--------DGL----
4	V--LSPADKTNVKAAWGKVGAHAGYGAEALE----RMFLSFPTTKTYFPHF-DLSHGSAV------KGHGKKVAKALSAVHLDDMPNALSALS MVHLTPEEKSAVTALWGKVNVDEV-GGEALGRLVSRLLVVYPWTQRFFESFGDLSTPDAVMGNPKVKAHGKKV--------LGAFSDGL----
5	V--LSPADKTNVKAAWGKVGAHAGYGAEALE----RMFLSFPTTKTYFPHF-DLSHGSAVKGHGKKVAKALSAVHLDDMPNALSALS MVHLTPEEKSAVTALWGKVNVDEV-GGEALGRLVSRLLVVYPWTQRFFESFGDLSTPDAVMGNPK--VKAHGKKVLGAFSDGL----
6	V--LSPADKTNVKAAWGKVGAHAGYGAEALE----RMFLSFPTTKTYFPHF-DLSHGSAVKGHGKKVAKALSAVHLDDMPNALSALS MVHLTPEEKSAVTALWGKVNVDEV-GGEALGRLVSRLLVVYPWTQRFFESFGDLSTPDAVMGNPK--VKA----HGKKVLGAFSDGL
7	V--LSPADKTNVKAAWGKVGAHAGYGAEALER----MFLSFPTTKTYFPHF-DLSHGSA------VKGHGKKVAKALSAVHLDDMPNALSALS MVHLTPEEKSAVTALWGKVNVDEV-GGEALGRLVSRLLVVYPWTQRFFESFGDLSTPDAVMGNPKVKAHGKKVLGAFS--------DGL----
8	V--LSPADKTNVKAAWGKVGAHAGYGAEALER----MFLSFPTTKTYFPHF-DLSHGSA------VKGHGKKVAKALSAVHLDDMPNALSALS MVHLTPEEKSAVTALWGKVNVDEV-GGEALGRLVSRLLVVYPWTQRFFESFGDLSTPDAVMGNPKVKAHGKKV--------LGAFSDGL----
9	V--LSPADKTNVKAAWGKVGAHAGYGAEALER----MFLSFPTTKTYFPHF-DLSHGSAV------KGHGKKVAKALSAVHLDDMPNALSALS MVHLTPEEKSAVTALWGKVNVDEV-GGEALGRLVSRLLVVYPWTQRFFESFGDLSTPDAVMGNPKVKAHGKKVLGAFS--------DGL----
10	V--LSPADKTNVKAAWGKVGAHAGYGAEALER----MFLSFPTTKTYFPHF-DLSHGSAV------KGHGKKVAKALSAVHLDDMPNALSALS MVHLTPEEKSAVTALWGKVNVDEV-GGEALGRLVSRLLVVYPWTQRFFESFGDLSTPDAVMGNPKVKAHGKKV--------LGAFSDGL----
11	V--LSPADKTNVKAAWGKVGAHAGYGAEALER----MFLSFPTTKTYFPHF-DLSHGSAVKGHGKKVAKALSAVHLDDMPNALSALS MVHLTPEEKSAVTALWGKVNVDEV-GGEALGRLVSRLLVVYPWTQRFFESFGDLSTPDAVMGNPK--VKAHGKKVLGAFSDGL----
12	V--LSPADKTNVKAAWGKVGAHAGYGAEALER----MFLSFPTTKTYFPHF-DLSHGSAVKGHGKKVAKALSAVHLDDMPNALSALS MVHLTPEEKSAVTALWGKVNVDEV-GGEALGRLVSRLLVVYPWTQRFFESFGDLSTPDAVMGNPK--VKA----HGKKVLGAFSDGL
13	V--LSPADKTNVKAAWGKVGAHAGYGAEALERM----FLSFPTTKTYFPHF-DLSHGSA------VKGHGKKVAKALSAVHLDDMPNALSALS MVHLTPEEKSAVTALWGKVNVDEV-GGEALGRLVSRLLVVYPWTQRFFESFGDLSTPDAVMGNPKVKAHGKKVLGAFS--------DGL----
14	V--LSPADKTNVKAAWGKVGAHAGYGAEALERM----FLSFPTTKTYFPHF-DLSHGSA------VKGHGKKVAKALSAVHLDDMPNALSALS MVHLTPEEKSAVTALWGKVNVDEV-GGEALGRLVSRLLVVYPWTQRFFESFGDLSTPDAVMGNPKVKAHGKKV--------LGAFSDGL----
15	V--LSPADKTNVKAAWGKVGAHAGYGAEALERM----FLSFPTTKTYFPHF-DLSHGSAV------KGHGKKVAKALSAVHLDDMPNALSALS MVHLTPEEKSAVTALWGKVNVDEV-GGEALGRLVSRLLVVYPWTQRFFESFGDLSTPDAVMGNPKVKAHGKKVLGAFS--------DGL----
16	V--LSPADKTNVKAAWGKVGAHAGYGAEALERM----FLSFPTTKTYFPHF-DLSHGSAV------KGHGKKVAKALSAVHLDDMPNALSALS MVHLTPEEKSAVTALWGKVNVDEV-GGEALGRLVSRLLVVYPWTQRFFESFGDLSTPDAVMGNPKVKAHGKKV--------LGAFSDGL----
17	V--LSPADKTNVKAAWGKVGAHAGYGAEALERM----FLSFPTTKTYFPHF-DLSHGSAVKGHGKKVAKALSAVHLDDMPNALSALS MVHLTPEEKSAVTALWGKVNVDEV-GGEALGRLVSRLLVVYPWTQRFFESFGDLSTPDAVMGNPK--VKAHGKKVLGAFSDGL----
18	V--LSPADKTNVKAAWGKVGAHAGYGAEALERM----FLSFPTTKTYFPHF-DLSHGSAVKGHGKKVAKALSAVHLDDMPNALSALS MVHLTPEEKSAVTALWGKVNVDEV-GGEALGRLVSRLLVVYPWTQRFFESFGDLSTPDAVMGNPK--VKA----HGKKVLGAFSDGL

Table 7.1: Test case: all 18 maximal alignments.

Chapter 8

Dynamic Programming and Learning Dynamics

Throughout Chapter 7, we have assumed the system dynamics to be deterministic. This greatly simplified the discussion, still encompassed a large number of interesting applications, and did not suffer from the computational problems which are often encountered when dealing with non-deterministic dynamical systems—simply because in general the number of possible (random) paths tends to grow exponentially with the size of the state space $\mathcal{X}$ and the number of time steps. Nevertheless, when studying the optimal control of dynamical systems, one easily finds examples where the dynamics itself is subject to uncertainty. It certainly seems an interesting challenge to study also these systems from the view-point of the theory of lower previsions, and to know in what cases those computational problems can be overcome. As a initial step in that direction, we investigate in this chapter the applicability of Bellman's dynamical programming algorithm to a very simple type of dynamical system, namely with

- finite state space $\mathcal{X}$, and
- finite control space $\mathcal{U}$.

First, we shall review the most important results already achieved in this field, and identify their shortcomings. Then, in an attempt to remedy those shortcomings, we shall suggest a different model, and try to find all optimal paths based on the notions introduced in Chapter 6 and Chapter 7.

8.1 Introduction

One of the most important models describing non-deterministic finite-state systems are Markov decision processes, which are essentially controlled Markov chains—note that Markov chains were originally devised to study natural language texts; see Markov [55]. Markov decision processes model the uncertainty about the dynamics through so-called *transition probabilities*. Assuming the reward for each transition under each control action to be known, an optimal control is then obtained by maximising the expected reward; this comes down to maximising expected utility, as explained in Section 6.1 on p. 268. This maximisation problem, and many variants thereof, can be solved efficiently using dynamic programming techniques; see for instance Bertsekas [8] for an excellent overview.

Already early in the development of the theory of Markov decision processes it was recognised that the transition probabilities themselves are often subject to uncertainty, simply because they are often hard to measure in practice. To deal with the lack of information about the transition probabilities, two solutions have been suggested and studied in the literature:

(i) learning—we update our knowledge about the transition probabilities as we observe transitions; see for instance Bellman [3], Martin [56], and Satia and Lave [68].

(ii) sets—we only assume that the transition probabilities belong to some convex set; see for instance Wolfe and Dantzig [93], Satia and Lave [68], White and Eldeib [90], Givan, Leach, and Dean [36], Harmanec [41], and Kozine and Utkin [52].

Both solutions have their drawbacks. The learning-based solution relies heavily on prior information about the transition probabilities. If this prior information is incorrect, the optimal control law can be subject to serious bias in the initial phase of the control process. A drawback of the set-based solution is that it does not involve learning, and ignores possibly useful information that is available in many practical problems. Moreover, it has a problematic relation with optimality: working with a set of transition probabilities, we can only associate a lower and upper expected reward, *i.e.*, an interval for the expected reward, with each control law. Almost all authors therefore have considered only maximin or maximax solutions. They

develop algorithms, based on dynamic programming, to find control laws that either maximise the minimal expected gain (pessimistic, maximin), or that maximise the maximal expected gain (optimistic, maximax), ignoring possibly optimal control laws in between.

One notable exception is provided by Harmanec [41], who suggests a dynamic programming algorithm to calculate the set of all maximal elements with respect to a partial preference order which is based on comparing intervals, *i.e.*, interval dominance (introduced in Section 6.6 on p. 292). In that way, not only the extreme solutions are recovered. However, Harmanec [41] did not question in what sense his proposed dynamic programming method leads to optimal policies. As we have already argued, we should approach the problem from the opposite side: we *first* define a notion of optimality and investigate whether the dynamic programming argument holds or not for this notion of optimality, instead of blindly "generalising" Bellman's algorithm.

We have shown in Chapter 7 that the dynamic programming argument holds if our notion of optimality satisfies two conditions: (i) the principle of optimality, and (ii) insensitivity with respect to the omission of non-optimal elements. Unfortunately, the principle of optimality is not satisfied when using the partial ordering, namely interval dominance, suggested by Harmanec [41]: this follows from the counterexample of Section 7.2.6 on p. 313. Hence, Harmanec's [41] algorithm actually does *not* result in *optimal* control laws in the sense of maximality with respect to the suggested partial ordering. In Chapter 7, we suggested a different partial order for deterministic systems with uncertain gain, which does satisfy the principle of optimality and the insensitivity property. However, as is also noted by Harmanec [41], this partial order does not simply generalise to non-deterministic systems. As we shall see in Section 8.3, the reason is that in Markov decision processes there is so-called act-state dependence.

Our primary goal is to combine the learning-based solution with the set-based solution in order to overcome the problems from which each of these methods suffer separately. Basically, we wish to update the set of transition probabilities based on observations of previous transitions. A Markov decision process can be considered as a collection of independent multinomial sampling models (see Martin [56]), and we have a well-developed tool for updating imprecise prior information about multinomial sampling models:

the imprecise Dirichlet model, which is due to Walley [87]. Before doing this, we first need to generalise preference to the case of act-state dependence. We then show that there are fairly general conditions under which the principle of optimality and insensitivity with respect to omission of non-optimal elements still hold.

The main result of this quest shall be that, under fairly general assumptions, which are satisfied if we invoke the imprecise Dirichlet model, we can apply dynamic programming to find, not only maximin or maximax, but the set of all optimal control laws. In doing so, we make only very weak assumptions about the transition probabilities, and we can incorporate learning about them. Unfortunately, for the learning approach to work, the control laws must depend on the complete state history (as with its classical counterpart; see Bertsekas [8]), and therefore, a direct implementation of the suggested dynamic programming algorithm will only be feasible for problems with relatively small state spaces and small control spaces.

Section 8.2 is concerned with the definition and properties of conditional lower previsions, which we shall need extensively further on. In Section 8.3 we motivate a new partial preference order which allows for act-state dependence. In Section 8.4 we precisely define the class of dynamical systems under study and describe how to compare control laws according to the preference order introduced in Section 8.3. Our main result is in Section 8.5, where we state conditions for the principle of optimality to hold. Together with the insensitivity property, which is almost trivially satisfied for the systems under study, this means that we can construct a dynamic programming algorithm. In order to demonstrate that the conditions for which the principle of optimality holds are not overly restrictive, we show in Section 8.6 that they are satisfied in case of simultaneous learning and optimal control of a Markov decision process by means of an imprecise Dirichlet model. In Section 8.7 we present a numerical example.

8.2 Conditional Lower Previsions

8.2.1 Definition

Let us consider two random variables, say X and Y. For the sake of simplicity, and as we do not need the more general case, we shall assume that they can

only assume a finite number of values: $\mathcal{X}$ and $\mathcal{Y}$ are finite sets. Consequently, all random quantities involved are bounded, and hence, are gambles.

The *conditional lower prevision* $\underline{P}(f|y)$ of a gamble f on X conditional on $y \in \mathcal{Y}$ is defined as the supremum buying price for f, conditional on the observation of the value y of Y; $\underline{P}(f|y)$ is the highest price s such that for any $t < s$, we are willing to pay t after observing $Y = y$, but prior to observation of X, if we are guaranteed to receive $f(x)$ when observing $X = x$. Mathematically, $\underline{P}(\bullet|\bullet)$ is a real-valued mapping defined on some subset of

$$\mathcal{L}(X) \times \mathcal{Y} = \{(f, y)\colon f \in \mathcal{L}(X),\ y \in \mathcal{Y}\}.$$

Only to simplify the notation in the proof of Theorem 8.1 below, we shall assume that the domain of $\underline{P}(\bullet|\bullet)$ is finite. All results extend straightforwardly to the general case.

Fixing y, we can view $\underline{P}(\bullet|y)$ as a lower prevision on X. We shall say that $\underline{P}(\bullet|\bullet)$ *separately avoids sure loss* if $\underline{P}(\bullet|y)$ avoids sure loss for each $y \in \mathcal{Y}$. Similarly, we shall say that $\underline{P}(\bullet|\bullet)$ is *separately coherent* if $\underline{P}(\bullet|y)$ is coherent for each $y \in \mathcal{Y}$. Note that Walley [86, Section 6.2, pp. 289–293] has a slightly different notion of separate coherence. Our definition of separate coherence is not as general as Walley's, but it is much simpler, and it suffices for the purpose of this work.

As we have explained in Section 4.1 on p. 95 ff., if $\underline{P}(\bullet|\bullet)$ separately avoids sure loss, then for each y in $\mathcal{Y}$, the natural extension of $\underline{P}(\bullet|y)$ to the set $\mathcal{L}(X)$ of all gambles on X exists, and is real-valued. We shall denote this natural extension by $\underline{E}(\bullet|y)$. Note that $\underline{E}(\bullet|\bullet)$ is now a separately coherent conditional lower prevision on all of $\mathcal{L}(X)\times\mathcal{Y}$. We shall call it the *separate natural extension* of $\underline{P}(\bullet|\bullet)$.

It is convenient to view the separate natural extension $\underline{E}(\bullet|\bullet)$ in a slightly different way, namely, as a $\mathcal{L}(X, Y)$–$\mathcal{L}(Y)$-mapping, which we shall denote by $\underline{E}(\bullet|Y)$:

$$\underline{E}(f(X, Y)|Y)(y) := \underline{E}(f(X, y)|y), \tag{8.1}$$

for any gamble $f(X, Y)$ in $\mathcal{L}(X, Y)$. Here, $f(X, y)$ denotes a gamble on X by fixing the value y of Y in f, *i.e.*, $f(X, y)(x) := f(x, y)$.

In case of n variables $X_1, \ldots, X_n$, each of the conditional lower previsions, resp. $\underline{P}(\bullet|x_1)$ defined on a subset of $\mathcal{L}(X_2)$ for each x_1 in $\mathcal{X}_1$, $\underline{P}(\bullet|x_1x_2)$ defined on a subset of $\mathcal{L}(X_3)$ for each (x_1, x_2) in $\mathcal{X}_1 \times \mathcal{X}_2, \ldots$, and $\underline{P}(\bullet|x_1 \ldots x_{n-1})$

defined on a subset of $\mathscr{L}(X_n)$ for each $(x_1, \ldots, x_{n-1})$ in $\mathscr{X}_1 \times \cdots \times \mathscr{X}_{n-1}$, extends—through the method described above—resp. to a $\mathscr{L}(X_1, X_2)$–$\mathscr{L}(X_1)$-mapping, a $\mathscr{L}(X_1, X_2, X_3)$–$\mathscr{L}(X_1, X_2)$-mapping, ..., and a $\mathscr{L}(X_1, \ldots, X_n)$–$\mathscr{L}(X_1, \ldots, X_{n-1})$-mapping. Concatenating all of these mappings, we end up with a collection of $\mathscr{L}(X_1, \ldots, X_n)$–$\mathscr{L}(X_1, \ldots, X_i)$-mappings ($i \in \{1, \ldots, n-1\}$), which correspond in fact to the following conditional lower previsions, each defined on $\mathscr{L}(X_1, \ldots, X_n)$:

$$\underline{\mathbb{E}}(f|X_1 \ldots X_{n-1}) := \underline{E}(f|X_1 \ldots X_{n-1}) \tag{8.2a}$$

$$\underline{\mathbb{E}}(f|X_1 \ldots X_{n-2}) := \underline{E}(\underline{\mathbb{E}}(f|X_1 \ldots X_{n-1})|X_1 \ldots X_{n-2}) \tag{8.2b}$$

$$= \underline{E}(\bullet|X_1 \ldots X_{n-2}) \circ \underline{E}(\bullet|X_1 \ldots X_{n-1})(f)$$

$$\underline{\mathbb{E}}(f|X_1 \ldots X_{n-3}) := \underline{E}(\underline{\mathbb{E}}(f|X_1 \ldots X_{n-2})|X_1 \ldots X_{n-3}) \tag{8.2c}$$

$$= \underline{E}(\bullet|X_1 \ldots X_{n-3}) \circ \underline{E}(\bullet|X_1 \ldots X_{n-2}) \circ \underline{E}(\bullet|X_1 \ldots X_{n-1})(f)$$

$$\vdots$$

$$\underline{\mathbb{E}}(f|X_1) := \underline{E}(\underline{\mathbb{E}}(f|X_1X_2)|X_1) \tag{8.2d}$$

$$= \underline{E}(\bullet|X_1) \circ \underline{E}(\bullet|X_1X_2) \circ \cdots \circ \underline{E}(\bullet|X_1X_2 \ldots X_{n-1})(f)$$

for any gamble f on $(X_1, \ldots, X_n)$. We shall call these conditional lower previsions the *marginal extensions* of the conditional lower previsions $\underline{P}(\bullet|\bullet)$, $\underline{P}(\bullet|\bullet\,\bullet)$, ..., and $\underline{P}(\bullet|\bullet\ldots\bullet)$. In the classical theory of probability Eq. (8.2) is Bayes rule. It generalises Walley's marginal extension [86, Section 6.7, pp. 313–314] for the special case we study.

Note that, when for instance x_1 is fixed in Eq. (8.2d), it follows that $\underline{\mathbb{E}}(\bullet|x_1)$ is a coherent lower prevision on $\mathscr{L}(X_2, \ldots, X_n)$. This simply follows from the separate coherence of each of the mappings $\underline{E}(\bullet|X_1)$, $\underline{E}(\bullet|X_1X_2)$, ..., and $\underline{E}(\bullet|X_1X_2 \ldots X_{n-1})$. So, Eq. (8.2d) defines a separately coherent conditional lower prevision. In the same way, it follows that all of the marginal extensions are separately coherent.

8.2.2 Marginal Extension Theorem

We now prove that Eq. (8.2) agrees with the conditional lower prevision obtained by application of the axioms of rationality (Axiom 3.1 on p. 49) on the original assessments $\underline{P}(\bullet|\bullet)$, $\underline{P}(\bullet|\bullet\bullet)$, ..., and $\underline{P}(\bullet|\bullet\ldots\bullet)$, as in Theorem 4.3 on p. 96 for the unconditional case. This works simply as follows: for any g

in dom $\underline{P}(\bullet|x_1 \dots x_i)$, we are marginally disposed to accept the gain gamble

$$\Big[g(X_{i+1}) - \underline{P}(g(X_{i+1})|x_1 \dots x_i)\Big] I_{(X_1,\dots,X_i)=(x_1,\dots,x_i)}.$$

Indeed, if the outcome of $(X_1, \dots, X_i)$ is not equal to $(x_1, \dots, x_i)$, then this gain gamble has zero gain, which is *marginally acceptable*: we're willing to accept it if we also receive an arbitrary small amount of strictly positive utility along with the gamble. If the outcome is $(x_1, \dots, x_i)$, then we are disposed to pay any price strictly less than $\underline{P}(g(X_{i+1})|x_1 \dots x_i)$ for the gamble $g(X_{i+1})$, so $g(X_{i+1}) - \underline{P}(g(X_{i+1})|x_1 \dots x_i) + \epsilon$ is acceptable for any $\epsilon > 0$, or equivalently, $g(X_{i+1}) - \underline{P}(g(X_{i+1})|x_1 \dots x_i)$ is marginally acceptable.

Now, the axioms of rationality imply that if a collection of gambles is marginally acceptable, then so must be any non-negative linear combination of them, and so must be any gamble that is point-wise larger than such a sum. So, fix for instance x_1 in $\mathcal{X}_1$, and let f be any gamble on $(X_1, \dots, X_n)$. If for some choice of $\lambda_{g,x_1,\dots,x_i} \geq 0$ and α in $\mathbb{R}$, it holds that

$$\begin{aligned}
&[f(X_1, X_2, \dots, X_n) - \alpha] I_{X_1=x_1} \\
&\qquad \geq \sum_{(x_2,\dots,x_n)} \sum_{i=1}^{n-1} \sum_{g \in \operatorname{dom} \underline{P}(\bullet|x_1 \dots x_i)} \\
&\qquad\qquad \lambda_{g,x_1,\dots,x_i} \Big[g(x_{i+1}) - \underline{P}(g(X_{i+1})|x_1 \dots x_i)\Big] I_{(X_1,X_2,\dots,X_i)=(x_1,x_2,\dots,x_i)}
\end{aligned}$$

or equivalently, if

$$\begin{aligned}
&f(x_1, X_2, \dots, X_n) - \alpha \\
&\qquad \geq \sum_{i=1}^{n-1} \sum_{g \in \operatorname{dom} \underline{P}(\bullet|x_1 X_2 \dots X_i)} \lambda_{g,x_1,X_2,\dots,X_i} \Big[g(X_{i+1}) - \underline{P}(g(X_{i+1})|x_1 X_2 \dots X_i)\Big],
\end{aligned}$$

then the gamble $[f(X_1, X_2, \dots, X_n) - \alpha] I_{X_1=x_1}$ should also be marginally acceptable, or equivalently, the lower prevision of f, conditional on x_1, should be at least α. The idea of the theorem below, and of natural extension in general, is to maximise α subject to the constraints implied by the axioms of rationality.

Let's mention that Eq. (8.2) can also be given an interpretation as the smallest coherent lower prevision that is a behavioural extension of the original

assessments, however, the precise formulation of this result is not straightforward, and therefore we shall simply accept Eq. (8.2) as *the* natural extension of the original assessments. The theorem below supports that choice. We refer to Walley's book [86, Chapters 7–8] for an in depth discussion of the conceptual difficulties encountered when dealing with conditional lower previsions. Note that the equation below is essentially an instance of Walley's much more general definition of natural extension; see Walley [86, Section 8.1, pp. 408–415]. In case that the conditional lower previsions $\underline{P}(\bullet|\bullet\ldots\bullet)$ are independent of their conditioning random variables (*i.e.*, if $\underline{P}(\bullet|x_1)$ is independent of the value of x_1 in $\mathcal{X}_1$, $\underline{P}(\bullet|x_1x_2)$ is independent of the values of x_1 in $\mathcal{X}_1$ and x_2 in $\mathcal{X}_2$, *etc.*), then Theorem 8.1 also proves that our marginal extension agrees with the forward irrelevant product, see De Cooman and Miranda [22, p. 454, Eq. (4) and Theorem 1]. De Cooman and Zaffalon [25, p. 118, Theorem A.1] proved a stronger version of Theorem 8.1 in case of two conditioning random variables.

Theorem 8.1. *Let $k \in \{1,\ldots,n-1\}$, let f be any gamble on $(X_1,\ldots,X_n)$, and let $(x_1,\ldots,x_k)$ be any element of $\mathcal{X}_1\times\cdots\times\mathcal{X}_k$. The marginal extension of f conditional on $(x_1,\ldots,x_k)$, that is, $\underline{\mathbb{E}}(f(x_1,\ldots,x_k,X_{k+1},\ldots,X_n)|x_1\ldots x_k)$, is equal to the maximum achieved by α subject to the constraints*

$$f(x_1,x_2,\ldots,x_n)-\alpha \geq \sum_{i=k}^{n-1}\sum_{g\in\operatorname{dom}\underline{P}(\bullet|x_1\ldots x_i)} \lambda_{g,x_1,\ldots,x_i}\left[g(x_{i+1})-\underline{P}(g|x_1\ldots x_i)\right] \quad (8.3)$$

for all $(x_{i+1},\ldots,x_n)$ in $\mathcal{X}_{i+1}\times\cdots\times\mathcal{X}_n$, where each $\lambda_{g,x_1,\ldots,x_i}$ may vary over the set of non-negative real numbers.

Proof. It suffices to prove the case $k=1$: the general case then follows simply by considering $(X_1,\ldots,X_k)$ as a single variable.

If Eq. (8.3) is satisfied, then, since $\underline{\mathbb{E}}(f(x_1,X_2,\ldots,X_n)|x_1)$ corresponds to a coherent lower prevision when x_1 is fixed,

$$\begin{aligned}&\underline{\mathbb{E}}(f(x_1,X_2,\ldots,X_n)|x_1)-\alpha\\ &\quad\geq \sum_{i=k}^{n-1}\underline{\mathbb{E}}\left(\sum_{g\in\operatorname{dom}\underline{P}(\bullet|x_1\ldots x_i)}\lambda_{g,x_1,X_2,\ldots,X_i}\left[g(X_{i+1})-\underline{P}(g(X_i)|x_1X_2\ldots X_i)\right]\middle|x_1\right).\end{aligned}$$

The term for $i = n-1$ is bounded from below by zero:

$$\begin{aligned}
&\underline{\mathbb{E}}\Bigg(\sum_{g \in \operatorname{dom} \underline{P}(\bullet|x_1 \ldots x_{n-1})} \lambda_{g,x_1,X_2,\ldots,X_{n-1}} \Big[g(X_n) - \underline{P}(g(X_n)|x_1 X_2 \ldots X_{n-1}) \Big] \Bigg| x_1 \Bigg) \\
&\quad = \underline{E}(\bullet|x_1) \circ \underline{E}(\bullet|x_1 X_2) \circ \cdots \circ \underline{E}(\bullet|x_1 X_2 \ldots X_{n-1}) \\
&\qquad \Bigg(\sum_{g \in \operatorname{dom} \underline{P}(\bullet|x_1 \ldots x_i)} \lambda_{g,x_1,X_2,\ldots,X_{n-1}} \Big[g(X_n) - \underline{P}(g(X_n)|x_1 X_2 \ldots X_{n-1}) \Big] \Bigg) \\
&\quad = \underline{E}(\bullet|x_1) \circ \underline{E}(\bullet|x_1 X_2) \circ \cdots \circ \underline{E}(\bullet|x_1 X_2 \ldots X_{n-2}) \\
&\qquad \Bigg(\sum_{g \in \operatorname{dom} \underline{P}(\bullet|x_1 \ldots x_i)} \lambda_{g,x_1,X_2,\ldots,X_{n-1}} \\
&\qquad\qquad \underline{E}\Big(g(X_n) - \underline{P}(g(X_n)|x_1 X_2 \ldots X_{n-1}) \Big| x_1 X_2 \ldots X_{n-1} \Big) \Bigg) \\
&\quad \geq 0,
\end{aligned}$$

since we defined $\underline{E}(\bullet|x_1 X_2 \ldots X_{n-1})$ exactly as the separate natural extension of $\underline{P}(\bullet|x_1 X_2 \ldots X_{n-1})$. In the same way, it follows that all other terms, for $i \in \{1, \ldots, n-2\}$, are non-negative as well. Hence, we find that

$$\underline{\mathbb{E}}(f(x_1, X_2, \ldots, X_n)|x_1) - \alpha \geq 0,$$

and therefore, $\underline{\mathbb{E}}(f(x_1, X_2, \ldots, X_n)|x_1)$ must be at least as large as the maximum achieved by α under the given constraints.

To prove the converse inequality, fix any $\epsilon > 0$. Consider again $i = n-1$. By Theorem 4.3 on p. 96, for each $(x_2, \ldots, x_{n-1})$ in $\mathcal{X}_2 \times \cdots \times \mathcal{X}_{n-1}$ and each g in $\operatorname{dom} \underline{P}(\bullet|x_1 \ldots x_{n-1})$, we may choose $\lambda^{\epsilon}_{g,x_1,x_2,\ldots,x_{n-1}} \geq 0$ such that

$$\begin{aligned}
&\underline{E}(f(x_1, x_2, \ldots, x_{n-1}, X_n)|x_1 x_2 \ldots x_{n-1}) - \epsilon \\
&\qquad\qquad \leq \alpha + \sum_{g \in \operatorname{dom} \underline{P}(\bullet|x_1 \ldots x_i)} \lambda^{\epsilon}_{g,x_1,x_2,\ldots,x_{n-1}} \underline{P}(g(X_n)|x_1 x_2 \ldots x_{n-1})
\end{aligned}$$

and at the same time

$$f(x_1, x_2, \ldots, x_{n-1}, X_n) \geq \alpha + \sum_{g \in \operatorname{dom} \underline{P}(\bullet|x_1 \ldots x_{n-1})} \lambda^{\epsilon}_{g,x_1,x_2,\ldots,x_{n-1}} g(X_n),$$

so,

$$\sum_{g\in\operatorname{dom}\underline{P}(\bullet|x_1\dots x_{n-1})} \lambda^{\epsilon}_{g,x_1,x_2,\dots,x_{n-1}} \Big[g(X_n) - \underline{P}(g(X_n)|x_1x_2\dots x_{n-1})\Big]$$
$$\leq f(x_1,x_2,\dots,x_{n-1},X_n) - \underline{E}(f(x_1,x_2,\dots,x_{n-1},X_n)|x_1x_2\dots x_{n-1}) + \epsilon$$

Similarly, for $i = n-2$, we may choose $\lambda^{\epsilon}_{g,x_1,x_2,\dots,x_{n-2}} \geq 0$ such that

$$\sum_{g\in\operatorname{dom}\underline{P}(\bullet|x_1\dots x_{n-2})} \lambda^{\epsilon}_{g,x_1,x_2,\dots,x_{n-2}} \Big[g(X_{n-1}) - \underline{P}(g(X_{n-1})|x_1x_2\dots x_{n-2})\Big]$$
$$\leq f_{n-1}(x_1,x_2,\dots,x_{n-2},X_{n-1}) - \underline{E}(f_{n-1}(x_1,x_2,\dots,X_{n-1})|x_1x_2\dots x_{n-2}) + \epsilon,$$

where we choose $f_{n-1}(x_1,x_2,\dots,x_{n-2},X_{n-1})$ such that it cancels with a term in the previous expression, namely $\underline{E}(f(x_1,x_2,\dots,X_{n-1},X_n)|x_1x_2\dots X_{n-1})$. For $i = n-3$, we choose $\lambda^{\epsilon}_{g,x_1,x_2,\dots,x_{n-3}} \geq 0$ such that

$$\sum_{g\in\operatorname{dom}\underline{P}(\bullet|x_1\dots x_{n-3})} \lambda^{\epsilon}_{g,x_1,x_2,\dots,x_{n-3}} \Big[g(X_{n-2}) - \underline{P}(g(X_{n-2})|x_1x_2\dots x_{n-3})\Big]$$
$$\leq f_{n-2}(x_1,x_2,\dots,x_{n-3},X_{n-2}) - \underline{E}(f_{n-2}(x_1,x_2,\dots,X_{n-2})|x_1x_2\dots x_{n-3}) + \epsilon,$$

choosing $f_{n-2}(x_1,x_2,\dots,x_{n-3},X_{n-2})$ such that it cancels with a term in the previous expression, namely $\underline{E}(f_{n-1}(x_1,\dots,x_{n-3},X_{n-2},X_{n-1})|x_1\dots x_{n-3}X_{n-2})$. Continuing this process until $i = 1$ and summing all inequalities, we recover that

$$\sum_{i=1}^{n-1} \sum_{g\in\operatorname{dom}\underline{P}(\bullet|x_1\dots x_i)} \lambda^{\epsilon}_{g,x_1,\dots,x_i} \Big[g(x_{i+1}) - \underline{P}(g|x_1\dots x_i)\Big]$$
$$\leq f(x_1,\dots,x_n) - \underline{E}(f_2(x_1,X_2)|x_1) + (n-1)\epsilon,$$

for every $\epsilon > 0$. Expressing f_2 directly as a function of f, *i.e.*, tracing back its definition, we find that

$$f_2(x_1,X_2|x_1)$$
$$= \underline{E}(\bullet|x_1) \circ \underline{E}(\bullet|x_1X_2) \circ \cdots \circ \underline{E}(\bullet|x_1X_2\dots X_{n-1})(f(x_1,X_2,\dots,X_n))$$
$$= \underline{\mathbb{E}}(f(x_1,X_2,\dots,X_n)|x_1),$$

which establishes the desired inequality. Indeed, for every ϵ, the constraints are satisfied by choosing $\lambda^{\epsilon}_{g,x_1,\ldots,x_i}$'s as constructed above, and choosing $\alpha = \underline{\mathbb{E}}(f(x_1, X_2, \ldots, X_n)|x_1) - (n-1)\epsilon$. Therefore, the maximum achieved by α under the given constraints must be at least $\underline{\mathbb{E}}(f(x_1, X_2, \ldots, X_n)|x_1) - (n-1)\epsilon$. Since this holds for any $\epsilon > 0$, the maximum is actually at least $\underline{\mathbb{E}}(f(x_1, X_2, \ldots, X_n)|x_1)$. □

8.3 $\underline{P}$-Maximality under Partial Act-State Dependence

In this section we generalise $\underline{P}$-maximality to the case in which there is (partial) act-state dependence. This will allow us to model simultaneous learning and optimal control. The analysis that follows may seem overly complicated and unnecessary, but these ideas are nevertheless essential to explain under what conditions dynamic programming fails when we simultaneously learn and act.

Let X be a combination of two random variables Ξ and Θ, *i.e.*, $X = (\Xi, \Theta)$ and $\mathcal{X} = \boldsymbol{\Xi} \times \boldsymbol{\Theta}$. Assume that actions $a \in A$ do not influence the value of Θ. So, our beliefs about Θ can be modelled by a coherent extended lower prevision $\underline{P}$ on some linear subspace of $\mathcal{R}(\Theta)$, independent of the action a we take. For each action $a \in A$ and each $\theta \in \boldsymbol{\Theta}$, suppose that our (act-dependent) beliefs about Ξ are modelled through a conditional extended lower prevision $\underline{P}_a(\bullet|\theta)$ defined on some linear subspace of $\mathcal{R}(\Xi)$.

We assume that for all actions a and all possible values of θ, the random quantity $J_a(\Xi, \theta)$, as an element of $\mathcal{R}(\Xi)$, belongs to $\operatorname{dom} \underline{P}_a$, and that for all actions a, the random quantity $\underline{P}_a(J_a|\Theta)$, as an element of $\mathcal{R}(\Theta)$, belongs to $\operatorname{dom} \underline{P}$.

Definition 8.2. For any two actions a and b in A, we say that *a is strictly preferred to b with respect to $\underline{P}$ and $\underline{P}_\bullet(\bullet|\Theta)$* if

$$\underline{P}(\underline{P}_a(J_a|\Theta) - \overline{P}_b(J_b|\Theta)) > 0. \tag{8.4}$$

Let's explain how Eq. (8.4) establishes a strict preference of action a over action b. First, note that since there is act-state independence with respect to Θ, taking an action does not influence the value θ of Θ, so whatever action

we take, the value of Θ is the same. Moreover, it does not matter whether we observe Θ prior to taking an action, or after taking an action. However, we cannot say anything about Ξ prior to taking an action (at this point, we could only model it using the vacuous lower prevision on Ξ). So, it does matter whether we observe Ξ prior to taking an action, or after taking an action, and the outcome of Ξ is expected to depend on the action we take. We're now ready for a precise formulation:

If Eq. (8.4) is satisfied, then we are willing to pay a strictly positive price prior to the observation of Ξ and Θ in order to engage in the two-stage gamble that consists of taking action a and gaining $J_a(\theta, \xi)$ after observation of $\Theta = \theta$ and $\Xi = \xi$, and then taking action b and losing $J_b(\theta, \xi')$ after a second observation of $\Theta = \theta$ and $\Xi = \xi'$. Indeed:

- Using the behavioural interpretation of $\underline{P}$, Eq. (8.4) says that we are willing to pay a strictly positive price prior to observation of Θ in order to receive $\underline{P}_a(J_a|\theta) - \overline{P}_b(J_b|\theta)$, if θ turns out to be the value of Θ, and this independent of the action we take. Hence, it also holds that for some $\epsilon > 0$, we are, prior to observing Θ, willing to pay a strictly positive price in order to receive $\underline{P}_a(J_a|\theta) - \epsilon$ and to lose $\overline{P}_b(J_b|\theta) + \epsilon$ if θ has been observed.

- Suppose now θ has been observed. Then, using the behavioural interpretation of $\underline{P}_a(J_a|\theta)$, for any $\epsilon > 0$ we are willing to lose $\underline{P}_a(J_a|\theta) - \epsilon$ prior to observation of Ξ, in order to take action a and receive $J_a(\xi, \theta)$ after observation of $\Xi = \xi$.

- But, we are also willing to take action b and lose $J_b(\xi', \theta)$ after observation of $\Xi = \xi'$, if we receive $\overline{P}_b(J_b|\theta) + \epsilon$ prior to observation of Ξ.

Combining all these dispositions, we conclude that prior to any observation of Ξ and Θ, we are willing to pay a strictly positive price in order to take action a and receive $J_a(\xi, \theta)$, and then to take action b and lose $J_b(\xi', \theta)$. Let's emphasise again that we can take the same value θ of Θ because the action we take has no influence on the value of this variable.

If there is full act-state dependence, we can identify Ξ with X, and we recover the ordering used by Harmanec [41] in the context of imprecise Markov decision processes: $\underline{P}_a(J_a) > \overline{P}_b(J_b)$. This corresponds to interval dominance, *i.e.*, weak preference without point-wise ordering; see Definition 6.25 on p. 292.

In case of full act-state independence, we can identify Θ with X, recovering $\underline{P}(f_a - f_b) > 0$; we used this strict preference relation before (again, if we ignore point-wise dominance), see Definition 6.4 on p. 278. This order is stronger than the interval ordering, so it leads to a smaller set of optimal actions: this makes sense according to the principle that the stronger our beliefs, the smaller the set of optimal actions, if we view act-state independence as an additional piece of information.

8.4 Imprecise Statistical Decision Processes

We now introduce dynamical systems with uncertain dynamics described by conditional lower previsions. These systems, which we term *imprecise statistical decision processes*, include Markov decision processes and generalise them to imprecise probabilities. However, we do not assume the Markov condition to hold *a priori* because predictions about the dynamics of the system must be allowed to depend on the full system history, if we are to learn about the dynamics based on observations of the behaviour of the system in the past.

8.4.1 States, Controls and Control Laws

Let $\mathcal{X}$ denote the finite set of *states* the system can assume, and let $\mathcal{U}$ denote the finite set of *controls* we can apply. The variable that represents the system state at time k is denoted by X_k, and a particular value of X_k is denoted by x_k. We assume that there is a time N beyond which we are not interested in dynamics of the system. Consider the system at time k. We can imagine

- observing $X_k = x_k$,
- applying the control $\mu_k(x_k) \in \mathcal{U}$ and observing $X_{k+1} = x_{k+1}$,
- applying the control $\mu_{k+1}(x_k x_{k+1}) \in \mathcal{U}$ and observing $X_{k+2} = x_{k+2}$,
- *etc.*,
- applying the control $\mu_{N-1}(x_k x_{k+1} \dots x_{N-1}) \in \mathcal{U}$ and observing $X_N = x_N$.

This control operation is characterised by a finite sequence of functions $\pi_k = (\mu_k, \mu_{k+1}, \dots, \mu_{N-1})$, where $\mu_\ell \colon \mathcal{X}^{\ell-k+1} \to \mathcal{U}$. We call π_k a *control law* from time k. We denote the set of all control laws from time k by Π_k.

With each control law $\pi_k \in \Pi_k$ we can associate a *gain gamble from time ℓ after observation of $x_k \dots x_{\ell-1}$* (with $\ell \geq k$; if $\ell = k$ then $x_k \dots x_{\ell-1}$ is assumed to be an empty sequence, *i.e.*, there is no observation, and we may also write $J_{\pi_k}(x_k, \dots, x_N)$),

$$J_{\pi_k(x_k \dots x_{\ell-1})}(x_\ell, \dots, x_N) = \sum_{q=\ell}^{N-1} g_q(x_q, \mu_q(x_k \dots x_q), x_{q+1}) + g_N(x_N) \tag{8.5}$$

It is interpreted as a gamble on $(X_\ell, \dots, X_N)$. Each transition incurs a gain: starting at time q in state x_q, applying control $u_q \in \mathcal{U}$ and arriving in state $x_{q+1} \in \mathcal{X}$, we receive an amount $g_q(x_q, u_q, x_{q+1})$ of linear utility. Arriving in the final state x_N at time N, we receive an additional gain $g_N(x_N)$. Observe that $J_{\pi_k(x_k \dots x_{\ell-1})}$ depends on π_k only through $\mu_\ell(x_k \dots x_{\ell-1} X_\ell)$, $\dots$, $\mu_{N-1}(x_k \dots x_{\ell-1} X_\ell \dots X_{N-1})$. This sequence, which corresponds to the control law π_k after observation of $x_k \dots x_{\ell-1}$, is denoted by $\pi_k(x_k \dots x_{\ell-1})$.

Our goal is to find optimal control laws, that is, control laws that maximise their corresponding gain gamble. In order to do so, we construct a strict partial order on gain gambles, as in Eq. (8.4). This order is derived from conditional lower previsions that describe the uncertain dynamics of the system.

8.4.2 A Learning Model for Uncertain Dynamics

A simple way to describe uncertain dynamics, including learning, is as follows. Suppose at time k we select π_k, and applying π_k up to time ℓ ($\ell \geq k$) we observe $x_k \dots x_\ell$. We can now model our knowledge about the state at time $\ell + 1$ by a lower prevision on some finite subset of $\mathcal{L}(X_{\ell+1})$, conditional on $x_k \dots x_\ell$, and depending on the control history $\mu_k(x_k)$, $\dots$, $\mu_{\ell-1}(x_k \dots x_{\ell-1})$ and the current control $\mu_\ell(x_k \dots x_\ell)$. The lower previsions may depend on the full system history, and not only on the current control and state as is the case with Markov decision processes. This allows us to adapt our model according to observations of the system history, and hence, to incorporate learning the system dynamics.

As in Section 8.3, we separate those variables Θ which are not influenced by the control law. Hence, we describe the dynamics by a lower prevision $\underline{P}$ on some finite subset of $\mathcal{L}(\Theta)$, and conditional lower previsions $\underline{P}_{\pi_k}(\bullet | x_k \dots x_\ell \theta)$ on some finite subset of $\mathcal{L}(X_{\ell+1})$, for each $\pi_k \in \Pi_k$, each $k \leq \ell < N$, each state

sequence $x_k \dots x_\ell$ and each value of θ.[1] The conditional lower prevision is allowed to depend on the control law π_k, but the parameters θ are assumed not to be influenced by the control law.

At first sight, the separation of act-state independent variables may appear to be merely a technical matter. But in fact, from Theorem 8.5 it will follow that this separation is *essential* to make the principle of optimality work when the dynamics is described by an imprecise probability model. If we do not separate those variables, we naturally arrive at the weaker ordering used by Harmanec [41] which does not satisfy the principle of optimality.

How can we identify act-state independent variables? Looking at the example invoking the imprecise Dirichlet model for learning dynamics at the end of Section 8.5, these variables naturally arise as the hyper-parameters of the model, because they only model prior information. Thus in general, modelling learning by an imprecise hierarchical model, the hyper-parameters of the model, which are commonly used to represent prior information, are a natural choice for act-state independent variables. The remaining variables, in particular the states at different time points, will usually be act-state dependent.

The conditional lower previsions $\underline{P}_{\pi_k}(\bullet|x_k \dots x_\ell\theta)$ combine, through separate natural extension and the marginal extension theorem, to

$$\underline{\mathbb{E}}_{\pi_k}(\bullet|x_k \dots x_\ell\theta) = \underline{E}_{\pi_k}(\bullet|x_k \dots x_\ell\theta) \circ \underline{E}_{\pi_k}(\bullet|x_k \dots x_\ell X_{\ell+1}\theta) \circ \cdots \\ \cdots \circ \underline{E}_{\pi_k}(\bullet|x_k \dots x_\ell X_{\ell+1} \dots X_{N-1}\theta) \quad (8.6)$$

on $\mathcal{L}(X_{\ell+1}, \dots, X_N)$, as in Eq. (8.2). We can now use Eq. (8.4) to compare control laws after observation of a state sequence. Of course, after such observation it only makes sense to compare control laws with the same control history. Let $\Pi_k(x_k \dots x_\ell, u_k \dots u_{\ell-1})$ denote the set of those elements π_k of Π_k for which

$$\pi_k(x_k) = u_k, \quad \pi_k(x_k x_{k+1}) = u_{k+1}, \quad \dots, \quad \pi_{\ell-1}(x_k \dots x_{\ell-1}) = u_{\ell-1}. \quad (8.7)$$

It is convenient to identify $\Pi_k(x_k)$ with Π_k.

The natural extension of the unconditional lower prevision $\underline{P}$ to the set of all gambles on Θ is denoted by $\underline{\mathbf{E}}_{\underline{P}}$.

Definition 8.3. Let $\pi_k, \rho_k \in \Pi_k(x_k \dots x_\ell, u_k \dots u_{\ell-1})$. We say that π_k is preferred

[1] A better but very heavy notation would be $\underline{P}_{\mu_k(x_k)\dots\mu_\ell(x_k\dots x_\ell)}(\bullet|x_k \dots x_\ell\theta)$.

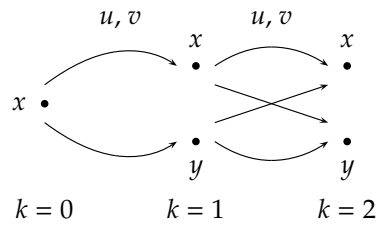


Figure 8.1: A simple sequential decision process

to ρ_k after observation of state sequence $x_k \dots x_\ell$ and application of control sequence $u_k \dots u_{\ell-1}$, and we write $\pi_k >_{x_k \dots x_\ell, u_k \dots u_{\ell-1}} \rho_k$, if

$$\underline{\mathbf{E}}_{\underline{P}}(\underline{\mathbb{E}}_{\pi_k}(J_{\pi_k(x_k \dots x_\ell)} | x_k \dots x_\ell \Theta) - \overline{\mathbb{E}}_{\rho_k}(J_{\rho_k(x_k \dots x_\ell)} | x_k \dots x_\ell \Theta)) > 0. \tag{8.8}$$

Observe that, once $x_k \dots x_\ell$ and $u_k \dots u_{\ell-1}$ are fixed, the ordering depends on π_k and ρ_k only through $\pi_k(x_k \dots x_\ell)$ and $\rho_k(x_k \dots x_\ell)$. It is easy to show that $>_{x_k \dots x_\ell, u_k \dots u_{\ell-1}}$ is a strict partial order. Using Eq. 6.5 on p. 278, we obtain an optimality criterion for control laws by selecting as optimal the set of those actions which are maximal with respect to the partial order of Eq. (8.8).

Definition 8.4. A control law $\pi_k \in \Pi_k$ is said to be *optimal* if it is maximal in $\Pi_k(x_k)$ with respect to $>_{x_k}$ for each each $x_k \in \mathcal{X}$. Let $k \leq \ell < N - 1$. The control law π_k is said to be *optimal from time* ℓ if it is maximal in $\Pi_k(x_k \dots x_\ell, \mu_k(x_k) \dots \mu_{\ell-1}(x_k \dots x_{\ell-1}))$ with respect to the partial order $>_{x_k \dots x_\ell, \mu_k(x_k) \dots \mu_{\ell-1}(x_k \dots x_{\ell-1})}$ for each state sequence $x_k \dots x_\ell$.

Does this definition make sense? By assumption, Π_k is finite, and the existence of maximal control laws with respect to $>_{x_k \dots x_\ell, \mu_k(x_k) \dots \mu_{\ell-1}(x_k \dots x_{\ell-1})}$ is easy to prove, as is the existence of control laws which are simultaneously maximal with respect to $>_{x_k \dots x_\ell, \mu_k(x_k) \dots \mu_{\ell-1}(x_k \dots x_{\ell-1})}$ for all state sequences $x_k \dots x_\ell$.

8.5 The Principle of Optimality

Consider the sequential decision process depicted in Fig. 8.1. At each time k we can choose between two actions, u and v. We shall make no assumption on the connection between actions and dynamics, nevertheless, it may be convenient for the reader to assume such a connection. Consider the control

law π_0 which applies v at time 0, and u if $x_1 = x$ and v if $x_1 = y$ at time 1:

$$\begin{aligned}\mu_0(x) &= v,\\ \mu_1(xx) &= u,\\ \mu_1(xy) &= v.\end{aligned}$$

The principle of optimality stipulates that if π_0 belongs to the set of optimal control laws from time 0, then the control law $\pi_0(x)$, which applies u if $x_1 = x$ and v if $x_1 = y$ at time 1, should belong to the set of optimal control laws from time 1. As a consequence, we can significantly reduce the complexity of calculating the set of optimal control laws. To see how this works, assume that for instance ρ_1, specified by $\nu_1(x) = v$ and $\nu_1(y) = u$, is not optimal from time 1. Using the principle of optimality, σ_0 and σ_0', specified by

$$\begin{aligned}\kappa_0(x) &= u, & \kappa_0'(x) &= v\\ \kappa_1(xx) &= \nu_1(x) = v, & \kappa_1'(xx) &= \nu_1(x) = v,\\ \kappa_1(xy) &= \nu_1(y) = u, & \kappa_1'(xy) &= \nu_1(y) = u,\end{aligned}$$

cannot be optimal from time 0, because otherwise ρ_1 should have to be optimal by the principle of optimality. Hence, when we already know the optimal control laws from time $\ell + 1$, we can use this information in order to reduce the search space when looking for optimal control laws from time ℓ. Of course, we can do this only if reducing the search space does not change the set of optimal elements we eventually end up with: our notion of optimality must be *insensitive to the omission of non-optimal elements*. Observe that the number of control laws grows exponentially with the length of the paths under consideration, but by the principle of optimality and the insensitivity property we do not need to consider most of them. In this way, we arrive at an exponential speedup in the search for the set of all optimal control laws. Writing this down in a formal way, we arrive at a generalisation of Bellman's equation.

Fortunately, the insensitivity property holds in all cases where the search space is finite and optimality is induced by a partial ordering: it suffices that every non-optimal element is dominated by an optimal element. But this is immediate, as the set of all possible actions is finite. The principle of optimality is more difficult to establish. The following theorem states

sufficient conditions under which the principle of optimality holds for the imprecise statistical decision problems under study.

Theorem 8.5 (Principle of Optimality)**.** *Let $k < N$ and $\pi_k \in \Pi_k$. For any $k \leq \ell < N$, it holds that if π_k is optimal from time ℓ then it is optimal from time $\ell + 1$, whenever all of the following conditions are satisfied:*

- *The conditional lower previsions $E_{\pi_k}(\bullet|x_k \dots x_\ell\theta)$ are linear, for all $k \leq \ell < N$, all values of θ, and all state sequences $x_k \dots x_\ell$.*
- *$\underline{\mathbf{E}}_{\underline{P}}$ is vacuous, that is, there is a subset T of Θ such that $\underline{\mathbf{E}}_{\underline{P}}(f(\Theta)) = \inf_{\theta\in T} f(\theta)$ for any gamble f on Θ.*
- *For any $x_{\ell+1} \in \mathcal{X}$ it holds that*

$$\underline{\mathbf{E}}_{\underline{P}}(E_{\pi_k}(I_{X_{\ell+1}=x_{\ell+1}}|x_k \dots x_\ell\Theta)) > 0. \tag{8.9}$$

Proof. First, observe that for any control law $\rho_k = (\nu_k, \dots, \nu_{N-1})$ it holds that the marginal extension Eq. (8.6) is linear, and

$$\begin{aligned}\mathbb{E}_{\rho_k}(J_{\rho_k(x_k\dots x_\ell)}|x_k \dots x_\ell\theta) = E_{\rho_k}(g(x_\ell, \nu_\ell(x_k \dots x_\ell), X_{\ell+1})|x_k \dots x_\ell\theta) \\ + E_{\rho_k}\Big(\mathbb{E}_{\rho_k}(J_{\rho_k(x_k\dots x_\ell X_{\ell+1})}|x_k \dots x_\ell X_{\ell+1}\theta) \\ \Big|x_k \dots x_\ell\theta\Big)\end{aligned} \tag{8.10}$$

We prove the theorem by contraposition. Assume that $\pi_k = (\mu_k, \dots, \mu_{N-1})$ is not optimal from time $\ell + 1$. Then there must be a state sequence $x_k \dots x_{\ell+1}$ and a control law $\rho_k = (\nu_k, \dots, \nu_{N-1})$ which is preferred to π_k after observation of $x_k \dots x_{\ell+1}$ and application of control sequence $\mu_k(x_k) \dots \mu_\ell(x_k \dots x_\ell)$,

$$\begin{aligned}\nu_k(x_k) &= \mu_k(x_k), \\ &\dots, \\ \nu_\ell(x_k \dots x_\ell) &= \mu_\ell(x_k \dots x_\ell),\end{aligned} \tag{8.11}$$

and

$$\inf_{\theta\in T}(\mathbb{E}_{\rho_k}(J_{\rho_k(x_k\dots x_{\ell+1})}|x_k \dots x_{\ell+1}\theta) - \mathbb{E}_{\pi_k}(J_{\pi_k(x_k\dots x_{\ell+1})}|x_k \dots x_{\ell+1}\theta)) > 0, \tag{8.12}$$

Since Eq. (8.11) and Eq. (8.12) only depend on ρ_k through $\rho_k(x_k \dots x_{\ell+1})$,

we are free to choose the remaining components of ρ_k. For instance, choose

$$\begin{aligned} v_{\ell+1}(x_k \dots x_\ell X_{\ell+1}) &= \mu_{\ell+1}(x_k \dots x_\ell X_{\ell+1}), \\ &\dots, \\ v_{N-1}(x_k \dots x_\ell X_{\ell+1} \dots X_{N-1}) &= \mu_{N-1}(x_k \dots x_\ell X_{\ell+1} \dots X_{N-1}), \end{aligned} \tag{8.13}$$

whenever $X_{\ell+1} \neq x_{\ell+1}$. But, for this choice of ρ_k, it holds that ρ_k is also preferred to π_k after observation of only $x_k \dots x_\ell$ and application of the controls $\mu_k(x_k) \dots \mu_\ell(x_k \dots x_{\ell-1})$.

Indeed, this statement follows if we can prove that

$$\inf_{\theta \in T} (\mathbb{E}_{\rho_k}(J_{\rho_k(x_k \dots x_\ell)} | x_k \dots x_\ell \theta) - \mathbb{E}_{\pi_k}(J_{\pi_k(x_k \dots x_\ell)} | x_k \dots x_\ell \theta)) > 0. \tag{8.14}$$

By Eq. (8.11) it holds that

$$E_{\pi_k}(\bullet | x_k \dots x_\ell \theta) = E_{\rho_k}(\bullet | x_k \dots x_\ell \theta)$$

since $E_{\rho_k}(\bullet | x_k \dots x_\ell \theta)$ only depends on ρ_k through $\rho_k(x_k \dots x_\ell)$. Using this equality when applying Eq. (8.10) on both π_k and ρ_k, we find that Eq. (8.14) is equivalent to

$$\begin{aligned} \inf_{\theta \in T} (E_{\pi_k}(\mathbb{E}_{\rho_k}(J_{\rho_k(x_k \dots x_\ell X_{\ell+1})} | x_k \dots x_\ell X_{\ell+1} \theta) \\ - \mathbb{E}_{\pi_k}(J_{\pi_k(x_k \dots x_\ell X_{\ell+1})} | x_k \dots x_\ell X_{\ell+1} \theta) | x_k \dots x_\ell \theta)) > 0. \end{aligned}$$

By Eq. (8.13), and again since $E_{\pi_k}(\mathbb{E}_{\rho_k}(J_{\rho_k(x_k \dots x_\ell X_{\ell+1})} | x_k \dots x_\ell X_{\ell+1} \theta))$ only depends on ρ_k through $\rho_k(x_k \dots x_\ell X_{\ell+1})$, this is equivalent to

$$\begin{aligned} \inf_{\theta} (E_{\pi_k}(I_{X_{\ell+1} = x_{\ell+1}} [\mathbb{E}_{\rho_k}(J_{\rho_k(x_k \dots x_{\ell+1})} | x_k \dots x_{\ell+1} \theta) \\ - \mathbb{E}_{\pi_k}(J_{\pi_k(x_k \dots x_{\ell+1})} | x_k \dots x_{\ell+1} \theta)] | x_k \dots x_\ell \theta)) > 0, \end{aligned}$$

where $I_{X_{\ell+1} = x_{\ell+1}}$ is the indicator function of the singleton $\{x_{\ell+1}\}$. But this strict inequality follows from Eq. (8.9), Eq. (8.12) and the linearity of $E_{\pi_k}(\bullet | x_k \dots x_\ell \theta)$. The proof is established. □

Roughly, Theorem 8.5 states that the principle of optimality holds if all the imprecision is concentrated in the state-independent part of the model,

and if this imprecise part is of the vacuous type: θ is only known to belong to some set $T \subseteq \Theta$. It may appear that imprecision is more or less left out of the picture by the requirement that the conditional lower previsions should be linear. This is not the case: whenever the imprecise model can be described by a *set* of precise models

$$\{E_{\pi_k}(\bullet|x_k \dots x_\ell\theta) \colon \theta \in T\} \tag{8.15}$$

and these precise models are connected through a conditioning parameter θ (more precisely, a conditioning random variable Θ) as in Eq. (8.15), the principle of optimality applies when using the preference order Eq. (8.8). Imprecise probability models are often expressed in terms of sets of precise models. The theorem tells us that we should look for an act-state independent variable which parametrises this set. If this is possible, we can apply dynamic programming.

8.6 Invoking the Imprecise Dirichlet Model

The conditions of Theorem 8.5 are satisfied when we use an imprecise Dirichlet model (introduced by Walley [87]) in order to represent learning the system dynamics. In this model the conditional linear previsions are given by

$$E_{\pi_k}(f|x_k \dots x_\ell\theta) = \sum_{x_{\ell+1}\in\mathcal{X}} f(x_{\ell+1}) \frac{s\theta^{\mu_\ell(x_k\dots x_\ell)}_{x_\ell x_{\ell+1}} + n^{\mu_\ell(x_k\dots x_\ell)}_{x_\ell x_{\ell+1}}(x_k \dots x_\ell, \pi_k)}{s + N^{\mu_\ell(x_k\dots x_\ell)}_{x_\ell}(x_k \dots x_\ell, \pi_k)} \tag{8.16}$$

for any gamble f on $X_{\ell+1}$, and the imprecise (vacuous) unconditional lower prevision is given by

$$\underline{P}(g) = \inf_{\substack{\theta^u_{xy}\geq\epsilon,\\ \sum_{y\in\mathcal{X}}\theta^u_{xy}=1}} g(\theta), \tag{8.17}$$

for all gambles g on Θ, where $\epsilon > 0$ is an arbitrary small strictly positive real number (less than $\frac{1}{|\mathcal{X}|}$). Let's briefly explain what these expressions mean, and how we arrive at them.

We use $n^u_{xy}(x_k \dots x_\ell, \pi_k)$ to denote the number of transitions from state x to state y by applying control u, in the sequence $x_k \dots x_\ell$ subject to control law π_k, and $N^u_x(x_k \dots x_\ell, \pi_k)$ denotes the number of transitions that start in state x

and apply control u, in the sequence $x_k \dots x_\ell$ under control law π_k:

$$n^u_{xy}(x_k, \pi_k) = 0, \tag{8.18}$$

$$n^u_{xy}(x_k \dots x_{\ell+1}, \pi_k) = n^u_{xy}(x_k \dots x_\ell, \pi_k) + \begin{cases} 1, & \text{if } x_\ell = x, x_{\ell+1} = y \text{ and } \mu_\ell(x_k \dots x_\ell) = u \\ 0, & \text{otherwise} \end{cases} \tag{8.19}$$

and,

$$N^u_x(x_k \dots x_\ell, \pi_k) = \sum_{y \in \mathcal{X}} n^u_{xy}(x_k \dots x_\ell, \pi_k). \tag{8.20}$$

Equation (8.16) is the predictive lower prevision on $X_{\ell+1}$ which arises from an independent product of precise Dirichlet models on the transition probabilities from state x_ℓ applying $\mu_\ell(x_k \dots x_\ell)$ after having observed $x_k \dots x_\ell$ subject to control law π_k [56]. We assume that observation of transitions from one state do not influence our knowledge about transitions from another state. This motivates the use of an independent product of Dirichlet models, each model modelling transitions from a particular state.

The hyper-parameters of these models are s and θ^u_{xy}, for each x, $y \in \mathcal{X}$ and $u \in \mathcal{U}$. The hyper-parameter $s > 0$ determines the adaptivity of the model (lower s means faster learning), and the hyper-parameters $\theta^u_{xy} \geq \epsilon$, $\sum_{y \in \mathcal{X}} \theta^u_{xy} = 1$, determine the prior transition probabilities from state x to state y applying control u. Eq. (8.17) follows then from the assumption that we know (almost) nothing about the transition probabilities *a priori*. Hence, we use the vacuous lower prevision $\underline{P}$ on $\mathcal{L}(\Theta)$, where we use Θ as a notation for the collection of all hyper-parameters Θ^u_{xy} for x, $y \in \mathcal{X}$ and $u \in \mathcal{U}$. The hyper-parameters s and Θ, which represent prior information about the dynamics of the system, are obviously not influenced by the control law π_k: they are act-state independent variables. The vacuous lower prevision is chosen such that *a priori* the lower probability of any transition is at least ϵ. A completely vacuous model for the hyper-parameters Θ is obtained by setting $\epsilon = 0$, but, we must choose $\epsilon > 0$ to ensure that Eq. (8.9) holds.

8.7 A Numerical Example

Consider again the Markov decision process depicted in Figure 8.1. Recall that at each time k we can choose between two actions, u and v. Transition

probabilities are denoted as θ^v_{yx} (the probability from state y to state x when taking action v), and the reward associated with this transition is denoted by r^v_{yx} (so, for instance, $g(y,v,x) = r^v_{yx}$, and $h(x) = h(y) = 0$). Initially, all transition probabilities are known to be at least $\frac{1}{10}$, and we precisely know the rewards:

$$r^u_{xx} = r^u_{yx} = 1 \qquad r^v_{xx} = r^v_{yx} = 2$$
$$r^u_{xy} = r^u_{yy} = 1.5 \qquad r^v_{xy} = r^v_{yy} = 0.75$$

Intuitively, it is clear that insufficient information is available in order to construct a unique optimal feedback. However, suppose we are in state x at time $k = 0$, take action v and end up in state x at time $k = 1$. Then it seems reasonable to assume that when we select action v again, the probability that we end up in x again is higher than the probability of ending up in y. In fact, the reward associated with this transition, r^v_{xx}, is the highest possible reward. Even if we do not know precisely the value of θ^v_{xx}, after observing the transition from state x at time k to state x at $k+1$ under action v, we obtain, through the imprecise Dirichlet model (hyper-parameter $s = 1$), a sufficiently narrow probability interval for θ^v_{xx} in order to ensure that we will end up with the highest possible reward by taking action v from state x at time $k = 1$. This demonstrates the possible benefit of learning.

Let's verify this result, and apply Bellman's dynamic programming algorithm to obtain all globally optimal feedback controls.

8.7.1 Conditional Expected Gains After Observations

Assuming a Dirichlet prior with parameters s and θ, we have, by Eq. (8.10),

$$\begin{aligned}\mathbb{E}_{\pi_k}(J_{\pi_k(x)}|x\theta) = {} & \theta^{\pi_k(x)}_{xx}\left(r^{\pi_k(x)}_{xx} + \mathbb{E}_{\pi_k}(J_{\pi_k(xx)}|xx\theta)\right) \\ & + \theta^{\pi_k(x)}_{xy}\left(r^{\pi_k(x)}_{xy} + \mathbb{E}_{\pi_k}(J_{\pi_k(xy)}|xy\theta)\right) \end{aligned} \tag{8.21}$$

$$\mathbb{E}_{\pi_k}(J_{\pi_k(xx)}|xx\theta) = \begin{cases} \frac{s\theta^{\pi_k(xx)}_{xx}+1}{s+1} r^{\pi_k(xx)}_{xx} + \frac{s\theta^{\pi_k(xx)}_{xy}}{s+1} r^{\pi_k(xx)}_{xy}, & \text{if } \pi_k(x) = \pi_k(xx) \\ \theta^{\pi_k(xx)}_{xx} r^{\pi_k(xx)}_{xx} + \theta^{\pi_k(xx)}_{xy} r^{\pi_k(xx)}_{xy}, & \text{otherwise.} \end{cases} \tag{8.22}$$

$$\mathbb{E}_{\pi_k}(J_{\pi_k(xy)}|xy\theta) = \theta^{\pi_k(xy)}_{yx} r^{\pi_k(xy)}_{yx} + \theta^{\pi_k(xy)}_{yy} r^{\pi_k(xy)}_{yy} \tag{8.23}$$

8.7.2 Dynamic Programming

Optimal Control Laws After Observing xx

The feedback π_k is optimal from xx if it holds for all π'_k that

$$\sup_{\theta_{\bullet\bullet} \geq \frac{1}{10}} \left[\mathbb{E}_{\pi_k}(J_{\pi_k(xx)}|xx\theta) - \mathbb{E}_{\pi_k}(J_{\pi'_k(xx)}|xx\theta) \right] \geq 0. \tag{8.24}$$

- $\pi_k(x) = u$ & $\pi_k(xx) = u$ is not optimal. Indeed:

  - Consider $\pi'_k(x) = v$ & $\pi'_k(xx) = v$. Then, Eq. (8.24) becomes

$$\sup_{\theta_{\bullet\bullet} \geq \frac{1}{10}} \left[\frac{s\theta_{xx}^{\pi_k(xx)} + 1}{s+1} r_{xx}^{\pi_k(xx)} - \frac{s\theta_{xx}^{\pi'_k(xx)} + 1}{s+1} r_{xx}^{\pi'_k(xx)} + \frac{s\theta_{xy}^{\pi_k(xx)}}{s+1} r_{xy}^{\pi_k(xx)} - \frac{s\theta_{xy}^{\pi'_k(xx)}}{s+1} r_{xy}^{\pi'_k(xx)} \right] \geq 0,$$

    or equivalently,

$$\sup_{\theta_{\bullet\bullet} \geq \frac{1}{10}} \left[\frac{s\theta_{xx}^{u} + 1}{s+1} r_{xx}^{u} - \frac{s\theta_{xx}^{v} + 1}{s+1} r_{xx}^{v} + \frac{s\theta_{xy}^{u}}{s+1} r_{xy}^{u} - \frac{s\theta_{xy}^{v}}{s+1} r_{xy}^{v} \right] \geq 0,$$

    or equivalently,

$$\sup_{\substack{t \in [\frac{1}{10}, \frac{9}{10}] \\ t' \in [\frac{1}{10}, \frac{9}{10}]}} \left[\frac{st + 1}{s+1} r_{xx}^{u} - \frac{st' + 1}{s+1} r_{xx}^{v} + \frac{s(1-t)}{s+1} r_{xy}^{u} - \frac{s(1-t')}{s+1} r_{xy}^{v} \right] \geq 0,$$

    Equivalently, with $s = 1$,

$$\max \left\{ \begin{array}{l} \dfrac{11r_{xx}^{u} + 9r_{xy}^{u}}{20} - \dfrac{11r_{xx}^{v} + 9r_{xy}^{v}}{20}, \\ \dfrac{19r_{xx}^{u} + r_{xy}^{u}}{20} - \dfrac{11r_{xx}^{v} + 9r_{xy}^{v}}{20}, \\ \dfrac{11r_{xx}^{u} + 9r_{xy}^{u}}{20} - \dfrac{19r_{xx}^{v} + r_{xy}^{v}}{20}, \\ \dfrac{19r_{xx}^{u} + r_{xy}^{u}}{20} - \dfrac{19r_{xx}^{v} + r_{xy}^{v}}{20} \end{array} \right\} \geq 0,$$

which reduces to

$$\max\left\{\frac{11\times 1+9\times 1.5}{20}-\frac{11\times 2+9\times 0.75}{20},\right.$$
$$\frac{19\times 1+1.5}{20}-\frac{11\times 2+9\times 0.75}{20},$$
$$\frac{11\times 1+9\times 1.5}{20}-\frac{19\times 2+0.75}{20},$$
$$\left.\frac{19\times 1+1.5}{20}-\frac{19\times 2+0.75}{20}\right\}\geq 0,$$

But, this inequality is not satisfied. Therefore, $\pi_k(x)=u$ & $\pi_k(xx)=u$ is not optimal.

- $\pi_k(x)=u$ & $\pi_k(xx)=v$ is not optimal. Indeed:
  - Consider $\pi'_k(x)=v$ & $\pi'_k(xx)=v$. Then, Eq. (8.24) becomes

$$\sup_{\theta^{\bullet}_{\bullet}\geq\frac{1}{10}}\left[\theta^{\pi_k(xx)}_{xx}r^{\pi_k(xx)}_{xx}-\frac{s\theta^{\pi'_k(xx)}_{xx}+1}{s+1}r^{\pi'_k(xx)}_{xx}\right.$$
$$\left.+\theta^{\pi_k(xx)}_{xy}r^{\pi_k(xx)}_{xy}-\frac{s\theta^{\pi'_k(xx)}_{xy}}{s+1}r^{\pi'_k(xx)}_{xy}\right]\geq 0,$$

or equivalently,

$$\sup_{\theta^{\bullet}_{\bullet}\geq\frac{1}{10}}\left[\theta^{v}_{xx}r^{v}_{xx}-\frac{s\theta^{v}_{xx}+1}{s+1}r^{v}_{xx}+\theta^{v}_{xy}r^{v}_{xy}-\frac{s\theta^{v}_{xy}}{s+1}r^{v}_{xy}\right]\geq 0,$$

or equivalently,

$$\sup_{t\in[\frac{1}{10},\frac{9}{10}]}\left[tr^{v}_{xx}-\frac{st+1}{s+1}r^{v}_{xx}+(1-t)r^{v}_{xy}-\frac{s(1-t)}{s+1}r^{v}_{xy}\right]\geq 0,$$

Equivalently, with $s=1$,

$$\max\left\{\frac{r^{v}_{xx}+9r^{v}_{xy}}{10}-\frac{11r^{v}_{xx}+9r^{v}_{xy}}{20},\right.$$
$$\left.\frac{9r^{v}_{xx}+r^{v}_{xy}}{10}-\frac{19r^{v}_{xx}+r^{v}_{xy}}{20}\right\}\geq 0,$$

which reduces to

$$\max\left\{\frac{1\times 2+9\times 0.75}{10}-\frac{11\times 2+9\times 0.75}{20},\right.$$
$$\left.\frac{9\times 2+1\times 0.75}{10}-\frac{19\times 2+1\times 0.75}{20}\right\}\geq 0,$$

But, this inequality is not satisfied. Therefore, $\pi_k(x)=u$ & $\pi_k(xx)=v$ is not optimal.

- $\pi_k(x)=v$ & $\pi_k(xx)=v$ is optimal. Indeed,
  - For $\pi'_k(x)=v$ & $\pi'_k(xx)=v$, Eq. (8.24) is trivially satisfied.
  - Consider $\pi'_k(xx)=u$ & $\pi'_k(x)=u$. Then, with $s=1$, Eq. (8.24) becomes

$$\max\left\{-\frac{11\times 1+9\times 1.5}{20}+\frac{11\times 2+9\times 0.75}{20},\right.$$
$$-\frac{19\times 1+1.5}{20}+\frac{11\times 2+9\times 0.75}{20},$$
$$-\frac{11\times 1+9\times 1.5}{20}+\frac{19\times 2+0.75}{20},$$
$$\left.-\frac{19\times 1+1.5}{20}+\frac{19\times 2+0.75}{20}\right\}\geq 0,$$

    This inequality is satisfied.
  - Consider $\pi'_k(x)=v$ & $\pi'_k(xx)=u$. Then it is similarly checked that Eq. (8.24) is satisfied.
  - Consider $\pi'_k(x)=u$ & $\pi'_k(xx)=v$. Again, it is similarly checked that Eq. (8.24) is satisfied.
- And finally, after similar considerations, it follows that that $\pi_k(x)=v$ & $\pi_k(xx)=u$ is optimal too.

Optimal Control Laws After Observing xy

In a similar way, it can be shown that

- $\pi_k(x)=u$ & $\pi_k(xy)=u$ is not optimal.
- $\pi_k(x)=u$ & $\pi_k(xy)=v$ is not optimal.

- $\pi_k(x) = v$ & $\pi_k(xy) = u$ is optimal.
- $\pi_k(x) = v$ & $\pi_k(xy) = v$ is optimal.

Applying The Principle of Optimality

By the above results, it follows that after observation of the initial state x, the only possibly optimal controls laws are

- $\pi_k(x) = v$ & $\pi_k(xx) = u$ & $\pi_k(xy) = u$.
- $\pi_k(x) = v$ & $\pi_k(xx) = u$ & $\pi_k(xy) = v$.
- $\pi_k(x) = v$ & $\pi_k(xx) = v$ & $\pi_k(xy) = u$.
- $\pi_k(x) = v$ & $\pi_k(xx) = v$ & $\pi_k(xy) = v$.

Indeed, these are exactly the control laws which are both optimal after observation of xx, and after observation of xy, so by the principle of optimality, and insensitivity to omission of non-optimal elements, any control law that is optimal after observation of x must belong to this class—note that a control law π_k belongs to this class if and only if $\pi_k(x) = v$.

It turns out that all of these control laws are optimal after observation of x: for any π_k and π'_k such that $\pi_k(x) = \pi'_k(x) = v$, it holds that

$$\sup_{\theta_{\bullet\bullet} \geq \frac{1}{10}} \left[\mathbb{E}_{\pi_k}(J_{\pi_k(x)}|xx\theta) - \mathbb{E}_{\pi_k}(J_{\pi'_k(x)}|xx\theta) \right] \geq 0,$$

using Eq. (8.21). So, all these control laws are incomparable after observation of only x. This simply means that we don't have enough initial information to further discriminate between them.

8.7.3 Result

We conclude that a control law π_k is optimal (after observation of the initial state x) if and only if $\pi_k(x) = v$. We have already argued that v is, intuitively, a good choice as an initial control: in the given example, this is the only way to learn about the probability θ^v_{xx} of a transition that has the highest reward $r^v_{xx} = 2$ associated to it.

We have also demonstrated how the principle of optimality can be invoked to sequentially reduce the set of candidate optimal control laws. In the example, the number of candidates was halved after only one step.

Chapter 9

Conclusion

From the introductory chapters, Chapters 3–6, the most important conclusion is that coherent lower previsions are belief models (Chapters 3–5), which generalise many of the existing models for uncertainty, and which naturally lead to a theory of robust optimality (Chapter 6). They allow us in Chapters 7–8 to study dynamical systems whose uncertain gain or uncertain dynamics cannot be described by the classical theory of probability, for instance because insufficient information is available in order to identify a probability measure.

The main conclusion of Chapter 7 is that the method of dynamic programming can in principle be extended to deterministic systems with an uncertain gain, where the uncertainty about the gain is modelled by a lower prevision. We have demonstrated how the principle of optimality, together with the insensitivity property, yields an efficient recursive algorithm in order to calculate optimal paths. Basically, it reduces the global optimisation problem which requires a search over the space of all possible paths, to a sequence of $N - k$ (where k is the initial time, and N is the time horizon) local optimisation problems requiring only a search over the control space $\mathcal{U}$. In this way, the principle of optimality yields an exponential speedup in determining optimal paths. This was nicely demonstrated by the sequence alignment algorithm (Fig. 7.7 on p. 328).

But our general study of what conditions a generalised notion of optimality should satisfy for the Bellman approach to work is of some interest in itself too. In particular, besides an obvious extension of the well-known principle of optimality, another condition emerges that relates to the nature

of the optimality operators *per se*: the optimality of a path should be invariant under the omission of non-optimal paths from the set of paths under consideration. If optimality is induced by a strict partial ordering of paths, then this second condition is satisfied whenever the existence of dominating optimal paths for non-optimal ones is guaranteed.

Another important observation is that, contrary to $\underline{P}$-maximality and $\mathcal{M}$-maximality, the dynamic programming method cannot be used to solve optimisation problems corresponding to $\underline{P}$-maximinity, $\underline{P}$-maximaxity, and weak $\underline{P}$-maximality: for these notions the principle of optimality does not hold in general.

From Chapter 8, where we have investigated finite-state dynamical systems with uncertain dynamics described by conditional lower previsions, we conclude that Bellman's dynamic programming algorithm still works if the lower previsions describing the dynamics satisfy a very particular structural property: all the imprecision must be concentrated in the state-independent part of the model, and this imprecise part must be of the vacuous type. It is quite remarkable that the separation of act-state independent beliefs from act-state dependent beliefs is essential for the dynamic programming approach to work.

However, due to the fact that the control laws must depend on the full system history in order to allow learning about the system dynamics, we must repeat the algorithm at each time step for all possible system histories, and not simply for all possible states as in the case without learning. As a result, the algorithm still needs an exponential time, but even so, the search space has been exponentially reduced. This is inevitable also in the classical approach, even when considering sufficient statistics; see for instance Bertsekas [8].

On the other hand, the learning approach, using the imprecise Dirichlet model, leads to more determinate beliefs as time increases. Hence, with longer time horizon the incomparability of control laws will be less likely, and the size of the set of optimal control laws will tend to stabilise. In this way, it is less prone to the problem of huge, exponentially growing sets of optimal elements, as is often experienced with the method proposed in Harmanec [41], and the non-learning method discussed in Chapter 7.

Appendix A

The Extended Real Numbers

In this appendix we recall the definition and elementary properties of the extended real calculus. For the sake of completeness, the proofs are given too; apparently, the properties of the extended real number system on which we rely in this work, are rather hard to find in the literature.

A.1 Definitions

Definition A.1. The set $\mathbb{R}^*$ of *extended real numbers* is defined by $\mathbb{R} \cup \{-\infty, +\infty\}$.

Definition A.2. The addition "+" on $\mathbb{R}$ is extended to $\mathbb{R}^*$ as follows:

$$\begin{aligned} -\infty + (-\infty) &= -\infty, & +\infty + (+\infty) &= +\infty, \\ a + (-\infty) = -\infty + a &= -\infty, & a + (+\infty) = +\infty + a &= +\infty, \qquad \text{if } a \in \mathbb{R}. \end{aligned}$$

We call a sum of extended real numbers *well defined* if it cannot be reduced to $+\infty + (-\infty)$ or $-\infty + (+\infty)$.

As usual, "$a+(+\infty)$" is abbreviated to "$a+\infty$", and "$a+(-\infty)$" is abbreviated to "$a - \infty$". We also write, for instance, "$\{\sum_{i=1}^n a_n \text{ w.d.}: a_1 \in A_1, \ldots, a_n \in A_n\}$" as an abbreviation for the set "$\{\sum_{i=1}^n a_n : \sum_{i=1}^n a_n \text{ well defined, and } a_1 \in A_1, \ldots, a_n \in A_n\}$", where $A_1, \ldots, A_n$ are subsets of $\mathbb{R}^*$.

Definition A.3. The *well defined sum* of two subsets A and B of $\mathbb{R}^*$ is defined by

$$A + B = \{a + b \text{ w.d.}: a \in A, b \in B\}.$$

Definition A.4. The multiplication "$\times$" on $\mathbb{R}$ is extended to $\mathbb{R}^*$ as follows:

$$\begin{aligned}
&-\infty \times -\infty = +\infty \times +\infty = +\infty, \quad -\infty \times +\infty = +\infty \times -\infty = -\infty, \\
&\quad a \times -\infty = -\infty \times a = -\infty, \qquad a \times +\infty = +\infty \times a = +\infty, \quad \text{if } a > 0, \\
&\quad a \times -\infty = -\infty \times a = +\infty, \qquad a \times +\infty = +\infty \times a = -\infty, \quad \text{if } a < 0, \\
&\quad a \times -\infty = -\infty \times a = 0, \qquad a \times +\infty = +\infty \times a = 0, \qquad \text{if } a = 0.
\end{aligned}$$

Definition A.5. The ordering $\leq$ on $\mathbb{R}$ is extended to $\mathbb{R}^*$ by defining $-\infty \leq a$ and $a \leq +\infty$ for any $a \in \mathbb{R}^*$.

Definition A.6. The equivalence relation $=$ on $\mathbb{R}$ is extended to $\mathbb{R}^*$ by defining $-\infty = -\infty$ and $+\infty = +\infty$.

We shall write "$x = \pm\infty$" as an abbreviation of "$x = -\infty$ or $x = +\infty$", where x denotes any extended real number.

A.2 Properties

Proposition A.7. *The addition "+" on $\mathbb{R}^*$ is commutative and associative, the multiplication "$\times$" on $\mathbb{R}^*$ is commutative and associative.*

Proof. Immediate. □

Lemma A.8. *For any non-zero real number λ, and any extended real numbers a_1, ..., a_n, it holds that $\sum_{i=1}^n a_i$ is well defined if and only if $\sum_{i=1}^n \lambda a_i$ is well defined, and in such a case*

$$\lambda \sum_{i=1}^{n} a_i = \sum_{i=1}^{n} \lambda a_i$$

Proof. Immediate. □

Lemma A.9. *Let a, b, c and d be sums of extended real numbers. Let A and B be subsets of $\mathbb{R}^*$. The following statements hold.*

(i) If $a + b$ is well defined then a and b are well defined. Conversely, if a is not well defined or b is not well defined then $a + b$ is not well defined.

(ii) If a and b are well defined and $a - b$ is not well defined, then it can only be that $a = b = \pm\infty$.

(iii) *"$a + b - a$ well defined" implies "$a \in \mathbb{R}$, b well defined and $a + b - a = b$". Conversely, "$a \in \mathbb{R}$ and b well defined" implies "$a + b - a$ well defined and $a + b - a = b$".*

(iv) *"$a \geq b + c$ whenever a and $b + c$ are well defined" is equivalent to "$a - b \geq c$ whenever $a - b$ and c are well defined". Hence, also "$a + b \geq c$ whenever $a + b$ and c are well defined" is equivalent to "$a \geq c - b$ whenever a and $c - b$ are well defined".*

(v) *If "$c \geq d$ whenever c and d are well defined", then "$a \geq b + c$ whenever a and $b + c$ are well defined" implies that "$a \geq b + d$ whenever a and $b + d$ are well defined".*

(vi) *"$a = b + c$ whenever a and $b + c$ are well defined" is equivalent to "$a - b = c$ whenever $a - b$ and c are well defined".*

(vii) $\sup A = \sup(A \setminus \{-\infty\})$ *and* $\inf A = \inf(A \setminus \{+\infty\})$.

(viii) $\sup(A + B) = \sup A + \sup B$ *whenever the right hand side is well defined.*

Proof. (i)&(ii). Immediate from the definition of well defined.

(iii). If $a + b - a$ is well defined then by (i) a and b are well defined. Also, a must be a real number since $a + b - a$ would reduce to $+\infty - \infty$ otherwise. In all three cases $b = \pm\infty$, or b real, the equality follows. The other implication is proven in a similar way.

(iv). If a, b or c is not well defined then the equivalence is trivial by (i). Therefore we can assume without loss of generality that a, b and c are well defined.

Assume that $a - b$ is well defined. We show that under the assumption "$b + c$ well defined implies $a \geq b + c$", $a - b \geq c$ holds.

If $b + c$ is not well defined, then we have to consider the following cases.

(Ia) $b = -c = -\infty$. In this case, $a - b = +\infty$, and $c = +\infty$ too, so $a - b \geq c$ holds.

(Ib) $b = -c = +\infty$. In this case, $a - b = -\infty$, and $c = -\infty$ too, so $a - b \geq c$ holds.

If, on the other hand, $b + c$ is well defined, then we know that $a \geq b + c$. We have to consider the following cases.

(IIa) $(b + c) - b$ not well defined, $b + c = b = +\infty$. We have that $a \geq b + c = +\infty$, which implies that also $a = +\infty$, so $a - b$ is not well defined; we reached a contradiction, which means that this case cannot occur.

(IIb) $(b + c) - b$ not well defined, $b + c = b = -\infty$. Since $a - b$ is well defined by assumption, $a > -\infty$, and therefore $a - b = +\infty$ and $a - b \geq c$ holds.

(IIc) $(b + c) - b$ well defined. Then $b \in \mathbb{R}$. If $c = -\infty$ then the inequality $a - b \geq c$ is obvious. If $c = +\infty$ then $a \geq b + c = +\infty$ and the inequality $a - b \geq c$ holds. If $c \in \mathbb{R}$ and $a = -\infty$ then we have a contradiction since $a \geq b + c$, so this case cannot occur. If $c \in \mathbb{R}$ and $a \in \mathbb{R}$ then the inequality $a - b \geq c$ follows from the usual real calculus. Finally, if $c \in \mathbb{R}$ and $a = +\infty$ then the inequality $a - b \geq c$ is obvious.

Conversely, assume that $b + c$ is well defined. We show that under the assumption "$a - b$ well defined implies $a - b \geq c$", $a \geq b + c$ holds.

If $a - b$ is not well defined, then we have to consider the following cases.

(Ia) $a = b = +\infty$. In this case, $b + c = +\infty$, and $a = +\infty$ too, so $a \geq b + c$ holds.

(Ib) $a = b = -\infty$. In this case, $b + c = -\infty$, and $a = -\infty$ too, so $a \geq b + c$ holds.

If, on the other hand, $a - b$ is well defined, then we know that $a - b \geq c$. We have to consider the following cases.

(IIa) $(a-b)+b$ not well defined, $a-b = -b = -\infty$. We have that $-\infty = a-b \geq c$, which implies that also $c = -\infty$, so $b + c$ is not well defined; we reached a contradiction, which means that this case cannot occur.

(IIb) $(a - b) + b$ not well defined, $a - b = -b = +\infty$. Since $b + c$ is well defined by assumption, $c < +\infty$, and therefore $b + c = -\infty$ and $a \geq b + c$ holds.

(IIc) $(a - b) + b$ well defined. Then $b \in \mathbb{R}$. If $a = +\infty$ then the inequality $a \geq b + c$ is obvious. If $a = -\infty$ then $-\infty = a - b \geq c$, so $c = -\infty$, hence $b + c = -\infty$, and the inequality $a \geq b + c$ holds. If $a \in \mathbb{R}$ and $c = +\infty$ then we have a contradiction since $a - b \geq c$ so this cannot occur. If $a \in \mathbb{R}$ and $c \in \mathbb{R}$ then the inequality $a \geq b + c$ follows from the usual real calculus. Finally, if $a \in \mathbb{R}$ and $c = -\infty$ then the inequality $a \geq b + c$ is obvious.

The second equivalence follows simply by replacing b by $-b$ in the first equivalence and using commutativity of the addition on $\mathbb{R}^*$.

(v). Assume c and d are well defined. If a or b is not well defined then the equivalence is trivial by (i). Therefore we can assume without loss of generality that a and b are well defined.

If $b = -\infty$ or $c = -\infty$, then $b = -\infty$ or $d = -\infty$, and the statement trivially holds. Without loss of generality, we may thus assume that b and c are strictly larger than $-\infty$. In particular, we only need to consider cases in which $b + c$ is well defined.

If $b = +\infty$ or $c = +\infty$, then $a = +\infty$ (whenever $b + c$ is well defined, which is the case) and the statement trivially holds. If both b and c are real, again the statement trivially holds, even though d may be $-\infty$.

(vi). If a, b or c is not well defined then the equivalence is trivial by (i). Therefore we can assume without loss of generality that a, b and c are well defined.

The equivalence follows from (iv). Indeed,

$$
\begin{aligned}
& (b + c \text{ w.d.} \implies a = b + c) \\
\iff & (b + c \text{ w.d.} \implies (a \geq b + c \text{ and } b + c \geq a)) \\
\iff & (b + c \text{ w.d.} \implies a \geq b + c) \text{ and } (b + c \text{ w.d.} \implies b + c \geq a) \\
\iff & (a - b \text{ w.d.} \implies a - b \geq c) \text{ and } (a - b \text{ w.d.} \implies c \geq a - b) \\
\iff & (a - b \text{ w.d.} \implies (a - b \geq c \text{ and } c \geq a - b)) \\
\iff & (a - b \text{ w.d.} \implies a - b = c)
\end{aligned}
$$

(vii). This is follows from the fact that $\sup \emptyset = -\infty$ and $\inf \emptyset = +\infty$.

(viii). If $\sup A = -\infty$ then (a) $A = \emptyset$, in which case $A + B = \emptyset$, or (b) $A = \{-\infty\}$, in which case $A + B = \emptyset$ or $A + B = \{-\infty\}$. So the proposition holds if $\sup A = -\infty$ or $\sup B = -\infty$ (by commutativity of the addition and symmetry).

If $\sup A = +\infty$ then B must contain extended real numbers strictly larger than $-\infty$ (otherwise $\sup A + \sup B$ would not be well defined). For any such number $b \in B$, $b > -\infty$, we have that $\sup A + b = +\infty$, whence $\sup\{a + b; a \in A, b \in B, a, b > -\infty\} = +\infty$, and consequently, $\sup\{a + b; a \in A, b \in B, a + b \text{ well defined}\} = \sup(A + B) = +\infty$. So the proposition holds if $\sup A = +\infty$ or $\sup B = +\infty$ (by commutativity of the addition and symmetry).

If $\sup A$ and $\sup B$ are real numbers, then the property follows from the continuity of the addition. □

Bibliography

[1] Martin Aigner. *Combinatorial Theory*. Springer, Berlin, 1997. Reprint of the 1979 edition. Originally published as Vol. 234 of the "Grundlehren der mathematischen Wissenschaften".

[2] F. J. Anscombe and R. J. Aumann. A definition of subjective probability. *Annals of Mathematical Statistics*, 34(1):199–205, March 1963.

[3] Richard Bellman. A problem in the sequential design of experiments. *Sankhya*, 16:221–229, 1956.

[4] Richard Bellman. *Dynamic Programming*. Princeton University Press, Princeton, 1957.

[5] S. A. Benner, M. A. Cohen, and G. H. Gonnet. Amino-acid substitution during functionally constrained divergent evolution of protein sequences. *Protein Engineering*, 7(11):1323–1332, 1994.

[6] J. O. Berger. The robust Bayesian viewpoint. In J. B. Kadane, editor, *Robustness of Bayesian Analyses*, pages 63–144. Elsevier Science, Amsterdam, 1984.

[7] Jakob Bernoulli. *Ars Conjectandi*. Basel, 1713.

[8] Dimitri P. Bertsekas. *Dynamic Programming and Stochastic Control*, volume 125 of *Mathematics in Science and Engineering*. Academic Press, 1976.

[9] K.P.S. Bhaskara Rao and M. Bhaskara Rao. *Theory of Charges, a Study of Finitely Additive Measures*. Academic Press, London, 1983.

[10] M. Chevé and R. Congar. Optimal pollution control under imprecise environmental risk and irreversibility. *Risk Decision and Policy*, 5:151–164, 2000.

[11] Gustave Choquet. Theory of capacities. *Annales de l'Institut Fourier*, 5:131–295, 1954.

[12] Lucio Crisma, Patrizia Gigante, and Pietro Millossovich. A notion of coherent prevision for arbitrary random numbers. *Quaderni del Dipartimento di Matematica Applicata alle Scienze Economiche Statistiche e Attuariali "Bruno de Finetti"*, 5, 1997.

[13] Lucio Crisma, Patrizia Gigante, and Pietro Millossovich. A notion of coherent prevision for arbitrary random quantities. *Journal of the Italian Statistical Society*, 6(3):233–243, 1997.

[14] Gaston Darboux. Mémoire sur les fonctions discontinues. *Annales Scientifiques de l'École Normale Supérieure (2^e Série)*, IV:57–112, 1875.

[15] M. O. Dayhoff, R. M. Schwartz, and B. C. Orcutt. A model of evolutionary change in proteins. In M. O. Dayhoff, editor, *Atlas of protein sequence and structure*, volume 5, pages 345–352. National biomedical research foundation, 1978.

[16] Marquis de Condorcet. *Essai sur l'Application de l'Analyse à la Probabilité des Décisions Rendues à la Pluralité des Voix*. L'Imprimerie Royale, Paris, 1785.

[17] Gert de Cooman. Possibility theory I: the measure- and integral-theoretic groundwork. *International Journal of General Systems*, 25:291–323, 1997.

[18] Gert de Cooman. Possibility theory II: conditional possibility. *International Journal of General Systems*, 25:325–351, 1997.

[19] Gert de Cooman. Possibility theory III: possibilistic independence. *International Journal of General Systems*, 25:353–371, 1997.

[20] Gert de Cooman. Belief models: an order theoretic analysis. In G. de Cooman, T. L. Fine, and T. Seidenfeld, editors, *ISIPTA '01 – Proceedings of the Second International Symposium on Imprecise Probabilities and Their Applications*, pages 93–103, Maastricht, 2001. Shaker Publishing.

[21] Gert de Cooman and Dirk Aeyels. Supremum preserving upper probabilities. *Information Sciences*, 118:173–212, 1999.

[22] Gert de Cooman and E. Miranda. A weak law of large numbers for coherent lower previsions. In *Proceedings of the Tenth International Conference IPMU 2004 (Information Processing and Management of Uncertainty in Knowledge-Based Systems, 4–9 July 2004, Perugia, Italy)*, volume 1, pages 451–458, Rome, Italy, 2004. Editrice Università La Sapienza.

[23] Gert de Cooman and Matthias C. M. Troffaes. Dynamic programming for deterministic discrete-time systems with uncertain gain. International Journal of Approximate Reasoning (in press).

[24] Gert de Cooman and Matthias C. M. Troffaes. Dynamic programming for discrete-time systems with uncertain gain. In Jean-Marc Bernard, Teddy Seidenfeld, and Marco Zaffalon, editors, *ISIPTA '03 – Proceedings of the Third International Symposium on Imprecise Probabilities and Their Applications*, pages 162–176. Carleton Scientific, July 2003.

[25] Gert de Cooman and Marco Zaffalon. Updating beliefs with incomplete observations. *Artificial Intelligence*, 159:75–125, 2004.

[26] Bruno de Finetti. La prévision: ses lois logiques, ses sources subjectives. *Annales de l'Institut Henri Poincaré*, 7:1–68, 1937.

[27] Bruno de Finetti. *Theory of Probability: A Critical Introductory Treatment*. Wiley, New York, 1974–5. Two volumes.

[28] Dieter Denneberg. *Non-additive Measure and Integral*. Kluwer, Dordrecht, 1994.

[29] Didier Dubois and Henri Prade. *Théorie des possibilité*. Masson, Paris, 1985.

[30] N. Dunford and J. T. Schwartz. *Linear Operators*. John Wiley & Sons, New York, 1957.

[31] Nelson Dunford. Integration in general analysis. *Transactions of the American Mathematical Society*, 37(3):441–453, May 1935.

[32] A. Elofsson. A study on protein sequence alignment quality. *Proteins-structure function and genetics*, 46(3):330–339, 2002.

[33] Scott Ferson, Vladik Kreinovich, Lev Ginzburg, Davis S. Myers, and Kari Sentz. Constructing probability boxes and Dempster-Shafer structures. Technical Report SAND2002–4015, Sandia National Laboratories, January 2003.

[34] Itzhak Gilboa and David Schmeidler. Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics*, 18(2):141–153, 1989.

[35] F. J. Giron and S. Rios. Quasi-Bayesian behaviour: A more realistic approach to decision making? In J. M. Bernardo, J. H. DeGroot, D. V. Lindley, and A. F. M. Smith, editors, *Bayesian Statistics*, pages 17–38. University Press, Valencia, 1980.

[36] Robert Givan, Sonia Leach, and Thomas Dean. Bounded-parameter Markov decision processes. *Artificial Intelligence*, 122:71–109, 2000.

[37] G. G. Gould. Intergration over vector-valued measures. *Proceedings of the London Mathematical Society*, 15(3):193–225, 1965.

[38] Gabriele H. Greco. Sur la mesurabilité d'une fonction numérique par rapport à une famille d'ensembles. *Rend. Sem. Mat. Univ. Padova*, 65:163–176, 1981.

[39] X. Gu and W. H. Li. Estimation of evolutionary distances under stationary and nonstationary models of nucleotide substitution. *Proceedings of the National Academy of Sciences of the United States of America*, 95(11):5899–5905, 1998.

[40] Paul Richard Halmos. *Measure Theory*. Springer, New York, 1974.

[41] David Harmanec. Generalizing Markov decision processes to imprecise probabilities. *Journal of Statistical Planning and Inference*, 105(1):199–213, June 2002.

[42] T. H. Hildebrandt. On bounded functional operations. *Transactions of the American Mathematical Society*, 36(4):868–875, October 1934.

[43] T. H. Hildebrandt. *Introduction to the Theory of Integration*. Academic Press, London, 1963.

[44] Richard B. Holmes. *Geometric Functional Analysis and Its Applications*. Springer, New York, 1975.

[45] Hugo Janssen. *Een ordetheoretische en behavioristische studie van possibilistische processen*. PhD thesis, Ghent University, October 1999.

[46] D. T. Jones, W. R. Taylor, and J. M. Thornton. The rapid generation of mutation data matrices from protein sequences. *Computer Applications in the Biosciences*, 8(3):275–282, 1992.

[47] Daniel Kahneman and Amos Tversky. Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–291, March 1979.

[48] Olav Kallenberg. *Foundations of Modern Probability*. Probability and Its Applications. Springer, second edition, 2002.

[49] John L. Kelley. *General Topology*. American Book, New York, 1955.

[50] A. Kolmogoroff. Untersuchungen über den Integralbegriff. *Mathematische Annalen*, 103:654–696, 1930.

[51] Heinz König. *Measure and Integration: An Advanced Course in Basic Procedures and Applications*. Springer, Berlin, 1997.

[52] Igor O. Kozine and Lev V. Utkin. Interval-valued finite Markov chains. *Reliable Computing*, 8:97–113, 2002.

[53] Volker Krätschmer. Coherent lower previsions and Choquet integrals. *Fuzzy Sets and Systems*, 138:469–484, 2003.

[54] Isaac Levi. *The Enterprise of Knowledge. An Essay on Knowledge, Credal Probability, and Chance*. MIT Press, Cambridge, 1983.

[55] A. A. Markov. Primer statisticheskogo issledovaniya nad tekstom "Evgeniya Onegina", illyustriruyuschij svyaz' ispytanij v cep'. *Izvestija Imp. Akad. nauk*, SPb, VI seriya(3):153–162, 1913.

[56] J. J. Martin. *Bayesian Decision Theory and Markov Chains*. John Wiley & Sons, New York, 1967.

[57] E. H. Moore and H. L. Smith. A general theory of limits. *American Journal of Mathematics*, 44(2):102–121, April 1922.

[58] David W. Mount. *Bioinformatics. Genome and Sequence Analysis*. Cold Spring Harbor Laboratory Press, New York, 2001.

[59] Tobias Müller and Martin Vingron. Modeling amino acid replacement. *Journal of Computational Biology*, 7(6):761–776, 2000.

[60] S. B. Needleman and C. D. Wunsch. An efficient method applicable to the search for similarities in the amino acid sequences of two proteins. *Journal of Molecular Biology*, 48:443–453, 1970.

[61] Craig G. Nevill-Manning, Cecil N. Huang, and Douglas L. Brutlag. Pairwise protein sequence alignment using Needleman-Wunsch and Smith-Waterman algorithms, 1997. personal communication.

[62] H. T. Nguyen, N. T. Nguyen, and T. Wang. On capacity functionals in interval probabilities. *International Journal of Uncertainty, Fuzziness, and Knowledge-Based Systems*, 5(3):359–377, June 1997.

[63] Blaise Pascal. *Pensées*. Maxi-Livres, Paris, 2001. Unfinished work, published posthumously from collected fragments. First incomplete edition: Port-Royal, 1670. First complete reproduction: Michaut, Basle, 1896.

[64] W. R. Pearson. Empirical statistical estimates for sequence similarity searches. *Journal of Molecular Biology*, 276(1):71–84, 1998.

[65] Frank P. Ramsey. Truth and probability. In R. B. Braithwaite, editor, *Foundations of Mathematics and other Logical Essays*, chapter VII, pages 156–198. Routledge and Kegan Paul, London, 1931. Published posthumously.

[66] B. Riemann. Ueber die Darstellbarkeit einer Function durch eine trigonometrische Reihe. *Abhandlungen der Königlichen Gesellschaft der Wissenschaften zu Göttingen*, 13:87–131, 1868. Originally delivered in 1854 by Riemann as part of his habilitation at the University of Göttingen.

[67] Sheldon M. Ross. *Stochastic Processes*. John Wiley & Sons, New York, 2nd edition, 1996.

[68] Jay K. Satia and Jr. Roy E. Lave. Markovian decision processes with uncertain transition probabilities. *Operations Research*, 21(3):728–740, 1973.

[69] Leonard J. Savage. *The Foundations of Statistics*. Dover, New York, 1972. Second revised edition.

[70] Eric Schechter. *Handbook of Analysis and Its Foundations*. Academic Press, San Diego, 1997.

[71] M. J. Schervish, T. Seidenfeld, J. B. Kadane, and I. Levi. Extensions of expected utility theory and some limitations of pairwise comparisons. In Jean-Marc Bernard, Teddy Seidenfeld, and Marco Zaffalon, editors, *ISIPTA '03 – Proceedings of the Third International Symposium on Imprecise Probabilities and Their Applications*, pages 496–510. Carleton Scientific, July 2003.

[72] T. Seidenfeld, M. J. Schervish, and J. B. Kadane. A representation of partially ordered preferences. *The Annals of Statistics*, 23:2168–2217, 1995.

[73] Amartya Sen. Social choice theory: A re-examination. *Econometrica*, 45(1):53–89, January 1977.

[74] Glenn Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, 1976.

[75] Glenn Shafer. Allocations of probability. *The Annals of Probability*, 7(5):827–839, October 1979.

[76] Matthias C. M. Troffaes. Efficient and robust global amino acid sequence alignment with uncertain evolutionary distances. In *Proceedings of the Tenth International Conference IPMU 2004 (Information Processing and Management of Uncertainty in Knowledge-Based Systems)*, volume 1, pages 571–578, July 2004.

[77] Matthias C. M. Troffaes. Learning and optimal control of imprecise Markov decision processes by dynamic programming using the imprecise Dirichlet model. In Miguel Lopéz-Díaz, María Á. Gil, Przemyslaw Grzegorzewski, Olgierd Hyrniewicz, and Jonathan Lawry, editors, *Soft Methodology and Random Information Systems*, pages 141–148, Berlin, 2004. Springer.

[78] Matthias C. M. Troffaes and Gert de Cooman. Extension of coherent lower previsions to unbounded random variables. In *Proceedings of the Ninth International Conference IPMU 2002 (Information Processing and Management of Uncertainty in Knowledge-Based Systems)*, pages 735–42, Annecy, France, July 2002. ESIA – Université de Savoie.

[79] Matthias C. M. Troffaes and Gert de Cooman. Lower previsions for unbounded random variables. In Przemyslaw Grzegorzewski, Olgierd Hryniewicz, and Mariá Ángeles Gil, editors, *Soft Methods in Probability, Statistics and Data Analysis*, Advances in Soft Computing, pages 146–155, New York, September 2002. Physica-Verlag.

[80] Matthias C. M. Troffaes and Gert de Cooman. Extension of coherent lower previsions to unbounded random variables. In B. Bouchon-Meunier, L. Foulloy, and R. R. Yager, editors, *Intelligent Systems for Information Processing: From Representation to Applications*, pages 277–288. North-Holland, November 2003.

[81] Lev V. Utkin and Sergey V. Gurov. Imprecise reliability for some new lifetime distribution classes. *Journal of Statistical Planning and Inference*, 105(1):215–232, June 2002.

[82] Giuseppe Vitali. Sul problema della misura dei gruppi di punti di una retta, 1905.

[83] J. von Neumann and O. Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, 1944.

[84] Detlof von Winterfeldt and Ward Edwards. *Decision Analysis and Behavioral Research*. Cambridge University Press, Cambridge, 1986.

[85] Peter Walley. Coherent lower (and upper) probabilities. Technical Report 22, University of Warwick, Coventry, 1981.

[86] Peter Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London, 1991.

[87] Peter Walley. Inferences from multinomial data: Learning about a bag of marbles. *Journal of the Royal Statistical Society*, 58(1):3–34, 1996.

[88] Peter Walley and Gert de Cooman. A behavioral model for linguistic uncertainty. *Information Sciences*, 134:1–37, 2001.

[89] Pei-Zhang Wang. Fuzzy contactibility and fuzzy variables. *Fuzzy Sets and Systems*, 8(1):81–92, June 1982.

[90] Chelsea C. White and Hany K. Eldeib. Markov decision processes with imprecise transition probabilities. *Operations Research*, 42(4):739–749, 1994.

[91] Stephen Willard. *General Topology*. Addison-Wesley, 1970.

[92] P. M. Williams. Indeterminate probabilities. In M. Przelecki, K. Szaniawski, and R. Wojcicki, editors, *Formal Methods in the Methodology of Empirical Sciences*, pages 229–246. Reidel, Dordrecht, 1976.

[93] Philip Wolfe and G. B. Dantzig. Linear programming in a Markov chain. *Operations Research*, 10:702–710, 1961.

[94] William Henry Young. A general theory of integration. *Philosophical Transactions of the Royal Society of London*, 204(A):211–252, 1905.

[95] Lofti A. Zadeh. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1:3–28, 1978.