BAYESIAN STATISTICS 6, pp.000–000
J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith (Eds.)

Simplifying Complex Designs: Bayes Linear Experimental Design for Grouped Multivariate Exchangeable Systems

SIMON C. SHAW and MICHAEL GOLDSTEIN
University of Durham, UK

SUMMARY

We adopt a Bayes linear approach to tackle design problems with many variables cross-classified in many ways. We investigate designs where we wish to sample individuals belonging to different groups, exploiting the powerful properties of the adjustment of second-order exchangeable vectors. The types of information we gain by sampling are identified with the orthogonal canonical directions. We show how we may express these directions in terms of the different factors of the model. This allows us to solve a series of lower dimensional problems, through which we may identify the different aspects of our adjusted beliefs with the different aspects of the choice of design, leading both to qualitative insights and quantitative guidance for the optimal choice of design.

Keywords: BAYES LINEAR METHODS; SECOND-ORDER EXCHANGEABILITY; CANONICAL DIRECTIONS; CANONICAL RESOLUTIONS; OPTIMAL DESIGN.

1. INTRODUCTION

Experimental design with many variables cross-classified in many ways is a challenging problem for Bayes analysis. Typically, we want to make inferences about various combinations of the elements of the model, but it is unlikely that all combinations will be equally important. Different objectives will lead to different choices of optimum (for example in terms of variance reduction for the quantities of interest). Often, it is difficult for formal design criteria to capture all of our aims, so that we need to have heuristic insights into the types of information that each design conveys. Further, introducing fully specified prior beliefs over a complex model may be a difficult elicitation problem, and may also make the optimal choice of design highly complex computationally. Therefore, our aims are to i) simplify the elicitation requirements ii) tame the computational problems for the design choice iii) provide qualitative insights into the effectiveness of the different choices of design. We argue that a Bayes linear approach provides a platform for achieving these aims. We proceed as follows. We motivate our approach through a simple example, before briefly reviewing Bayes linear methods. We then formalise the ideas exhibited in the first example, showing how we may simplify many designs of interest by decomposing the design into smaller subspaces, over which the adjustment has the same qualitative features. We conclude with an example illustrating how we make the simplifications.

2. EXAMPLE 1 - LEARNING ABOUT MEAN COMPONENTS

Suppose that we have two collections of individuals, on each of whom we wish to make the same two measurements. Thus, we could think of a clinical trial where each patient receives one of two treatments. Denote the measurements for the ith individual in the gth treatment group by the vector $X_{gi} = [X_{g1i}, X_{g2i}]^T$. We consider that individuals receiving the same treatment are second-order exchangeable and co-exchangeable across treatments, defined as follows. Let $\mathcal{B}_i, i = 1, 2, \ldots$ be a (potentially) infinite sequence of random vectors, then the collection $\mathcal{B}_1, \mathcal{B}_2, \ldots$ is second-order exchangeable if the second-order specification is invariant under permutation, i.e. if

$$E(\mathcal{B}_i) = \mu \qquad Var(\mathcal{B}_i) = V \qquad Cov(\mathcal{B}_i, \mathcal{B}_j) = C \qquad \forall i \neq j \tag{1}$$

Goldstein (1986) derived the following representation theorem for second-order exchangeable random vectors. We may introduce the vector $\mathcal{M}(\mathcal{B})$, representing the underlying mean components, and the mutually uncorrelated vectors $\mathcal{R}_1(\mathcal{B}), \mathcal{R}_2(\mathcal{B}), \ldots$ representing the individual residuals from the mean, and write $\mathcal{B}_i = \mathcal{M}(\mathcal{B}) + \mathcal{R}_i(\mathcal{B})$ for each i. Two second-order exchangeable collections, $\mathcal{B}_1^*$ and $\mathcal{B}_2^*$ are termed co-exchangeable if the joint second-order specification is invariant under permutation, i.e. if each collection is second-order exchangeable and $Cov(\mathcal{B}_{1i}^*, \mathcal{B}_{2j}^*)$ does not depend on i or j. Suppose that $\mathcal{B}_1^*$, $\mathcal{B}_2^*$, $\ldots$ are a sequence of co-exchangeable exchangeable systems, where we write $\mathcal{B}_{ij}^* = \mathcal{M}(\mathcal{B}_i^*) + \mathcal{R}_j(\mathcal{B}_i^*)$. Goldstein (1986) showed that all of the residual vectors, $\mathcal{R}_j(\mathcal{B}_i^*)$, are mutually uncorrelated and are also uncorrelated with all of the underlying mean component vectors, $\mathcal{M}(\mathcal{B}_i^*)$. Second-order exchangeability therefore produces model prior specification through a small number of specifications over observable quantities, whilst also not placing too harsh a constraint upon our beliefs. Approaches via symmetric attitudes to the data may be found in Dawid (1988).

We use the representation theorem to write $X_{gi} = \mathcal{M}(X_g) + \mathcal{R}_i(X_g)$, where $Cov(\mathcal{M}(X_g), \mathcal{R}_j(X_h)) = 0$, $\forall g, h, j$ and $Cov(\mathcal{R}_i(X_g), \mathcal{R}_j(X_h)) = 0$, $\forall g, h, i, j$ (except when both $g = h$ and $i = j$). We then need to specify $Cov(\mathcal{M}(X_g), \mathcal{M}(X_h)) = C_{gh}$ and $Var(\mathcal{R}_i(X_g)) = E_g$. Suppose that we make the following specification for these covariance and variance matrices:

$$Cov(\mathcal{M}(X_g), \mathcal{M}(X_h)) = \gamma_{gh} \begin{pmatrix} 9 & 4 \\ 4 & 9 \end{pmatrix} \qquad Var(\mathcal{R}_i(X_g)) = \begin{pmatrix} 15 & 6 \\ 6 & 15 \end{pmatrix} \tag{2}$$

where $\gamma_{gh} = 1$ if $g = h$, and $\gamma_{gh} = \gamma$ otherwise, for $-1 \leq \gamma \leq 1$. From each group, we wish to take a sample of size n. Construct the vector of sample means $\overline{X}_{(n)} = [\overline{X}_{(n)1}^T, \overline{X}_{(n)2}^T]^T$ where $\overline{X}_{(n)g} = \frac{1}{n}\sum_{i=1}^n X_{gi}$, and use this to learn about linear combinations, $l_{11}\mathcal{M}(X_{11}) + l_{12}\mathcal{M}(X_{12}) + l_{21}\mathcal{M}(X_{21}) + l_{22}\mathcal{M}(X_{22})$, of the underlying mean quantities. Thus, we may be interested in the differences between treatments for a particular measurement, the overall treatment effects, the differences between the measurements, an overall average, and so on.

We shall use Bayes linear fitting (see section 3). For a general quantity $\mathcal{B}$, and observed data vector $\mathcal{D}$, we may obtain the adjusted variance $Var_{\mathcal{D}}(\mathcal{B})$. Equivalently, the adjusted variance is the posterior variance if the joint distribution of all of our quantities is multivariate normal. Simple, scale-free, measures of the effect of an adjustment upon $\mathcal{B}$ are given by the resolution, $R_{\mathcal{D}}(\mathcal{B})$, or equivalently, $RR_{\mathcal{D}}(\mathcal{B})$, the resolution ratio:

$$R_{\mathcal{D}}(\mathcal{B}) = 1 - \frac{Var_{\mathcal{D}}(\mathcal{B})}{Var(\mathcal{B})} \qquad RR_{\mathcal{D}}(\mathcal{B}) = \frac{Var(\mathcal{B}) - Var_{\mathcal{D}}(\mathcal{B})}{Var_{\mathcal{D}}(\mathcal{B})} \tag{3}$$

Thus, if $R_{\mathcal{D}}(\mathcal{B})$ is near zero, then, relative to our prior knowledge about $\mathcal{B}$, we do not expect the sample to be informative for $\mathcal{B}$. A value of $R_{\mathcal{D}}(\mathcal{B})$ close to one suggests that the sample

is expected to be highly informative for $\mathcal{B}$. Likewise, if $RR_{\mathcal{D}}(\mathcal{B})$ is very large, then we expect the sample to be highly informative. Typically, we may use such a summary to aid the choice of optimal sample sizes for design problems where variance reduction of certain quantities of interest is viewed as a benefit, balanced against the cost of the sample.

We shall focus attention upon the sum, $S_g = \mathcal{M}(X_{g1}) + \mathcal{M}(X_{g2})$, and the difference, $D_g = \mathcal{M}(X_{g1}) - \mathcal{M}(X_{g2})$, of the variables from each group. Form the collections $S = \{S_1, S_2\}$ and $D = \{D_1, D_2\}$. Let $\langle S \rangle$, $\langle D \rangle$ be the collections of linear combinations of the elements of S, D respectively. It can be checked that every element in S is uncorrelated with every element in D. We consider first the effect of the adjustment of elements in $\langle S \rangle$ by $\overline{X}_{(n)}$, for a sample of size n. Table 1 summarises the adjustment.

Table 1. *Resolutions for the components of $\langle S \rangle$, adjusting by $\overline{X}_{(n)}$.*

quantity	resolution (sample size n)
$S_+ = S_1 + S_2$	$\lambda_{+(n)} = \frac{13n(1+\gamma)}{13n(1+\gamma)+21}$
$S_- = S_1 - S_2$	$\lambda_{-(n)} = \frac{13n(1-\gamma)}{13n(1-\gamma)+21}$
$l_1S_1 + l_2S_2$	$\frac{(1+\gamma)(l_1+l_2)^2}{2(l_1^2+2\gamma l_1 l_2+l_2^2)}\lambda_{+(n)} + \frac{(1-\gamma)(l_1-l_2)^2}{2(l_1^2+2\gamma l_1 l_2+l_2^2)}\lambda_{-(n)}$

The two combinations S_+ and S_- are uncorrelated and the resolution for each element, $l_1S_1 + l_2S_2$, of $\langle S \rangle$ may be found as a weighted average of the resolutions for these two quantities. For all n, the largest resolution is $\lambda_{+(n)}$ corresponding to quantities proportional to S_+, and the smallest resolution is $\lambda_{-(n)}$ corresponding to quantities proportional to S_-. Hence, we expect to learn most about quantities in $\langle S \rangle$ that are highly correlated with S_+, and least for those quantities highly correlated with S_-. Notice that since this feature does not depend upon the sample size, the underlying features of the adjustment within $\langle S \rangle$ remain the same for all sample sizes. Observe the role played by the parameter γ. As $\gamma \to 1$, we learn progressively less about the differences between the means of the variables across treatments. Notice that $\gamma = 1$ corresponds to a correlation of one between the treatments. Similarly, as $\gamma \to 0$, then $\lambda_{-(n)} \to \lambda_{+(n)}$, and we learn equally in the two directions. A value of $\gamma = 0$ corresponds to the two treatments being uncorrelated.

We now consider the effect of the adjustment for the elements of $\langle D \rangle$. Table 2 summarises the adjustment.

Table 2. *Resolutions for the components of $\langle D \rangle$, adjusting by $\overline{X}_{(n)}$.*

quantity	resolution (sample size n)
$D_+ = D_1 + D_2$	$\mu_{+(n)} = \frac{5n(1+\gamma)}{5n(1+\gamma)+9}$
$D_- = D_1 - D_2$	$\mu_{-(n)} = \frac{5n(1-\gamma)}{5n(1-\gamma)+9}$
$l_1D_1 + l_2D_2$	$\frac{(1+\gamma)(l_1+l_2)^2}{2(l_1^2+2\gamma l_1 l_2+l_2^2)}\mu_{+(n)} + \frac{(1-\gamma)(l_1-l_2)^2}{2(l_1^2+2\gamma l_1 l_2+l_2^2)}\mu_{-(n)}$

The key point to observe here is that the adjustment over $\langle D \rangle$ is essentially the same as that over $\langle S \rangle$, with λ replaced by μ, so that each of the comments that we made when discussing $\langle S \rangle$ remain valid for elements in $\langle D \rangle$.

We remark further that the ratio $RR_{\overline{X}_{(n)}}(S_+)/RR_{\overline{X}_{(n)}}(S_-) = \{1+\gamma\}/\{1-\gamma\}$ remains the same for all choices of matrices in (2). Additionally, note that $RR_{\overline{X}_{(n)}}(D_+)/RR_{\overline{X}_{(n)}}(D_-) = RR_{\overline{X}_{(n)}}(S_+)/RR_{\overline{X}_{(n)}}(S_-)$ so that this ratio remains constant across the two collections. Changing the matrices in (2) only changes S_+, S_-, D_+, D_- to four other quantities sharing the same features as those described above, with the ratio of resolution ratios for these quantities still given by $\{1+\gamma\}/\{1-\gamma\}$.

If we consider a general element $Y = l_1\mathcal{M}(X_{11}) + l_2\mathcal{M}(X_{12}) + l_3\mathcal{M}(X_{21}) + l_4\mathcal{M}(X_{22})$ and wish to examine the effect of the adjustment upon it, then:

$$R_{\overline{X}_{(n)}}(Y) = \frac{Cov(Y,S_+)^2\lambda_{+(n)} + Cov(Y,S_-)^2\lambda_{-(n)} + Cov(Y,D_+)^2\mu_{+(n)} + Cov(Y,D_-)^2\mu_{-(n)}}{Cov(Y,S_+)^2 + Cov(Y,S_-)^2 + Cov(Y,D_+)^2 + Cov(Y,D_-)^2}$$

so that each resolution may be expressed as a weighted average of the resolutions of the four quantities we discussed previously.

Hence, we decompose the structure of the design to separate the problem into uncorrelated spaces that share the same properties. Within each space we find uncorrelated directions that summarise the types of information we expect to learn, both for quantities within that space, and also for a general quantity when the directions from each space are collected together. We see how such decompositions ease the task of picking effective designs. For example, if our primary goal is to choose sample sizes to achieve specific variance reductions, then examination of these spaces shows how we may assess the effects of different sample sizes. These properties apply to a wide class of complex design problems as we now explain.

3. BAYES LINEAR METHODS AND CANONICAL STRUCTURE

In situations where meaningful full prior specification is difficult, the Bayes linear approach may be used as an alternative approach, based on partial prior specification. The elicitation is simplified through the need for only a second-order specification. An overview of the methodology may be found in Farrow & Goldstein (1993), whilst Goldstein (1994, 1997) concentrates upon foundational aspects. In a typical analysis, we observe a data collection $\mathcal{D} = [\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_s]^T$, and wish to evaluate the effect of this vector upon the expectations and variances of a vector $\mathcal{B} = [\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_r]^T$ of interest. For any $\mathcal{B}$, the adjusted expectation, $E_{\mathcal{D}}(\mathcal{B})$, and adjusted variance, $Var_{\mathcal{D}}(\mathcal{B})$, are calculated as:

$$E_{\mathcal{D}}(\mathcal{B}) = E(\mathcal{B}) + Cov(\mathcal{B},\mathcal{D})Var(\mathcal{D})^{\dagger}(\mathcal{D} - E(\mathcal{D})) \quad (4)$$

$$Var_{\mathcal{D}}(\mathcal{B}) = Var(\mathcal{B}) - Cov(\mathcal{B},\mathcal{D})Var(\mathcal{D})^{\dagger}Cov(\mathcal{D},\mathcal{B}) \quad (5)$$

where $Var(\mathcal{D})^{\dagger}$ is the Moore-Penrose generalised inverse of $Var(\mathcal{D})$. Note that whilst no distributional assumptions are made, the adjustment of expectations and variances corresponds to the respective conditionals if the joint probability distribution of all the quantities is multivariate normal. For any finite set, $\tilde{\mathcal{B}}$, of quantities, we let $\langle \tilde{\mathcal{B}} \rangle$ represent the space of all linear combinations of the elements of $\tilde{\mathcal{B}}$. We say that two spaces, $\langle \tilde{\mathcal{B}} \rangle$, $\langle \hat{\mathcal{B}} \rangle$ are orthogonal, written $\langle \tilde{\mathcal{B}} \rangle \perp \langle \hat{\mathcal{B}} \rangle$, if every element of $\tilde{\mathcal{B}}$ is uncorrelated with every element of $\hat{\mathcal{B}}$.

Goldstein (1981) considered the effect of the adjustment of $\langle \mathcal{B} \rangle$, and showed that for any $\mathcal{A} \in \langle \mathcal{B} \rangle$, there was a set of canonical directions $Z = (Z_1, \ldots, Z_r) \subseteq \langle \mathcal{B} \rangle$ with corresponding

canonical resolutions $1 \geq \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_r \geq 0$ which satisfied the following properties. The Z_j have expectation 0, are mutually uncorrelated, and are scaled to have prior variance 1. Each $\mathcal{A}$ may be expressed as $\mathcal{A} = \sum_{j=1}^{r} Cov(\mathcal{A}, Z_j) Z_j$, and the resolution expressed as

$$R_{\mathcal{D}}(\mathcal{A}) = 1 - \frac{\sum_{j=1}^{r}(1-\lambda_j) Cov(\mathcal{A}, Z_j)^2}{Var(\mathcal{A})} \tag{6}$$

Hence, constrained by being uncorrelated with $(Z_1, \ldots, Z_j)$, Z_{j+1} is the element of $\langle \mathcal{B} \rangle$ maximising the resolution. Thus, we expect to learn most about elements of $\langle \mathcal{B} \rangle$ having strong correlations with the directions with large resolutions.

As the canonical directions pinpoint the types of information we may receive by sampling, the strength of this information being quantified by the canonical resolutions, we can gain insights into the benefits of different designs by comparing the canonical structures and determining which is most informative for the most important directions of interest. Goldstein & Wooff (1997) adopt such an approach for choosing sample sizes in balanced designs.

Goldstein & Wooff (1998) consider the canonical structure for the underlying mean components brought about by a second-order exchangeable sample of size n. The canonical directions are the same for all sample sizes, and if λ is a canonical resolution for a sample of size 1 corresponding to direction W, then $\lambda_{(n)} = n\lambda/\{n\lambda + (1-\lambda)\}$ is a canonical resolution for a sample of size n corresponding to the same direction. As the canonical directions do not depend upon the sample size, the underlying qualitative features of the adjustment are not affected by sample size. Therefore, it is straightforward to use the values $\lambda_{(n)}$ to simplify design questions for which the sample size has to be determined to achieve specific variance reductions of combinations of the underlying mean components. In the next section, we shall generalise this property for a wide class of complex designs. Goldstein & Wooff (1998) also showed that in a predictive analysis for future observables, equivalent results apply. There are also similar implications for predictive analysis in this paper.

4. GROUPED MULTIVARIATE EXCHANGEABLE SYSTEMS

4.1. *Specification of the system*

We consider designs where the data can be classified as coming from one of g_0 groups. For each individual, we wish to measure the same set of v_0 variables. Let $X_{gi} = [X_{g1i} \ldots X_{gv_0i}]^T$ be the vector of measurements for the ith individual in the gth group. We consider that individuals within the same group are second-order exchangeable, and that they are co-exchangeable across groups. Thus, following Goldstein (1986), we may write each X_{gi} as the sum of an underlying mean component vector, $\mathcal{M}(X_g)$, and an uncorrelated residual vector specific to the individual, $\mathcal{R}_i(X_g)$, that is $X_{gi} = \mathcal{M}(X_g) + \mathcal{R}_i(X_g)$. We make the following specifications for the $\mathcal{M}(X_g)$'s and the $\mathcal{R}_i(X_g)$'s. From these we may deduce the variance and covariance specifications for the X_{gi}'s.

$$Cov(\mathcal{M}(X_g), \mathcal{M}(X_h)) = \alpha_{gh} C \;\forall g, h \tag{7}$$

$$Cov(\mathcal{M}(X_g), \mathcal{R}_j(X_h)) = 0 \;\forall g, h, j \tag{8}$$

$$Cov(\mathcal{R}_i(X_g), \mathcal{R}_j(X_h)) = \begin{cases} \beta_g E & \forall g = h, i = j \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

where C, E are general $v_0 \times v_0$ positive semi-definite matrices with (v,w)th entries $(C)_{vw} = c_{vw}$, $(E)_{vw} = e_{vw}$ respectively. Let A be the $g_0 \times g_0$ matrix with (g,h)th entry $(A)_{gh} = \alpha_{gh}$, and let B be the $g_0 \times g_0$ diagonal matrix with (g,g)th entry $(B)_{gg} = \beta_g$. By letting

$l = [l_1 \dots l_{g_0}]^T$ and considering $Var(\sum_{g=1}^{g_0} l_g \mathcal{M}(X_{gv})) = c_{vv} l^T A l$ for $c_{vv} \neq 0$ we see that A must be positive semi-definite. For simplicity of exposition, in this paper we shall assume that the matrices of interest are of full rank. Specifically, C, E, A are positive definite and for each g, $\beta_g > 0$. As in Goldstein & Wooff (1998), if we do not have invertibility, we obtain corresponding results over the linear span of the columns of the matrices that we construct below.

Second-order co-exchangeability, as introduced by Goldstein (1986), requires that we specify $Cov(\mathcal{M}(X_g), \mathcal{M}(X_h)) = C_{gh}$ and $Var(\mathcal{R}_i(X_g)) = E_g$. All other components are uncorrelated. Notice that the results that we shall develop are concerned with designs where the covariance matrices for the underlying mean components are proportional, and the residual variance matrices are also proportional. This reflects the judgement that to the level of specification we are willing to make, beliefs over the residual components across groups are exchangeable up to a scale factor, and there is a generalised form of scaled exchangeability for the relationships between the mean components.

4.2. *Adjusting the mean components*

We wish to take a sample of size $n_g > 0$ from the gth group. Collect the sample sizes together into the matrix $N = diag(n_1, \dots, n_{g_0})$. As demonstrated in Goldstein & Wooff (1998), the resulting sample mean vectors $\overline{X}_{(ng)g} = \frac{1}{n_g} \sum_{i=1}^{n_g} X_{gi}$ are Bayes linear sufficient for adjusting the underlying mean components. From the vector of mean components specific to the gth group, $\mathcal{M}(X_g) = [\mathcal{M}(X_{g1}) \dots \mathcal{M}(X_{gv_0})]^T$, we form the column vector of all the mean components, $\mathcal{M}(X) = [\mathcal{M}(X_1)^T \dots \mathcal{M}(X_{g_0})^T]^T$. We let $\langle \mathcal{M}(X) \rangle$ denote the collection of all linear combinations of the elements of $\mathcal{M}(X)$. We now consider separately the analysis of variables and of groups, as follows.

Definition 1. *The underlying canonical variable directions are defined as the columns of the matrix $Y = [Y_1 \dots Y_{v_0}]$ solving the generalised eigenvalue problem $CY = (C + E)Y\Phi$, where $\Phi = diag(\phi_1, \dots, \phi_{v_0})$ is the matrix of eigenvalues. Y is chosen so that $Y^T C Y = I$, $Y^T (C + E) Y \Phi = I$. The ordered eigenvalues $1 > \phi_1 \geq \dots \geq \phi_{v_0} > 0$ are termed the underlying canonical variable resolutions.*

(We are able to choose Y in the stated form through standard results on simultaneous diagonalisation of matrices, see for example Theorem $VI.1.15$ of Stewart & Sun (1990)). To motivate this definition, consider adjusting the mean components of a single group, g, from a sample, of size n_g, drawn purely from that group. In this case, the canonical directions are given by $(1/\sqrt{\alpha_{gg}}) Y_s^T \mathcal{M}(X_g)$, for $s = 1, \dots, v_0$, with corresponding canonical resolutions given by $n_g \alpha_{gg} \phi_s / \{n_g \alpha_{gg} \phi_s + (1 - \phi_s)\beta_g\}$.

Definition 2. *The underlying canonical group directions are defined as the columns of the matrix $W = [W_1 \dots W_{g_0}]$ solving the generalised eigenvalue problem $AW = (A + N^{-1}B)W\Psi$, where $\Psi = diag(\psi_1, \dots, \psi_{g_0})$ is the matrix of eigenvalues. W is chosen so that $W^T A W = I$, $W^T (A + N^{-1}B) W \Psi = I$. The ordered eigenvalues $1 > \psi_1 \geq \dots \geq \psi_{g_0} > 0$ are termed the underlying canonical group resolutions.*

To motivate this definition, consider adjusting the mean components relating to the vth variable from a sample of size n_g from the gth group for $g = 1, \dots g_0$ which measures only the vth variable for each individual. The canonical directions for this adjustment are given by $(1/\sqrt{c_{vv}}) W_d^T \mathcal{M}(X_v)$, for $d = 1, \dots, v_0$, where $\mathcal{M}(X_v) = [\mathcal{M}(X_{1v}) \dots \mathcal{M}(X_{g_0 v})]^T$ is the

vector of mean components specific to the vth variable. The corresponding canonical resolutions are given by $c_{vv}\psi_d/\{c_{vv}\psi_d + (1-\psi_d)e_{vv}\}$.

We now show that the canonical variable and group analysis completely determine the adjustment of the full collection $\mathcal{M}(X)$ by the vector of sample means $\overline{X}_{(N)}$, where $\overline{X}_{(N)} = [\overline{X}^T_{(n_1)1} \ldots \overline{X}^T_{(n_{g_0})g_0}]^T$. We have the following theorem; the proof is in the appendix.

Theorem 1. *The adjustment of $\mathcal{M}(X)$ by $\overline{X}_{(N)}$ satisfies the following properties*

1. There exist v_0 g_0-dimensional uncorrelated subspaces, $\langle \mathcal{Z}_1 \rangle, \ldots, \langle \mathcal{Z}_{v_0} \rangle$ of $\langle \mathcal{M}(X) \rangle$. For each s, $\mathcal{Z}_s = \{Y_s^T \mathcal{M}(X_1), \ldots, Y_s^T \mathcal{M}(X_{g_0})\}$.

2. The canonical directions for the adjustment in each $\langle \mathcal{Z}_s \rangle$ share the same co-ordinate representation, and are given by the columns of the matrix $Z_s = [Z_{1s} \ldots Z_{g_0 s}]$, where $Z_{ds} = W_d^T \mathcal{Z}_s$.

3. The collection $Z = \{Z_{ds}\}$ for $d = 1, \ldots, g_0$, $s = 1, \ldots, v_0$, are the canonical directions of the complete adjustment with canonical resolutions given by

$$\lambda_{ds} = \frac{\psi_d \phi_s}{\psi_d \phi_s + (1-\psi_d)(1-\phi_s)}$$

4. The resolution ratio for the canonical directions is given by:

$$RR_{\overline{X}_{(N)}}(Z_{ds}) = \frac{\lambda_{ds}}{1-\lambda_{ds}} = \frac{\psi_d}{1-\psi_d} \times \frac{\phi_s}{1-\phi_s}$$

This theorem tells us many things about our adjustment. Most importantly, it illustrates how we may simplify the design. We break down the $v_0 g_0 \times v_0 g_0$ problem into one problem of size $v_0 \times v_0$ and another of size $g_0 \times g_0$. Not only is this a great advantage computationally, but the two problems also have interpretable forms. The $v_0 \times v_0$ problem consists of finding the underlying canonical variable structure; the $g_0 \times g_0$ of finding the underlying canonical group structure. Notice that beyond being positive definite, no other symmetry requirements are placed on either A or B, whilst there is also no requirement of balance.

Observe that if $\phi_s > \phi_{s'}$, then $\lambda_{ds} > \lambda_{ds'}$. Likewise, if $\psi_d > \psi_{d'}$, then $\lambda_{ds} > \lambda_{d's}$. Changing the sample size only effects W and Ψ, so that the impact of changing the sample size may be easily seen. Thus, we can see how we may choose the sample sizes to optimise many different design criteria in terms of quantities of interest in the $\mathcal{M}(X)$; for example we may choose N's to learn about the most important group contrasts. Notice also that we can easily assess the sensitivity of the design to the proportionality parameters in a similar way by merely looking at the resulting impact on W and Ψ.

The elegant features of the adjustment of exchangeable vectors shown in Goldstein & Wooff (1998) are also found in the balanced design:

Corollary 1. *If $N = nI$, then the canonical directions are the same for all n, and if $\lambda_{ds(1)} = \psi_{d(1)}\phi_s/\{\psi_{d(1)}\phi_s + (1-\psi_{d(1)})(1-\phi_s)\}$ are the canonical resolutions for a sample of size $n = 1$, so $\Psi_{(1)}$ solves $AW = (A+B)W\Psi_{(1)}$, then the canonical resolutions and resolution ratio for general n are given by:*

$$\lambda_{ds(n)} = \frac{n\lambda_{ds(1)}}{(n-1)\lambda_{ds(1)} + 1} \qquad RR_{\overline{X}_{(n)}}(Z_{ds}) = n \times \frac{\psi_{d(1)}}{1-\psi_{d(1)}} \times \frac{\phi_s}{1-\phi_s} \qquad (10)$$

Thus, in a similar vein, we may use (10) to simplify design problems for choosing sample sizes to achieve specific variance reductions. For example, we have the following corollary:

Corollary 2. *The sample size n in each group required to achieve a proportionate variance reduction of κ for Z_{ds}, is $n \geq \{\kappa/(1-\kappa)\}\{(1-\lambda_{ds(1)})/\lambda_{ds(1)}\}$. If the minimal canonical resolution for $n = 1$ is $\lambda_{min} = \psi_{min}\phi_{min}/\{\psi_{min}\phi_{min} + (1-\psi_{min})(1-\phi_{min})\}$, then a sample size of $\{\kappa/(1-\kappa)\}\{(1-\lambda_{min})/\lambda_{min}\} = \{\kappa/(1-\kappa)\}\{(1-\psi_{min})/\psi_{min}\}\{(1-\phi_{min})/\phi_{min}\}$, rounded up, in each group is the minimum sample which is sufficient to achieve a proportionate variance reduction of κ for every element of $\langle \mathcal{M}(X) \rangle$.*

5. EXAMPLE 2 - ILLUSTRATING THE THEOREM

We illustrate the theorem computationally, with an example similar in spirit to Example 1. We wish to take the same two measurements from each individual, in a trial where each individual receives one of four treatments. We consider that individuals receiving the same treatment are exchangeable, and that they are co-exchangeable across treatment groups. In line with the development of the previous section, we consider situations where the prior covariance matrices for the underlying mean components are proportional, and the residual covariance matrices are also proportional. Thus, in the notation of that section we have $v_0 = 2$, $g_0 = 4$ and the underlying proportional positive definite matrices we specify as

$$C = \begin{pmatrix} 9 & 4 \\ 4 & 9 \end{pmatrix} \qquad E = \begin{pmatrix} 15 & 6 \\ 6 & 15 \end{pmatrix} \tag{11}$$

We then specify the positive definite matrices of proportionality constants as:

$$A = \begin{pmatrix} 0.98 & 0.79 & 0.70 & 0.74 \\ 0.79 & 0.87 & 0.69 & 0.72 \\ 0.70 & 0.69 & 0.71 & 0.67 \\ 0.74 & 0.72 & 0.67 & 0.77 \end{pmatrix} \qquad B = \begin{pmatrix} 0.83 & 0 & 0 & 0 \\ 0 & 0.61 & 0 & 0 \\ 0 & 0 & 0.74 & 0 \\ 0 & 0 & 0 & 0.57 \end{pmatrix} \tag{12}$$

Notice that we specify the same problem over the variables as in Example 1. However, we have altered the specifications over the group space in two ways. Firstly, we have doubled the number of groups, and secondly, we have removed the symmetry in the relationships between the groups that existed in the first example. The important points to note however, are that the spaces $\langle \mathcal{Z}_1 \rangle$, and $\langle \mathcal{Z}_2 \rangle$ will have the same form as in Example 1, namely the collection of sums of the variables and differences between the variables for each group, and that the adjustment over these two spaces is essentially the same and dependent only upon the matrices A and B and the sample sizes chosen. For simplicity of exposition, we shall consider here only a balanced design, that is $n_g = n \; \forall g$.

The first step is to find the underlying canonical variable directions by solving $CY = (C+E)Y\Phi$. We find that

$$Y = \begin{pmatrix} 0.1961 & 0.3162 \\ 0.1961 & -0.3162 \end{pmatrix} \qquad \Phi = \begin{pmatrix} 0.3824 & 0 \\ 0 & 0.3571 \end{pmatrix} \tag{13}$$

Let $S_g = \mathcal{M}(X_{g1}) + \mathcal{M}(X_{g2})$ and $D_g = \mathcal{M}(X_{g1}) - \mathcal{M}(X_{g2})$, and create the uncorrelated vector-spaces $\langle \mathcal{Z}_1 \rangle$, $\langle \mathcal{Z}_2 \rangle$ where $\mathcal{Z}_1 = \{0.1961S_1, 0.1961S_2, 0.1961S_3, 0.1961S_4\}$, and $\mathcal{Z}_2 = \{0.3162D_1, 0.3162D_2, 0.3162D_3, 0.3162D_4\}$.

We now find the underlying canonical group directions by solving $AW = (A+B)W\Psi$. For the balanced design, we know that the canonical directions are the same for all sample sizes, so we need only solve for a sample size of one in each group. We find that

$$W = \begin{pmatrix} 0.2505 & 2.0529 & 0.6152 & 0.0970 \\ 0.3297 & -0.2167 & -2.5522 & -0.4804 \\ 0.2432 & -0.8584 & 0.5641 & 2.9189 \\ 0.3330 & -1.2055 & 1.4953 & -2.3654 \end{pmatrix} \tag{14}$$

and $\Psi = diag(0.8161, 0.1695, 0.1471, 0.0940)$. The columns of W are the coordinates of the directions in the $\mathcal{Z}_i$ that are the canonical directions, both within the $\langle \mathcal{Z}_i \rangle$, and, when collected together, for the $\langle \mathcal{M}(X) \rangle$. For example, the direction $(0.2505, 0.3297, 0.2432, 0.3330)^T$ is the most informative direction for the spaces $\langle \mathcal{Z}_1 \rangle$ and $\langle \mathcal{Z}_2 \rangle$. We now apply the theorem to join the two stages together to find the canonical structure for the overall adjustment. The outcome is shown in Table 3.

Table 3. *Canonical directions and resolutions for the mean components.*

Direction	Z_{11}	Z_{12}	Z_{21}	Z_{22}	Z_{31}	Z_{32}	Z_{41}	Z_{42}
Resolution $(n = 1)$	0.7331	0.7114	0.1122	0.1018	0.0965	0.0874	0.0603	0.0545
Component								
$\mathcal{M}(X_{11})$	0.0491	0.0792	0.4026	0.6492	0.1207	0.1946	0.0190	0.0307
$\mathcal{M}(X_{12})$	0.0491	-0.0792	0.4026	-0.6492	0.1207	-0.1946	0.0190	-0.0307
$\mathcal{M}(X_{21})$	0.0647	0.1043	-0.0425	-0.0685	-0.5005	-0.8071	-0.0942	-0.1519
$\mathcal{M}(X_{22})$	0.0647	-0.1043	-0.0425	0.0685	-0.5005	0.8071	-0.0942	0.1519
$\mathcal{M}(X_{31})$	0.0477	0.0769	-0.1683	-0.2714	0.1106	0.1784	0.5724	0.9230
$\mathcal{M}(X_{32})$	0.0477	-0.0769	-0.1683	0.2714	0.1106	-0.1784	0.5724	-0.9230
$\mathcal{M}(X_{41})$	0.0653	0.1053	-0.2364	-0.3812	0.2932	0.4728	-0.4639	-0.7480
$\mathcal{M}(X_{42})$	0.0653	-0.1053	-0.2364	0.3812	0.2932	-0.4728	-0.4639	0.7480

Thus, the most informative direction is given by Z_{11}, and the second canonical direction by Z_{12}. The structure of these are shown below:

$$Z_{11} = 0.0491\mathcal{M}(X_{11}) + 0.0491\mathcal{M}(X_{12}) + \ldots + 0.0653\mathcal{M}(X_{41}) + 0.0653\mathcal{M}(X_{42}) \quad (15)$$
$$= 0.2505(0.1961S_1) + \ldots + 0.3330(0.1961S_4) \quad (16)$$

$$Z_{12} = 0.0792\mathcal{M}(X_{11}) - 0.0792\mathcal{M}(X_{12}) + \ldots + 0.1053\mathcal{M}(X_{41}) - 0.1053\mathcal{M}(X_{42}) \quad (17)$$
$$= 0.2505(0.3162D_1) + \ldots + 0.3330(0.3162D_4) \quad (18)$$

The corresponding resolutions for a sample of size one in each group are given by

$$\lambda_{11(1)} = 0.7331 = \frac{0.8161 \times 0.3824}{0.8161 \times 0.3824 + (1 - 0.8161)(1 - 0.3824)} \quad (19)$$

$$\lambda_{12(1)} = 0.7114 = \frac{0.8161 \times 0.3571}{0.8161 \times 0.3571 + (1 - 0.8161)(1 - 0.3571)} \quad (20)$$

The resolutions for general n may be found from the formula shown in Corollary 1. For example, $\lambda_{11(n)} = 0.7331n/\{0.7331(n-1) + 1\}$.

Notice that there are several small eigenvalues. We expect to learn slowly about the corresponding quantities. If particular interest was held in these quantities, for example Z_{41} is approximately contrasting the third and the fourth groups, we may apply Corollary 2 to investigate required sample sizes to achieve proportionate variance reductions for quantities of interest. For example, since $\psi_{min} = 0.0940$ and $\phi_{min} = 0.3571$, sample sizes of 18, 157, 330 in each group will guarantee a proportionate variance reduction of 50 percent, 90 percent and 95 percent, respectively, for every element of $\langle \mathcal{M}(X) \rangle$.

ACKNOWLEDGEMENTS

The first named author is supported by a grant from the UK's EPSRC. The calculations in Section 5 were carried out using the Bayes linear computing package $[B/D]$, which is available at the internet site *http://fourier.dur.ac.uk:8000/stats/bd/home.html*

REFERENCES

Dawid, A. P. (1988). Symmetry models and hypotheses for structured data layouts. *J. Roy. Statist. Soc. B* **50**, 1–34 (with discussion).

Farrow, M. and Goldstein, M. (1993). Bayes linear methods for grouped multivariate repeated measurements studies with applications to crossover trials. *Biometrika* **80**, 39–59.

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

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

Goldstein, M. (1996). Revising exchangeable beliefs: subjectivist foundations for the inductive argument. *Aspects of Uncertainty: a Tribute to D. V. Lindley* (P. R. Freeman and A. F. M. Smith, eds.). Chichester: Wiley, 201–222.

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

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

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

Searle, S. R., Casella, G. and McCulloch, C. E. (1992). *Variance Components*. Chichester: Wiley.

Stewart, G. and Sun, J.-g. (1990). *Matrix perturbation theory*. New York: Academic Press.

APPENDIX: PROOF OF THEOREM

Goldstein (1981) shows that the canonical directions and resolutions may be calculated from the eigenstructure of the resolution transform for the adjustment. In our case, the resolution transform is given by

$$T = Var(\mathcal{M}(X))^{-1}Cov(\mathcal{M}(X), \overline{X}_{(N)})Var(\overline{X}_{(N)})^{-1}Cov(\overline{X}_{(N)}, \mathcal{M}(X)) \qquad (21)$$

Calculating the desired variances and covariances from equations (7), (8), and (9), we find that $Var(\mathcal{M}(X)) = Cov(\mathcal{M}(X), \overline{X}_{(N)}) = (A \otimes C)$ and $Var(\overline{X}_{(N)}) = \{(A \otimes C) + (N^{-1}B \otimes E)\}$, where $(R \otimes S)$ denotes the direct product of R and S. For further details and properties of the direct product see Searle *et al.* (1992). Since A, C, N, B, E are positive definite, then we have invertibility of $\{(A \otimes C) + (N^{-1}B \otimes E)\}$ and $(A \otimes C)$, and so we have that $T = \{(A \otimes C) + (N^{-1}B \otimes E)\}^{-1}(A \otimes C)$. We find the matrix of eigenvectors, $\mathcal{Z}$, and the corresponding diagonal matrix of eigenvalues, Λ, of T by solving the equivalent generalised eigenvalue probem $(A \otimes C)\mathcal{Z} = \{(A \otimes C) + (N^{-1}B \otimes E)\}\mathcal{Z}\Lambda$. Let (Φ, Y), (Ψ, W) be the pairs of matrices of eigenvalues and eigenvectors respectively solving the two generalised eigenvalue problems $CY = (C + E)Y\Phi$ and $AW = (A + N^{-1}B)W\Psi$. Consider

$$(A \otimes C)(W \otimes Y)\{(I - \Psi) \otimes (I - \Phi)\} = AW(I - \Psi) \otimes CY(I - \Phi) \qquad (22)$$

Notice from the solution of our two generalised problems above, we may write $CY(I - \Phi) = EY\Phi$ and $AW(I - \Psi) = N^{-1}BW\Psi$. Substituting these into (22), we find that:

$$(A \otimes C)(W \otimes Y)\{(I - \Psi) \otimes (I - \Phi)\} = N^{-1}BW\Psi \otimes EY\Phi \qquad (23)$$

$$= (N^{-1}B \otimes E)(W \otimes Y)(\Psi \otimes \Phi) \qquad (24)$$

Adding $(A \otimes C)(W \otimes Y)(\Psi \otimes \Phi)$ to both sides gives:

$$(A \otimes C)(W \otimes Y)\{(\Psi \otimes \Phi) + (I - \Psi) \otimes (I - \Phi)\} \\ = \{(A \otimes C) + (N^{-1}B \otimes E)\}(W \otimes Y)(\Psi \otimes \Phi) \qquad (25)$$

Note that since $0 < \psi_d < 1 \; \forall d = 1, \ldots, g_0$ and $0 < \phi_s < 1 \; \forall s = 1, \ldots, v_0$, we may invert $\{(\Psi \otimes \Phi) + (I - \Psi) \otimes (I - \Phi)\}$ so that:

$$(A \otimes C)\mathcal{Z} = \{(A \otimes C) + (N^{-1}B \otimes E)\}\mathcal{Z}\Lambda \qquad (26)$$

where $\mathcal{Z} = (W \otimes Y)$ and $\Lambda = (\Psi \otimes \Phi)\{(\Psi \otimes \Phi) + (I - \Psi) \otimes (I - \Phi)\}^{-1}$. The canonical directions are then given by $Z_{ds} = (W_d \otimes Y_s)^T \mathcal{M}(X)$, the corresponding canonical resolutions being given by the diagonal elements of Λ. Orthogonality is verified by noting $Cov((W_d \otimes Y_s)^T \mathcal{M}(X), (W_{d'} \otimes Y_{s'})^T \mathcal{M}(X)) = (W^T A W)_{dd'}(Y^T C Y)_{ss'}$. This gives us parts *3* and *4* of the theorem. Parts *1* and *2* follow by observing that we may arrange the eigenvectors as the matrix $[W \otimes Y_1 \ldots W \otimes Y_{v_0}]$.