The Annals of Statistics
1989, Vol. 17, No. 3, 1315–1324

MINIMAXITY OF RANDOMIZED OPTIMAL DESIGNS[1]

BY PETER M. HOOPER

University of Alberta

A strategy of randomizing a design that is L-optimal for an ideal model is shown to possess a minimax property under a larger, more realistic model. This generalizes a result of Li.

1. Introduction. Experimental randomization is generally regarded as a source of robustness against model inadequacies; see Wu (1981) and references therein. Li (1983) showed that any nonrandomized design can be improved in a minimax sense by randomization. More precisely, he considered the performance of estimators under violations of an ideal unit-effects model and showed that the maximum risk is reduced when the design is suitably randomized. His main assumption was that the enlarged model (ideal model plus violations) is invariant under a group of permutations of the experimental units. "Suitable" randomization was defined as the uniform distribution on the group.

This result suggests that randomizing a design that is optimal under the ideal model might produce a randomized design that is minimax under the enlarged model. Li (1983) proved that in certain situations this is indeed the case. For this second result, he restricted attention to linear estimators and quadratic loss functions. Under the ideal model this leads to the family of design criteria known as linear criteria, or L-criteria, which includes the well known A-criterion.

The purpose of the present paper is to generalize and simplify Li's second result. The generalization involves a weakening of model assumptions. The simplification results from considering the enlarged model first and the ideal model second. The structure of the enlarged model is described by a group of permutations, e.g., the group related to a simple orthogonal block structure [Nelder (1965)]. The ideal model is obtained by averaging over the group and then choosing a least favorable distribution.

2. Statement of the theorem. Consider a comparative experiment in which treatments are assigned to experimental units and univariate responses are observed. There are n experimental units indexed by a set Δ. By ordering Δ in some way, we may identify Δ with $\{1, \ldots, n\}$. It is assumed that treatment and unit effects are additive and that a known linear model is suitable for the treatment effects. Thus each potential treatment is identified with a known vector $x \in \mathbb{R}^p$ and the treatment effect is $x'\beta$ for some unknown $\beta \in \mathbb{R}^p$. The treatment-effects model might involve nontrivial assumptions, such as a

Received September 1987; revised August 1988.
[1]Research supported in part by the Natural Sciences and Engineering Research Council of Canada.
AMS 1980 *subject classification.* 62K05.
Key words and phrases. A-optimality, design robustness, experimental randomization, L-optimality.

regression model for a quantitative treatment factor. Sacks and Ylvisaker (1984) studied designs that are robust against inadequacies in the treatment-effects model. This problem will not be considered here.

A nonrandomized design is an assignment of treatments to units. It is useful to adopt a somewhat redundant description of the set of designs by splitting the assignment into two stages: First assign treatments to labels $1, \ldots, n$; then assign labels to actual units. There is a one-to-one correspondence between the set of possible first stage assignments and a set $\mathscr{X}$ of $n \times p$ matrices $X \equiv (x_1, \ldots, x_n)'$. The effect of the treatment assigned to label i is $x_i'\beta$. Each second stage assignment is equivalent to a permutation G of Δ defined by: Label i is assigned to unit $G^{-1}i$. The permutations G will be restricted to a subgroup $\mathscr{G}$ describing the structure of the experimental material. It is assumed that every first stage assignment $X \in \mathscr{X}$ can be combined with every second stage assignment $G \in \mathscr{G}$, so that $\mathscr{X} \times \mathscr{G}$ defines the set of all nonrandomized designs under consideration.

For each $G \in \mathscr{G}$ let $\bar{G}$ be the $n \times n$ permutation matrix defined by

$$(\bar{G}u)_i = u_{G^{-1}i}, \qquad u \in \mathbb{R}^n,\ i = 1, \ldots, n.$$

It is easily verified that $G \to \bar{G}$ is an isomorphism between $\mathscr{G}$ and a group $\bar{\mathscr{G}}$ of permutation matrices. We will suppress the bar and identify $(\mathscr{G}, G)$ with $(\bar{\mathscr{G}}, \bar{G})$. Notice that (X, G) and (HX, HG) define the same design for all $H \in \mathscr{G}$.

For a given design (X, G), let $\mathbf{y}_i$ denote the response from unit $G^{-1}i$. (All random variables will be expressed in boldface notation.) The treatment associated with $\mathbf{y}_i$ depends on X while the unit associated with $\mathbf{y}_i$ depends on G. Since treatment and unit effects are assumed additive, we have

$$\mathbf{y}_i = x_i'\beta + \mathbf{u}_{G^{-1}i}, \qquad i = 1, \ldots, n, \tag{2.1}$$

where $x_i'\beta$ is the treatment effect and $\mathbf{u}_{G^{-1}i}$ is the unit effect. In vector notation (2.1) becomes

$$\mathbf{y} = X\beta + G\mathbf{u}. \tag{2.2}$$

Here $\beta \in \mathbb{R}^p$ is an unknown vector of parameters and $\mathbf{u}$ is an unobservable random vector whose distribution is independent of (X, G).

This model can be expanded slightly to incorporate technical errors arising from variability in treatment application. Let $\mathscr{L}(\cdot)$ denote the distribution of its argument. Consider a model

$$\mathbf{y} = X\beta + \mathbf{t} + G\tilde{\mathbf{u}}, \tag{2.3}$$

where $\mathbf{t}$ and $\tilde{\mathbf{u}}$ are independent and $\mathscr{L}(G\mathbf{t}) = \mathscr{L}(\mathbf{t})$ for all $G \in \mathscr{G}$. Defining $\mathbf{u} = G^{-1}\mathbf{t} + \tilde{\mathbf{u}}$ reduces (2.3) to (2.2).

Suppose the experiment is designed to estimate $L\beta$, where L is a known $q \times p$ matrix. We restrict attention to linear estimators $A\mathbf{y}$, $A \in \mathbb{R}^{q \times n}$, and adopt the quadratic loss function

$$\|A\mathbf{y} - L\beta\|^2 = (A\mathbf{y} - L\beta)'(A\mathbf{y} - L\beta).$$

Using randomization if desired, consider the problem of choosing a design (X, G) and corresponding estimator $A\mathbf{y}$ to minimize the maximum risk. A distribution on $\mathscr{X} \times \mathscr{G} \times \mathbb{R}^{q \times n}$ will be called a *strategy*. Let $\mathscr{P}_u$ be a family of distributions

on $\mathbb{R}^n$ representing prior information about the unit effects. A strategy $\mathscr{L}(\mathbf{X}, \mathbf{G}, \mathbf{A})$ will be called $\mathscr{P}_u$-*minimax* if it minimizes

$$\sup E\|\mathbf{A}(\mathbf{X}\beta + \mathbf{G}\mathbf{u}) - L\beta\|^2, \tag{2.4}$$

where the supremum is over $(\beta, \mathscr{L}(\mathbf{u})) \in \mathbb{R}^p \times \mathscr{P}_u$ and $(\mathbf{X}, \mathbf{G}, \mathbf{A})$ is distributed independently of $\mathbf{u}$.

A $\mathscr{P}_u$-minimax strategy depends only on the second-order model determined by $\mathscr{P}_u$. This will be described by functions $\psi: \mathbb{R}^n \to [0, \infty]$ defined by

$$\psi(a) = E(a'\mathbf{u})^2.$$

Put $\Psi = \{\psi: \mathscr{L}(\mathbf{u}) \in \mathscr{P}_u\}$. Observe that $E\|\mathbf{u}\|^2 < \infty$ for all $\mathscr{L}(\mathbf{u}) \in \mathscr{P}_u$ if and only if $\psi(a) < \infty$ for all $(\psi, a) \in \Psi \times \mathbb{R}^n$ and, in this case, $\psi(a) = a'(E\mathbf{u}\mathbf{u}')a$ defines a bijection between ψ and $E\mathbf{u}\mathbf{u}'$.

Suppose $\mathscr{G}$ describes the basic structure of the experimental material. The principal example is the automorphism group defined by a block structure; see Section 3. The second-order model ψ contains further, perhaps more tentative, assumptions about the unit effects. It is reasonable that the more detailed description be compatible with the basic description. The following assumption makes this idea precise.

The usual action of $\mathscr{G}$ on the set of distributions on $\mathbb{R}^n$, $\mathscr{L}(\mathbf{u}) \to \mathscr{L}(G\mathbf{u})$, determines a corresponding action on the set of ψ-functions, $\psi \to \psi G'$, where

$$\psi G'(a) = \psi(G'a) = E(a'G\mathbf{u})^2.$$

It is assumed that Ψ is $\mathscr{G}$-invariant, i.e.,

$$\psi \in \Psi \quad \text{implies} \quad \psi G' \in \Psi \quad \text{for all } G \in \mathscr{G}. \tag{2.5}$$

This will be the case if $\mathscr{P}_u$ is $\mathscr{G}$-invariant, i.e., $\mathscr{L}(\mathbf{u}) \in \mathscr{P}_u$ implies $\mathscr{L}(G\mathbf{u}) \in \mathscr{P}_u$ for all $G \in \mathscr{G}$.

Let $\mathbf{H}$ be distributed uniformly on $\mathscr{G}$ and independently of $\mathbf{u}$. The uniform distribution is characterized by the invariance property

$$\mathbf{H} \sim G\mathbf{H} \sim \mathbf{H}G \quad \text{for all } G \in \mathscr{G}. \tag{2.6}$$

For $\psi \in \Psi$ define $\bar{\psi}: \mathbb{R}^n \to [0, \infty]$ by $\bar{\psi}(a) = E(a'\mathbf{H}\mathbf{u})^2$ or, more briefly,

$$\bar{\psi} = E\psi\mathbf{H}'.$$

Notice that $\bar{\psi}$ is $\mathscr{G}$-invariant:

$$\bar{\psi}G' = \bar{\psi} \quad \text{for all } G \in \mathscr{G}. \tag{2.7}$$

Consider the set

$$\left\{a \in \mathbb{R}^n: \sup_{\psi \in \Psi} \bar{\psi}(a) < \infty\right\}. \tag{2.8}$$

It is easily verified that (2.8) is a $\mathscr{G}$-invariant subspace. Let Q denote the corresponding projection matrix. Thus Q is an $n \times n$ symmetric idempotent matrix satisfying $Q = GQG'$ for all $G \in \mathscr{G}$ and $Q\mathbb{R}^n$ equals the set (2.8).

Let $\mathscr{S}_n$ be the set of $n \times n$ nonnegative definite symmetric matrices. For $\psi \in \Psi$ put

$$V_\psi = E(\mathbf{QHuu'H'Q})$$

and observe that $a'V_\psi a = \bar{\psi}(Qa)$ so $V_\psi \in \mathscr{S}_n$ and V_ψ depends only on $\bar{\psi}$. Furthermore we have $V_\psi = QV_\psi$ and $V_\psi = GV_\psi G'$ for all $G \in \mathscr{G}$. Recall the usual partial ordering on $\mathscr{S}_n$: $V_1 \le V_2$ if $V_2 - V_1 \in \mathscr{S}_n$. Let $\operatorname{tr} V$ denote the trace of V.

The following assumption is nontrivial and important: There exists a matrix $V \in S_n$ that is least favorable relative to $\{V_\psi : \psi \in \Psi\}$ in the sense that

$$\begin{aligned} &V_\psi \le V \quad \text{for all } \psi \in \Psi, \\ &\sup_{\psi \in \Psi} \operatorname{tr} V_\psi = \operatorname{tr} V. \end{aligned} \tag{2.9}$$

It is not hard to show that V is uniquely determined (see Lemma 3), $V = QV$ and $V = GVG'$ for all $G \in \mathscr{G}$. Condition (2.9) is a generalization of (3.5) in Li (1983). See Li (1983), page 233 for an example where the condition fails.

Suppose there exists a solution (X_0, A_0) to the problem:

$$\begin{aligned} &\text{minimize } \operatorname{tr} AVA' \text{ subject to} \\ &AX = L, \qquad AQ = A \quad \text{and} \quad (X, A) \in \mathscr{X} \times \mathbb{R}^{q \times n}. \end{aligned} \tag{2.10}$$

The problem is usually posed in the following context. Given $(X, G) \in \mathscr{X} \times \mathscr{G}$, consider the model

$$\mathbf{y} = X\beta + G(\eta + \mathbf{e}), \tag{2.11}$$

where η is an unknown vector in $(I_n - Q)\mathbb{R}^n$ and $\mathbf{e}$ is an unobservable random vector with $E\mathbf{e} = 0$ and $E\mathbf{ee}' = V$. We will refer to this as the *ideal* model. Notice that $G\eta$ is an unknown vector in $(I_n - Q)\mathbb{R}^n$ and $E(G\mathbf{e})(G\mathbf{e})' = GVG' = V$, so the ideal model depends only on X, not G. We have $AX = L$ and $AQ = A$ if and only if $A\mathbf{y}$ is an unbiased estimator for $L\beta$ under (2.11). The risk of an unbiased estimator $A\mathbf{y}$ is $\operatorname{tr} AVA'$. Thus X_0 is an L-optimal design under the ideal model and $A_0\mathbf{y}$ is the corresponding best linear unbiased estimator.

If X is compact, as is the case in most applications, then a solution to (2.10) exists. To see this, let $A_X\mathbf{y}$ be the best linear unbiased estimator of $L\beta$ under the ideal model (2.11) and observe that $\operatorname{tr} A_X V A'_X$ is a continuous function of X.

The matrix V may be replaced by $c_1V + c_2(I_n - Q)$, $c_1 > 0$, without changing the solution to (2.10). In particular, if $V \propto Q$, then we may take $V = I_n$.

The main result can now be stated. The proof is given in Section 4.

THEOREM. *If Ψ is $\mathscr{G}$-invariant, V satisfies* (2.9) *and* (X_0, A_0) *is a solution to* (2.10), *then the strategy* $\mathscr{L}(X_0, \mathbf{H}, A_0)$ *is* $\mathscr{P}_u$*-minimax.*

3. Application. The theorem reduces the problem of finding a $\mathscr{P}_u$-minimax strategy to one of finding an L-optimal design under an ideal model. The latter problem has been studied extensively; see Silvey (1980), Li (1983), Hedayat and Majumbar (1985) and references therein. To apply the theorem one must specify the relevant aspects of the second-order model for the unit effects, namely $\mathscr{G}$, Q

and V. The choice of Q and V is usually made easier by the following considerations.

Put $\mathbf{W} = E\{\mathbf{Huu'H'} \mid \mathbf{u}\}$, i.e.,

$$\mathbf{W} = (\#\mathcal{G})^{-1} \sum_{H \in \mathcal{G}} H\mathbf{uu'}H'.$$

Notice that

$$\mathbf{W} = G\mathbf{W}G' \quad \text{for all } G \in \mathcal{G}. \tag{3.1}$$

For many groups $\mathcal{G}$ that are transitive on the label set Δ, condition (3.1) implies that the spectral decomposition of $\mathbf{W}$ has the form

$$\mathbf{W} = \sum_j (\operatorname{tr} Q_j\mathbf{W}/\operatorname{tr} Q_j)Q_j, \tag{3.2}$$

where the Q_j are nonzero symmetric idempotent matrices, summing to I_n, with $Q_j = GQ_jG'$ for all $G \in \mathcal{G}$; see Speed (1987). The important point is that the Q_j depend only on $\mathcal{G}$ and hence are, in theory, known. The subspaces $Q_j\mathbb{R}^n$ are called *strata*.

If (3.2) holds, then

$$\bar{\psi}(a) = Ea'\mathbf{W}a = \sum_j \xi_{\psi j} a'Q_j a,$$

where

$$\xi_{\psi j} = E\big(\operatorname{tr} Q_j\mathbf{W}/\operatorname{tr} Q_j\big) = E\big(\|Q_j\mathbf{Hu}\|^2/\operatorname{tr} Q_j\big). \tag{3.3}$$

Define

$$\xi_j = \sup_{\psi \in \Psi} \xi_{\psi j} \quad \text{and} \quad J = \{j: \xi_j < \infty\}. \tag{3.4}$$

Observe that

$$Q = \sum_{j \in J} Q_j \quad \text{and} \quad V_\psi = \sum_{j \in J} \xi_{\psi j} Q_j. \tag{3.5}$$

The only possible choice for V is

$$V = \sum_{j \in J} \xi_j Q_j. \tag{3.6}$$

Condition (2.9) holds if and only if

$$\sup_{\psi \in \Psi} \sum_{j \in J} \xi_{\psi j} = \sum_{j \in J} \xi_j \tag{3.7}$$

or, equivalently, the vector $(\xi_j,\ j \in J)$ can be approximated by $(\xi_{\psi j},\ j \in J)$ with $\psi \in \Psi$. Roughly speaking, this means that maximum variability can occur simultaneously in all strata where variability is bounded.

The matrix V is determined, up to a scale factor, by ξ_j/ξ_k, $j, k \in J$. If we thought that the maximum variability was the same for all strata indexed by $j \in J$, then we would take $V \propto Q$ and, in the ideal model (2.11), we could replace

V by I_n. If J contains only a single element, then we have $V \propto Q$ by definition and (3.7) is trivially satisfied.

Specifying the set J is effectively the same as assigning fixed effects to certain unit factors in the ideal model. For example, in units-within-blocks models $\mathscr{G}$ is a wreath-product group (independently permute units within each block, then permute blocks) and there are three strata: mean, interblock and intrablock, indexed by $j = 0$, 1 and 2. If we thought that variability between blocks was much larger than variability within blocks, then we might take $\xi_0 = \xi_1 = \infty$ and $\xi_2 < \infty$. This leads to an ideal model with fixed block effects and many optimal-design results are available. On the other hand, it might be more realistic to take $\xi_2 < \xi_1 < \infty$. This leads to an ideal model with random block effects and little about optimality is known; see Rosenberger (1986) for a review of the literature on combined interblock–intrablock estimators.

In criss-cross models such as Latin squares, where units are arranged in rows and columns, $\mathscr{G}$ is a direct product group (independently permute rows and columns) and there are four strata: mean, row, column and row $\times$ column interaction, indexed by $j = 0$, 1, 2 and 12. Taking $\xi_0 = \xi_1 = \xi_2 = \infty$ and $\xi_{12} < \infty$ leads to an ideal model with fixed row and column effects.

The most commonly used models for unit effects are the simple orthogonal block structures of Nelder (1965). Units-within-blocks models and criss-cross models are the simplest nontrivial structures in this class. General structures are obtained by iterating and combining the nesting and crossing operations. The factorial dispersion models described by Speed and Bailey (1982, 1987) and Speed (1987) form a generalization of Nelder's class. In these models the label set Δ is a cross product of sets Δ_f, $f \in F$, where F is a set of block factors and Δ_f indexes the levels of factor f. There is a partial order less than or equal to on F, where $f_1 \leq f_2$ means that factor f_1 is nested within factor f_2. Bailey, Praeger, Rowley and Speed (1983) derived the largest group preserving the block structure: $\tilde{\mathscr{G}}$ is the generalized wreath product of $\tilde{\mathscr{G}}_f$, $f \in F$, where $\tilde{\mathscr{G}}_f$ is the group of all permutations of Δ_f. If we take $\mathscr{G} = \tilde{\mathscr{G}}$, then (3.2) holds and there is a general algorithm for computing the Q_j. A simpler version of the algorithm, given in Nelder (1965), suffices for most applications.

Other there is secondary structure in the experimental material not accounted for by the factorial dispersion model, e.g., location in blocks. For the sake of simplicity, it may be reasonable to ignore secondary structure when considering optimality questions; however, it is desirable that treatment assignments be approximately balanced with respect to the secondary structure for all possible designs. The potential for serious imbalance throws in doubt the conditional relevance of the randomization. Approximate balance can sometimes be ensured by using a minimax strategy with restricted randomization.

Suppose $\mathscr{G}$ is the generalized wreath product of $\mathscr{G}_f$, $f \in F$, where $\mathscr{G}_f$ is a subgroup of $\tilde{\mathscr{G}}_f$. Theorem C of Bailey, Praeger, Rowley and Speed (1983) implies that $\mathscr{G}$ produces the same decomposition (3.2) as $\tilde{\mathscr{G}}$ if and only if each $\mathscr{G}_f$ is 2-transitive on Δ_f. Suppose this is the case. Let $\tilde{\mathbf{H}}$ and $\mathbf{H}$ be distributed uniformly on $\tilde{\mathscr{G}}$ and $\mathscr{G}$, respectively. Suppose $\mathscr{P}_u$ ignores the secondary unit structure so that (2.5) holds for $G \in \tilde{\mathscr{G}}$. Assume, without loss of generality, that

$\mathscr{X}$ is $\tilde{\mathscr{G}}$-invariant. Suppose $\mathscr{L}(X_0, \tilde{\mathbf{H}}, A_0)$ is a $\mathscr{P}_u$-minimax strategy. It follows that $\mathscr{L}(\tilde{G}X_0, \tilde{\mathbf{H}}, A_0\tilde{G}')$ is $\mathscr{P}_u$-minimax for each $\tilde{G} \in \tilde{\mathscr{G}}$. The above mentioned application of Theorem C then implies that $\mathscr{L}(\tilde{G}X_0, \mathbf{H}, A_0\tilde{G}')$ is $\mathscr{P}_u$-minimax for all $\tilde{G} \in \tilde{\mathscr{G}}$. It is sometimes possible to choose $\tilde{G}$ and $\mathscr{G}$ so that treatment assignments are approximately balanced with respect to the secondary structure for all possible realizations of $\mathbf{H}$; see Bailey (1986).

If the secondary structure is deemed sufficiently important, then an even smaller group $\mathscr{G}$ may be chosen. This complicates the choice of Q and V. The decomposition (3.2) may no longer be available. Even if (3.2) does hold, the number of strata will be increased.

Li (1983), Theorem 3.1, considered models with the following structure. Suppose the experimental units are arranged in b blocks, $\Delta = \Delta_1 \cup \cdots \cup \Delta_b$, and that each block has a unique structure. It makes sense to permute units within each block but not to permute blocks. Suppose $\mathscr{G} = \mathscr{G}_1 \times \cdots \times \mathscr{G}_b$, where $\mathscr{G}_k$ is a group acting transitively on Δ_k and permutations are defined by $Gi = G_k i$ for $i \in \Delta_k$, $G = (G_1, \ldots, G_b)$. Thus each block is randomized independently of the others. Suppose $\mathscr{G}_k$ determines a spectral decomposition of the form (3.2) for the kth block, $k = 1, \ldots, b$. In each block one of the strata is the one-dimensional block-mean stratum. Let Q_{wb} be the orthogonal projection matrix onto the $(n - b)$-dimensional within-blocks subspace. It is not hard to show that $Q_{wb}\mathbf{W}Q_{wb}$ admits a spectral decomposition of the form (3.2). The strata are all those determined by $\mathscr{G}_k$, $k = 1, \ldots, b$, with the exception of the b block-mean strata. If it is thought that variability between blocks is very large compared with variability within blocks, then it is reasonable to restrict Q so that $QQ_{wb} = Q$. This reduces the problem of choosing Q and V to that previously considered. Li assumed, for each block, a simple orthogonal block structure with all factors crossed, i.e., no nesting. He further assumed that exactly one ξ_j is finite for each block, namely the ξ_j corresponding to the highest-order interaction.

Li's Example 4 provides an interesting application. There is one block with eight experimental units classified by three factors, each at two levels. Let factors 1, 2 and 3 correspond to rows, columns and layers. Applying the algorithm and notation of Nelder (1965), one finds there are eight strata indexed by 0, 1, 2, 3, 12, 13, 23 and 123. Li defines two families $\mathscr{P}_{u1}$ and $\mathscr{P}_{u2}$—determined by G_1 and G_2 in his notation—and two nonrandomized designs d_1 and d_2. It is straightforward to show that, up to a scale factor, we have $\xi_0 = \xi_1 = \xi_2 = \xi_3 = \infty$, $\xi_{13} = \xi_{23} = \sigma^2 < \infty$ under both families, $\xi_{123} = \sigma^2$, $\xi_{12} = 1 + \sigma^2$ under $\mathscr{P}_{u1}$ and $\xi_{123} = 1 + \sigma^2$, $\xi_{12} = \sigma^2$ under $\mathscr{P}_{u2}$. It follows that

$$Q = Q_{12} + Q_{13} + Q_{23} + Q_{123},$$

$$V = \begin{cases} V_1 \equiv Q_{12} + \sigma^2 Q & \text{under } \mathscr{P}_{u1}, \\ V_2 \equiv Q_{123} + \sigma^2 Q & \text{under } \mathscr{P}_{u2}. \end{cases}$$

The designs d_1 and d_2 are Youden hypercubes and hence are A-optimal under (2.11) with $V \propto Q$ [Cheng (1978), Theorem 5.1]. Now the best linear unbiased estimator of the treatment difference is based on stratum 123 using d_1 and

stratum 12 using d_2. Thus d_1, but not d_2, is A-optimal for $V = V_1$ while d_2, but not d_1, is A-optimal for $V = V_2$. The theorem shows that a $\mathscr{P}_{ui}$-minimax strategy is obtained by randomizing d_i.

The minimaxity theory helps clarify the relationship between randomization and optimality. Two weaknesses in the present theory should be noted, however. First, the optimality criterion ignores the desirability of a valid estimate of error for each stratum; see Bailey and Rowley (1987). Second, the restriction to linear estimators is undesirable when estimates are combined from several strata.

4. Proof. The proof of the theorem is based on three lemmas. The proofs of Lemmas 1 and 2 are straightforward and so are omitted.

LEMMA 1. *Let* $\mathbf{z}$ *and* $\mathbf{w}$ *be jointly distributed with* $\mathbf{z} \in \mathbb{R}^p$ *and* $\mathbf{w} \in \mathbb{R}$. *If*

$$\sup_{\beta \in \mathbb{R}^p} E(\mathbf{z}'\beta + \mathbf{w})^2 < \infty,$$

then $\mathbf{z} = 0$ *with probability* 1.

LEMMA 2. *If* $\mathbf{a}$ *is a random n-vector with* $\sup_{\psi \in \Psi} E\bar{\psi}(\mathbf{a}) < \infty$, *then* $P\{\mathbf{a} \in Q\mathbb{R}^n\} = 1$.

LEMMA 3. *If* (2.9) *holds and* $\mathbf{A}$ *is a random* $q \times n$ *matrix, then* $\sup_{\psi \in \Psi} E(\operatorname{tr} \mathbf{A}V_\psi \mathbf{A}') = E(\operatorname{tr} \mathbf{A}V\mathbf{A}')$.

PROOF. Let $\lambda_1 \geq \cdots \geq \lambda_n$ and $\lambda_{\psi 1} \geq \cdots \geq \lambda_{\psi n}$ denote the ordered eigenvalues of V and V_ψ and let r be the rank of V. Recall that $\operatorname{tr} V = \Sigma\lambda_i$. Assumption (2.9) implies that $\lambda_{\psi i} \leq \lambda_i$ for $i = 1, \ldots, n$, $\sup_{\psi \in \Psi} \lambda_{\psi r} = \lambda_r > 0$ and $\inf_{\psi \in \Psi}(\operatorname{tr} V - \operatorname{tr} V_\psi)/\lambda_{\psi r} = 0$. Using standard methods, one may verify that if $\lambda_{\psi r} > 0$, then

$$(\operatorname{tr} \mathbf{A}V\mathbf{A}')/\lambda_1 \leq (\operatorname{tr} \mathbf{A}V_\psi \mathbf{A}')/\lambda_{\psi r} \tag{4.1}$$

and

$$0 \leq \operatorname{tr} \mathbf{A}V\mathbf{A}' - \operatorname{tr} \mathbf{A}V_\psi \mathbf{A}' \leq (\operatorname{tr} \mathbf{A}V\mathbf{A}')(\operatorname{tr} V - \operatorname{tr} V_\psi)/\lambda_{\psi r}. \tag{4.2}$$

The conclusion of the lemma follows from (4.1) if $E(\operatorname{tr} \mathbf{A}V\mathbf{A}') = \infty$ and from (4.2) if $E(\operatorname{tr} \mathbf{A}V\mathbf{A}') < \infty$. □

PROOF OF THEOREM. The main assumptions are given at (2.5), (2.9) and (2.10). Let $\mathscr{L}(\mathbf{X}, \mathbf{G}, \mathbf{A})$ be a strategy with bounded risk M defined as

$$M \equiv \sup E\|(\mathbf{AX} - L)\beta + \mathbf{AGu}\|^2 < \infty,$$

where the supremum is over $(\beta, \mathscr{L}(\mathbf{u})) \in \mathbb{R}^p \times \mathscr{P}_u$ and $(\mathbf{X}, \mathbf{G}, \mathbf{A})$ is distributed independently of $\mathbf{u}$. If no such strategy exists, then all strategies are trivially minimax. Lemma 1 shows that $M < \infty$ implies

$$\mathbf{AX} = L \quad \text{with probability 1.} \tag{4.3}$$

Let $\mathbf{a}_i'$ denote the ith row of $\mathbf{A}$ and observe that

$$M = \sup E\|\mathbf{AGu}\|^2 = \sup_{\psi\in\Psi} \sum_{i=1}^{q} E\psi\mathbf{G}'(\mathbf{a}_i).$$

By (2.5) we have

$$M = \sup_{\psi\in\psi} \sum_{i=1}^{q} E\psi H'\mathbf{G}'(\mathbf{a}_i)$$

for all $H \in \mathscr{G}$. Thus

$$\begin{aligned} M &= (\#\mathscr{G})^{-1} \sum_{H\in\mathscr{G}} \sup_{\psi\in\Psi} \sum_{i=1}^{q} E\psi H'\mathbf{G}'(\mathbf{a}_i) \\ &\geq \sup_{\psi\in\Psi} \sum_{i=1}^{q} E\left\{(\#\mathscr{G})^{-1} \sum_{H\in\mathscr{G}} \psi H'\mathbf{G}'(\mathbf{a}_i)\right\} \\ &= \sup_{\psi\in\Psi} \sum_{i=1}^{q} E\bar{\psi}\mathbf{G}'(\mathbf{a}_i) \\ &= \sup_{\psi\in\Psi} \sum_{i=1}^{q} E\bar{\psi}(\mathbf{a}_i). \end{aligned}$$

The last step uses (2.7).

Next, Lemma 2 and $M < \infty$ show that $P\{\mathbf{a}_i \in Q\mathbb{R}^n\} = 1$ for $i = 1, \ldots, q$. Thus

$$\mathbf{A}Q = \mathbf{A} \quad \text{with probability 1} \tag{4.4}$$

and

$$\begin{aligned} M &\geq \sup_{\psi\in\Psi} \sum_{i=1}^{q} E\bar{\psi}(Q\mathbf{a}_i) \\ &= \sup_{\psi\in\Psi} \sum_{i=1}^{q} E\mathbf{a}_i'V_\psi\mathbf{a}_i \\ &= \sup_{\psi\in\Psi} E \operatorname{tr}\mathbf{A}V_\psi\mathbf{A}'. \end{aligned}$$

By Lemma 3, (2.10), (4.3) and (4.4) we have

$$\begin{aligned} M &\geq E \operatorname{tr}\mathbf{A}V\mathbf{A}' \\ &\geq \operatorname{tr} A_0VA_0' \\ &= \sup_{\psi\in\Psi} \operatorname{tr} A_0V_\psi A_0' \\ &= \sup E\|(A_0X_0 - L)\beta + A_0\mathbf{Hu}\|^2. \end{aligned}$$

This completes the proof. □

Acknowledgments. I thank the referees for their helpful comments.

REFERENCES

BAILEY, R. A. (1986). Randomization, constrained. *Encyclopedia of Statistical Sciences* (S. Kotz and N. L. Johnson, eds.) **7** 524–530. Wiley, New York.

BAILEY, R. A. and ROWLEY, C. A. (1987). Valid randomization. *Proc. Roy. Soc. London. Ser. A* **410** 105–124.

BAILEY, R. A., PRAEGER, C. E., ROWLEY, C. A. and SPEED, T. P. (1983). Generalized wreath products of permutation groups. *Proc. London Math. Soc.* (*3*) **47** 69–82.

CHENG, C.-S. (1978). Optimal designs for the elimination of multi-way heterogeneity. *Ann. Statist.* **6** 1262–1272.

HEDAYAT, A. S. and MAJUMBAR, D. (1985). Families of A-optimal block designs for comparing test treatments with a control. *Ann. Statist.* **13** 757–767.

LI, K.-C. (1983). Minimaxity for randomized designs: Some general results. *Ann. Statist.* **11** 225–239.

NELDER, J. A. (1965). The analysis of randomized experiments with orthogonal block structure. I: Block structure and the null analysis of variance. *Proc. Roy. Soc. London Ser. A* **283** 147–162.

ROSENBERGER, J. L. (1986). Recovery of interblock information. *Encyclopedia of Statistical Sciences* (S. Kotz and N. L. Johnson, eds.) **7** 654–656. Wiley, New York.

SACKS, J. and YLVISAKER, D. (1984). Some model robust designs in regression. *Ann. Statist.* **12** 1324–1348.

SILVEY, S. D. (1980). *Optimal Design*. Chapman and Hall, New York.

SPEED, T. P. (1987). What is an analysis of variance (with discussion)? *Ann. Statist.* **15** 885–941.

SPEED, T. P. and BAILEY, R. A. (1982). On a class of association schemes derived from lattices of equivalence relations. In *Algebraic Structures and Applications* (P. Schultz, C. E. Praeger and R. P. Sullivan, eds.) 55–74. Dekker, New York.

SPEED, T. P. and BAILEY, R. A (1987). Factorial dispersion models. *Internat. Statist. Rev.* **55** 261–277.

WU, C.-F. (1981). On the robustness and efficiency of some randomized designs. *Ann. Statist.* **9** 1168–1177.

DEPARTMENT OF STATISTICS
AND APPLIED PROBABILITY
UNIVERSITY OF ALBERTA
EDMONTON, ALBERTA
CANADA T6G 2G1