Finite second-order exchangeability and the adjustment of beliefs

Simon C Shaw & Michael Goldstein
Department of Mathematical Sciences
University of Durham, UK

March 27, 2002

ABSTRACT

We consider the adjustment of a finite sequence of second-order exchangeable vectors, and show that the adjustment shares the same powerful properties as in the infinite case discussed by Goldstein & Wooff (1998). The types of information we gain by sampling are identified with the orthogonal canonical directions and we show that the canonical directions for the adjustment of the population mean collection are the same for each sample size. These canonical directions share the same co-ordinate representation for all population sizes $N > 2$ enabling simple comparisons between both the effects of different sample sizes and of different population sizes. If the finite sequence may be embedded in an infinite sequence, we compare the adjustment of the population mean collection with the underlying population mean collection in the infinite case showing that the differences are quantitative rather than qualitative in that the respective canonical directions share, up to a scale-factor, the same co-ordinate structure. The adjusted variances for the canonical directions in the finite sequence may be obtained from those in the infinite sequence via multiplication of a finite population correction.

Keywords: Bayes linear methods; canonical directions; canonical resolutions; finite second-order exchangeability; finite population correction.

1 Introduction

Goldstein & Wooff (1998) investigated second-order exchangeable belief adjustment when the sampling population was judged to be infinite. In sampling n independent and identically distributed (iid) quantities, the variance of the sample average is $(1/n)$ times the variance of a single such quantity and as such is is often easy to gauge the relationship between sample size and variance reduction of the population mean. Goldstein & Wooff (1998) show that for second-order belief adjustment there is a natural generalisation when we adjust beliefs about a vector of population means, preserving the simplicity of the relationship between sample size and variance reduction.

In practice however, the assumption of a sequence being potentially infinite is a modelling simplification, for it is usually possible to give an upper bound on the length of the sequence under consideration. However, it is not always easy or straightforward to specify accurately the upper bound and so it is often easier to proceed with the simplification that the sequence was of potentially infinite length.

As with any modelling assumption, we would like to investigate the consequences of the infinite approximation and how it effects both our modelling and learning. In sampling n iid quantities from a population of size N, the link between the finite and the infinite is through the multiplication of the (infinite) variance by the finite population correction $(1-(n/N))$ (see Barnett (1974; p26)). In this paper we show that a similar relationship holds when we perform a second-order belief adjustment for a vector of population means drawn from a finite population compared to the infinite assumption.

This approach assumes that the finite second-order exchangeable sequence may be embedded in or extended to an infinite sequence of similarly defined collections. Of course, not every finite sequence may be embedded into a longer, possibly infinite, sequence and so the direct study of finite sequences has a broader scope than the study of infinite sequences for it will include sequences that have no analogue in the infinite framework.

We proceed as follows. We recall the representation theorem for finite second-oder exchangeable beliefs and discuss the geometry underpinning the Bayes linear approach. The resolution transform is the tool used to analyse the relationships between a collection of Bayes linear belief adjustments and we show that the transform for the population mean collection, relating to a population of size N, has essentially the same form whatever the sample size. We then show that the results of this analysis may be obtained by studying the population structure of a population of size 2, enabling us to draw simple comparisons between different population sizes. Finally, we consider the relationship between the finite reality and the infinite assumption, showing that the qualitative features of the adjustment of the respective population mean collections are the same and the quantitative differences may be obtained via the use of a finite population correction. The theory is illustrated by an example.

2 Second-order exchangeable beliefs

We wish to make a series of measurements on a sample of individuals; our interests lie in traits common to the individuals and so we elect to make the same series of measurements on each individual. We gather these together as the collection $\mathcal{C} = \{X_1, X_2, \ldots\}$, finite or infinite, where each X_v is a real valued function of the quantities that will be measured on the individuals. The collection for each individual is generated from $\mathcal{C}$ by letting $\mathcal{C}_i = \{X_{1i}, X_{2i}, \ldots\}$ be the measurements for the ith individual. The full population collection is formed as the union of all of the elements in all of the individual collections, $\mathcal{C}_i$, and denoted by $\mathcal{C}^*$.

The prior means, variance and covariance for each pair of quantities are specified directly. Thus, expectation is treated as a primitive quantity; de Finetti (1974) provides a detailed explanation of this approach. We regard a sequence of quantities as second-order exchangeable if our first and second-order beliefs about the sequence are unaffected by permuting the order of the sequence.

Definition 1 *(Goldstein (1986a)) The collection of measurements $\mathcal{C}$ is second-order exchangeable over the full collection $\mathcal{C}^*$ if*

$$E(X_{vi}) = m_v \ \forall v, i; \tag{1}$$
$$Cov(X_{vi}, X_{wi}) = d_{vw} \ \forall v, w, i; \tag{2}$$
$$Cov(X_{vi}, X_{wj}) = c_{vw} \ \forall v, w, i \neq j. \tag{3}$$

Thus, as Goldstein (1986a) emphasises, irrespective of the total number of individuals in the population, all that is required is the cogitation of two individuals with all other specifications following from the perceived symmetries in the population. Definition 1 applies whether the full collection $\mathcal{C}^*$ is finite or infinite. The definition of second-order exchangeability involves a much more limited and achievable prior specification than that for full exchangeability, see for example Lad (1996; Definition 5.14), which although it reduces the prior specification burden is still far more detailed than we would ever reasonably be able to make. Goldstein (1986a) derives the representation theorems for second-order exchangeable beliefs for the separate cases when the sampling population, $\mathcal{C}^*$, is infinite and when it is finite. These representations are constructed directly from the specifications (1) - (3) and so are consistent with our beliefs about observables. The finite population representation theorem is given below; the infinite population representation theorem may be deduced in the limit $N \rightarrow \infty$, the limit being in mean square.

Theorem 1 *(Goldstein (1986a)) If the population collection consists of N individuals, that is $\mathcal{C}^* = \cup_{i=1}^{N} \mathcal{C}_i$, and $\mathcal{C}$ is second-order exchangeable over $\mathcal{C}^*$, then we may introduce the further collections of random quantities $\mathcal{M}^{[N]}(\mathcal{C}) = \{\mathcal{M}^{[N]}(X_1),\ \mathcal{M}^{[N]}(X_2),\ \ldots\}$, and, for each $i = 1, \ldots, N$, $\mathcal{R}_i^{[N]}(\mathcal{C}) = \{\mathcal{R}_i^{[N]}(X_1),\ \mathcal{R}_i^{[N]}(X_2),\ \ldots\}$, and write*

$$X_{vi} \quad = \quad \mathcal{M}^{[N]}(X_v) + \mathcal{R}_i^{[N]}(X_v), \tag{4}$$

where $\mathcal{M}^{[N]}(X_v) = (1/N)\sum_{i=1}^{N} X_{vi}$. The collections $\mathcal{M}^{[N]}(\mathcal{C})$ and $\mathcal{R}_i^{[N]}(\mathcal{C})$ satisfy the following relationships

$$E(\mathcal{M}^{[N]}(X_v)) \quad = \quad m_v \ \forall v; \tag{5}$$

$$E(\mathcal{R}_i^{[N]}(X_v)) \quad = \quad 0 \ \forall v, i; \tag{6}$$

$$Cov(\mathcal{M}^{[N]}(X_v), \mathcal{M}^{[N]}(X_w)) \quad = \quad c_{vw} + \frac{1}{N}(d_{vw} - c_{vw}) \ \forall v, w; \tag{7}$$

$$Cov(\mathcal{M}^{[N]}(X_v), \mathcal{R}_j^{[N]}(X_w)) \quad = \quad 0 \ \forall v, w, j; \tag{8}$$

$$Cov(\mathcal{R}_i^{[N]}(X_v), \mathcal{R}_j^{[N]}(X_w)) \quad = \quad \begin{cases} \frac{N-1}{N}(d_{vw} - c_{vw}) & \text{if } i = j \ \forall v, w; \\ -\frac{1}{N}(d_{vw} - c_{vw}) & \text{otherwise.} \end{cases} \tag{9}$$

This representation theorem expresses each individual measurement as the sum of a population mean quantity and a residual from this mean quantity. $\mathcal{M}^{[N]}(\mathcal{C})$ is the population mean collection for the individuals and $\mathcal{R}_i^{[N]}(\mathcal{C})$ the discrepancy for the ith individual from the overall mean. Notice that, from equation (8), the population mean collection is uncorrelated with each individual residual collection, $\mathcal{R}_i^{[N]}(\mathcal{C})$. The residual collections, see equation (9), are correlated (to order $(1/N)$) but in the limit $N \rightarrow \infty$ are uncorrelated. This limit provides the situation of infinite second-order exchangeability discussed by Goldstein & Wooff (1998). The population mean collection are denoted by $\mathcal{M}(\mathcal{C})$ and the residual collection for the ith individual is denoted by $\mathcal{R}_i(\mathcal{C})$. These two collections are unobservable, with $\mathcal{M}(\mathcal{C})$ acting as underlying population mean collection and $\mathcal{R}_i(\mathcal{C})$ the ith discrepancy from the underlying average. Contrast this with the observability of $\mathcal{M}^{[N]}(\mathcal{C})$ and each $\mathcal{R}_i^{[N]}(\mathcal{C})$.

We may extend Theorem 1 over the natural linear space by letting $\langle\mathcal{C}\rangle$ denote the collection of finite linear combinations, $\mathcal{X} = \sum_u \alpha_u X_{v_u}$, where $\{v_1, v_2, \ldots\}$ is a general finite subset of integers, of elements of $\mathcal{C}$. For each individual i, we construct the value of $\mathcal{X}$, and denote this by $\mathcal{X}_i = \sum_u \alpha_u X_{v_u i}$. Let $\langle\mathcal{C}_i\rangle$, $\langle\mathcal{M}^{[N]}(\mathcal{C})\rangle$, $\langle\mathcal{R}_i^{[N]}(\mathcal{C})\rangle$, for each $i = 1, \ldots, N$, be, respectively, the collection of finite linear combinations of elements of $\mathcal{C}_i$, $\mathcal{M}^{[N]}(\mathcal{C})$ and

$\mathcal{R}_i^{[N]}(\mathcal{C})$, then for each $\mathcal{X} \in \langle\mathcal{C}\rangle$, for each individual i, we have

$$\mathcal{X}_i = \mathcal{M}^{[N]}(\mathcal{X}) + \mathcal{R}_i^{[N]}(\mathcal{X}), \tag{10}$$

where $\mathcal{M}^{[N]}(\mathcal{X}) = \sum_u \alpha_u \mathcal{M}^{[N]}(X_{v_u})$ and $\mathcal{R}_i^{[N]}(\mathcal{X}) = \sum_u \alpha_u \mathcal{R}_i^{[N]}(X_{v_u})$.

3 Geometric representations; Bayes linear methods

De Finetti (1974; Sections 2.8, 4.17) formulated a geometric interpretation of an individual's current expectations. For a general collection of random quantities $\mathcal{B} = \{\mathcal{B}_1, \mathcal{B}_2, \ldots\}$, of interest, construct the space $\langle\mathcal{B}\rangle$ of finite linear combinations of the elements of $\mathcal{B}$ with the unit constant, $\mathcal{B}_0$, added. We view $\langle\mathcal{B}\rangle$ as a vector space by considering each $\mathcal{B}_v$ as a vector with linear combinations of random quantities represented as the corresponding linear combination of vectors. The geometric framework is added by forming the inner product space $[\mathcal{B}]$ from the minimal closure of $\langle\mathcal{B}\rangle$ by imposing the following inner product and norm for $\mathcal{A}$, $\mathcal{A}^\dagger \in \langle\mathcal{B}\rangle$,

$$(\mathcal{A}, \mathcal{A}^\dagger) = Cov(\mathcal{A}, \mathcal{A}^\dagger); \tag{11}$$

$$\| \mathcal{A} \|^2 = Var(\mathcal{A}). \tag{12}$$

We restrict $\mathcal{B}$ to elements with finite prior variance and the inner product is formed over the closure of the equivalence classes of random quantities which differ by a constant. Goldstein (1981; p108) follows the convention of standardising every quantity $\mathcal{A}$ by subtracting its prior mean. Goldstein (1986b; p200) shows how the standardised inner product representation may be linked to the (unstandardised) random quantities via a projection operation. For two subspaces, $[\mathcal{B}^\dagger]$ and $[\mathcal{B}^\ddagger]$, if every element of the collection $\mathcal{B}^\dagger$ is uncorrelated with every element of $\mathcal{B}^\ddagger$ then $[\mathcal{B}^\dagger]$ and $[\mathcal{B}^\ddagger]$ are said to be orthogonal, written $[\mathcal{B}^\dagger] \perp [\mathcal{B}^\ddagger]$.

Suppose we are to receive the values of a data collection $\mathcal{D} = (\mathcal{D}_1, \mathcal{D}_2, \ldots)$ where the inner product $[\mathcal{B} \cup \mathcal{D}]$ is specified. In a Bayes linear analysis, we are interested in what happens to the elements of the geometric structure following the receipt of $\mathcal{D}$. For a random quantity $\mathcal{A} \in \langle\mathcal{B}\rangle$, the element $\mathcal{A}^\ddagger \in [\mathcal{D}]$ which minimises $\| \mathcal{A} - \mathcal{A}^\ddagger \|$ is termed the adjusted expectation of $\mathcal{A}$ given $\mathcal{D}$, written $E_{\mathcal{D}}(\mathcal{A})$. Thus, $E_{\mathcal{D}}(\mathcal{A})$ is the orthogonal projection of $\mathcal{A}$ into $[\mathcal{D}]$. The squared orthogonal distance from $\mathcal{A}$ to $[\mathcal{D}]$ is termed the adjusted variance of $\mathcal{A}$ given $\mathcal{D}$, written $Var_{\mathcal{D}}(\mathcal{A})$. Thus, we may decompose $\mathcal{A}$ into the sum of two uncorrelated components, $\mathcal{A} = E_{\mathcal{D}}(\mathcal{A}) + \{\mathcal{A} - E_{\mathcal{D}}(\mathcal{A})\}$. Taking the variance we have

$$Var(\mathcal{A}) = Var(E_{\mathcal{D}}(\mathcal{A})) + Var_{\mathcal{D}}(\mathcal{A}). \tag{13}$$

$Var(E_{\mathcal{D}}(\mathcal{A}))$ is the variability of $\mathcal{A}$ which is accounted for by the variability in $\mathcal{D}$ and is often called the resolved variance, $RVar_{\mathcal{D}}(\mathcal{A})$. Intuitively, the observation of $\mathcal{D}$ is expected to be informative for $\mathcal{A}$ if $RVar_{\mathcal{D}}(\mathcal{A})$ is large relative to $Var(\mathcal{A})$. Thus, the resolution,

$$R_{\mathcal{D}}(\mathcal{A}) = \frac{RVar_{\mathcal{D}}(\mathcal{A})}{Var(\mathcal{A})} = \frac{Var(\mathcal{A}) - Var_{\mathcal{D}}(\mathcal{A})}{Var(\mathcal{A})}, \tag{14}$$

is a simple, scale free, measure of the impact of the adjustment upon $\mathcal{A}$. If $R_{\mathcal{D}}(\mathcal{A})$ is near zero, then, relative to our prior knowledge about $\mathcal{A}$, we do not expect the Bayes linear analysis of the data to be informative for $\mathcal{A}$. A value of $R_{\mathcal{D}}(\mathcal{A})$ close to one suggests that the analysis is expected to be highly informative.

An overview of the Bayes linear methodology may be found in Goldstein (1999), whilst Goldstein (1994, 1997) concentrates upon foundational aspects. As Hartigan (1969; p447) points out, if $\mathcal{B} \cup \mathcal{D}$ is multivariate normal then, for each $\mathcal{A} \in \langle \mathcal{B} \rangle$, $E_{\mathcal{D}}(\mathcal{A}) = E(\mathcal{A}|\mathcal{D})$ and $Var_{\mathcal{D}}(\mathcal{A}) = Var(\mathcal{A}|\mathcal{D})$.

We return now to exchangeable beliefs discussed in Section 2. We consider the observation of a sample of $n \leq N$ exchangeable collections, which we label for convenience, $\mathcal{C}_1, \ldots, \mathcal{C}_n$ and we let $\mathcal{C}(n) = \cup_{i=1}^{n} \mathcal{C}_i$. We want to use this data to revise our beliefs over the mean collection, $\mathcal{M}^{[N]}(\mathcal{C})$. We are also interested in learning about future collections of individuals and we shall comment upon this in Section 10. For each $\mathcal{X} \in \langle \mathcal{C} \rangle$, denote the average of the first n values by $\mathcal{S}_n(\mathcal{X}) = (1/n) \sum_{i=1}^{n} \mathcal{X}_i$. The collection of sample averages are denoted by $\mathcal{S}_n(\mathcal{C}) = \{\mathcal{S}_n(X_1), \mathcal{S}_n(X_2), \ldots\}$. We have the following theorem; the proof is in the appendix.

Theorem 2 *If $\mathcal{C}$ is second-order exchangeable over $\mathcal{C}^* = \cup_{i=1}^{N} \mathcal{C}_i$, then $\mathcal{S}_n(\mathcal{C})$ is Bayes linear sufficient for $\mathcal{C}(n)$ for adjusting the collection $\mathcal{M}^{[N]}(\mathcal{C})$.*

A detailed discussion of Bayes linear sufficiency may be found in Goldstein & O'Hagan (1996); the meaning is that having performed a linear fit on $\mathcal{S}_n(\mathcal{C})$, we can obtain no further reduction in adjusted variance by linear fitting the mean collection on the full sample $\mathcal{C}(n)$.

4 Coherency relations between different sample sizes in finite sampling

The resolution transform for general collections $\mathcal{B}$ given a data collection $\mathcal{D}$ for each $\mathcal{A} \in [\mathcal{B}]$ is defined to be $T_{\mathcal{D}}(\mathcal{A}) = E_{\mathcal{B}}\{E_{\mathcal{D}}(\mathcal{A})\}$. Thus, for each $\mathcal{A} \in [\mathcal{B}]$, $T_{\mathcal{D}}(\mathcal{A})$ is the point in $[\mathcal{B}]$ which is closest to $E_{\mathcal{D}}(\mathcal{A})$. Goldstein (1981) shows that $T_{\mathcal{D}}$ is a bounded, self-adjoint operator on $[\mathcal{B}]$, satisfying for each $\mathcal{A} \in [\mathcal{B}]$,

$$Var_{\mathcal{D}}(\mathcal{A}) \quad = \quad Var(\mathcal{A}) - Cov(\mathcal{A}, T_{\mathcal{D}}(\mathcal{A})). \tag{15}$$

Hence, substituting equation (15) into equation (14), we have that

$$R_{\mathcal{D}}(\mathcal{A}) \quad = \quad \frac{Cov(\mathcal{A}, T_{\mathcal{D}}(\mathcal{A}))}{Var(\mathcal{A})}. \tag{16}$$

If either of the collections $\mathcal{B}$ or $\mathcal{D}$ are finite dimensional, then Goldstein (1981) showed that we may extract for $T_{\mathcal{D}}$ a set of eigenvectors, $Z = \{Z_1, Z_2, \ldots\}$, which form a basis for $[\mathcal{B}]$ and corresponding ordered eigenvalues, $1 \geq \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_r \geq 0$. If both the collections $\mathcal{B}$ and $\mathcal{D}$ are infinite then, unless a certain compactness condition holds, the eigenstructure of $T_{\mathcal{D}}$ may be more complicated (see Goldstein (1981; p114)). The eigenvalue λ_j is termed the jth canonical resolution, and the corresponding eigenvector Z_j is termed the jth canonical direction. The canonical resolutions and directions satisfy the following properties. The Z_j have expectation 0, are mutually uncorrelated, and are scaled to have prior variance 1. As Z forms a basis for $[\mathcal{B}]$, each $\mathcal{A} \in [\mathcal{B}]$ may be expressed as $\mathcal{A} = \sum_{j=1}^{r} Cov(\mathcal{A}, Z_j) Z_j$. In addition, however, we may use equations (15) and (16) to express the adjusted variance and resolution for each $\mathcal{A} \in [\mathcal{B}]$ as a linear combination of the adjusted variances and resolutions, respectively, of the Z_i as follows:

$$Var_{\mathcal{D}}(\mathcal{A}) \;=\; \sum_{j=1}^{r} (1 - \lambda_j) Cov(\mathcal{A}, Z_j)^2; \quad R_{\mathcal{D}}(\mathcal{A}) \;=\; \frac{\sum_{j=1}^{r} \lambda_j Cov(\mathcal{A}, Z_j)^2}{Var(\mathcal{A})}. \tag{17}$$

Hence, constrained by being uncorrelated with $(Z_1, \ldots, Z_j)$, Z_{j+1} is the element of $[\mathcal{B}]$ maximising the resolution. Thus, we expect to learn most about elements of $[\mathcal{B}]$ having strong correlations with the directions with large resolutions. The canonical directions thus identify the types of information that we expect to gain by sampling, with the quantification of how much we learn in each direction being provided by the canonical resolutions. Thus, we may gain insights into the benefits of different experimental designs by comparing the canonical structures across different choices of $\mathcal{D}$. Goldstein & Wooff (1997) adopt such an approach for choosing sample sizes in balanced designs.

Goldstein & Wooff (1998) examine the adjustment of beliefs based upon second-order exchangeable samples drawn from infinite populations. Suppose that a sample of size n is drawn and denote the resolution transform for the mean collection $\mathcal{M}(\mathcal{C})$, based on n observations, $\mathcal{C}(n)$, by $T_n(\cdot) = E_{\mathcal{M}(\mathcal{C})}\{E_{\mathcal{C}(n)}(\cdot)\}$. They demonstrated strong coherence relationships between the adjustments based upon samples of different sizes. The canonical directions are the same for all sample sizes, and if λ is a canonical resolution for a sample of size 1 with corresponding canonical direction W, then $\lambda_{(n)} = n\lambda/\{(n-1)\lambda + 1\}$ is the canonical resolution corresponding to the canonical direction W for a sample of size n. The underlying qualitative features of the adjustment are not effected by sample size and it is straightforward to use the values $\lambda_{(n)}$ to simplify design questions for which the sample size has to be determined which will ensure specific variance reductions for combinations of the underlying mean components which are of interest.

We now show that there are similar coherence relationships obtained between adjustments based upon second-order exchangeable samples drawn from finite rather than infinite populations. Denote the resolution transform for the mean collection $\mathcal{M}^{[N]}(\mathcal{C})$, based on n observations, $\mathcal{C}(n) = \bigcup_{r=1}^{n} \mathcal{C}_r$, by $T_n^{[N]}(\cdot) = E_{\mathcal{M}^{[N]}(\mathcal{C})}\{E_{\mathcal{C}(n)}(\cdot)\}$. We have the following theorem, the proof is in the appendix.

Theorem 3 *The eigenvectors of $T_n^{[N]}$ are the same for each n. If $Y^{[N]}$ is an eigenvector of $T_1^{[N]}$ with corresponding eigenvalue $\lambda^{[N]}$, then the corresponding eigenvalue $\lambda_{(n)}^{[N]}$ for $T_n^{[N]}$ is*

$$\lambda_{(n)}^{[N]} \quad = \quad \frac{n(N-1)\lambda^{[N]}}{(n-1)N\lambda^{[N]} + (N-n)}. \tag{18}$$

Theorem 3 shows that for the adjustment of $[\mathcal{M}^{[N]}(\mathcal{C})]$ by $\mathcal{C}(n)$, the canonical directions $Z_1^{[N]} = Var(Y_1^{[N]})^{-\frac{1}{2}}Y_1^{[N]}$, $Z_2^{[N]} = Var(Y_2^{[N]})^{-\frac{1}{2}}Y_2^{[N]}$, ..., where each $Y_s^{[N]}$ is an eigenvector of $T_n^{[N]}$, are the same for each n, and are termed the canonical directions induced by (finite) exchangeability. The coherence relations discovered for second-order exchangeable infinite populations by Goldstein & Wooff (1998) are also present in the modelling of second-order exchangeable finite populations. Notice that from equation (11), for any sample size n, the adjusted variance for any $\mathcal{M}^{[N]}(\mathcal{X}) \in [\mathcal{M}^{[N]}(\mathcal{C})]$ is given by

$$Var_{\mathcal{C}(n)}(\mathcal{M}^{[N]}(\mathcal{X})) \quad = \quad \sum_i \frac{(N-n)(1-\lambda_i^{[N]})}{(n-1)N\lambda_i^{[N]} + (N-n)} Cov(\mathcal{M}^{[N]}(\mathcal{X}), Z_i^{[N]})^2, \tag{19}$$

and the resolution may be similarly expressed. Thus, in an identical manner to Goldstein & Wooff (1998), we see how we may exploit equation (18) to simplify design problems for which we are required to choose the sample size to achieve a specified variance reduction in elements of interest in the $[\mathcal{M}^{[N]}(\mathcal{C})]$. We have the following corollary.

Corollary 1 *Suppose that $Y^{[N]}$ is an eigenvector of $T_1^{[N]}$ with eigenvalue $\lambda^{[N]} > 0$. Then the sample size n required to achieve a proportionate variance reduction of κ for $Y^{[N]}$, that is so that $Var_{\mathcal{C}(n)}(Y^{[N]}) \leq (1-\kappa)Var(Y^{[N]})$, is $n \geq \{\kappa(1-\lambda^{[N]})\}/\{\lambda^{[N]}(1-\kappa)+(1/N)(1-\lambda^{[N]})\}$.*

If the minimal eigenvalue of $T_1^{[N]}$ is $\lambda_{min}^{[N]}$, then to achieve a proportionate variance reduction of κ for every element of $[\mathcal{M}^{[N]}(\mathcal{C})]$ requires a sample size, rounded up, of $\{\kappa(1-\lambda_{min}^{[N]})\}/\{\lambda_{min}^{[N]}(1-\kappa)+(1/N)(1-\lambda_{min}^{[N]})\}$.

5 Extendible second-order exchangeable sequences

The use of finite second-order exchangeability is desirable within the subjective framework as it acknowledges the necessarily finite nature of our actual second-order exchangeability judgements. The finite sequence could form part of a larger (possibly infinite) sequence of second-order exchangeable individuals or the sequence cannot be embedded in any longer sequence of second-order exchangeable collections. The limit to the sequence length may be theoretical: our second-order judgements may be coherent for a population of size N_1, but incoherent for a population of size $N_2 > N_1$. For example, if there is a $c_{vv} < 0$, then equation (7) shows that we could obtain a negative variance in the representation theorem. Alternatively, the limit may be conceptual: the sequence is of length N and it makes no sense to think of a longer sequence. An example that covers both of these scenarios is when we consider drawing a sample without replacement from an urn containing two balls, one marked 0 and the other marked 1. This example was discussed by Diaconis (1977) in highlighting that de Finetti's representation theorem for (fully) exchangeable sequences required an infinite sequence. In the second-order setting, letting X denote the value on the ball, we could assess for each individual i, $E(X_i) = 1/2$, $Var(X_i) = 1/4$ and for differing individuals i, j, $Cov(X_i, X_j) = -1/4$. Mathematically, the only coherent second-order exchangeable sequence having these prior beliefs must have $N = 2$ since, from equation (7), $Var(\mathcal{M}^{[N]}(X)) = (2-N)/4N < 0$ if $N > 2$. Conceptually, the sequence has to be of length two: there are only two balls in the urn.

In many situations however, the population size may not be known and so we seek methods to compare the effect of the population size on our calculations. We make the following definition.

Definition 2 *Suppose that the collection of measurements $\mathcal{C}$ is second-order exchangeable over $\mathcal{C}_N^* = \cup_{i=1}^N \mathcal{C}_i$. The population $\mathcal{C}_N^*$ is M-extendible if $\mathcal{C}$ is second-order exchangeable over $\mathcal{C}_{N+M}^* = \cup_{i=1}^{N+M} \mathcal{C}_i$.*

Thus, our consideration for each case and our considerations between each pair of cases is the same in the two sequences. We remark that this definition is purely theoretical. We may view infinite second-order exchangeability as corresponding to the assumption of M-extendibility for all $M > N$. Recall that for $\mathcal{C}$ to be second-order exchangeable over $\mathcal{C}^*$, all we require is the consideration of two cases with all other cases following by symmetry. This observation allows us to regard any second-order exchangeable sequence as having been extended from a second-order collection of length two. We now examine the relationship between the adjustment of the mean collections $[\mathcal{M}^{[2]}(\mathcal{C})]$ and $[\mathcal{M}^{[N]}(\mathcal{C})]$ given a collection of second-order exchangeable measurements which are $(N-2)$-extendible and a sample, of size one, drawn from the populations.

Theorem 4 *Suppose that $Y^{[2]} = \sum_u \xi_u \mathcal{M}^{[2]}(X_{v_u})$, where $\{v_1, v_2, \ldots\}$ is a general finite subset of integers, is an eigenvector of $T_1^{[2]}$, with eigenvalue $\lambda^{[2]}$. Then $Y^{[N]} = \sum_u \xi_u \mathcal{M}^{[N]}(X_{v_u})$ is an eigenvector of $T_1^{[N]}$, with eigenvalue*

$$\lambda^{[N]} = \frac{2(N-1)\lambda^{[2]} + (2-N)}{N}. \tag{20}$$

The eigenvectors of $T_1^{[2]}$ and $T_1^{[N]}$ thus share the same co-ordinate representation, with easily modified eigenvalues. By combining Theorem 3 and Theorem 4, we have the following corollary.

Corollary 2 *If $Y^{[2]} = \sum_u \xi_u \mathcal{M}^{[2]}(X_{v_u})$ is an eigenvector of $T_1^{[2]}$, with eigenvalue $\lambda^{[2]}$, then $Y^{[N]} = \sum_u \xi_u \mathcal{M}^{[N]}(X_{v_u})$ is, for each $n \leq N$, an eigenvector of $T_n^{[N]}$ with eigenvalue*

$$\lambda_{(n)}^{[N]} = \frac{2n(N-1)^2\lambda^{[2]} - n(N-1)(N-2)}{2(n-1)N(N-1)\lambda^{[2]} - (n-2)N(N-1)}. \tag{21}$$

Hence, the canonical directions for the adjustment of $[\mathcal{M}^{[N]}(\mathcal{C})]$ by $\mathcal{C}(n)$ have, up to a scale factor to ensure a prior variance of one, the same co-ordinate representation for all N and n, with simply modified canonical resolutions. Thus, the qualitative information provided by the adjustment remains the same for all possible sequence lengths and all possible sample sizes and the quantitative information is easy to compare across this, via equation (21). Thus, not only is it straightforward to compare the effect of changing the sample size for learning about $[\mathcal{M}^{[N]}(\mathcal{C})]$, but we now see that it is straightforward to compare the differences between learning about the corresponding quantities in differing $[\mathcal{M}^{[N]}(\mathcal{C})]$s. For example, if N_1 and N_2 are feasible sequence lengths we have that

$$\frac{N_1}{N_1 - 1}(1 - \lambda^{[N_1]}) = \frac{N_2}{N_2 - 1}(1 - \lambda^{[N_2]}). \tag{22}$$

Since the canonical resolutions share the same co-ordinate representation for all feasible choices of N it is straightforward to compare $R_{\mathcal{C}(n)}(\mathcal{M}^{[N_1]}(\mathcal{X}))$ with $R_{\mathcal{C}(n)}(\mathcal{M}^{[N_2]}(\mathcal{X}))$ for any $\mathcal{X} \in \langle\mathcal{C}\rangle$ and so gauge the effect of the population size on our adjustments.

Note also the computational advantage. In order to compare all sample sizes for all sequence lengths, we need only to consider the transform for a sample of size one and a sequence of length two.

It should be emphasised that the work in this section remains valid for the two types of second-order exchangeable sequences of length two. The first when we have only finite extendibility (for whatever reason) and the second when we have infinite extendibility. We now proceed by assuming we have infinite extendibility and showing the links between infinite second-order exchangeable sequences and finite second-order exchangeable sequences.

6 Linking finite and infinite adjustments

Suppose that we have a sequence of two second-order exchangeable collections $\mathcal{C}_1$ and $\mathcal{C}_2$ and that the sequence is M-extendible for all M, so that there is an infinite collection $\mathcal{C}_\infty^* = \cup_{i=1}^{\infty}\mathcal{C}_i$ over which $\mathcal{C}$ is second-order exchangeable. We have the following theorem.

Theorem 5 *If $Y^{[2]} = \sum_u \xi_u \mathcal{M}^{[2]}(X_{v_u})$, where $\{v_1, v_2, \ldots\}$ is a general finite subset of integers, is an eigenvector of $T_1^{[2]}$, with eigenvalue $\lambda^{[2]}$, then $Y = \sum_u \xi_u \mathcal{M}(X_{v_u})$ is an eigenvector of T_1, with eigenvalue*

$$\lambda = 2\lambda^{[2]} - 1. \tag{23}$$

By combining the results of Theorem 3 of Goldstein & Wooff (1998) and Theorem 5 we have the following corollary.

Corollary 3 *If* $Y^{[2]} = \sum_u \xi_u \mathcal{M}^{[2]}(X_{v_u})$ *is an eigenvector of* $T_1^{[2]}$*, with eigenvalue* $\lambda^{[2]}$*, then* $Y = \sum_u \xi_u \mathcal{M}(X_{v_u})$ *is, for each* n*, an eigenvector of* T_n *with eigenvalue*

$$\lambda_{(n)} = \frac{2n\lambda^{[2]} - n}{2(n-1)\lambda^{[2]} - (n-2)}. \tag{24}$$

Thus, the eigenvectors of T_n share the same co-ordinate representation to those of $T_1^{[2]}$ and from Corollary 2 these share the same co-ordinate representation as those of $T_n^{[N]}$. Hence, the qualitative information for the adjustment of $[\mathcal{M}(\mathcal{C})]$ by $\mathcal{C}(n)$ is the same as for the adjustment of $[\mathcal{M}^{[N]}(\mathcal{C})]$ by $\mathcal{C}(n)$. The difference between the infinite assumption and the finite reality is quantitative rather than qualitative. Notice that by comparing equation (24) with equation (21) we see that

$$\lambda_{(n)} = \lim_{N \to \infty} \lambda_{(n)}^{[N]}. \tag{25}$$

and by comparing the proofs of Theorem 4 and Theorem 5, we have for any $\mathcal{X} \in \langle \mathcal{C} \rangle$,

$$R_{\mathcal{C}(n)}(\mathcal{M}(\mathcal{X})) = \lim_{N \to \infty} R_{\mathcal{C}(n)}(\mathcal{M}^{[N]}(\mathcal{X})). \tag{26}$$

Thus, the adjustments of the finite and infinite sequences coalesce as $N \to \infty$ and the adjustment of $[\mathcal{M}(\mathcal{C})]$ by $\mathcal{C}(n)$ may be interpreted as the limit of the adjustment of $[\mathcal{M}^{[N]}(\mathcal{C})]$ by $\mathcal{C}(n)$. All of the coherency conditions derived by Goldstein & Wooff (1998) are driven by second-order exchangeability rather than the infinite assumption.

We have expressed all our canonical resolutions in terms of $\lambda^{[2]}$. This emphasises the symmetry in our second-order exchangeable beliefs that we really do only need to consider the beliefs between two individuals. Further, it allows us to include the theory for both infinitely extendible and finitely extendible sequences together. In many applications, for example population sampling, the convenient assumption is that N is infinite. As we have explained, in the context of the second-order adjustment of population vectors, this assumption leads to a quantitative difference and we would like to understand this difference. Using equations (21) and (24), we may write, for any infinitely extendible second-order exchangeable sequence of length N,

$$\lambda_{(n)}^{[N]} = \frac{\lambda_{(n)}}{\lambda_{(N)}}. \tag{27}$$

By expressing $\lambda_{(N)}$ in terms of $\lambda_{(n)}$, equation (27) may be rearranged as

$$\lambda_{(n)}^{[N]} = \lambda_{(n)} + \frac{n}{N}(1 - \lambda_{(n)}). \tag{28}$$

Thus, we could consider the quantitative difference between infinite and finite modelling to be that for the finite case we need a finite model correction term for the canonical resolutions; this correction term is $(n/N)(1 - \lambda_{(n)})$. Notice the multiplication by the sampling fraction, (n/N). If $Y_s^{[N]}$ is an eigenvector of $T_n^{[N]}$ with eigenvalue $\lambda_{(n)s}^{[N]}$ and Y_s, $\lambda_{(n)s}$ the corresponding eigenvector and eigenvalue for T_n, then from equation (16) and using (28) we have that

$$Var_{\mathcal{C}(n)}(Y_s^{[N]}) = (1 - (n/N))Var_{\mathcal{C}(n)}(Y_s). \tag{29}$$

The multiplier $(1 - (n/N))$ is the same for each eigenvector of $T_n^{[N]}$ and may be recognised as the finite population correction. In the same way of sampling n iid quantities from a population of size N, the link between the finite and infinite populations when we perform a second-order belief adjustment for a vector of population means is by multiplication of the adjusted variances for each of the canonical directions by the finite population correction to obtain the adjusted variance for each of the canonical directions in the finite setting. As the finite population correction attaches itself to each eigenvector of T_n, the sampling fraction, (n/N), provides a simple 'rule of thumb' for assessing the validity of the infinite approximation to the finite judgement; the smaller the value of (n/N), the greater the validity of the approximation.

If we are interested in a given direction in $[\mathcal{M}^{[N]}(\mathcal{C})]$, which does not correspond to a canonical resolution of $T_n^{[N]}$ then we may directly compare the adjustment of $\mathcal{M}^{[N]}(\mathcal{X})$ and $\mathcal{M}(\mathcal{X})$ by $\mathcal{C}(n)$ for any $\mathcal{X} \in \langle\mathcal{C}\rangle$. Suppose that $Z_s = Y_s = \sum_u \xi_u \mathcal{M}(X_{i_u})$ is a canonical direction of T_n, with canonical resolution $\lambda_{(n)s}$. Hence, the ξ_us have been chosen so that $Var(Y_s) = 1$. Then, by Theorem 5 and Theorem 4, $Z_s^{[N]} = a_s Y_s^{[N]}$ is a canonical direction of $T_n^{[N]}$ with canonical resolution $\lambda_{(n)s}^{[N]}$. a_s is chosen to ensure that $Z_s^{[N]}$ has prior variance 1, so that $a_s^{-2} = Var(Y_s^{[N]})$. Now $Z_s \in [\mathcal{M}(\mathcal{C})]$ so $Z_s = \mathcal{M}(\mathcal{Z}_s)$ for some $\mathcal{Z}_s \in \langle\mathcal{C}\rangle$. From the second-order specifications and the corresponding representation theorems (see Goldstein (1986a)) we may write, for any $\mathcal{X} \in \langle\mathcal{C}\rangle$,

$$Cov(\mathcal{M}^{[N]}(\mathcal{Z}_s), \mathcal{M}^{[N]}(\mathcal{X})) = Cov(\mathcal{M}(\mathcal{Z}_s), \mathcal{M}(\mathcal{X})) + \frac{1}{N} Cov(\mathcal{R}_i(\mathcal{Z}_s), \mathcal{R}_i(\mathcal{X})). \qquad (30)$$

Hence, $a_s = \{1 + (1/N)Var(\mathcal{R}_i(\mathcal{Z}_s))\}^{-\frac{1}{2}}$, and using equation (16) we may write, for any $\mathcal{X} \in \langle\mathcal{C}\rangle$,

$$Var_{\mathcal{C}(n)}(\mathcal{M}^{[N]}(\mathcal{X})) = \left(1 - \frac{n}{N}\right) \sum_s (1 - \lambda_{(n)s}) \frac{\{Cov(\mathcal{M}(\mathcal{Z}_s), \mathcal{M}(\mathcal{X})) + (1/N)Cov(\mathcal{R}_i(\mathcal{Z}_s), \mathcal{R}_i(\mathcal{X}))\}^2}{\{1 + (1/N)Var(\mathcal{R}_i(\mathcal{Z}_s))\}}. \qquad (31)$$

Equation (31) thus expresses the adjusted variance for any $\mathcal{M}^{[N]}(\mathcal{X}) \in [\mathcal{M}^{[N]}(\mathcal{C})]$, having observed a sample of size n, for any (potentially) infinitely extendible second-order exchangeable sequence of length N in terms of relationships and adjustments of quantities purely in the infinite sequence. By noting that,

$$Var_{\mathcal{C}(n)}(\mathcal{M}(\mathcal{X})) = \sum_s (1 - \lambda_{(n)s}) \{Cov(\mathcal{M}(\mathcal{Z}_s), \mathcal{M}(\mathcal{X}))\}^2, \qquad (32)$$

it is straightforward to compare the effect of the infinite approximation for any sample size n and any sequence length N. Notice the rather simple dependence upon $(1/N)$. By letting $N \to \infty$, it is easy to confirm equation (26). Thus, not only do we know that qualitatively $T_n^{[N]}$ and T_n provide the same information, but also that the quantitative differences are straightforward to calculate via equations (31) and (32), providing an easy way to assess the differences between the more realistic modelling framework of finite second-order exchangeability, where the difficulty may lie in determining N, and the convenient use of infinite second-order exchangeability.

7 A note on the prediction of future individuals

Suppose that we wish to consider the effect of observing n individuals for predicting the values for a further r individuals who are second-order exchangeable with those in the sample. This adjustment is driven completely by the relationships between the individuals and as such we may consider the collection of individuals, $\mathcal{C}^*_{n+r}$, to form a finite second-order exchangeable sequence of length $n+r$. Whether of not $\mathcal{C}^*_{n+r}$ is extendible or not is irrelevant for this prediction. Goldstein & Wooff (1998; Section 8) consider prediction in the case when $\mathcal{C}^*_{n+r}$ is M-extendible for all M, so that the total population is potentially infinite. They show that the canonical directions of the predictive adjustment share, up to a scale factor, the same co-ordinate representation as for the adjustment of the underlying mean components. All that is relevant for the prediction is the finite sequence $\mathcal{C}^*_{n+r}$ and so these results cross over in the case of the total population being finite rather than infinite. Indeed, we may view the qualitative similarity between prediction and learning about the underlying mean components in the infinite sequence to be a consequence of the qualitative similarity between finite and infinite learning as discussed in Section 8.

Acknowledgements

The first named author was supported by a grant from the UK's EPSRC.

References

Barnett, V. (1974). *Elements of sampling theory.* London: The English Universities Press.

de Finetti, B. (1974). *Theory of Probability*, Volume 1. London: Wiley.

Diaconis, P. (1977). Finite forms of de Finetti's theorem on exchangeability. *Synthese 36*, 271–281.

Goldstein, M. (1981). Revising previsions: a geometric interpretation (with discussion). *J. R. Statist. Soc. B 43*, 105–130.

Goldstein, M. (1986a). Exchangeable belief structures. *J. Amer. Statist. Ass. 81*, 971–976.

Goldstein, M. (1986b). Separating beliefs. In P. K. Goel and A. Zellner (Eds.), *Bayesian Inference and Decision Techniques*, pp. 197–215. Amsterdam: North Holland.

Goldstein, M. (1990). Influence and Belief Adjustment. In R. M. Oliver and J. Q. Smith (Eds.), *Influence Diagrams, Belief Nets and Decision Analysis*, pp. 143–174. Chichester: Wiley.

Goldstein, M. (1994). Revising exchangeable beliefs: subjectivist foundations for the inductive argument. In P. Freeman and A. F. M. Smith (Eds.), *Aspects of Uncertainty: A Tribute to D. V. Lindley*, pp. 201–222. Chichester: Wiley.

Goldstein, M. (1997). Prior inferences for posterior judgements. In M. L. D. Chiara, K. Doets, D. Mundici, and J. van Benthem (Eds.), *Structures and Norms in Science, Tenth International Congress of Logic, Methodology and Philosophy of Science, Florence, August 1995*, pp. 55–71. Dordrecht: Kluwer Academic Publishers.

Goldstein, M. (1999). Bayes linear analysis. In S. Kotz, C. B. Read, and D. L. Banks (Eds.), *Encyclopedia of Statistical Sciences Update Volume 3*, pp. 29–34. Chichester: Wiley.

Goldstein, M. and A. O'Hagan (1996). Bayes Linear Sufficiency and Systems of Expert Posterior Assessments. *J. R. Statist. Soc. B 58*, 301–316.

Goldstein, M. and D. A. Wooff (1997). Choosing sample sizes in balanced experimental designs: a bayes linear approach. *The Statistician 46*, 1–17.

Goldstein, M. and D. A. Wooff (1998). Adjusting exchangeable beliefs. *Biometrika 85*, 39–54.

Hartigan, J. A. (1969). Linear Bayesian methods. *J. R. Statist. Soc. B 31*, 446–454.

Lad, F. (1996). *Operational Subjective Statistical Methods.* New York: Wiley.

Appendix: Proof of theorems

Proof of Theorem 2 - The collection of individuals $\mathcal{C}(n)$ form a finite second-order exchangeable sequence of length n and applying Theorem 1 allows us to express each $X_{vi} \in \mathcal{C}(n)$ as $X_{vi} = \mathcal{S}_n(X_v) + \mathcal{T}_i^{[n]}(X_v)$. The collection of residuals of the ith sampled individual from the collection of sample means is $\mathcal{T}_i^{[n]}(\mathcal{C}) = \{\mathcal{T}_i^{[n]}(X_1),\ \mathcal{T}_i^{[n]}(X_2),\ \ldots\}$. For each $i = 1, \ldots, n$, the representation theorem gives $[\mathcal{S}_n(\mathcal{C})] \perp [\mathcal{T}_i^{[n]}(\mathcal{C})]$ and we may easily check that $[\mathcal{M}^{[N]}(\mathcal{C})] \perp [\mathcal{T}_i^{[n]}(\mathcal{C})]$. Noting that $[\mathcal{C}_i] \subset [\mathcal{S}_n(\mathcal{C})] \cup [\mathcal{T}_i^{[n]}(\mathcal{C})]$ we adjust $[\mathcal{M}^{[N]}(\mathcal{C})]$ by $\mathcal{C}(n)$ in two stages, first by $\mathcal{T}^{[n]}(\mathcal{C}) = \cup_{i=1}^{n} \mathcal{T}_i^{[n]}(\mathcal{C})$ and then by $\mathcal{S}_n(\mathcal{C})$. The stated orthogonalities (see Goldstein (1990)) yield the result. □

Proof of Theorem 3 - The proof proceeds in a similar way to the proof of Theorem 3 of Goldstein & Wooff (1998; p53). $Y^{[N]}$ is an eigenvector of $T_1^{[N]}$ with eigenvalue $\lambda^{[N]}$ if and only if, for all $\mathcal{X} \in [\mathcal{C}]$,

$$(1-\lambda^{[N]})Cov(Y^{[N]}, \mathcal{M}^{[N]}(\mathcal{X})) \quad = \quad Cov(\mathcal{R}_1^{[N]}(U_{Y^{[N]}}), \mathcal{R}_1^{[N]}(\mathcal{X})), \tag{33}$$

where $E_{\mathcal{C}(1)}(Y^{[N]}) = \mathcal{M}^{[N]}(U_{Y^{[N]}}) + \mathcal{R}^{[N]}(U_{Y^{[N]}})$ for some $U_{Y^{[N]}} \in [\mathcal{C}]$. Similarly, $Y^{[N]}$ is an eigenvector of $T_n^{[N]}$, with eigenvalue μ, if and only if, for some $Z_{Y^{[N]}} \in [\mathcal{C}]$,

$$E_{\mathcal{C}(n)}(Y^{[N]}) \quad = \quad \mu Y^{[N]} + \frac{1}{n}\sum_{i=1}^{n} \mathcal{R}_i^{[N]}(Z_{Y^{[N]}}), \tag{34}$$

or equivalently, if and only if, for all $\mathcal{X} \in [\mathcal{C}]$

$$(1-\mu)Cov(Y^{[N]}, \mathcal{M}^{[N]}(\mathcal{X})) \quad = \quad Cov(\frac{1}{n}\sum_{i=1}^{n} \mathcal{R}_i^{[N]}(Z_{Y^{[N]}}, \frac{1}{n}\sum_{i=1}^{n} \mathcal{R}_i^{[N]}(\mathcal{X})). \tag{35}$$

Then as:

$$Cov(\frac{1}{n}\sum_{i=1}^{n} \mathcal{R}_i^{[N]}(U), \frac{1}{n}\sum_{i=1}^{n} \mathcal{R}_i^{[N]}(V)) \quad = \quad \frac{(N-n)}{n(N-1)} Cov(\mathcal{R}_1^{[N]}(U), \mathcal{R}_1^{[N]}(V)) \tag{36}$$

for all $U, V \in [\mathcal{C}]$, we have that $U_{Y^{[N]}}$ satisfies (33), with eigenvalue $\lambda^{[N]}$, if and only if

$$Z_{Y^{[N]}} = \frac{n(N-1)}{(n-1)N\lambda^{[N]} + (N-n)} U_{Y^{[N]}} \tag{37}$$

satisfies (35) with eigenvalue $\mu = \lambda_{(n)}^{[N]}$. □

Proof of Theorem 4 - From the proof to Theorem 3, we have that $Y^{[2]}$ is an eigenvector of $T_1^{[2]}$ if and only if, for all $\mathcal{X} \in [\mathcal{C}]$

$$(1-\lambda^{[2]})Cov(Y^{[2]}, \mathcal{M}^{[2]}(\mathcal{X})) \quad = \quad Cov(\mathcal{R}_1^{[2]}(U_{Y^{[2]}}), \mathcal{R}_1^{[2]}(\mathcal{X})), \tag{38}$$

where

$$E_1(Y^{[2]}) \;=\; \mathcal{M}^{[2]}(U_{Y^{[2]}}) + \mathcal{R}_1^{[2]}(U_{Y^{[2]}}) \;=\; \lambda^{[2]} Y^{[2]} + \mathcal{R}_1^{[2]}(U_{Y^{[2]}}). \tag{39}$$

Similarly, $Y^{[N]}$ is an eigenvector of $T_1^{[N]}$ if and only if, for all $\mathcal{X} \in [\mathcal{C}]$

$$(1-\lambda^{[N]}) Cov(Y^{[N]}, \mathcal{M}^{[N]}(\mathcal{X})) \;=\; Cov(\mathcal{R}_1^{[N]}(U_{Y^{[N]}}), \mathcal{R}_1^{[N]}(\mathcal{X})), \tag{40}$$

where

$$E_1(Y^{[N]}) \;=\; \mathcal{M}^{[N]}(U_{Y^{[N]}}) + \mathcal{R}_1^{[N]}(U_{Y^{[N]}}) \;=\; \lambda^{[N]} Y^{[N]} + \mathcal{R}_1^{[N]}(U_{Y^{[N]}}). \tag{41}$$

Now, from equation (9), we have

$$Cov(\mathcal{R}_i^{[N]}(X_v), \mathcal{R}_i^{[N]}(X_w)) \;=\; \frac{2(N-1)}{N} Cov(\mathcal{R}_i^{[2]}(X_v), \mathcal{R}_i^{[2]}(X_w)) \tag{42}$$

so that

$$Cov(\mathcal{R}_1^{[N]}(U_{Y^{[N]}}), \mathcal{R}_1^{[N]}(\mathcal{X})) \;=\; \frac{2(N-1)}{N} Cov(\mathcal{R}_1^{[2]}(U_{Y^{[2]}}), \mathcal{R}_1^{[2]}(\mathcal{X})). \tag{43}$$

Similarly, from equation (7), we have

$$\begin{aligned} &Cov(\mathcal{M}^{[N]}(X_v), \mathcal{M}^{[N]}(X_w) \;= \\ &\qquad Cov(\mathcal{M}^{[2]}(X_v), \mathcal{M}^{[2]}(X_w) + \frac{2-N}{N} Cov(\mathcal{R}_i^{[2]}(X_v), \mathcal{R}_i^{[2]}(X_w)). \end{aligned} \tag{44}$$

From equation (41) we have that

$$Cov(Y^{[N]}, \mathcal{M}^{[N]}(\mathcal{X})) \;=\; \frac{1}{\lambda^{[N]}} Cov(\mathcal{M}^{[N]}(U_{Y^{[N]}}), \mathcal{M}^{[N]}(\mathcal{X})), \tag{45}$$

so that, by equation (44), we have

$$Cov(Y^{[N]}, \mathcal{M}^{[N]}(\mathcal{X})) \;= Cov(Y^{[2]}, \mathcal{M}^{[2]}(\mathcal{X})) + \frac{2-N}{N\lambda^{[N]}} Cov(\mathcal{R}_1^{[2]}(U_{Y^{[2]}}), \mathcal{R}_1^{[2]}(\mathcal{X})). \tag{46}$$

By substituting equations (43) and (46) into equation (40), we have that $Y^{[N]}$ is an eigenvector of $T_1^{[N]}$ if and only if, for all $\mathcal{X} \in [\mathcal{C}]$

$$(1-\lambda^{[N]}) Cov(Y^{[2]}, \mathcal{M}^{[2]}(\mathcal{X})) \;=\; \frac{N\lambda^{[N]} - (2-N)}{N\lambda^{[N]}} Cov(\mathcal{R}_1^{[2]}(U_{Y^{[2]}}), \mathcal{R}_1^{[2]}(\mathcal{X})). \tag{47}$$

We thus have that $U_{Y^{[2]}}$ satisfies equation (38), with eigenvalue $\lambda^{[2]}$, if and only if $U_{Y^{[N]}} = \alpha U_{Y^{[2]}}$ satisfies equation (47), with eigenvalue $\lambda^{[N]}$ where

$$\frac{N\lambda^{[N]} - (2-N)}{N\lambda^{[N]}(1-\lambda^{[N]})} \alpha \;=\; \frac{1}{1-\lambda^{[2]}}. \tag{48}$$

Using equations (41) and (39) we have that

$$\alpha \;=\; \frac{\lambda^{[N]}}{\lambda^{[2]}}. \tag{49}$$

Substituting equation (49) into equation (48) and rearranging gives

$$\lambda^{[N]} = \frac{2(N-1)\lambda^{[2]} + (2-N)}{N}, \tag{50}$$

and hence the result. □

Proof of Theorem 5 - By noting that,

$$\begin{aligned} Cov(\mathcal{M}(X_v), \mathcal{M}(X_w)) &= Cov(\mathcal{M}^{[2]}(X_v), \mathcal{M}^{[2]}(X_w)) - \\ &\qquad Cov(\mathcal{R}_i^{[2]}(X_v), \mathcal{R}_i^{[2]}(X_w)); \qquad (51) \\ Cov(\mathcal{R}_i(X_v), \mathcal{R}_i(X_w)) &= 2Cov(\mathcal{R}_i^{[2]}(X_v), \mathcal{R}_i^{[2]}(X_w)), \qquad (52) \end{aligned}$$

the result follows in a similar way to Theorem 4. □