Probability II (MATH 2647)

Lecturer Dr. O. Hryniv
email Ostap.Hryniv@durham.ac.uk
office CM309

http://maths.dur.ac.uk/stats/courses/ProbMC2H/Probability2H.html

or via DUO

This term we shall consider:

- Review of Core A Probability
- Elements of convergence and integration (*)
- Generating functions
- Markov chains
- Random walks (*)
- ...

(*) important if you are going to do Math Finance!

Lectures: Tuesday 11, Wednesday 11 in CM101 **weeks 1-10,20**

Problems Classes: Thursday 12 in CM101 **weeks 4,6,8,10**

Tutorials: according to your timetable **weeks 3,5,7,9,20**

Exam: 2 hours, 6 questions (3+3)

> *"Credit will be given for the best TWO answers from Section A and the best TWO answers from Section B. Questions in Section B carry ONE and a HALF times as many marks as those in Section A. Use of electronic calculators is forbidden."*

subject to confirmation for this year!

Homeworks: set on Wednesday,
due in **CM116** on Thursday the week after

exact deadline will be communicated via email!

☞ **the best preparation for the exam!** data!

☞ All course materials available from the **course webpage**!

Sample space & Events

Sample space Ω is a collection of all possible outcomes of a probabilistic experiment;

Event is a collection of possible outcomes, ie., a subset of the sample space.

E.g., the *impossible* event $\varnothing$, the *certain* event Ω;

also, if $A \subset \Omega$ and $B \subset \Omega$ are events, one considers $A \cup B$ (**A or B**), $A \cap B$ (**A and B**), $A^c \equiv \Omega \setminus A$ (**not A**), $A \setminus B$ (**A but not B**).

$\boldsymbol{\sigma}$-FIELDS

Definition

Let $\mathcal{F}$ be a collection of subsets of Ω. We shall call $\mathcal{F}$ a **$\boldsymbol{\sigma}$-field** if it has the following properties:

1. $\varnothing \in \mathcal{F}$;
2. if $A_1, A_2, \dots \in \mathcal{F}$, then $\bigcup_{k=1}^{\infty} A_k \in \mathcal{F}$;
3. if $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$.

Remark If condition 2. is replaced with

$$2'. \qquad \text{if } A_1,\, A_2 \in \mathcal{F}, \text{ then } A_1 \cup A_2 \in \mathcal{F};$$

equivalently, with

$$2''. \qquad \text{if } A_1,\, A_2,\, \dots,\, A_m \in \mathcal{F}, \text{ then } \bigcup_{k=1}^{m} A_k \in \mathcal{F};$$

the collection $\mathcal{F}$ is called a **field**.

Remark

Condition 2. in the definition above is important:
indeed, in a standard coin-flipping experiment let

$$A_k = \{\text{first 'head' occurs on } k\text{th flip}\} ;$$

then

$$\{\text{a 'head' observed}\} = \bigcup_{k=1}^{\infty} A_k .$$

This also explains why one needs to study infinite sample spaces.

Probability distribution

Definition

Let Ω be a sample space, and $\mathcal{F}$ be a σ-field of events in Ω.
A **probability distribution** P on $(\Omega, \mathcal{F})$ is a collection of *numbers* $\mathrm{P}(A)$, $A \in \mathcal{F}$, possessing the following properties:

- A1 for every event $A \in \mathcal{F}$, $\mathrm{P}(A) \geq 0$;
- A2 $\mathrm{P}(\Omega) = 1$;
- A3 for any pair of *incompatible* events A and B, $\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B)$;
- A4 for any *countable* collection $A_1, A_2, \ldots$ of *mutually incompatible* events,

$$\mathrm{P}\Big(\bigcup_{k=1}^{\infty} A_k\Big) = \sum_{k=1}^{\infty} \mathrm{P}(A_k).$$

recall: events A, B are **incompatible** if $A \cap B = \varnothing$.

Probability space

Definition
A *probability space* is a triple $(\Omega, \mathcal{F}, \mathsf{P})$, where Ω is a sample space, $\mathcal{F}$ is a σ-field of events in Ω, and $\mathsf{P}(\,\cdot\,)$ is a probability measure on $(\Omega, \mathcal{F})$.

In what follows we shall **always assume** that some probability space $(\Omega, \mathcal{F}, \mathsf{P})$ is fixed.

Why do we care?

- To avoid pathological events which cannot be assigned a probability
  Do not worry, no such problems in **this** course!
- To avoid "paradoxes" (e.g., Bertrand's paradox)
- ...

Bertrand's paradox

Suppose a chord of a circle is chosen at random. What is the probability that the chord is longer than a side of an equilateral triangle inscribed in the circle?

Different types of "randomness":

"random endpoints" / "random radius" / "random midpoint"

give different results!

Some properties

The following properties are immediate from the above axioms:

P1 for any pair of events A, B in Ω we have

$$\mathsf{P}(B \setminus A) = \mathsf{P}(B) - \mathsf{P}(A \cap B)\,,$$

$$\mathsf{P}(A \cup B) = \mathsf{P}(A) + \mathsf{P}(B \setminus A);$$

in particular, $\mathsf{P}(A^{\mathsf{c}}) = 1 - \mathsf{P}(A)$;

P2 if events A, B in Ω are such that $\varnothing \subseteq A \subseteq B \subseteq \Omega$, then

$$0 = \mathsf{P}(\varnothing) \leq \mathsf{P}(A) \leq \mathsf{P}(B) \leq \mathsf{P}(\Omega) = 1\,.$$

P3 if A_1, A_2, ..., A_n are events in Ω, then

$$\mathsf{P}\Big(\bigcup_{k=1}^{n} A_k\Big) \leq \sum_{k=1}^{n} \mathsf{P}(A_k)$$

with the inequality becoming an **equality** if these events are **mutually incompatible**;

Some properties [cont'd]

Probability measure is **continuous** along **monotone** sequences:

If $A_1 \subseteq A_2 \subseteq \dots$ are events and $A = \lim_{k\to\infty} A_k \equiv \bigcup_{k=1}^{\infty} A_k$, then

$$\mathsf{P}(A) = \lim_{k\to\infty} \mathsf{P}(A_k).$$

If $B_1 \supseteq B_2 \supseteq \dots$ are events and $B = \lim_{k\to\infty} B_k \equiv \bigcap_{k=1}^{\infty} B_k$, then

$$\mathsf{P}(B) = \lim_{k\to\infty} \mathsf{P}(B_k).$$

Example:
A coin with success probability $p \in (0,1)$ is tossed repeatedly. Let S_k be the event "no successes in the first k tosses". Then $S_k \to S$, where S is the event "no successes were observed".

As the sequence $\{S_k\}_{k\geq 1}$ is **monotone** with $\mathsf{P}(S_k) = (1-p)^k$, we get

$$\mathsf{P}(S) = \lim_{k\to\infty} \mathsf{P}(S_k) = 0.$$

Independence of events

Definition

We say that events A and B are *independent* if

$$\mathsf{P}\big(A \cap B\big) = \mathsf{P}(A)\,\mathsf{P}(B)\,.$$

Definition

A collection $A_1, A_2, \ldots$ of events is (mutually) *independent* if its **every finite** sub-collection is independent,
ie., for all $k \in \mathbb{N}$ and all $i_1, \ldots, i_k \geq 1$,

$$\mathsf{P}\Big(\bigcap_{\ell=1}^{k} A_{i_\ell}\Big) = \prod_{\ell=1}^{k} \mathsf{P}\big(A_{i_\ell}\big)\,.$$

Notion of *independence* is the key distinction between probability and (real) analysis (measure theory)!

Example:

A coin showing heads with probability p is tossed repeatedly.
Let R be the event "sub-word TT occurs before sub-word HH",

$$R = \{ TT, HTT, THTT, \dots \} .$$

Assuming independence of individual outcomes, find $\mathsf{P}(R)$.

Consider $r_k = \{ \text{"}TT \text{ first occurs at kth toss and no } HH \text{ observed"} \}$.
Then $\mathsf{P}(r_k) = p^m q^{k-m}$ with $q = 1 - p$ and $m = \lfloor k/2 \rfloor - 1$, implying $\mathsf{P}(R) = (1+p)q^2/(1-pq)$.

Conditional probability

Definition

The *conditional probability* of event A given event B such that $\mathsf{P}(B) > 0$, is

$$\mathsf{P}(A \mid B) \stackrel{\text{def}}{=} \frac{\mathsf{P}(A \cap B)}{\mathsf{P}(B)}\,.$$

It follows that

$$\mathsf{P}(A \cap B) = \mathsf{P}(A)\,\mathsf{P}(B \mid A) = \mathsf{P}(B)\,\mathsf{P}(A \mid B)\,;$$

in particular, $\mathsf{P}(A \mid B) = \mathsf{P}(A)$ if A and B are independent.

Law of total probability

If events $B_1, \ldots, B_n$ form a *partition* of Ω, then for any event A,

$$\mathrm{P}(A) = \sum_{k=1}^{n} \mathrm{P}(A \cap B_k) = \sum_{k=1}^{n} \mathrm{P}(B_k) \cdot \mathrm{P}(A \,|\, B_k)\,.$$

Example:
One independently rolls twenty standard six-sided dice. What is the probability that the sum of observed results is divisible by 6?

Example (Renewals)

A diligent janitor replaces a light bulb the instant it burns out. Suppose:

- that the first bulb is put in at time 0
- that the lifetimes of different light bulbs are independent
- and have the same distribution.

Let $A = \{\text{a light bulb was replaced at time } n\}$, with $r_n = \mathrm{P}(A)$;

$B_k = \{\text{the first light bulb was replaced at time } k\}$, $f_k = \mathrm{P}(B_k)$.

The law of total probability then implies

$$r_n = \mathrm{P}(A) = \sum_{k=1}^{n} \mathrm{P}(B_k) \cdot \mathrm{P}(A \mid B_k) = \sum_{k=1}^{n} f_k r_{n-k}\,.$$

Sequential Bayes formula

If events A, B, C are such that $\mathrm{P}(B \cap C) > 0$, then

$$\mathrm{P}\big(A \cap B \mid C\big) = \mathrm{P}\big(A \mid B \cap C\big)\,\mathrm{P}\big(B \mid C\big)\,.$$

Equivalently,

$$\mathrm{P}\big(A \cap B \cap C\big) = \mathrm{P}\big(A \mid B \cap C\big)\,\mathrm{P}\big(B \mid C\big)\,\mathrm{P}\big(C\big)\,.$$

Example:
Consecutive winnings of a game, X_k, are iid random variables with

$$\mathrm{P}(X=1) = 1 - \mathrm{P}(X=-1) = p\,,$$

so the total winning is $S_0 = 0$, $S_n = \sum_{k=1}^{n} X_k$. Consider the events

$$A = \big\{S_n = 2 \text{ for some } n > 0\big\} = \big\{\text{win } 2\big\}$$

$$B = \big\{S_n = 1 \text{ for some } n > 0\big\} = \big\{\text{win } 1\big\}\,, \qquad C = \big\{S_0 = 0\big\}\,.$$

By sequential Bayes,

$$\begin{aligned}\mathrm{P}\big(A \mid C\big) = \mathrm{P}\big(A \cap B \mid C\big) &= \mathrm{P}\big(A \mid B \cap C\big)\,\mathrm{P}\big(B \mid C\big)\\ &\overset{\textbf{!!}}{=} \mathrm{P}\big(1 \rightsquigarrow 2\big)\,\mathrm{P}\big(0 \rightsquigarrow 1\big) \overset{\textbf{!?}}{=} \mathrm{P}(B \mid C)^2\,.\end{aligned}$$

In other words, $\mathrm{P}_0(\{\text{win } 2\}) = \mathrm{P}_0(\{\text{win } 1\})^2$.

Random variables

Definition

An integer-valued random variable X is a map $\Omega \to \mathbb{Z}$;

its *distribution* is given by

$$p_k = \mathsf{P}(X = k)\,, \qquad k \in \mathbb{Z}\,;$$

its *expectation* $\mathsf{E}X$ is the number

$$\mathsf{E}X = \sum_{k \in \mathbb{Z}} k\, p_k \in [-\infty, +\infty]\,.$$

Independence of random variables

Definition

Random variables X and Y are **independent**, if for **all** A, $B \subseteq \mathbb{R}$,

$$\mathsf{P}(X \in A, Y \in B) = \mathsf{P}(X \in A)\,\mathsf{P}(Y \in B)\,.$$

Key property:

If X and Y are *independent* rv's in $\mathbb{Z}$, and f and g are real functions, then $f(X)$ and $g(Y)$ are **independent rv's**.

Example:

If X, Y are independent integer-valued rv's, then

$$\mathsf{E}\big(s^{X+Y}\big) = \mathsf{E}\big(s^{X}\big)\mathsf{E}\big(s^{Y}\big)\,.$$

[similarly, for any **countable** collection of rv's]

Conditional expectation

If an event A is such that $\mathsf{P}(A) > 0$, then the *conditional expectation* of a random variable X *given* A is

$$\mathsf{E}\big(X \mid A\big) = \sum_{k \in \mathbb{Z}} k\mathsf{P}\big(X = k \mid A\big) .$$

Partition theorem for expectations

If events B_1, B_2, form a partition of Ω, then

$$\mathsf{E}(X) = \sum_{k=1}^{\infty} \mathsf{E}(X \mid B_k) \cdot \mathsf{P}(B_k) ,$$

provided all expressions on the right are finite.

Example:
Let random variables X_1, X_2, ... be independent identically distributed, with values in $\{-K, \dots, K\}$. Put $S_n = \sum_{k=1}^{n} X_k$.

Let, further, $N \geq 1$ be an integer valued random variable independent of the sequence X_k, $k \geq 1$. Assume that $\mathsf{E}N < \infty$ and denote $S_N = X_1 + \cdots + X_N$. Find $\mathsf{E}S_N$.

For every $n \in \mathbb{N}$, we have

$$\mathsf{E}\big(S_N \mid N = n\big) = n\mathsf{E}(X_1 \mid N = n) = n\mathsf{E}(X_1)$$

so that

$$\begin{aligned}\mathsf{E}\big(S_N\big) &= \sum_{n=1}^{\infty} \mathsf{E}\big(S_N \mid N = n\big)\, \mathsf{P}(N = n) \\ &= \mathsf{E}(X_1) \sum_{n=1}^{\infty} n\, \mathsf{P}(N = n) = \mathsf{E}(X_1)\, \mathsf{E}(N)\,.\end{aligned}$$

Exercise: Use the same approach to compute $\mathsf{Var}S_N$.