

APPROXIMATE INFORMATION MATRICES FOR ESTIMATING A GIVEN SET OF CONTRASTS

L.-J. Kao*, P. K. Yates, S. M. Lewis and A. M. Dean*

The Ohio State University and University of Southampton*

Abstract. This paper considers the construction of block designs for estimating given sets of treatment contrasts. Necessary and sufficient conditions are given for the form of the matrix M which minimizes the trace of $HM^{-1}H'$, where H is the contrast matrix. The application of this result in constructing highly efficient designs is illustrated.

Key words and phrases: A-criterion, approximate information matrix, block designs, contrasts.

1. Approximate Information Matrices

Let $\mathcal{D} = \mathcal{D}(v, b, k)$ be the class of block designs, not necessarily connected, having b blocks of size k and v treatment labels, where the i th label occurs r_i times in the design ($i = 1, \dots, v$). Let $H\tau$ be a vector of contrasts in the treatment parameters τ , where $\text{rank}(H) = h \leq v - 1$ and let \mathcal{D}_H be the subset of designs in \mathcal{D} for which $H\tau$ is estimable under the usual additive model. Let C_d be the information matrix of the design $d \in \mathcal{D}_H$ and define the set \mathcal{C}_H by $\mathcal{C}_H = \{C_d : d \in \mathcal{D}_H\}$.

A common objective in designing an experiment is to find a design $d^* \in \mathcal{D}_H$ for which the average variance of the least squares estimators of $H\tau$ is minimized or, equivalently, for which $\text{tr}(HC_d^-H')$ is a minimum over all $C_d \in \mathcal{C}_H$, where C_d^- is the Moore-Penrose generalized inverse of C_d and tr denotes trace. In general, this is a very difficult problem. Specific contrast matrices that have been considered in the literature include all pairwise comparisons (reviewed in Shah and Sinha (1989)), test treatment versus control contrasts (reviewed in Hedayat, Jacroux and Majumdar (1988), and also in Shah and Sinha (1989)), dual versus single contrasts (Gerami, Lewis, Majumdar and Notz (1994)), and general contrasts in completely randomized designs (Sinha (1970)).

Our approach is to minimize $\text{tr}(HM^{-1}H')$, for any given set of contrasts $H\tau$, over all matrices $M \in \mathcal{M}$ where the set \mathcal{M} contains \mathcal{C}_H and is defined below. Although this minimization may not result in a matrix M corresponding to an information matrix of a design $d \in \mathcal{D}_H$, we show in Section 2 that it can lead

to the identification of highly efficient designs. Accordingly, we call $M \in \mathcal{M}$ an *approximate information matrix*. Let the set \mathcal{M} be defined as

$$\mathcal{M} = \{M : M \text{ is a } v \times v, \text{ symmetric, non-negative definite matrix with } h \leq \text{rank}(M) \leq v - 1; M1_v = 0 \text{ and } \text{tr}(M) \leq c_{\max}\}, \tag{1}$$

where c_{\max} is the maximum trace of C_d over \mathbf{D}_H and where 1_v is a vector of v unit elements. Let x_1, \dots, x_v be a set of orthonormal eigenvectors of $M \in \mathcal{M}$ and let $\lambda_1, \dots, \lambda_v$ be the corresponding eigenvalues. On defining $\lambda_i^{-1} = 0$ if $\lambda_i = 0$, and writing M^- in terms of its spectral decomposition, it follows that

$$\text{tr}(HM^-H') = \text{tr} \left[H \left(\sum_{i=1}^v \lambda_i^{-1} x_i x_i' \right) H' \right] = \sum_{i=1}^v \lambda_i^{-1} \text{tr}(H x_i x_i' H') = \sum_{i=1}^v \lambda_i^{-1} \beta_i, \tag{2}$$

where $\beta_i = x_i' H' H x_i, 1 \leq i \leq v$. Using the Cauchy-Schwarz inequality,

$$\left(\sum_{i=1}^v \lambda_i^{-1} \beta_i \right) \left(\sum_{i=1}^v \lambda_i \right) \geq \left(\sqrt{\beta_1} + \dots + \sqrt{\beta_v} \right)^2 \tag{3}$$

with equality holding if and only if $\lambda_i = \alpha \beta_i^{1/2}$ for some constant $\alpha, 1 \leq i \leq v$. Therefore, writing $\sum \lambda_i = t$, and using (2) and (3), we have $\text{tr}(HM^-H') \geq t^{-1}(\sum_{i=1}^v \beta_i^{1/2})^2$ and $\text{tr}(HM^-H')$ is a minimum over \mathcal{M} if and only if M is such that (i) $\lambda_i = \alpha \beta_i^{1/2} (i = 1, \dots, v)$, where α is a constant such that $\sum \lambda_i = c_{\max}$, and (ii) $\sum_{i=1}^v \beta_i^{1/2}$ is a minimum over all sets of orthonormal vectors x_1, \dots, x_v in R^v . The following theorem characterizes $M \in \mathcal{M}$ so that (i) and (ii) are achieved.

Theorem. *Let \mathcal{M} be defined as in (1) and let $H\tau$ be a vector of contrasts in the treatment parameters τ where $\text{rank}(H) = h$. Let u_1, \dots, u_v be a set of orthonormal eigenvectors of $H'H$ and let $\theta_1, \dots, \theta_v$ be the corresponding eigenvalues. Then $\text{tr}(HM^*H') = \min\{\text{tr}(HM^-H') : M \in \mathcal{M}\}$ if and only if $M^* = \alpha \sum_{i=1}^v \sqrt{\theta_i} u_i u_i'$, where α is a constant such that $\text{tr}(M^*) = c_{\max}$.*

Proof. Let $H'H$ have $\ell + 1 \geq 2$ distinct eigenvalues, and let the multiplicity of the j th distinct eigenvalue be n_j . Without loss of generality we assume that the labelling of the eigenvectors u_1, \dots, u_v of $H'H$ is such that the eigenvalues are ordered $\theta_1 = \dots = \theta_{s_1} > \theta_{s_1+1} = \dots = \theta_{s_2} > \dots > \theta_{s_\ell+1} = \dots = \theta_v = 0$ where $s_j = \sum_{i=1}^j n_i$ for $j = 1, \dots, \ell$ and where $h = s_\ell$ and $v = s_{(\ell+1)}$. Define E_{j+1} to be the subspace of R^v spanned by $u_{s_j+1}, \dots, u_{s_{(j+1)}}$.

Sufficiency: Follows from the fact that, when M^* is as in the statement of the theorem, $\text{tr}(HM^*H')$ achieves the lower bound $B(H)$ of Gerami and Lewis (1992).

Necessity: Define β_i as in (2). Write $H'H$ in terms of its spectral decomposition. It then follows that $\beta_i = \sum_{j=1}^v \theta_j (u'_j x_i)^2$, $i = 1, \dots, v$. Let $\beta = P\theta$, where $\beta = (\beta_1, \dots, \beta_v)'$, $\theta = (\theta_1, \dots, \theta_v)'$ and P is a $v \times v$ matrix with (i, j) th element $(u'_j x_i)^2$.

Expressing each x_1, \dots, x_v in terms of the orthonormal basis u_1, \dots, u_v of R^v (and vice versa), it can be shown that P is doubly stochastic. Hence, from Theorem A.4 of Marshall and Olkin (1979, p. 20), $\beta = P\theta$ is majorized by θ . Theorem A.4.b of Marshall and Olkin (1979, p. 58) establishes that the function $-\sum_{i=1}^v \beta_i^{1/2}$ is strictly Schur convex on $(R^+)^v$. Hence, $\sum_{i=1}^v \theta_i^{1/2} \leq \sum_{i=1}^v \beta_i^{1/2}$ with equality if and only if $\beta = R\theta$, for some permutation matrix R . Without loss of generality take $R = I$, and $\beta = \theta$.

In order to find an M in the set \mathcal{M} defined in (1) whose eigenvectors satisfy (ii), we seek x_1, \dots, x_v so that the matrix $P = \{(u'_j x_i)^2\}$ satisfies $P\theta = \beta = \theta$. It is shown in the appendix that P has block diagonal form with $n_j \times n_j$ matrices P_j ($j = 1, \dots, \ell + 1$) on the diagonal. It follows from the definition of P that, for $i = s_j + 1, \dots, s_{(j+1)}$, the vector x_i is orthogonal to each vector $u_t \notin E_{j+1}$. Hence, $x_i \in E_{j+1}$. This holds for all $j = 1, \dots, \ell + 1$, and consequently (x_1, \dots, x_v) is an alternative complete set of orthonormal eigenvectors of $H'H$. Using (i), the matrix M which minimizes $\text{tr}(HM^{-1}H')$ is $M^* = \alpha \sum_{i=1}^v \sqrt{\theta_i} x_i x'_i = \alpha \sum_{i=1}^v \sqrt{\theta_i} u_i u'_i$ where α is such that $\text{tr}(M^*) = c_{\max}$.

Corollary. *If $d^* \in D_H$ has information matrix M^* defined in the theorem then d^* is A -optimal over D .*

2. Applications and Discussion

A direct application of the theorem and its corollary is the well-known result that a balanced block design is A -optimal for estimating a complete set of $v - 1$ orthonormal contrasts, or a complete set of pairwise comparisons.

For other sets of contrasts $H\tau$, the approximate information matrix M^* generally will not correspond to the information matrix of any design in D_H . However, it enables highly efficient designs to be located within the class of aligned designs. Aligned designs are defined by Lewis and Gerami (1994) to be designs whose information matrices have a set of orthonormal eigenvectors in common with $H'H$. The contrasts $H\tau$ are estimable in all aligned designs. In the class of non-aligned designs, estimability would need to be checked, for example, via $HC^{-1}C = H$.

The following problem, which was posed by a statistician at a pharmaceutical company, serves to illustrate our approach. In planning a trial to compare a very large number of stimuli, prior knowledge was used to give a preliminary ranking to the size of the effects of the stimuli, labelled $\tau_0, \dots, \tau_{v-1}$. An incomplete block

design was required to enable efficient estimation of the following subset of all pairwise comparisons:

$$\tau_i - \tau_{i+j} \quad \text{for } i = 0, \dots, v-1; \quad j = 1, \dots, p < v-2, \quad (4)$$

where $i+j$ is evaluated modulo v . Comparisons between the high and low ranked stimuli were included as well as those between stimuli ranked close together. For this set of contrasts $H'H$ is a circulant matrix with initial row $(2p, -1'_p, 0, -1'_p)$, and hence M^* is also circulant.

The class of aligned designs includes all designs having circulant information matrices. For equal treatment replication, the approximate concurrence matrix is $N^*N^{*'} = k(rI - M^*)$. It is sensible to consider designs with equal treatment replication for two reasons. Firstly, in this particular set of contrasts, each treatment effect occurs the same number of times with the same coefficients. Secondly, the diagonal elements of M^* are equal and equi-sized blocks are required. The entries in the concurrence matrix are adjusted to attain integer entries and to retain the symmetric circulant structure under the restriction that the row sums are all equal to rk . Guided by these restrictions, a shortlist of approximate concurrence matrices can be drawn up, not all of which will necessarily correspond to realizable designs.

For illustration, consider $v = 6$ treatments and the contrasts (4) with $p = 2$. For a design with $b = 12$ blocks of size $k = 3$, $H'H$ and M^* are circulant matrices with first rows as follows, $H'H : \{4 \ -1 \ -1 \ 0 \ -1 \ -1\}$, and $M^* : \{4 \ -0.9 \ -0.9 \ -0.4 \ -0.9 \ -0.9\}$. For any binary design, the diagonal elements of its information matrix are $k^{-1}r(k-1)$. Hence, using M^* and $k = 3, r = 6$, the approximate concurrence matrix $N^*N^{*'}$ is circulant with first row $N^*N^{*'}$: $\{6 \ -2.7 \ -2.7 \ -1.2 \ -2.7 \ -2.7\}$. The following shortlist of circulant matrices with integer entries and row sums $rk = 18$ may be drawn up as candidates for the concurrence matrix of an efficient design.

$$\begin{array}{ll} \text{(a)} \ \{ 6 \ 3 \ 3 \ 0 \ 3 \ 3 \} & \text{(b)} \ \{ 6 \ 2 \ 3 \ 2 \ 3 \ 2 \} \\ \text{(c)} \ \{ 6 \ 3 \ 2 \ 2 \ 2 \ 3 \} & \text{(d)} \ \{ 6 \ 4 \ 2 \ 0 \ 2 \ 4 \} \\ \text{(e)} \ \{ 6 \ 2 \ 4 \ 0 \ 4 \ 2 \} & \end{array}$$

Cyclic designs exist with concurrence matrices c , d and e . The generating blocks for each of the designs are (012) and (013) for c , two copies of (012) for d , and (012) and two copies of (024) for e . A comparison of $\text{tr}(HC_d^- H')$ with its minimum value $\text{tr}(HM^{*-} H') = 4.9494$ (which is equal to the bound $B(H)$ of Gerami and Lewis (1992)), gives efficiencies 0.991, 0.961 and 0.949, respectively. Group divisible designs exist with concurrence matrices a and b , see Table 1, and these both have efficiency 0.990.

Table 1. Group divisible designs for $b = 12, r = 6, v = 6, k = 3$

Design with concurrence matrix (a)						Design with concurrence matrix (b)																	
0	0	0	0	0	0	1	1	1	2	3	3	0	0	0	0	0	0	1	1	1	1	2	2
1	1	1	2	2	4	2	2	3	3	4	4	1	1	2	2	3	4	2	3	3	3	3	4
2	5	5	4	4	5	3	3	5	4	5	5	2	5	3	4	4	5	4	4	5	5	5	5

Although the above application is such that a class of equi-replicate designs is a reasonable choice, the same approach can be taken when unequal treatment replication is more appropriate. In this case the formula of Jones (1976) can be used to allocate replications amongst the treatments.

Further examples are given by Lewis and Gerami (1994), and Kao, Notz and Dean (1994). The theorem is less useful as a tool for identifying efficient designs outside the class of aligned designs, since search algorithms such as that of Jones and Eccleston (1980) already exist.

Acknowledgement

P. K. Yates was supported by a Science and Engineering Research Council research studentship.

Appendix: The form of P satisfying $P\theta = \theta$

Let θ have entries $\theta_1 = \dots = \theta_{s_1} > \theta_{s_1+1} = \dots = \theta_{s_2} > \dots > \theta_{s_{\ell}+1} = \dots = \theta_v = 0$, where $s_j = \sum_{i=1}^j n_i$ for $j = 1, \dots, \ell$ and $v = s_{(\ell+1)}$ and $\ell \geq 1$. Then $\theta = (\theta_1, \dots, \theta_v)' = (a_1 1'_{n_1} \dots a_\ell 1'_{n_\ell} 0 1'_{n_{(\ell+1)}})'$. Let $P = \{p_{ij}\}$ be any $v \times v$ doubly stochastic matrix P , with non-negative elements, satisfying $P\theta = \theta$. The i th element of $P\theta$ is

$$a_1 \sum_{j=1}^{s_1} p_{ij} + a_2 \sum_{j=s_1+1}^{s_2} p_{ij} + \dots + a_\ell \sum_{j=s_{(\ell-1)}+1}^{s_\ell} p_{ij},$$

where $s_r = \sum_{i=1}^r n_i$, for $r = 1, \dots, \ell$. Now $P\theta - \theta = 0$ by assumption. For $1 \leq i \leq s_1$, the i th element of $P\theta - \theta$ is

$$(a_2 - a_1) \sum_{j=s_1+1}^{s_2} p_{ij} + \dots + (a_\ell - a_1) \sum_{j=s_{(\ell-1)}+1}^{s_\ell} p_{ij} = 0.$$

Since $a_1 > a_2 > \dots > a_\ell > 0$, and since the elements of P are non-negative, it follows that $p_{i(s_1+1)}, \dots, p_{is_\ell}$ are all zero. Thus, P can be partitioned as $\{P_{ij}\}$, $i, j = 1, 2$, where P_{11} is $n_1 \times n_1$ and $P_{12} = 0$. Since P is doubly stochastic, the row sums of P_{11} are unity. Consequently $1'P_{11}1 = n_1$. Now P_{11} has n_1 columns with non-negative entries summing to at most 1.0. Thus each column sum must be exactly 1.0 and, therefore, $P_{21} = 0$. This implies that P_{22} is doubly stochastic,

and the above argument can be repeated using the fact that $P_{22}\theta_2^* = \theta_2^*$ where $\theta_2^* = (\theta_{s_1+1}, \dots, \theta_v)'$. Repeated application of the argument shows that P has block diagonal form with i th matrix P_{ii} on the diagonal of size $n_i \times n_i$ ($i = 1, \dots, \ell + 1$).

References

- Gerami, A. and Lewis, S. M. (1992). Comparing dual with single treatments in block designs. *Biometrika* **79**, 603-610.
- Gerami, A. and Lewis, S. M., Majumdar, D. and Notz, W. I. (1994). Efficient block designs for comparing dual with single treatments. Ohio State University Technical Report, 501. (Submitted for publication).
- Hedayat, A. S., Jacroux, M. and Majumdar, D. (1988). Optimum designs for comparing test treatments with controls (with discussion). *Statist. Sci.* **3**, 462-491.
- Jones, B. (1976). An algorithm for deriving optimal block designs. *Technometrics* **18**, 451-458.
- Jones, B. and Eccleston, J. A. (1980). Exchange and interchange procedures to search for optimal designs. *J. Roy. Statist. Soc. Ser. B* **42**, 238-243.
- Kao, L.-J., Notz, W. I. and Dean, A. M. (1994). Efficient designs for conditional main effects contrasts. Ohio State University Technical Report, 529. (Submitted for publication).
- Lewis, S. M. and Gerami, A. (1994). Finding efficient block designs for estimating specific treatment contrasts. Southampton University Preprint Series, 253. (Submitted for publication).
- Marshall, A. W. and Olkin, I. (1979). *Inequalities: Theory of Majorization and its Applications*. Academic Press, New York.
- Shah, K. R. and Sinha, B. K. (1989). *Theory of Optimal Designs*. Lecture Notes in Statist. **54**, Springer-Verlag, Berlin.
- Sinha, B. K. (1970). On the optimality of some designs. *Calcutta Statist. Assoc. Bull.* **19**, 1-22.

Department of Statistics, The Ohio State University, Columbus, OH, 43210, U.S.A.

Faculty of Mathematical Studies, The University of Southampton, Southampton, S09 5NH, England.

(Received November 1993; accepted March 1995)