Minimax Property For Randomized Designs - A Unified Approach*

by

Ker-Chau Li

Purdue University

Mimeograph Series # 81-44

Department of Statistics
Division of Mathematical Sciences

AMS Subject classifications: Primary 62K99, Secondary 62A05,62C20.
*This research was sponsored in part by the National Science Foundation under grant NO. MCS-78-25301, and was completed in partial fulfillment of requirements for the Doctor of Philosophy degree in the University of California, Berkeley, Department of Statistics.

Minimax Property For Randomized Designs - A Unified Approach*

by

Ker-Chau Li

Purdue University

SUMMARY

In many commonly-used experimental design situtations, where the assumed model is contaminated by some small function, which is assumed to be invariant under a suitable group of transformations, we first demonstrate a "stochastic" minimax property for many commonly-used randomization procedures, which generalizes and unifies some results of C.F. Wu (1980) on the same topic. The results we obtain do not rely on any special analytic properties of the loss function and the estimators. Next, under the loss function of the A-criterion, and restricting to the use of linear estimators only, we search for the randomized strategy (i.e., randomized design and estimator) which minimizes the maximal risk. The classical A-optimal designs for the contaminationless models, after being randomized in the obvious way, and the least squares estimators, are found to be optimal in many, but not all set-ups (some counterexamples are provided). As a simple consequence of our general approach, some randomization procedures generated by a group of very small order are shown to be as efficient as the common-used one, which usually is generated by a group of very large order.

AMS Subject classifications: Primary 62K99, Secondary 62A05,62C20
*This research was sponsored in part by the National Science Foundation under grant NO. MCS-78-25301, and was completed in partial fulfillment of requirements for the Doctor of Philosophy degree in the University of California, Berkeley, Department of Statistics.

KEY WORDS AND PHARSES: A-optimality, doubly transitive groups, G-uniform randomization, k-way heterogeneity, minimaxity, randomization procedures, randomized designs, robustness, stochastic minimaxity.

MINIMAX PROPERTY FOR RANDOMIZED DESIGNS — A UNIFIED APPROACH

1 Introduction

The role of randomization in the design of experiments has been discussed in numerous papers (see the references given in [10]). Recently this problem was re-examined by C. F. Wu (1980) from the aspect of model-robustness (c.f. [7]) which is concerned with the case when the true model departs from the assumed model by some small quantity. Besides proposing a new measure for comparing designs in terms of maximum squared bias, he was able to demonstrate a minimax property for the balanced completely randomized design in the case where no blocking is needed. In more complex design situations, fixing an orthogonal design (e.g., a balanced block design or Latin square design), he established that the complete randomization (in the obvious way) is minimax when the model is invariant under a suitable transformation group and when the loss function used is the A-criterion. These results, among many other clever ideas, stimulated our further investigation. (The reader was suggested to read his paper first for more interpretations about the problems to be considered here.)

First, we observe that Wu's results rely heavily on being able to get explicit expressions for the expected loss for the procedures to be considered. Thus the role of randomization tends to be obscured among the desirable properties of least squares estimators, orthogonality of design, squared loss, uniform weights for all pairwise contrasts, etc. As a particular consequence, the important question about the minimax choice of a randomization procedure for a given non-orthogonal design (even a BIBD) was left unsolved. Therefore, it is our first aim to clarify the role of randomization in the aspect of

model-robustness for a fixed but arbitrary design by the aid of a rigorous decision – theoretic treatment. We carefully distinguish the notion of a randomization procedure from that of a randomized design. After introducing the concept of G-uniform randomization (which is reduced to the complete randomization when the group G involved is the group of all permutations on the labels of the experimental units), we find that ("for any design") G-uniform randomization is minimax, without assumptions on any special analytic properties of estimators and loss functions. Thus the above special question, among many others, is solved. Furthermore, treating the risk as a random variable generated by the randomization procedure concerned, and recognizing that (since the loss function is often an approximation) the mean of the risk is sometimes misleading as a criterion for choosing a reasonable randomization procedure, we base our results on the stochastic ordering of the risks. For convenience, the terminology, "stochastic" minimaxity (or its obvious variations), will be understood to emphasize the stochastic ordering nature of the criterion concerned. Thus, in short, our results establish that the G-uniform randomization is "stochastically" minimax. Above all, not only estimation problems but also testing problems are solved simultaneously. The rigorous development is given in Section 2.

In Section 3, for the estimation case, we consider the problem of finding a minimax randomized strategy (i.e., design and estimator) under the same model-violation consideration as in Section 2. Since a stochastic minimax solution in the sense mentioned above does not exist, we restrict ourselves to the commonly used A-criterion. The use of the least squares estimators is not assumed. Instead, we only require the

estimators to be linear in the observations. It turns out that completely randomizing an A-optimal design and restricting to the use of least squares estimators is a minimax strategy for many (but not all) design set-ups. Our set-up is so general that it covers the case where K-way heterogeneity elimination is considered and interactions are assumed ($K \geqslant 1$). But, surprisingly, it does not cover the usual additive set-up where no interactions are present for $K \geqslant 3$. Counter-examples are provided.

Also, as a simple consequence, for many design problems, a randomization procedure *uniformly* generated by any *doubly transitive* group is found to be of the same efficiency (in terms of squared loss) as that generated by the group of all permutations. Moreover, for the block design case with equal block sizes the result is even more interesting. The common randomization practice is to randomize completely the units within blocks and then to randomize completely the entire blocks. But we find that the following randomization procedure, which is much simpler, is of the same efficiency in our setting: "uniformly randomizing the units within blocks according to any doubly transitive groups of permutations within blocks and then uniformly randomly *rotating* the entire blocks."

Details are discussed in the last part of Section 3. Section 4 is devoted to the proofs of the results in Section 3.

2 "Stochastic" Minimaxity

First, we shall give a very useful lemma. Some definitions are provided for clarity and the notations once given will be used throughout this chapter.

Let $F = \{f | f$ is a real function on the class of all probability measures on R, such that $f(p\mu_1+(1-p)\mu_2) \leq \max\{f(\mu_1),f(\mu_2)\}$, for any probability measures μ_1 and μ_2 and any p such that $0 \leq p \leq 1\}$. For any random variable X, we write $f(X)$ to denote the number $f(\mu)$, where μ is the probability measure for X. For any two random variables X and Y, we write $X \overset{st}{=} Y$ if X and Y have the same distribution. Thus if $X \overset{st}{=} Y$ and $f \in F$ then $f(X) = f(Y)$.

Let G be a group of permutations on $\{1,\ldots,N\}$ (N is a natural number henceforth considered fixed). Denote the order of G by g. For any permutation π and any $\underset{\sim}{x} \in R^N$, write $\underset{\sim}{x} = (x_1,\ldots,x_N)^t$ and $\pi(\underset{\sim}{x}) = (x_{\pi(1)},\ldots,x_{\pi(N)})^t$. Define $A = \{\pi((1,\ldots,N)^t) | \pi \in G\}$. Let P be the set of all probability measures on A and let ξ_o be the probability measure uniform on A (i.e., assigning equal masses to all elements of A). Denote the random element generated by a probability measure $\xi \in P$ by $\underset{\sim}{\xi}$ (thus each realization of $\underset{\sim}{\xi}$ belongs to A). This convention is adopted throughout this paper, e.g., ν generates $\underset{\sim}{\nu}$ and μ generates $\underset{\sim}{\mu}$. The following lemma is fundamental.

LEMMA 2.1 Let h be any real function on A. Then, for any $f \in F$, we have

$$\min_{\xi \in P} \max_{\pi \in G} f(h(\pi(\underset{\sim}{\xi}))) = f(h(\underset{\sim}{\xi}_o)) = \max_{\pi \in G} f(h(\pi(\underset{\sim}{\xi}_o))) .$$

Proof. For any $\pi \in G$ and $\xi \in \mathcal{P}$, let $\tilde{h}(\pi,\xi)$ be the probability measure for $h(\pi(\underset{\sim}{\xi}))$. Observe that for any $\pi \in G$, we have

$$\tilde{h}(\pi,\xi_0) = \frac{1}{g} \sum_{\pi' \in G} \tilde{h}(\pi',\xi) \ .$$

Thus we get $f(\tilde{h}(\pi,\xi_0)) \leq \underset{\pi' \in G}{\text{Max}}\, f(\tilde{h}(\pi',\xi))$, because $f \in \mathcal{F}$. Hence the lemma follows obviously. □

Further discussions on this lemma, with many commonly-used f in $\mathcal{F}$, will be provided following the proof of Theorem 2.1 below.

Next, the interaction among designs, models, estimators and the loss function under a permutation group will be demonstrated. Since we shall develop a rigorous and general set-up to cover most design problems where we allow contaminations to be present in the model and therefore robustness is naturally to be considered, many definitions are introduced in what follows. The reader may want to follow along reading Ex. 1 just below at the same time as the general development, to aid in understanding.

Suppose T treatments are to be assigned to N units which may or may not be blocked. A design is a function d from $\{1,\ldots,N\}$ to $\{1,\ldots,T\}$ with the understanding that unit i receives treatment $d(i)$. Let $\underset{\sim}{D}$ be the set of all designs. Most often, the probability structure of the observations $(Y_1,\ldots,Y_N)^t = \underset{\sim}{Y}$ depends not only on the design used but also on the "true state" of nature which is usually parametrized by the block

effects, treatment effects, unit effects (when contaminations are allowed), etc., as we now describe more precisely below. Let S be a set, considered to be the set of parameter values of all such parameters. Let $\mathbf{Q}$ be the set of all probability measures on R^N. A function M from $\underset{\sim}{D} \times S$ to $\mathbf{Q}$ is called a model. Thus, under model M, the probability measure for $\underset{\sim}{Y}$ will be $M(d,S)$ if d is the design we use and S is the true value of the parameters characterizing the "true state" of nature. Let A be the action space (so $A = R^P$ when we are estimating P parameters and $A = \{1,2\}$ in the testing hypothesis problems). A loss function L is a nonnegative function on $S \times A$. A (non-randomized) decision rule δ is a function from R^N to A. Without further discussion, we assume measurability for all functions in appropriate sense. A decision rule is said to be an estimator when $A = R^P$. Let $\underset{\sim}{\Delta}$ be the set of all δ. Any element in $\underset{\sim}{D} \times \underset{\sim}{\Delta}$ is called a strategy. For any permutation π, define $\pi(d)$ to be the design such that $\pi(d)(i) = d(\pi(i))$ for $i = 1,\ldots,N$. Write $\pi(\underset{\sim}{Y}) = (Y_{\pi(1)},\ldots,Y_{\pi(N)})^t$. For any $m \in \mathbf{Q}$, define $\pi(m)$ to be the probability measure for $\pi(\underset{\sim}{Y})$ when $\underset{\sim}{Y}$ has the probability measure m. Obviously, for any real function f on R^N, integrable under m and $\pi(m)$, we have

$$(1) \qquad E_m f(\pi(\underset{\sim}{Y})) = E_{\pi(m)} f(\underset{\sim}{Y}) \quad .$$

For any decision rule δ, let $\pi(\delta)$ be the decision rule defined by $\pi(\delta)(\underset{\sim}{y}) = \delta(\pi^{-1}(\underset{\sim}{y}))$ for any $\underset{\sim}{y} \in R^N$. A group-invariant structure on the model is one of the key assumptions in our study. Therefore, we introduce the following definitions.

Definition 2.1 For any group G of permutations on $\{1,\ldots,N\}$ and any transformation group G^* on S, a model M is said to be (G,G^*)-invariant

if there is a homomorphism from G onto G^* such that for any $\pi \in G$, we have $\pi(M(d,s)) = M(\pi(d),\pi^*(s))$ where π^* denotes the homomorphic image of π.

To have a better understanding of how G^* acts on S, we note that the block effects and treatment effects in s are the same as those in $\pi^*(s)$ for all the cases we are interested in (c.f. the examples below).

Definition 2.2 For any transformation group G^* on S, a loss function L is said to be G^*-invariant if $L(\pi^*(s),a) = L(s,a)$ for any $s \in S$, $a \in A$ and $\pi^* \in G^*$.

Definition 2.3 A *choice* of decision rules is a function $\underset{\sim}{c}$ from $\underset{\sim}{D}$ to $\underset{\sim}{\Delta}$, and $\underset{\sim}{c}$ is said to be G-invariant if $\underset{\sim}{c}$ and π commute for any $\pi \in G$; i.e., $\underset{\sim}{c}(\pi(d)) = \pi(\underset{\sim}{c}(d))$ for any $d \in \underset{\sim}{D}$. Denote the set of all choices of decision rules by $\underset{\sim}{C}$ and denote the set of all G-invariant choices of decision rules by $\underset{\sim}{C}_G$. We say a choice of estimators instead of a choice of decision rules when $A = R^p$.

A *randomized design* is a probability measure on $\underset{\sim}{D}$. Let $\underset{\sim}{\mathcal{D}}$ be the set of all randomized designs. An element in $\underset{\sim}{\mathcal{D}} \times \underset{\sim}{C}$ is called a *randomized strategy*. Let $G(d) = \{\pi(d) | \pi \in G\}$, for any $d \in \underset{\sim}{D}$. Define $\underset{\sim}{\mathcal{D}}_{G(d)} = \{\xi | \xi \in \underset{\sim}{\mathcal{D}}$ and ξ has support only on $G(d)\}$. A *randomization procedure generated by* G is a function ϕ from $\underset{\sim}{D}$ to $\underset{\sim}{\mathcal{D}}$ such that $\phi(d) \in \underset{\sim}{\mathcal{D}}_{G(d)}$. (The distinction between a randomized design and a randomization procedure generated by G is crucial.) For any $d \in \underset{\sim}{D}$, the random element generated by $\phi(d)$ is denoted by $\underset{\sim}{\phi}(d)$. Let $\underset{\sim}{\Phi}_G$ be the set of all randomization procedures ϕ generated by G. Write $\underset{\sim}{P}_G$ to denote the set of all probability measures on G. The uniform measure on G is denoted by ν_0. The following definition defines a very

useful randomization procedure which reduces to the usual notion of complete randomization for an appropriate G.

Definition 2.4 The element ϕ_G in $\underset{\sim}{\Phi}_G$, defined by letting $\phi_G(d)$ be the probability measure for the random element $\underset{\sim}{\nu}_o(d)$ for any $d \in \underset{\sim}{D}$, is called the G-*uniform* randomization (or, for convenience, the randomization procedure *uniformly* generated by G).

Under loss function L, model M and $s \in S$, define the risk $r(d,\delta;L,s,M)$ of a (non-randomized) strategy (d,δ) to be $E_{M(d,s)}L(s,\delta(\underset{\sim}{Y}))$, define the *stochastic* risk $r(\mu,\underset{\sim}{c};L,s,M)$ of a randomized strategy $(\mu,\underset{\sim}{c})$ to be the probability measure of the random variable $r(\underset{\sim}{\mu},\underset{\sim}{c}(\underset{\sim}{\mu});L,s,M)$. We are ready to demonstrate the main result in this section.

Theorem 2.1 Assume that the model M is (G,G^*)-invariant and the loss function L is G-invariant. Then, for any G-invariant choice of decision rules $\underset{\sim}{c}$ and any $f \in F$, we have

$$(2) \qquad \min_{\phi \in \underset{\sim}{\Phi}_G} \max_{s \in S} f(r(\underset{\sim}{\phi}(d_o),\underset{\sim}{c}(\phi(d_o));L,s,M) = \max_{s \in S} f(r(\underset{\sim}{\phi}_G(d_o),\underset{\sim}{c}(\underset{\sim}{\phi}_G(d_o));L,s,M))$$

for all $d_o \in \underset{\sim}{D}$.

Proof. For any $\mu \in \underset{\sim}{\mathcal{D}}_{G(d_o)}$ we write $r(\underset{\sim}{\mu},\underset{\sim}{c}(\underset{\sim}{\mu});L,s,M) = r(\underset{\sim}{\mu},\underset{\sim}{c}(\underset{\sim}{\mu});s)$ for convenience. Define $G^*(s_o) = \{\pi^*(s_o) | \pi^* \in G^*\}$ for any $s_o \in S$.

It is clear that we only need to show that for any $f \in F$ and any

$$s_0 \in S, \quad \min_{\mu \in \mathcal{D}_{\underset{\sim}{G}}(d_0)} \max_{s \in G^*(s_0)} f(r(\underset{\sim}{\mu}, \underset{\sim}{c}(\underset{\sim}{\mu}); s)) = \max_{s \in G^*(s_0)} f(r(\underset{\sim}{\phi}_G(d_0), \underset{\sim}{c}(\underset{\sim}{\phi}_G(d_0)); s)) .$$

First, we have

$$(2') \quad r(\pi(d), \underset{\sim}{c}(\pi(d)); \pi^*(s_0)) = r(d, \underset{\sim}{c}(d); s_0) .$$

This is because

$$\begin{aligned} r(\pi(d), \underset{\sim}{c}(\pi(d)); \pi^*(s_0)) &= E_{M(\pi(d), \pi^*(s_0))} L(\pi^*(s_0), \underset{\sim}{c}(\pi(d))(\underset{\sim}{Y})) \\ &= E_{M(\pi(d), \pi^*(s_0))} L(s_0, \underset{\sim}{c}(\pi(d))(\underset{\sim}{Y})) \quad \text{(Definition 2.2)} \\ &= E_{\pi(M(d, s_0))} L(s_0, \underset{\sim}{c}(\pi(d))(\underset{\sim}{Y})) \quad \text{(Definition 2.1)} \\ &= E_{M(d, s_0)} L(s_0, \underset{\sim}{c}(\pi(d))(\pi(\underset{\sim}{Y}))) \quad \text{(By (1))} \\ &= E_{M(d, s_0)} L(s_0, \pi(\underset{\sim}{c}(d))(\pi(\underset{\sim}{Y}))) \quad \text{(Definition 2.3)} \\ &= E_{M(d, s_0)} L(s_0, \underset{\sim}{c}(d)(\underset{\sim}{Y})) \\ &= r(d, \underset{\sim}{c}(d); s_0) . \end{aligned}$$

Next, in order to apply Lemma 2.1, we take h to be the function defined by $h(\pi(1,\ldots,N)) = r(\pi(d_0), \underset{\sim}{c}(\pi(d_0)); s_0)$ for any $\pi \in G$. Therefore, from the above computation, we have $r(\pi_2(d_0), \underset{\sim}{c}(\pi_2(d_0)); (\pi_1^{-1})^*(s_0)) = h(\pi_1(\pi_2(1,\ldots,N)))$ for any $\pi_1, \pi_2 \in G$. Hence for any $\nu \in \underset{\sim}{P}_G$, we get

$$r(\underset{\sim}{\nu}(d_o),\underset{\sim}{\zeta}(\underset{\sim}{\nu}(d_o));(\pi_1^{-1})^*(s_o)) \overset{st}{=} h(\pi_1(\underset{\sim}{\nu}(1,\ldots,N))) \ .$$

Finally, noting that any probability measure in $\mathcal{D}_{\sim G(d_o)}$ (P in Lemma 2.1, respectively) can be derived from some probability measure in $P_{\sim G}$ by the mapping which maps π in G to $\pi(d_o)$ ($\pi(1,\ldots,N)$, respectively) in $G(d_o)$ (A, respectively), we have

$$\begin{aligned}
&\min_{\mu\in\mathcal{D}_{\sim G(d_o)}} \max_{s\in G^*(s_o)} f(r(\underset{\sim}{\mu},\underset{\sim}{\zeta}(\underset{\sim}{\mu});s)) \\
&= \min_{\mu\in\mathcal{D}_{\sim G(d_o)}} \max_{\pi_1\in G} f(r(\underset{\sim}{\mu},\underset{\sim}{\zeta}(\underset{\sim}{\mu});(\pi_1^{-1})^*(s_o))) \\
&= \min_{\nu\in P_{\sim G}} \max_{\pi_1\in G} f(r(\underset{\sim}{\nu}(d_o),\underset{\sim}{\zeta}(\underset{\sim}{\nu}(d_o));(\pi_1^{-1})^*(s_o))) \\
&= \min_{\nu\in P_{\sim G}} \max_{\pi_1\in G} f(h(\pi_1(\underset{\sim}{\nu}(1,\ldots,N)))) \\
&= \min_{\xi\in P} \max_{\pi_1\in G} f(h(\pi_1(\underset{\sim}{\xi}))) \\
&= \max_{\pi_1\in G} f(h(\pi_1(\underset{\sim}{\xi}_o))) \qquad \text{(by Lemma 2.1)} \\
&= \max_{\pi_1\in G} f(h(\pi_1(\underset{\sim}{\nu}_o(1,\ldots,N)))) \\
&= \max_{\pi_1\in G} f(r(\underset{\sim}{\nu}_o(d_o),\underset{\sim}{\zeta}(\underset{\sim}{\nu}_o(d_o));(\pi_1^{-1})^*(s_o))) \\
&= \max_{s\in G^*(s_o)} f(r(\underset{\sim}{\phi}_G(d_o),\underset{\sim}{\zeta}(\underset{\sim}{\phi}_G(d_o));s))
\end{aligned}$$

The last equality follows from Definition 2.4. □

The above theorem demonstrates a minimax property, which we shall call "stochastic" minimaxity because of the reasons to be given below, for the G-uniform randomization ϕ_G.

For any $t \in R$ and any probability measure on R, let $f_t(\xi) = \xi((t,\infty))$. Observe that a random variable X is usually said to be stochastically at least as large as another random variable Y if $f_t(X) \geq f_t(Y)$ for any $t \in R$. It is clear that a maximal or minimal element of a set of random variables in the sense of this ordering may not exist even if the cardinality of the set is finite. For example, in Lemma 2.1, we are unable to establish that $\underset{\xi \in P}{\text{Min}} \underset{\pi \in G}{\text{Max}} h(\pi(\underset{\sim}{\xi})) \overset{st}{=} h(\underset{\sim}{\xi}_0)$ simply because of the non-existence of a maximal element for the class $\{h(\pi(\underset{\sim}{\xi}))|\pi \in G\}$ (c.f. [6] for simple properties of majorization theory). However, we observe that $f_t \in F$ for any $t \in R$. Thus, by taking f to be f_t in Lemma 2.1, the result we obtain preserves the genuine spirit of stochastic minimaxity. The "stochastic" minimaxity that Theorem 2.1 demonstrates is thereby clear. Moreover, we may take $f(\xi)$ to be the "median" of ξ (or any "quantile", all given a convenient definition if not unique) and it is easy to check that $f \in F$. Finally, by taking f to be the mean functional of random variables, i.e., $f(X) = EX$, our "stochastic"-minimaxity result is then described in terms of the commonly-used sense of the risk of a randomized strategy in decision theory.

It will become clearer after studying the examples given below that all the assumptions involved in Theorem 2.1 arise quite naturally and the notion of *the randomization procedures generated by* G is a precise description of the commonly used notion of randomizations for a fixed design if G is chosen appropriately. We note further that if the

"expectation" in the definition of the risk of a strategy is replaced by the "median" (or any "quantile", or the "mode", all given a convenient definition if not unique), then the above theorem still holds. The reason is that if we replace the "expectation" in (1) by "median" (respectively, "quantile", "mode", etc.), then the equality there still holds.

Example 1. No blocking.

Consider the following set-up:

$$(4) \qquad Y_u = \alpha_{d(u)} + x_u + \varepsilon_u , \quad u = 1,\ldots,N,$$

where d is the design, α_i is the i-th treatment effect $(i = 1,\ldots,T)$, x_u is the u-th unit effect, and $\underset{\sim}{\varepsilon} = (\varepsilon_1,\ldots,\varepsilon_N)^t$ is the random error vector.

When $x_u = 0$ for all u's and ε_u's are assumed to be uncorrelated and with mean 0 and common variances, this set-up reduces to the classical model and, therefore, no randomization is needed. However, most often contaminations are unavoidable. Thus we now consider the following situation.

Let G be the set of all permutations on $\{1,\ldots,N\}$, let $\underset{\sim}{X}$ be a subset of R^N invariant under any permutation in G and let $\underset{\sim}{E}$ be a subset of Q invariant under G, in the sense that $m \in \underset{\sim}{E}$ and $\pi \in G$ implies $\pi(m) \in \underset{\sim}{E}$. We assume that the unit effect $\underset{\sim}{x} = (x_1,\ldots,x_N)^t \in \underset{\sim}{X}$ and the random error $\underset{\sim}{\varepsilon}$ has the probability measure $m \in \underset{\sim}{E}$. For example, we may take $\underset{\sim}{X} = \{\underset{\sim}{x} | \|\underset{\sim}{x}\|^2 \leq k\}$ $(k \geq 0)$, $\underset{\sim}{X} = \{\underset{\sim}{x} | |x_i| \leq k,\ i = 1,\ldots,N\}$ or $\underset{\sim}{X} = \{\underset{\sim}{x} | \sum_1^N x_i = 0,\ |x_i| \leq k\}$. Write $\underset{\sim}{\alpha} = (\alpha_1,\ldots,\alpha_T)^t$. In terms of the notations we have introduced, we take $S = \{(\underset{\sim}{\alpha},\underset{\sim}{x},m) | \underset{\sim}{\alpha} \in R^T,\ \underset{\sim}{x} \in \underset{\sim}{X}$ and $m \in \underset{\sim}{E}\}$,

$M(d,(\underset{\sim}{\alpha},\underset{\sim}{x},m))$ = the probability measure of $\underset{\sim}{Y}$ defined by (4) where $\underset{\sim}{\varepsilon}$ has the probability measure m, and

$G^* = \{\pi^* | \pi^*$ is a transformation on S such that $\exists\ \pi \in G$ for which, for any $(\underset{\sim}{\alpha},\underset{\sim}{x},m) \in S$, $\pi^*(\underset{\sim}{\alpha},\underset{\sim}{x},m) = (\underset{\sim}{\alpha},\pi(\underset{\sim}{x}),\pi(m))\}$.

Thus M is (G,G^*)-invariant since the homomorphism required in Definition 2.1 is obviously identified.

We discuss the estimation problem first. Take $A = R^P$. By definition, a G^*-invariant loss function L satisfies the condition that $L(\pi^*((\underset{\sim}{\alpha},\underset{\sim}{x},m)),a) = L((\underset{\sim}{\alpha},\underset{\sim}{x},m),a)$ for any $a \in R^P$, i.e.,

$$(5) \qquad L((\underset{\sim}{\alpha},\pi(\underset{\sim}{x}),\pi(m)),a) = L((\underset{\sim}{\alpha},\underset{\sim}{x},m),a) \quad .$$

Especially, if L is such that $L((\underset{\sim}{\alpha},\underset{\sim}{x}',m'),a) = L((\underset{\sim}{\alpha},\underset{\sim}{x},m),a)$ for any $\underset{\sim}{\alpha} \in R^T$, $\underset{\sim}{x}'$ and $\underset{\sim}{x} \in \underset{\sim}{X}$, m' and $m \in \underset{\sim}{E}$, and $a \in R^P$ then L is G^*-invariant. In other words, if what we do estimate depends only on $\underset{\sim}{\alpha}$ and in no way on $\underset{\sim}{x}$ or m, then the loss function is G^*-invariant. For example, we could have $L((\underset{\sim}{\alpha},\underset{\sim}{x},m),a) = (A\underset{\sim}{\alpha} - a)^t(A\underset{\sim}{\alpha} - a)$ for some $p \times T$ matrix A. When a choice of estimators $\underset{\sim}{\zeta}$ is G-invariant, the estimator $\underset{\sim}{\zeta}(d)$ used under the design d and the estimator $\underset{\sim}{\zeta}(\pi(d))$ used under $\pi(d)$ have the relation that $\underset{\sim}{\zeta}(d)(\underset{\sim}{Y}) = \underset{\sim}{\zeta}(\pi(d))(\pi(\underset{\sim}{Y}))$. This is a natural way of choosing estimators. In particular, in the case of estimating p contrasts among α_i's, when we say that least squares estimators are used, we mean $\underset{\sim}{\zeta}(d)$ = the least squares estimator under design d and it is clear that such a $\underset{\sim}{\zeta}$ is G-invariant. Moreover, we may want to use some other robust estimators instead of least squares estimators, but still we may want the choice of estimators G-invariant. For any $d_0 \in \underset{\sim}{D}$, it is clear that $G(d_0)$ is the class of designs which have the same replication numbers as those of the

designs d_0. Thus $\phi(d_0)$, the image of d_0 under ϕ, a randomization procedure generated by G, is simply a randomized design with the same replication numbers as those of d_0. And $\phi_G(d_0)$ is the randomized design derived by first completely randomizing the unit labels and then assigning treatments according to d_0. In the case where the replication numbers of d_0 are the same for T treatments, $\phi_G(d_0)$ is usually called a balanced completely randomized design.

The meaning of Theorem 2.1 in this case is now clear. It means that after fixing the replication numbers the complete randomization is "stochastically" minimax. Note that in this theorem only the randomization procedures are compared and nothing is said about how to select suitable replication numbers, which we shall discuss in Section 3. In the case where

$$(6) \qquad \underset{\sim}{E} = \{N(0,\sigma^2 I_{N\times N}) \mid \sigma \in R\}$$

we may want to estimate σ^2 also. Since $\pi(m) = m$ for all $m \in \underset{\sim}{E}$, from (5) it is clear that if L depends only on $\underset{\sim}{\alpha}$ and σ^2 then it is G-invariant. Therefore the complete randomization is still "stochastically" minimax even if we want to estimate σ^2. This fact is also noticed by Wu [1980] where only the ordinary sense of risk is considered.

In problems of testing hypotheses, we take $A = \{1,2\}$ and assume that L take values in $\{0,1\}$. Then, similar results can be derived without difficulties.

Example 2. Block design set-up.

Suppose the $N = \sum_{b=1}^{B} N_b$ units are arranged into b blocks with sizes $N_1,\ldots,N_B$ respectively. Consider the following set-up:

$$Y_u = \alpha_{d(u)} + \beta_u + x_u + \varepsilon_u \quad , \qquad u = 1,\ldots,N \ ,$$

where d is the design, α_i is the i-th treatment effect, x_u is the u-th effect, ε_u is the random error, and $\beta_u = \beta_b$ when u is in the b-th block, where β_b is called the b-th block effect. The set $\underset{\sim}{X}$ of all possible $\underset{\sim}{x}$ and the set $\underset{\sim}{E}$ of all possible probability measures m which generate $\underset{\sim}{\varepsilon}$ will be assumed invariant under the group of all permutations within blocks. To write more precisely, in terms of our notations, we take

G = the set of all permutations within blocks,

$S = \{(\underset{\sim}{\alpha},\underset{\sim}{\beta},\underset{\sim}{x},m) \mid \underset{\sim}{\alpha} \in R^T,\ \underset{\sim}{\beta} = (\beta_1,\ldots,\beta_B)^t \in R^B,\ \underset{\sim}{x} \in \underset{\sim}{X} \text{ and } m \in \underset{\sim}{E}\}$,

$M(d,(\underset{\sim}{\alpha},\underset{\sim}{\beta},\underset{\sim}{x},m))$ = the probability measure of $\underset{\sim}{Y}$ defined as above with $\underset{\sim}{\varepsilon}$ having the probability measure m,

and $G^* = \{\pi^* \mid \pi^*$ is a transformation on S such that $\exists\, \pi \in G$ for which for any $(\underset{\sim}{\alpha},\underset{\sim}{\beta},\underset{\sim}{x},m)$, $\pi^*(\underset{\sim}{\alpha},\underset{\sim}{\beta},\underset{\sim}{x},m) = (\underset{\sim}{\alpha},\underset{\sim}{\beta},\pi(\underset{\sim}{x}),\pi(m))\}$.

Thus M is (G,G^*)-invariant.

It is simple to see that a loss function L is invariant if $L((\underset{\sim}{\alpha},\underset{\sim}{\beta},\pi(\underset{\sim}{x}),\pi(m)),a) = L((\underset{\sim}{\alpha},\underset{\sim}{\beta},\underset{\sim}{x},m),a)$ for each $\pi \in G$, $(\underset{\sim}{\alpha},\underset{\sim}{\beta},\underset{\sim}{x},m) \in S$ and $a \in A$ (where $A = R^p$, as before, if we are in an estimation problem). Thus, when what we are interested in (estimating) depends only on $\underset{\sim}{\alpha}$ or $\underset{\sim}{\beta}$ and is in no way dependent on $\underset{\sim}{x}$ or m, we would use an invariant L. The meaning for a choice of estimator $\underset{\sim}{c}$ to be G-invariant is similar

to that in Example 1. For each $d_o \in \underset{\sim}{D}$, it is clear that $G(d_o)$ is the class of designs which have the same replication number for each block as that of d_o. For any d_o, $\phi_G(d_o)$ has the usual sense of complete randomization within blocks. So by Theorem 2.1 we demonstrate a minimax property for the procedure of complete randomization within blocks. The usual choices of $\underset{\sim}{X}$ are the same as those given in Example 1, except for some flexibility in consideration about the identifiability of the parameters. But sometimes, we may want to take, for example, $\underset{\sim}{X} = \{\underset{\sim}{x} \mid |x_u| \leq k_b$ for $u \in$ b-th block, $b = 1,\ldots,B$ and $u = 1,\ldots,N\}$, where $k_1,\ldots,k_B$ are specified positive numbers.

Example 3. Block design set-up with equal block sizes.

Consider the design set-up as in the Example 2, with $N_1 = \ldots = N_B$. Let G_t' be the group of all permutations on $\{1,\ldots,B\}$. For any $\pi_t' \in G_t'$, let π_t be the permutation on $\{1,\ldots,N\}$ which maps the i-th unit in the b-th block to the i-th unit in the $\pi_t'(b)$-th block. Let G_t be the group of all such π_t. Define G^o to be the group generated by G_t and the group G of all permutations within blocks. Obviously, any $\pi_o \in G^o$ can be decomposed as $\pi_o = \pi_t \cdot \pi$ for a unique $\pi_t \in G_t$ and $\pi \in G$. Now consider the case where $\underset{\sim}{X}$ and $\underset{\sim}{E}$ are invariant under G^o. Then S and M are the same as in Example 2, but $G^{o*} = \{\pi_o^* \mid \pi_o^*$ is a transformation on S such that $\exists\ \pi_o \in G^o$ for which, for any $(\underset{\sim}{\alpha},\underset{\sim}{\beta},\underset{\sim}{x},m)$, $\pi_o^*(\underset{\sim}{\alpha},\underset{\sim}{\beta},\underset{\sim}{x},m) = (\underset{\sim}{\alpha},\pi_t(\underset{\sim}{\beta}),\pi_o(\underset{\sim}{x}),m)\}$. Clearly, M is now (G^o,G^{o*})-invariant and a loss function L is invariant if $L((\underset{\sim}{\alpha},\pi_t(\underset{\sim}{\beta}),\pi(\underset{\sim}{x}),\pi(m)),a) = L((\underset{\sim}{\alpha},\underset{\sim}{\beta},\underset{\sim}{x},m),a)$ for any $\pi \in G^o$. Hence if what we are interested in depends only on $\underset{\sim}{\alpha}$ and in no way on $\underset{\sim}{\beta}$, $\underset{\sim}{x}$ or m, then an invariant L should be imposed. (Note that in Example 2, an

invariant L can depend on $\underset{\sim}{\beta}$ also). For any $d_0 \in \underset{\sim}{D}$, $G^0(d_0)$ is the class of designs which have the same set of replication numbers as that of d_0. Note that $G^0(d_0)$ is larger than the $G(d_0)$ of Example 2 in general. We have no difficulty in seeing that $\phi_{G^0(d_0)}$ is simply the randomized design derived by completely randomizing units within blocks and then completely randomizing the entire blocks. This procedure is minimax in the sense described by Theorem 2.1. The commonly used $\underset{\sim}{X}$'s suggested in Example 1 are invariant under our G^0 and also under the G in Example 2. Thus for these $\underset{\sim}{X}$'s, a confusion may arise about which procedure is appropriate, since Theorem 2.1 guarantees minimaxity for both procedures. But this confusion is easily removed after realizing that the competing classes of procedures are different. Thus the procedure presented in this example is certainly the better for these $\underset{\sim}{X}$'s.

Example 4. Two-way inhomogeneity.

The $N = f_1 f_2$ units are now arranged in an $f_1 \times f_2$ array. Suppose the model is $Y_{ij} = \alpha_{d(i,j)} + \beta_i + \gamma_j + x_{(i,j)} + \varepsilon_{(i,j)}$, $i = 1,\ldots,f_1$ and $j = 1,\ldots,f_2$ where d is the design, α_i is the i-th treatment effect, β_i is the i-th column effect, γ_j is the j-th row effect, $x_{(i,j)}$ is the (i,j)-th unit effect, and ε_{ij}'s are random errors. Let G be the group generated by all row permutations and column permutations. Let the set $\underset{\sim}{X}$ of all possible $\underset{\sim}{x}$ and the set $\underset{\sim}{E}$ of all possible distributions of errors be invariant under G. Then one can construct S, M, G^* without difficulties. Also, one sees that $\phi_G(d_0)$ is derived by the usual procedure of randomizing the rows and columns completely. Thus our theorem gives a minimax property for this procedure.

3 A-Criterion

We now consider the problem of choosing a good randomized strategy. The results in the above section suggest that we only have to find a suitable non-randomized strategy to be G^*-uniformly randomized. But this may be misleading unless we have more assumptions about the set-up with which we are concerned. More precisely, we have the following lemma. (For ease of developments, we defer the proofs for all the assertions made in this section to the next section.)

Lemma 3.1 Assume that the model M is (G,G^*)-invariant and the loss function L is G-invariant. Then we have

$$(7)\qquad \min_{(\mu,\underset{\sim}{c})\in \underset{\sim}{D}\times\underset{\sim}{C}} \max_{s\in G^*(s_0)} Er(\mu,\underset{\sim}{c}(\mu);L,s,M) = \min_{(d,\underset{\sim}{c})\in \underset{\sim}{D}\times\underset{\sim}{C}_G} Er(\underset{\sim}{\phi}_G(d),\underset{\sim}{c}(\underset{\sim}{\phi}_G(d));L,s_0,M)$$

for any $s_0\in S$, and

$$(8)\qquad \min_{(\mu\times\underset{\sim}{c})\in \underset{\sim}{D}\times\underset{\sim}{C}} \max_{s\in S} Er(\mu,\underset{\sim}{c}(\mu);L,s,M) = \max_{s\in S} \min_{(d,\underset{\sim}{c})\in \underset{\sim}{D}\times\underset{\sim}{C}_G} Er(\underset{\sim}{\phi}_G(d),\underset{\sim}{c}(\underset{\sim}{\phi}_G(d));L,s,M) .$$

In fact, (.7) and (8) hold if we replace the "E" by any convex functional f (i.e., $f(Z) \leq \frac{f(X)+f(Y)}{2}$, where Z is the random variable which equals X or Y with probability ½ each). However, the mean functional probably is the only interesting one which is convex. For the median functional, which is not convex, (7) is in general not true. Because of these facts, we are unable to discuss the "*stochastic*" ordering of the risk.

Therefore, the "risk" in this section is in its usual sense. Note that the minimizers for what is on the right-hand side of the equality in (7) may depend on s_o. Thus we cannot change the order of "Max" and "Min" on the right-hand side of (8) without further assumptions. However, for many (but not all) design set-ups, when we are restricted to the use of linear estimators (not necessarily the least squares estimators) and the conventional squared loss functions are considered, there exist equalities similar to (7) and (8) and, moreover, we not only find an explicit expression for the last term in (8), but also are able to change the order of the "Max" and "Min" there, and thus reduce the problem of finding a minimax randomized strategy to the classical problem of finding an optimal design. More precisely, we consider the following set-up which generalizes the examples of the last section.

Suppose T treatments are to be assigned to the $N = \sum_{b=1}^{B} N_b$ units which are classified into B blocks, where N_b is the size of the b-th block. Within the block b, the $N_b = \prod_{i=1}^{n(b)} f_i^{(b)}$ units are arranged according to n(b) factors so that when $n(b) \geq 2$ they form an n(b)-dimensional hyper-rectangle of size $f_1^{(b)} x \ldots x f_{n(b)}^{(b)}$, where $f_i^{(b)}$ is the number of levels of the i-th factor in the b-th block, and when n(b)=1, the $N_b = f_1^{(b)}$ units are assumed to be of the same level. To avoid trivialities, we assume $f_i^{(b)} \geq 2$. The u-th unit, when it falls in the block b, is now labeled by $(i_1^{(b)},\ldots,i_j^{(b)},..,i_{n(b)}^{(b)})$, where $1 \leq i_j^{(b)} \leq f_j^{(b)}$. Thus for $1 \leq u \leq N$, we also write $u = (i_1^{(b)},\ldots,i_{n(b)}^{(b)})$ to denote its label without ambiguity. We consider the following model:

$$Y_u = \alpha_{d(u)} + \beta^{(b)} + \sum_{j=1}^{n(b)} \beta^{(b)}_{(i_1^{(b)},\ldots,\hat{i}_j^{(b)},\ldots,i_{n(b)}^{(b)})} + x_u + \varepsilon_u$$

where $u = (i_1^{(b)},\ldots,i_{n(b)}^{(b)})$, d is the design, α_i is the i-th treatment effect, $\beta^{(b)}$ is the b-th block effect, x_u is the u-th unit effect, ε_u is the random error for the u-th observation, and $\beta^{(b)}_{(i_1^{(b)},\ldots,\hat{i}_j^{(b)},\ldots,i_{n(b)}^{(b)})}$ is the interaction of all factors but the j-th factor in the b-th block with levels $i_1^{(b)},\ldots,i_{n(b)}^{(b)}$. Note that all the lower level interactions of factors in the same block are implicit in this model. This model is reduced to the set-up of Example 1 (2, 4, respectively) when $B = 1$ and $n(1) = 1$ ($n(b)=1$ for $b = 1,\ldots,B$; $B=1$ and $n(1)=2$, respectively). We shall assume that $\underset{\sim}{x}$ belongs to $\underset{\sim}{X}$ and $\underset{\sim}{\varepsilon}$ has mean 0 and a probability measure $m \in \underset{\sim}{E}$, where $\underset{\sim}{X}$ and $\underset{\sim}{E}$ are both invariant under a group G to be specified below.

For any b and j such that $1 \leq b \leq B$ and $1 \leq j \leq n(b)$, suppose $G_j'^{(b)}$ is a doubly transitive group on $\{1,\ldots,f_j^{(b)}\}$; i.e., for any k_1,k_2,k_3 and $k_4 \in \{1,\ldots,f_j^{(b)}\}$ such that $k_1 \neq k_2$ and $k_3 \neq k_4$, there exists some $\pi \in G_j'^{(b)}$ such that $\pi(k_1) = k_3$ and $\pi(k_2) = k_4$. For any $\pi' \in G_j'^{(b)}$, define a permutation π on $\{1,\ldots,N\}$ by letting $\pi(u) = u$ for $u \notin$ the b-th block, and $\pi(u) = \left(i_1^{(b)},\ldots,i_{j-1}^{(b)}, \pi'(i_j^{(b)}), i_{j+1}^{(b)},\ldots,i_{n(b)}^{(b)}\right)$ when $u = (i_1^{(b)},\ldots,i_{n(b)}^{(b)})$. Let G_j^b be the group of all such π derived from any π' in $G_j'^{(b)}$ in the above way. Then, we define

$$(9) \qquad G = \prod_{b=1}^{B} \prod_{j=1}^{n(b)} G_j^b \quad .$$

Suppose we are interested in estimating p contrasts among α_i's, using a squared loss function and restricted to the consideration of

linear estimators only. Let $\underset{\sim}{\beta}(b)$ be the vector of dimension $\sum_{j=1}^{n(b)} \prod_{i\neq j}^{n(b)} f_i^{(b)} \equiv q_b$ with the $\beta^{(b)}_{(i_1^{(b)},\ldots,\hat{i}_j^{(b)},\ldots,i_{n(b)}^{(b)})}$'s as its components. Let $q = \sum_{b=1}^{B} q_b + B$. Thus in terms of the notations in the last section, we may take $S = \{(\underset{\sim}{\alpha},\underset{\sim}{\beta},\underset{\sim}{x},m) | \underset{\sim}{\alpha}\in R^T, \underset{\sim}{\beta} = (\beta^{(1)},\ldots,\beta^{(B)},\beta^t_{(1)},\ldots,\beta^t_{(B)})^t \in R^q, \underset{\sim}{x}\in X$ and $m\in \underset{\sim}{E}\}$, $A = R^p$, and $L((\underset{\sim}{\alpha},\underset{\sim}{\beta},\underset{\sim}{x},m),\underset{\sim}{a}) = (A\underset{\sim}{\alpha} - \underset{\sim}{a})^t(A\underset{\sim}{\alpha} - \underset{\sim}{a})$ for some $p\times T$ matrix A (where $\underset{\sim}{a}\in A$). It is clear what the corresponding G^* should be to make our model (G,G^*)-invariant.

Let $\Delta_o = \{\delta | \delta$ is a linear transformation from the vector space R^N to $R^P\}$. Let $\underset{\sim}{C}^o = \{\underset{\sim}{c} | \underset{\sim}{c}\in \underset{\sim}{C}$ and $\underset{\sim}{c}(d)\in\Delta_o$ for all $d\in \underset{\sim}{D}\}$ and let $\underset{\sim}{C}^o_G = \underset{\sim}{C}^o \cap \underset{\sim}{C}_G$. The results in this section will be established after we derive an explicit form for $Er(\underset{\sim}{\phi}_G(d),\underset{\sim}{c}(\underset{\sim}{\phi}_G(d));L,s,M)$, where $\underset{\sim}{c}\in \underset{\sim}{C}^o_G$. Since any element $\delta\in\Delta_o$ can be represented by a $p\times N$ matrix, we use δ to denote a $p\times N$ matrix also.

Write $\underset{\sim}{\alpha}_d = (\alpha_{d(1)},\ldots,\alpha_{d(N)})^t$. Let $\underset{\sim}{\beta}$ be $E\underset{\sim}{Y}$ under $\underset{\sim}{\alpha} = 0$ and $\underset{\sim}{x} = 0$. For any $\underset{\sim}{x}\in R^N$, we define $\bar{\underset{\sim}{x}} = \frac{1}{g}\sum_{\pi\in G}\pi(\underset{\sim}{x})$, and for any symmetric matrix $\underset{\sim}{V}$ we define $\bar{\underset{\sim}{V}} = \frac{1}{g}\sum_{\pi\in G}\pi(\underset{\sim}{V})$, where $\pi(\underset{\sim}{V}) = (v_{\pi(i)\pi(j)})$ for $\underset{\sim}{V} = (v_{ij})$.

Lemma 3.2 For $\underset{\sim}{c}\in \underset{\sim}{C}^o_G$, we have

$$(10)\qquad Er(\underset{\sim}{\phi}_G(d),\underset{\sim}{c}(\underset{\sim}{\phi}_G(d));L,s,M) = \|A\underset{\sim}{\alpha} - \underset{\sim}{c}(d)[(\underset{\sim}{\alpha}_d + \underset{\sim}{\beta}) + \bar{\underset{\sim}{x}}]\|^2 + \text{trace } \underset{\sim}{c}(d)(\overline{(\underset{\sim}{x}-\bar{\underset{\sim}{x}})(\underset{\sim}{x}-\bar{\underset{\sim}{x}})^t} + \bar{\underset{\sim}{V}})\underset{\sim}{c}(d)^t,$$

where $s = (\underset{\sim}{\alpha},\underset{\sim}{\beta},\underset{\sim}{x},m)$, $\underset{\sim}{V}$ is the covariance matrix of $\underset{\sim}{\varepsilon}$ which has probability measure m, and $\|\cdot\|$ is the Euclidean norm. We refer to the first term on the right-hand side of the equation in (10) as the squared bias term and the second term as the variance term.

For any permutation group G', let $G'(u)$ be the orbit of u under G', i.e., $G'(u) = \{u' | u' = \pi(u)$ for some $\pi \in G'\}$. For any subset H of $\{1,\ldots,N\}$, define 1_H to be the vector in R^N whose i-th coordinate is 1 if $i \in H$ and is 0 otherwise. Let $W = \{\underset{\sim}{v} | \underset{\sim}{v} \in R^N, \underset{\sim}{v}^t 1_{G_j^b(u)} = 0$ for any $b \in \{1,\ldots,B\}$, $j \in \{1,\ldots,n(b)\}$ and $u \in$ block $b\}$. Notice that $1_{G_j^b(u)}$ is just the expectation of $\underset{\sim}{Y}$ when all parameters (i.e., $\underset{\sim}{\alpha}, \underset{\sim}{\beta}, \underset{\sim}{x}$) are set equal to zero except that $\beta^{(b)}_{(i_1,\ldots,\hat{i}_j,\ldots,i_{n(b)})}$ (for the case $n(b) \geq 2$) or $\beta^{(b)}$ (for the case $n(b) = 1$) is set equal to 1.

Lemma 3.3 A necessary condition for $\underset{s \in S}{\text{Max}}\, Er(\underset{\sim}{\phi}_G(d), \underset{\sim}{c}(\underset{\sim}{\phi}_G(d)); L, s, M)$ to be finite is that the column vectors of $\underset{\sim}{c}(d)^t$ belong to W and $\underset{\sim}{c}(d)\underset{\sim}{\alpha}_d = A\underset{\sim}{\alpha}$ for all $\underset{\sim}{\alpha} \in R^T$.

Let $\underset{\sim}{D}^*$ be the class of designs for which $A\underset{\sim}{\alpha}$ is estimable. Let $\underset{\sim}{C}_G^* = \{\underset{\sim}{c} | \underset{\sim}{c} \in \underset{\sim}{C}_G^o$ and $\underset{\sim}{c}(d)\underset{\sim}{\alpha}_d = A\underset{\sim}{\alpha}$ for all $\underset{\sim}{\alpha} \in R^T$ and $d \in \underset{\sim}{D}^*\}$. Now, for each $d \in \underset{\sim}{D}^*$, the expected risk in Lemma 3.2 is now simplified a little bit since the squared bias term is reduced to $\|\underset{\sim}{c}(d)\underset{\sim}{\bar{x}}\|^2$ only. The lemma below will show that this term is 0 and also give an explicit expression for the variance term.

The following notations are introduced for further development. For each $b \in \{1,\ldots,B\}$, let $\Lambda^b = \{\gamma | \gamma \subset \{1,\ldots,n(b)\}$ and $\gamma \neq \phi\}$ and $\Lambda_k^b = \{\gamma | \gamma \in \Lambda^b$ and $\#\gamma = k\}$. Write $G_\gamma^b = \sum_{j \in \gamma} G_j^b$, for any $\gamma \in \Lambda^b$ and $G\phi = \{I\}$ where I is the identity permutation. Given $\underset{\sim}{x} \in R^N$, define $\underset{\sim}{x}^\gamma = \frac{(-1)^k}{\#G_\gamma^b} \cdot \sum_{\pi \in G_\gamma^b} \pi(\underset{\sim}{x})$ if $\gamma \in \Lambda_k^b$. Define $\underset{\sim}{x}_{[b]} \in R^N$ to be the projection of $\underset{\sim}{x} \in R^N$ on the block b, (i.e., the u-th component of $\underset{\sim}{x}_{[b]}$ is equal to that of $\underset{\sim}{x}$ or 0, depending on whether u falls in the b-th block or not).

Define $\tilde{\underset{\sim}{X}}_{[b]} = \sum_{\gamma \in \Lambda^b} \overline{\underset{\sim}{X}_{[b]}}^{\gamma}$ (note that it is clear that $\overline{\underset{\sim}{X}_{[b]}}^{\gamma} = (\overline{\underset{\sim}{X}}^{\gamma})_{[b]}$, and thus no ambiguity will arise in our definition).

For any $N \times N$ matrix $\underset{\sim}{V}$ ($\underset{\sim}{v}_u$ is the u-th column vector of $\underset{\sim}{V}$), let $\underset{\sim}{V}^*_{[b]}$ be the $N \times N$ matrix whose u-th column vector is $\underset{\sim}{v}_{u[b]} + \tilde{\underset{\sim}{v}}_{u[b]}$. Define $\underset{\sim}{V}_{[b]} = ((\underset{\sim}{V}^*_{[b]})^t)^*_{[b]}$.

Lemma 3.4 For any $\underset{\sim}{v} \in W$, $\underset{\sim}{x} \in R^N$, and any symmetric $N \times N$ matrix $\underset{\sim}{V}$ we have $\underset{\sim}{v}^t \tilde{\underset{\sim}{x}} = 0$,

$$(11) \qquad \underset{\sim}{v}^t \overline{\underset{\sim}{x}\underset{\sim}{x}^t} \underset{\sim}{v} = \sum_{b=1}^{B} C_b \| \underset{\sim}{x}_{[b]} + \tilde{\underset{\sim}{x}}_{[b]} \|^2 \, \underset{\sim}{v}^t_{[b]} \underset{\sim}{v}_{[b]} ,$$

and

$$(12) \qquad \underset{\sim}{v}^t \bar{\underset{\sim}{V}} \underset{\sim}{v} = \sum_{b=1}^{B} C_b \, \text{trace}\,(\underset{\sim}{V}_{[b]}) \underset{\sim}{v}^t_{[b]} \underset{\sim}{v}_{[b]} ,$$

where C_b is defined by $C_b = [1 + \sum_{\gamma \in \Lambda^b} \prod_{j \in \gamma} (f_j^{(b)} - 1)^{-1}]$.

Combining the above results, we establish the following theorem, in which $\underset{\sim}{M}$ is the covariance matrix for the probability measure m.

Theorem 3.1 Suppose that G is of the form (9),

$$(13) \qquad \underset{(\underset{\sim}{x},m) \in \underset{\sim}{X} \times \underset{\sim}{E}}{\text{Max}} \; C_b \cdot \{ \| \underset{\sim}{x}_{[b]} + \tilde{\underset{\sim}{x}}_{[b]} \|^2 + \text{trace}(\underset{\sim}{M}_{[b]}) \} \text{ is the same for } b = 1, \ldots, B ,$$

and the maxima are achieved for each b by some $(\underset{\sim}{x}^*, m^*) \in \underset{\sim}{X} \times \underset{\sim}{E}$. Then,

$$(14) \qquad \underset{(\mu \times \underset{\sim}{c}) \in \underset{\sim}{\mathcal{D}} \times \underset{\sim}{C}^o}{\text{Min}} \; \underset{s \in S}{\text{Max}} \; Er(\mu, \underset{\sim}{c}(\mu); L, s, M) = \underset{s \in S}{\text{Max}} \; Er(\underset{\sim}{\phi}_G(d^o), \underset{\sim}{c}^o(\underset{\sim}{\phi}_G(d^o)); L, s, M)$$

where $\underset{\sim}{c}^o$ is the choice of least squares estimators, i.e., $\underset{\sim}{C}^o(d)Y$ equals

the least squares estimator for $A\underset{\sim}{\alpha}$ when $A\underset{\sim}{\alpha}$ is estimable under d, and equals any linear estimators when $A\underset{\sim}{\alpha}$ is not estimable under d, and d^0 is the solution of the following problem:

$$(14) \qquad \min_{d \in \underset{\sim}{D}^*} \operatorname{trace}(\underset{\sim}{c}^0(d)^t \underset{\sim}{c}^0(d)) \quad .$$

Note that $(d^0, \underset{\sim}{c}^0(d^0))$ is an optimal strategy when no contamination is allowed and the errors are homoscedastic and uncorrelated. Thus the minimax solutions for our problem depend on the solutions of the corresponding optimal design problems which have been studied in a wide literature (e.g., [8],[9],[2]). Once we find an optimal design d^0, we G-uniformly randomize d^0 and use the least squares estimators. This will achieve the best we can do according to our theorem. Now we assume that all main contrasts among the treatment effects are of equal interest. Thus, each main contrast among the treatment effects appear exactly once in some coordinate of $A\underset{\sim}{\alpha}$ and $P = \binom{T}{2}$. Any d^0 minimizing (14) has been called an A-optimal design in most literature. When B=1, it is clear that we do not even have to worry about the existence of $(\underset{\sim}{x}^*, \underset{\sim}{m}^*)$. Thus, for the set-up of Example 1, an A-optimal design, i.e., the balanced completely randomized design, together with the least squares estimator, is a minimax randomized strategy. For the two-way inhomogeneity set-up of Example 4, G-uniformly randomizing a generalized Youden design (Kiefer [1975]), which is A-optimal, and using the least squares estimators, is a minimax randomized strategy. For the set-up of Example 2, $n(b) = 1$ for each b, but $B \neq 1$, we need to show (13) holds and must show the existence of $(\underset{\sim}{x}^*, m^*)$ in order to apply Theorem 3.1. This is obviously the case for the $\underset{\sim}{X}$'s suggested in Example 3.1 when the errors are homosce-

dastic and uncorrelated. However, there are cases where (13) holds, but $(\underset{\sim}{x}^*,m^*)$ does not exist; e.g., for the case where N_b's are the same, we may take, for a fixed positive number k_o,

$$\underset{\sim}{X} = \bigcup_{b=1}^{B} \{\underset{\sim}{x} \Big| |x_u| \leq k_o \text{ for all } u \text{ in the b-th block and } x_u = 0 \text{ elsewhere}\}$$

and

$$\underset{\sim}{E} = \{m \mid m \text{ is a probabability measure on } R^N \text{ with mean 0 and covariance matrix } \sigma^2 I_{N\times N} \text{ for some } \sigma \in R\}.$$

In this case, the randomization within blocks only evidently is not enough. We must look for some other group, larger than the one given there, to do the randomization. One possible choice is the group given in Example 3, which is not of the form (9). This motivates the following consideration.

Assume that $n(1) = \ldots = n(B) \equiv n$ and $f_j^{(1)} = \ldots = f_j^{(B)} \equiv f_j$ for $j \in \{1,\ldots,n\}$. Let G_t' be a *transitive* group on $\{1,\ldots,B\}$ in the sense that for any $k_1,k_2 \in \{1,\ldots,B\}$, there exists some $\pi \in G_t'$ such that $\pi(k_1) = k_2$. For any $\pi' \in G_t'$, define a permutation π on $\{1,\ldots,N\}$ by $\pi(i_1^{(b)},\ldots,i_n^{(b)}) = (i_1^{(\pi'(b))},\ldots,i_n^{(\pi'(b))})$. Let G_t be the group of all such π. Now define a group G^o by

(15) G^o is the group of permutations generated by G_t and the G defined by (9).

Note that G is a normal subgroup of G^o but G^o is not a direct product of G_t and G because G_t and G do not commute. Thus the G^o in (15) is not in the form of (9). Now suppose $\underset{\sim}{X}$ and $\underset{\sim}{E}$ are invariant

under G^0. It is not hard to construct a suitable G^{0*} to make our model (G^0, G^{0*}) invariant. We have the following analogue of Theorem 3.1.

Theorem 3.2 For the G^0 defined by (15), we have

$$(16)\quad \min_{(\mu\times \underset{\sim}{C})\in \underset{\sim}{\mathcal{D}}\times \underset{\sim}{C}^0} \max_{s\in S} Er(\mu, \underset{\sim}{c}(\mu); L, s, M) = \max_{s\in S} Er(\underset{\sim}{\phi}_{G^0}(d^0), \underset{\sim}{c}^0(\underset{\sim}{\phi}_{G^0}(d^0)); L, s, M)$$

where d^0 and c^0 are defined as in Theorem 3.1.

Taking n=1, we solve the case in Example 3. Thus, for instance, a G^0-uniformly randomized BIBD (if it exists) together with the least squares estimator is a minimax randomized strategy (under restriction to linear estimators).

There are some interesting consequences for the two theorems in this section. Most often, as in the above examples, we take $G_j'^{(b)}$ (G_t' respectively) to be the group of *all* permutation on $\{1,\ldots,f_j^{(b)}\}$ ($\{1,\ldots,B\}$ respectively) to construct the group G or G^0. (We call any minimax solution under this choice of G or G^0 a *completely* randomized minimax strategy). However, we may take $G_j'^{(b)}$ (G_t' respectively) to be any doubly transitive group (transitive group, respectively). (Notice that if the model is invariant under G then it is invariant under any subgroup of G.) As a consequence, we can reduce tremendously the order of the group G (or G^0) to be used to randomize an A-optimal design and still obtain a minimax randomized strategy with the same efficiency as the completely randomized one. Evidently a cyclic group on $\{1,\ldots,B\}$ is a transitive group with the smallest order. The construction of a doubly transitive group on a set with n labels has been studied in the literature of the group theory. For the case when n is a prime or a

prime power, simple ways of construction have appeared in Burnside [1911]. The case $n=p$, a prime, is particularly simple. The group $\{\pi_1^i \pi_2^j \mid i=1,\ldots,P$ and $j=1,\ldots,P-1\}$ (where π_1 is the permutation $(1,2,\ldots,n)$, meaning $\pi_1(i) = i+1 (\text{mod } p)$ as usual, and $\pi_2 = (2,\alpha+1,\alpha^2+1,\ldots,\alpha^{p-2}+1)$ for any primitive root α of p) is doubly transitive and has the order only of $p(p-1)$, which is the smallest one can have. It is natural to ask whether it is possible to weaken the assumptions in Theorem 3.1 so that it admits the case when $G_j'^{(b)}$'s are *transitive* only. But unfortunately, this way may be impossible for most cases. For instance, in Example .1, take N=4. Then for a cyclic group, $\overline{\underset{\sim}{x}\underset{\sim}{x}^t}$ is of the form $\begin{pmatrix} a & b & c & b \\ b & a & b & c \\ c & b & a & b \\ b & c & b & a \end{pmatrix}$ and thus (. 11), which is crucial in proving Theorem 3.1, does not hold for some $\underset{\sim}{v}$ such that $\underset{\sim}{v}^t\underset{\sim}{1}=0$ (where $\underset{\sim}{1}=(1\ 1\ 1\ 1)^t$). Thus, the roles of randomization within blocks and randomization between blocks are quite different (this fact never seems to be mentioned in the existing literatures). For the block design case, a clear interpretation, based on a method of error decomposition, about why (11) and (12) hold, is given in Cheng and Li [1980].

Remark 1. For the set-up with which Theorem 3.2 is concerned, it is important to notice that factors from different blocks are different. Theorem 3.2 does not cover the case where only one block is available (with $n \geq 2$) but we are allowed to make B replications on the same block (residual effects from replicating are not assumed), since we now have the restrictions on the parameter space that $\beta^{(1)}_{(i_1^{(1)},\ldots,i_j^{(1)},\ldots,i^{(1)})} = \cdots = \beta^{(B)}_{(i_1^{(B)},\ldots,i_j^{(B)},\ldots,i_n^{(B)})}$ for all possible indices, and there is no suitable

G^{0*} to make the model (G^0,G^{0*})-invariant. Another reason why Theorem 3.2 is not applicable here is that the W for which Lemma 3.3 holds in the replication case is larger than the one defined there above Lemma 3.3, and Lemma 3.4 does not seem to hold for such a large W. Besides, even for an ordinary uncontaminated model, the literatures on optimal designs still did not solve the design problems where replications are concerned and $n \geqslant 2$.

Remark 2. Except for the cases demonstrated in the examples of the last section, there is still no optimal solution for (14) available, even for the A-criterion case. In the following discussion, we assume $B = 1$ and $n(1) \geqslant 3$. Cheng [1978b, 1980] showed that any Youden hyperrectangle is A-optimal for the uncontaminated additive model which assumes only the main factor effects are present. But not every Youden hyperrectangle is equally good when higher order interactions are existent (for some of them, some treatment contrasts are not even estimable). Thus, for the interaction model, one might provide a new class of designs which possess more properties of balance than the Youden hyperrectangles in order to achieve A-optimality. However, no solution is found up to the present. On the other hand, we may want to reduce our model so that only the main effects are present and thus, hopefully, might use Cheng's results to conclude that G-uniformly randomizing any Youden hyperrectangle is a minimax randomized strategy. But, unfortunately, this is not true in general. The model is (G,G*)-invariant for an obvious G*. However, the W for the reduced additive model is too large to work in Lemma 3.3.4. The following example shows that for the reduced model, G-uniformly

randomizing one Youden hyperrectangle is better than G-uniformly randomizing another Youden hyperrectangle. (Thus there is no general theorem analogous to Theorem 3.1 or Theorem 3.2 in the reduced model.) Take T=2, B=1, n(1) = 3, and $f_1^{(1)} = f_2^{(1)} = f_3^{(1)} = 2$. Thus each unit u is labeled by (i,j,k) where i,j,k = 1 or 2. Let $\underset{\sim}{Z}_1 = \{(1,1,1),(2,2,1),(1,1,2),(2,2,2)\}$ and $\underset{\sim}{Z}_2 = \{(1,1,1),(2,2,1),(1,2,2),(2,1,2)\}$. Consider the designs d_1 and d_2, where $d_1^{-1}(\{1\}) = \underset{\sim}{Z}_1$ and $d_2^{-1}(\{1\}) = \underset{\sim}{Z}_2$. It is clear that both designs are Youden hypercubes. Thus both are A-optimal for the uncontaminated model. Although d_2 is intuitively better than d_1 (since it possesses better symmetry properties), it is not necessarily so. This is demonstrated by the following consideration. First, let $\underset{\sim}{1} = (1,\ldots,1)^t \in R^8$ and $\underset{\sim}{x}$ be in R^8 such that $x_u = 1$ for $u \in \underset{\sim}{Z}$ and 0 otherwise. Let $\underset{\sim}{X} = \{\underset{\sim}{x}, \underset{\sim}{1}-\underset{\sim}{x}\}$. Suppose $\underset{\sim}{\varepsilon}=0$. Take $G_1'^{(1)} = G_2'^{(1)} = G_3'^{(1)} =$ the group of all permutations on {1,2}. Now $\underset{\sim}{X}$ is invariant under G. Suppose we are interested in estimating the main treatment contrast. For d_1, any linear estimator has the maximal mean squared error either $+\infty$ or 1, while for d_2, the maximal mean squared error is 0 for some obvious linear estimators. We further observe that for any $\pi \in G$, we get either $\pi(d_1) = d_1$ or $\pi(d_1)$ = the design derived from d_1 by exchanging the labels of the two treatments, and similarly for d_2. So randomization does not change the minimax mean squared error here. Thus, G-uniformly randomizing d_2 provides a randomized strategy better than G-uniformly randomizing d_1 does in this case, regardless of the fact that both of them are A-optimal. However, the following example demonstrates that d_1 is better than d_2 in some cases.

Define $\underset{\sim}{x}$ to be the vector in R^8 such that $x_u = 1$, for

$u \in \{(1,1,1),(2,2,1)\}$, -1, for $u \in \{(1,1,2),(2,2,2)\}$, and 0 otherwise. Define $\underset{\sim}{x}'$ to be the vector in R^8 such that $x'_u = 1$, for $u \in \{(1,2,1),(2,1,1)\}$, -1, for $u \in \{(1,2,2),(2,1,2)\}$, and 0 otherwise. Take $\underset{\sim}{X} = \{\underset{\sim}{x}, -\underset{\sim}{x}, \underset{\sim}{x}', -\underset{\sim}{x}'\}$ and $\underset{\sim}{\varepsilon} = 0$. The group G is the same as before. Clearly, $\underset{\sim}{X}$ is invariant under G. Again, we are interested in estimating the main treatment contrast. Now d_1 is better than d_2, since for d_1, the minimax mean squared error is 0, but it is 1 for d_2. Thus we conclude that G-uniformly randomizing d_1 provides a randomized strategy better than G-uniformly randomizing d_2 does for this case. (But for the interaction model, d_1 is useless since the main treatment contrast will not be estimable under d_1.) This investigation shows that minimax randomized strategies for the additive model with contamination depend on what $\underset{\sim}{X}$ is, under *three- or more-way* heterogeneity elimination set-ups. This phenomenon does not happen when $n(1) \leqslant 2$, as is proved by Theorem 3.1.

4 Proofs

We prove the assertions claimed in the last section. We omit the L and M below when it causes no ambiguity.

Proof of Lemma 3.1 We have

$$\begin{aligned} &\max_{s \in G^*(s_0)} Er(\underset{\sim}{\mu}, \underset{\sim}{c}(\underset{\sim}{\mu}); s) \\ &= \max_{\pi \in G} Er(\underset{\sim}{\mu}, \underset{\sim}{c}(\underset{\sim}{\mu}); \pi^*(s_0)) \\ &\geqslant \frac{1}{g} \sum_{\pi \in G} Er(\underset{\sim}{\mu}, \underset{\sim}{c}(\underset{\sim}{\mu}); \pi^*(s_0)) \end{aligned}$$

$$= E\,\frac{1}{g}\sum_{\pi\in G} r(\pi^{-1}(\underset{\sim}{\mu}),\pi^{-1}(\underset{\sim}{c}(\underset{\sim}{\mu}));s_o) \qquad \text{(by (2'))}$$

$$= \sum_{d\in \underset{\sim}{D}} \mu(d)\cdot\frac{1}{g}\sum_{\pi\in G} r(\pi(d),\pi(\underset{\sim}{c}(d));s_o)$$

$$= \sum_{d\in \underset{\sim}{D}} \mu(d)\ Er(\underset{\sim}{\phi}_G(d),\underset{\sim}{c}_d(\underset{\sim}{\phi}_G(d));s_o)$$

(where $\underset{\sim}{c}_d$ is any element in $\underset{\sim}{C}_G$ such that $\underset{\sim}{c}_d(d) = \underset{\sim}{c}(d)$)

$$\geqslant \min_{(d,\underset{\sim}{c})\in \underset{\sim}{D}\times\underset{\sim}{C}_G} Er(\underset{\sim}{\phi}_G(d),\underset{\sim}{c}(\underset{\sim}{\phi}_G(d));s_o)\ .$$

Thus (.7) is proved since the other direction is obvious. Also, (8) follows from (7). □

Proof of Lemma 3.2 We have

$$Er(\underset{\sim}{\phi}_G(d),\underset{\sim}{c}(\underset{\sim}{\phi}_G(d));s) = \frac{1}{g}\sum_{\pi\in G} r(\pi^{-1}(d),\underset{\sim}{c}(\pi^{-1}(d));s)$$

$$= \frac{1}{g}\sum_{\pi\in G} r(d,\underset{\sim}{c}(d);\pi^*(s))$$

(by Definition 2.3 and (2'))

$$= \frac{1}{g}\sum_{\pi\in G}\Big\{\|A\underset{\sim}{\alpha} - \underset{\sim}{c}(d)(\underset{\sim}{\alpha}_d + \underset{\sim}{B} + \pi(\underset{\sim}{x}))\|^2 + \text{trace}\ \underset{\sim}{c}(d)\pi(\underset{\sim}{V})\underset{\sim}{c}(d)^t\Big\}$$

$$= E\,\|A\underset{\sim}{\alpha} - \underset{\sim}{c}(d)(\underset{\sim}{\alpha}_d + \underset{\sim}{B} + \underset{\sim}{v}_o(\underset{\sim}{x}))\|^2 + \text{trace}\ \underset{\sim}{c}(d)\bar{\underset{\sim}{V}}\ \underset{\sim}{c}(d)^t$$

(v_o is the uniform measure on G, as before)

$$= \|A\underset{\sim}{\alpha} - \underset{\sim}{c}(d)(\underset{\sim}{\alpha}_d + \underset{\sim}{\beta} + \bar{\underset{\sim}{x}})\|^2 + \text{trace } \underset{\sim}{c}(d)\overline{(\underset{\sim}{x}-\bar{\underset{\sim}{x}})(\underset{\sim}{x}-\bar{\underset{\sim}{x}})^t}\,\underset{\sim}{c}(d)^t + \text{trace } \underset{\sim}{c}(d)\bar{\underset{\sim}{V}}\,\underset{\sim}{c}(d)^t \,. \qquad \square$$

Proof of Lemma 3.3 Suppose $\underset{s\in S}{\text{Max}}\, Er(\underset{\sim}{\phi}_G(d), \underset{\sim}{c}(\underset{\sim}{\phi}_G(d)); s) < \infty$.

First, take $s = (\underset{\sim}{\alpha}, \underset{\sim}{0}, \underset{\sim}{0}, m)$. By Lemma 3.2, we have

$$Er(\underset{\sim}{\phi}_G(d), \underset{\sim}{c}(\underset{\sim}{\phi}_G(d)); s) = \|A\underset{\sim}{\alpha} - \underset{\sim}{c}(d)\underset{\sim}{\alpha}_d\|^2 + \text{trace } \underset{\sim}{c}(d)\bar{\underset{\sim}{V}}\,\underset{\sim}{c}(d)^t$$

$$\geqslant \|A\underset{\sim}{\alpha} - \underset{\sim}{c}(d)\underset{\sim}{\alpha}_d\|^2 \,.$$

Thus it is clear that $A\underset{\sim}{\alpha} = \underset{\sim}{c}(d)\underset{\sim}{\alpha}_d$ for all $\underset{\sim}{\alpha} \in R^T$. Next, for any $b \in \{1,\ldots,B\}$, $j \in \{1,\ldots,n(b)\}$ and $u \in$ the b-th block, we want to show $\underset{\sim}{c}(d)1_{G_j^b(u)} = 0$. When $n(b) = 1$, we have $j=1$, $G_1^b(u)$ = the b-th block. Then, taking $s = (\underset{\sim}{0}, \underset{\sim}{\beta}, \underset{\sim}{0}, m)$ where $\beta^{(b)} \in R$ and all the other coordinates of $\underset{\sim}{\beta}$ are 0, and applying Lemma 3.2, we get $\underset{s\in S}{\text{Max}}\, Er(\underset{\sim}{\phi}_G(d), \underset{\sim}{c}(\underset{\sim}{\phi}_G(d)); s) \geqslant (\beta^{(b)})^2 \cdot \|\underset{\sim}{c}(d) \cdot 1_{G_1^b(u)}\|^2$. Hence it is clear that $\underset{\sim}{c}(d)1_{G_1^b(u)} = 0$. When $n(b) \geqslant 2$, $G_j^b(u) = \{u' \mid u'$ is in the b-th block and has the same level for each factor as u except for the j-th factor$\}$. Thus, taking $s = (\underset{\sim}{0}, \underset{\sim}{\beta}, \underset{\sim}{0}, m)$ where $\beta^{(b)}_{(i_1^{(b)},\ldots,i_j^{(b)},\ldots,i_{n(b)}^{(b)})} \in R$ and all the other coordinates of $\underset{\sim}{\beta}$ are 0, and using similar arguments, we get $\underset{\sim}{c}(d)1_{G_j^b(u)} = 0$. $\square$

From now on, for any permutation π on $\{1,\ldots,N\}$, write $\underset{\sim}{\pi}$ to denote its corresponding permutation matrix. Thus $\pi(\underset{\sim}{x}) = \underset{\sim}{\pi}\underset{\sim}{x}$, and let P_b be the projection matrix such that $P_b\underset{\sim}{x} = \underset{\sim}{x}_{[b]}$.

Proof of Lemma 3.4 We show (11) holds first, by establishing a sequence of lemmas, as follows.

Lemma 4.1 For $\underset{\sim}{v} \in W$, we have

$$\underset{\sim}{v}^t \overline{\underset{\sim\sim}{x}\underset{\sim\sim}{x}^t} \underset{\sim}{v} = \sum_{b=1}^{B} \underset{\sim[b]}{v}^t \overline{\underset{\sim[b]}{x}\underset{\sim[b]}{x}^t} \underset{\sim[b]}{v} .$$

Proof. It is clear that P_b commutes with any permutation matrix on R^N induced from any permutation in G. Thus, we have $\overline{\underset{\sim[b]}{x}\underset{\sim[b]}{x}^t} = P_b \overline{\underset{\sim\sim}{x}\underset{\sim\sim}{x}^t} P_b^t$. Therefore, by the definition of W, it suffices to show that for any u_o and u in different blocks b_o and b, the (u_o,u)-th cell of $\overline{\underset{\sim\sim}{x}\underset{\sim\sim}{x}^t}$ is equal to the $(u_o,\pi(u))$-th cell, where $\pi \in G_j^b$; for any $1 \leq j \leq n(b)$. Since $\underset{\sim}{\pi}\, \overline{\underset{\sim\sim}{x}\underset{\sim\sim}{x}^t}\, \underset{\sim}{\pi}^t = \pi(\frac{1}{g} \sum_{\pi' \in G} \underset{\sim}{\pi'}\, \underset{\sim\sim}{x}\underset{\sim\sim}{x}^t\, \underset{\sim}{\pi'}^t)\underset{\sim}{\pi}^t = \frac{1}{g} \sum_{\pi'' \in G} \underset{\sim}{\pi''}\, \underset{\sim\sim}{x}\underset{\sim\sim}{x}^t\, \underset{\sim}{\pi''}^t = \overline{\underset{\sim\sim}{x}\underset{\sim\sim}{x}^t}$, by comparing the (u_o,u)-th cells of the matrices on both ends of the above equalities, we get the desired result. □

By the above lemma, in order to prove Lemma 3.4, we may assume that $B = 1$ without loss of generality. We delete the block label b from all our notations for simplicity (e.g., $G_j = G_j^b$, $\Lambda = \Lambda^b$, $\underset{\sim}{\tilde{x}} = \underset{\sim[b]}{\tilde{x}}$, $n(b) = n$, etc.).

Lemma 4.2 $\forall\, \underset{\sim}{x}, \underset{\sim}{x}' \in R^N$ and $r \in \Lambda$, we have $\overline{\underset{\sim}{x} + \underset{\sim}{x}'}^r = \underset{\sim}{\bar{x}}^r + \underset{\sim}{\bar{x}}'^r$.

Proof. This is obvious, since each permutation is a linear mapping. □

Lemma 4.3 $\forall\, \gamma \in \Lambda$ and $j \in \{1,\ldots,n\}$, we have $\overline{\underset{\sim}{\bar{x}}^{\gamma}}^{\{j\}} = \underset{\sim}{\bar{x}}^{\gamma \cup \{j\}}$ if $j \notin \gamma$,
$= -\underset{\sim}{\bar{x}}^{\gamma}$ if $j \in \gamma$.

Proof. Straightforward. □

Lemma 4.4 For any $\gamma \in \Lambda$, we have $\overline{\underset{\sim}{x} + \underset{\sim}{x}^{\gamma}} = 0$.

Proof. We shall prove this lemma by mathematical induction on the cardinality of γ. When $\gamma \in \Lambda_1$, we take $\gamma = \{1\}$ without loss of generality. Then,

$$\underset{\sim}{\bar{x}}^{\{1\}} = \sum_{\gamma \in \Lambda} \underset{\sim}{\bar{x}}^{\gamma}(1) = \sum_{1 \in \gamma \in \Lambda} \underset{\sim}{\bar{x}}^{\gamma}(1) + \sum_{1 \notin \gamma \in \Lambda} \underset{\sim}{\bar{x}}^{\gamma}(1) = -\sum_{1 \in \gamma \in \Lambda} \underset{\sim}{\bar{x}}^{\gamma}$$

$$+ \sum_{1 \notin \gamma \in \Lambda} \underset{\sim}{\bar{x}}^{\gamma \cup \{1\}} = -\underset{\sim}{\bar{x}}^{\{1\}} - \sum_{1 \notin \gamma \in \Lambda} \underset{\sim}{\bar{x}}^{\gamma \cup \{1\}} + \sum_{1 \notin \gamma \in \Lambda} \underset{\sim}{\bar{x}}^{\gamma \cup \{1\}}$$

$$= -\underset{\sim}{\bar{x}}^{\{1\}} .$$

The first equality holds by Lemma 4.2, and the third by Lemma 4.3. Thus the case when the cardinality of γ is 1 is proved.

Suppose for any $\gamma \in \Lambda_k$ our lemma holds. Given some $\gamma \in \Lambda_k$ and $j \in \{1,\ldots,n\} - \gamma$, we now show our lemma holds for $\gamma \cup \{j\}$. By Lemma 4.3, we have $\underset{\sim}{\bar{x}}^{\gamma \cup \{j\}} = \underset{\sim}{\bar{\bar{x}}}^{\{j\}} = -\underset{\sim}{\bar{x}}^{\gamma}(j) = -\underset{\sim}{\bar{x}}^{\gamma \cup \{j\}}$. Note that the second equality follows by our induction hypothesis. The proof is thus complete. □

Lemma 4.5 For any $\underset{\sim}{v} \in W$, We have $\underset{\sim}{v}^t \overline{\underset{\sim\sim}{x}\underset{\sim\sim}{x}^t} \underset{\sim}{v} = \underset{\sim}{v}^t \overline{(\underset{\sim}{x}+\tilde{\underset{\sim}{x}})(\underset{\sim}{x}+\tilde{\underset{\sim}{x}})^t} \underset{\sim}{v}$.

Proof. By the same reasons as in the proof of Lemma 4.2, we have $\underset{\sim}{v}^t \overline{(\underset{\sim}{x}+\tilde{\underset{\sim}{x}})(\underset{\sim}{x}+\tilde{\underset{\sim}{x}})^t} \underset{\sim}{v} = \underset{\sim}{v}^t \overline{\underset{\sim\sim}{x}\underset{\sim\sim}{x}^t} \underset{\sim}{v} + \underset{\sim}{v}^t \overline{\underset{\sim\sim}{x}\tilde{\underset{\sim\sim}{x}}^t} \underset{\sim}{v} + \underset{\sim}{v}^t \overline{\tilde{\underset{\sim\sim}{x}}\underset{\sim\sim}{x}^t} \underset{\sim}{v} + \underset{\sim}{v}^t \overline{\tilde{\underset{\sim\sim}{x}}\tilde{\underset{\sim\sim}{x}}^t} \underset{\sim}{v}$. Thus, it suffices to show that $\underset{\sim}{v}^t \underset{\sim\sim}{\pi} \tilde{\underset{\sim}{x}} = 0$ for all $\pi \in G$, since this implies $\underset{\sim}{v}^t \overline{\tilde{\underset{\sim\sim}{x}}\underset{\sim\sim}{x}^t} \underset{\sim}{v} = \frac{1}{g} \sum_{\pi \in G} \underset{\sim}{v}^t \underset{\sim}{\pi} \tilde{\underset{\sim\sim}{x}} \underset{\sim\sim}{x}^t \underset{\sim}{\pi}^t \underset{\sim}{v} = 0$ and, similarly, $\underset{\sim}{v}^t \overline{\underset{\sim\sim}{x}\tilde{\underset{\sim\sim}{x}}^t} \underset{\sim}{v} = 0 = \underset{\sim}{v}^t \overline{\tilde{\underset{\sim\sim}{x}}\tilde{\underset{\sim\sim}{x}}^t} \underset{\sim}{v}$. Since it is clear that W is invariant under any $\pi \in G$, we need only to show that $\underset{\sim}{v}^t \underset{\sim}{\bar{x}}^{\gamma} = 0$ for all $\underset{\sim}{v} \in W$ and any $\gamma \in \Lambda$. When $\gamma = \{j\}$, it is quite

easy to see that $\bar{\underset{\sim}{x}}^{\{j\}}$ is some linear combination of $1_{G_j(u)}$'s and thus $\underset{\sim}{v}^t \bar{\underset{\sim}{x}}^{\{j\}} = 0$ by the definition of W. When $\gamma = \{j\} \cup \gamma'$ where $j \notin \gamma'$ and $\gamma' \in \Lambda$, by Lemma 4.3 we have $\underset{\sim}{v}^t \bar{\underset{\sim}{x}}^{\gamma} = \underset{\sim}{v}^t \overline{\bar{\underset{\sim}{x}}^{\gamma'}}^{\{j\}} = 0$. The last equality is seen to hold if one views $\bar{\underset{\sim}{x}}^{\gamma'}$ as $\underset{\sim}{x}$ and applies the result for the case $\gamma = \{j\}$ which we have just established. □

By Lemma 4.5 and Lemma 4.4, in order to establish (11), we need only consider the case where $\bar{\underset{\sim}{x}}^{\gamma} = 0$ for all $\gamma \in \Lambda$ and show that $\underset{\sim}{v}^t \overline{\underset{\sim}{x}\underset{\sim}{x}^t} \underset{\sim}{v} = c\|\underset{\sim}{x}\|^2 \|\underset{\sim}{v}\|^2$ where $c = 1 + \sum_{\gamma \in \Lambda} \prod_{j \in \gamma} (f_j - 1)^{-1}$. Thus, fixing a $u = (i_1, \ldots, i_n)$, where $i_j \in \{1, \ldots, f_j\}$, and letting $\underset{\sim}{w} = (w_1, \ldots, w_N)'$ be the u-th column of $\overline{\underset{\sim}{x}\underset{\sim}{x}^t}$, we shall establish the following equality for $\underset{\sim}{v} \in W$:

$$(17) \qquad \underset{\sim}{w}^t \underset{\sim}{v} = c\|\underset{\sim}{x}\|^2 v_u \quad .$$

For any $\gamma \in \Lambda$, define $U_\gamma = G^\gamma(u) - \bigcup_{j \in \gamma} G^{\gamma - \{j\}}(u)$. (Recall that $G^*(u)$ is the orbit of u under G^γ and similarly for $G^{\gamma-\{j\}}(u)$.) Evidently,

$$(18) \qquad 1_{U_{\gamma \cup \{j\}}} = \sum_{u' \in U_\gamma} 1_{G_j(u')} - 1_{U_\gamma} \quad ,$$

and

$$(19) \qquad U_\gamma \cap U_{\gamma'} = \phi \quad \text{for any } \gamma \neq \gamma' \quad .$$

The following lemma is crucial for us to get (11).

Lemma 4.6 For any $\gamma \in \Lambda_k$ and any $u', u'' \in U_\gamma$, we have $w_{u'} = w_{u''}$ and this constant is equal to $(-1)^k \frac{\|\underset{\sim}{x}\|^2}{N} \prod_{j \in \gamma} (f_j - 1)^{-1}$. Moreover, we have $w_u = \|\underset{\sim}{x}\|^2$.

Assuming the validity of the above lemma, we proceed to establish (11).

Write $\lambda_\gamma \equiv (-1)^k \frac{\|\underset{\sim}{x}\|^2}{N} \sum_{j\in\gamma} (f_j - 1)^{-1}$. By (19) and Lemma 4.6, we have $\underset{\sim}{w}^t \underset{\sim}{v} = \sum_{\gamma\in\Lambda} \lambda_\gamma \underset{\sim}{1}^t_{U_\gamma} \underset{\sim}{v} + \|\underset{\sim}{x}\|^2 \cdot v_u$. Hence it suffices to show that $\underset{\sim}{1}^t_{U_\gamma} \underset{\sim}{v} = (-1)^k v_u$ for any $\gamma \in \Lambda_k$. We shall prove this by mathematical induction. When $\gamma = \{j\}$, by the definition of W, we have $0 = \underset{\sim}{1}^t_{G_j(u)} \underset{\sim}{v} = \underset{\sim}{1}_{U_\gamma} + v_u$. Thus our assertion is true for k=1. Suppose it is true for $\gamma \in \Lambda_k$. For $j \notin \gamma$, by (18) we have

$$\underset{\sim}{1}^t_{U_\gamma \cup \{j\}} \underset{\sim}{v} = \sum_{u'\in U_\gamma} \underset{\sim}{1}^t_{G_j(u')} \underset{\sim}{v} - \underset{\sim}{1}^t_{U_\gamma} \underset{\sim}{v} = 0 - \underset{\sim}{1}^t_{U_\gamma} \underset{\sim}{v} = (-1)^{k+1} v_u.$$

The second equality is due to the fact $\underset{\sim}{v} \in W$ and the last equality holds because of our induction hypothesis. Hence (17) is now established and so is (11).

Next, we shall establish (12). Since $\underset{\sim}{V}$ is symmetric, let $V^{\frac{1}{2}}$ be any square root of $\underset{\sim}{V}$. Let $\underset{\sim}{e}_u$ be the u-th column vector of $V^{\frac{1}{2}}$. Hence by (11) we have

$$\underset{\sim}{v}^t \overline{\underset{\sim}{V}} \underset{\sim}{v} = \underset{\sim}{v}^t \overline{\sum_{u=1}^N \underset{\sim}{e}_u \underset{\sim}{e}_u^t} \underset{\sim}{v} = \sum_{u=1}^N \underset{\sim}{v}^t \overline{\underset{\sim}{e}_u \underset{\sim}{e}_u^t} \underset{\sim}{v} = \sum_{b=1}^B \sum_{u=1}^N c_b \|\underset{\sim}{e}_{u[b]} + \tilde{\underset{\sim}{e}}_{u[b]}\|^2 \cdot \underset{\sim}{v}^t_{[b]} \underset{\sim}{v}_{[b]} .$$

Thus, it suffices to show that $\operatorname{trace} \underset{\sim}{V}_{[b]} = \sum_{u=1}^N \|\underset{\sim}{e}_{u[b]} + \underset{\sim}{e}_{u[b]}\|^2$. Let L_b be the $N \times N$ matrix such that $L_b \underset{\sim}{x} = \underset{\sim}{x}_{[b]} + \tilde{\underset{\sim}{x}}_{[b]}$ for any $\underset{\sim}{x} \in R^N$. Now, we have

$$\sum_{u=1}^{N} \| \underset{\sim}{e}_{u[b]} + \tilde{\underset{\sim}{e}}_{u[b]} \|^2 = \text{trace} \sum_{u=1}^{N} (\underset{\sim}{e}_{u[b]} + \tilde{\underset{\sim}{e}}_{u[b]})(\underset{\sim}{e}_{u[b]} + \tilde{\underset{\sim}{e}}_{u[b]})^t$$

$$= \text{trace} \sum_{u=1}^{N} L_b \underset{\sim}{e}_u \underset{\sim}{e}_u^t L_b^t = \text{trace } L_b V L_b^t$$

$$= \text{trace } \underset{\sim}{V}_{[b]} L_b^t = \text{trace } L_b (\underset{\sim}{V}^*_{[b]})^t_{[b]} = \underset{\sim}{V}_{[b]} .$$

Hence (12) is established. Finally, it is clear that $\underset{\sim}{v}^t \bar{\underset{\sim}{x}} = 0$ since $\bar{\underset{\sim}{x}}$ is some linear combination of $\underset{\sim}{1}_{G_j b(u)}$'s. The proof of Lemma 3.4 is thus complete. □

Proof of Lemma 4.6 Since each G_j' is doubly transitive, we see that for any u' and $u'' \in U_{\{j\}}$ (which implies that $u \neq u''$) there exists a π such that $\pi u = u$ and $\pi u' = \pi u''$. Now, since $\overline{\underset{\sim\sim}{x}\underset{\sim}{x}^t} = \pi \overline{\underset{\sim\sim}{x}\underset{\sim}{x}^t} \underset{\sim}{\pi}^t$, by comparing their (u,u')-th cells, we get $w_{u'} = w_{u''}$. Similar argument leads to the conclusion that $w_{uu'} = w_{uu''}$ for any u' and $u'' \in U_\gamma$, due to the existence of some π for which $\pi(u) = u$ and $\pi(u') = u''$. We now compute the constant for each U_γ.

First, the diagonal elements of $\overline{\underset{\sim\sim}{x}\underset{\sim}{x}^t}$ are the same due to the transitity of G. Hence they are equal to $\frac{1}{N} \|\underset{\sim}{x}\|^2$ since

$$\text{trace } \overline{\underset{\sim\sim}{x}\underset{\sim}{x}^t} = \frac{1}{g} \sum_{\pi \in G} \text{trace } \overline{\pi(\underset{\sim}{x}) \pi(\underset{\sim}{x})^t} = \frac{1}{g} \sum_{\pi \in G} \|\underset{\sim}{x}\|^2 = \|\underset{\sim}{x}\|^2 .$$

Next, let $\underset{\sim}{Z} = \underset{\sim}{\nu}_0(\underset{\sim}{x})$ (recall that $\underset{\sim}{\nu}_0$ is the uniform distribution on G). Clearly, we have $E\underset{\sim\sim}{Z}\underset{\sim}{Z}^t = \overline{\underset{\sim\sim}{x}\underset{\sim}{x}^t}$ and thus $w_{u'} = EZ_u Z_{u'}$. By the assumption that $\bar{\underset{\sim}{x}}^{\{j\}} = 0$, we get $\sum_{\pi \in G_j} Z_{\pi(u)} = 0$. Thus, $-Z_u^2 = \sum_{u' \in U\{j\}} Z_u Z_{u'}$.

After taking expectation on both sides, we obtain

$-w_u = \sum_{u' \in U_{\{j\}}} w_{u'} = (f_j - 1) w_{u'}$. Thus $w_{u'} = -(f_j-1)^{-1} w_u = -(f_j-1)^{-1} \cdot N^{-1} \|\underset{\sim}{x}\|^2$ as desired.

Next, supposing our lemma is true for some $\gamma \in \Lambda_k$, we shall show that it is also true for $\gamma \cup \{j\}$ with $j \notin \gamma$. For any $u' \in U_{\gamma \cup \{j\}}$, it is easy to find a $\pi_1 \in G^{\gamma}$ and a $\pi_2 \in G_j$ such that $\pi_1(u) \in U_{\gamma}$ and $\pi_2 \pi_1(u) = u'$. Now, $\underset{\sim}{\bar{x}}^{\{j\}} = 0$ implies that $\sum_{\pi \in G_j} Z_{\pi(\pi_1(u))} = 0$ and thus we obtain $Z_u \cdot Z_{\pi_1(u)} = -\sum_{\pi \in G_j \setminus \{I\}} Z_u \cdot Z_{\pi(\pi_1(u))}$. Taking the expectations on both sides and noticing that $\pi\pi_1(u) \in U_{\gamma \cup \{j\}}$ for any $\pi \in G_j \setminus \{I\}$, we get $w_{\pi_1(u)} = -(f_j-1) w_{u'}$. Thus

$$w_{u'} = -(f_j-1)^{-1} w_{\pi_1(u)} = -(f_j-1)^{-1} \cdot \|\underset{\sim}{x}\|^2 \; N^{-1}(-1)^k \prod_{i \in \gamma} (f_i-1)^{-1}$$

$$= \|\underset{\sim}{x}\|^2 \cdot N^{-1} \cdot (-1)^{k+1} \prod_{i \in \gamma \cup \{j\}} (f_i-1)^{-1}$$

as desired. □

Proof of Theorem 3.1 First, by Lemma 3.2, Lemma 3.3 and Lemma 3.4, for $d \in \underset{\sim}{D}^*$ and $\underset{\sim}{c} \in \underset{\sim}{C}_G$, we have

$$(20) \qquad Er(\underset{\sim}{\phi}_G(d), \underset{\sim}{c}(\underset{\sim}{\phi}_G(d)); s) = \text{trace } \underset{\sim}{c}(d) \, \overline{(\underset{\sim}{x}-\underset{\sim}{\bar{x}})(\underset{\sim}{x}-\underset{\sim}{\bar{x}})^t} \, \underset{\sim}{c}(d)^t + \sum_{b=1}^{B} c_b \, \text{trace}(\underset{\sim}{M}_{[b]}) \, \text{trace } \underset{\sim}{c}(d) \, P_b^t P_b \, \underset{\sim}{c}(d)^t ,$$

where P_b is a projection as before.

By an argument similar to the proof of Lemma 4.5, we have

$$(21) \qquad \underset{\sim}{c}(d) \, \overline{(\underset{\sim}{x}-\underset{\sim}{\bar{x}})(\underset{\sim}{x}-\underset{\sim}{\bar{x}})^t} \, \underset{\sim}{c}(d)^t = \underset{\sim}{c}(d) \, \overline{\underset{\sim}{x}\underset{\sim}{x}^t} \, \underset{\sim}{c}(d)^t .$$

Thus, by (20), (21) and (11), we have, for $d \in \underset{\sim}{D}^*$ and $c \in \underset{\sim}{C}_G^*$,

$$(22) \qquad Er(\underset{\sim}{\phi}_G(d),\underset{\sim}{c}(\underset{\sim}{\phi}_G(d));s) = \sum_{b=1}^{B} c_b(\|\underset{\sim}{x}_{[b]} + \tilde{\underset{\sim}{x}}_{[b]}\|^2 + \mathrm{trace}(\underset{\sim}{M}_{[b]})) \cdot \mathrm{trace}\ \underset{\sim}{c}(d)\ P_b^t P_b\ \underset{\sim}{c}(d)^t .$$

Next, from the proof of Lemma 3.1, it is clear that after replacing $\underset{\sim}{C}$ with $\underset{\sim}{C}^o$ and $\underset{\sim}{C}_G$ with $\underset{\sim}{C}_G^o$ in (8), the equality there still holds. Thus we have

$$\min_{(\mu,\underset{\sim}{c})\in \underset{\sim}{D}\times\underset{\sim}{C}^o} \max_{s\in S} Er(\mu,\underset{\sim}{c}(\mu);s) = \max_{s\in S} \min_{(d,c)\in \underset{\sim}{D}\times\underset{\sim}{C}_G^o} Er(\underset{\sim}{\phi}_G(d),\underset{\sim}{c}(\underset{\sim}{\phi}_G(d));s)$$

$$\geqslant \max_{s\in S} \min_{(d,\underset{\sim}{c})\in \underset{\sim}{D}^*\times\underset{\sim}{C}_G^*} \sum_{b=1}^{B} c_b(\|\underset{\sim}{x}_{[b]} + \tilde{\underset{\sim}{x}}_{[b]}\|^2 + \mathrm{trace}(\underset{\sim}{M}_{[b]})) \cdot \mathrm{trace}\ \underset{\sim}{c}(d)\ P_b^t P_b\ \underset{\sim}{c}(d)^t \geqslant \min_{(d,c)\in \underset{\sim}{D}^*\times\underset{\sim}{C}_G^*} \sum_{b=1}^{B} c_b(\|\underset{\sim}{x}^*_{[b]} + \tilde{\underset{\sim}{x}}^*_{[b]}\|^2 + \mathrm{trace}(\underset{\sim}{M}^*_{[b]})) \cdot \mathrm{trace}\ \underset{\sim}{c}(d)\ P_b^t P_b\ \underset{\sim}{c}(d)^t$$

$$= K \cdot \min_{(d,c)\in \underset{\sim}{D}^*\times\underset{\sim}{C}_G^*} \mathrm{trace}\ \underset{\sim}{c}(d)\underset{\sim}{c}(d)^t = k\ \mathrm{trace}\ \underset{\sim}{c}^o(d^o)\underset{\sim}{c}^o(d^o)^t ,$$

where k is the common value in (13).

On the other hand, by (11), and our assumption about the existence of (x^*,m^*), we have, for any $s \in S$,

$$Er(\underset{\sim}{\phi}_G(d^o),\underset{\sim}{c}^o(\underset{\sim}{\phi}_G(d^o));s) \leqslant \sum_{b=1}^{B} c_b\|\underset{\sim}{x}^*_{[b]} + \tilde{\underset{\sim}{x}}^*_{[b]}\|^2\ \mathrm{trace}\ \underset{\sim}{c}(d)\ P_b^t P_b\ \underset{\sim}{c}(d)^t \underset{\sim}{c}(d)$$

$$= K \cdot \mathrm{trace}\ \underset{\sim}{c}(d)\underset{\sim}{c}(d)^t .$$

Hence $\max_{s\in S} Er(\underset{\sim}{\phi}_G(d^o), \underset{\sim}{c}^o(\underset{\sim}{\phi}_G(d^o));s) \leq K \cdot \text{trace } \underset{\sim}{c}(d)\underset{\sim}{c}(d)^t$. □

Proof of Theorem 3.2 Let g^* be the order of G^*. Define

$\bar{\underset{\sim}{x}}^* = \frac{1}{g^*} \sum_{\pi\in G^*} \pi(\underset{\sim}{x})$ for $\forall\, \underset{\sim}{x}\in R^N$ and $\bar{\underset{\sim}{V}}^* = \frac{1}{g^*} \sum_{\pi\in G^*} \pi(\underset{\sim}{V})$ for any symmetric $N\times N$ matrix V. By treating $(\underset{\sim}{x}-\bar{\underset{\sim}{x}}^*)(\underset{\sim}{x}-\bar{\underset{\sim}{x}}^*)$ as $\underset{\sim}{V}$ we get the definition for $\overline{(\underset{\sim}{x}-\bar{\underset{\sim}{x}}^*)(\underset{\sim}{x}-\bar{\underset{\sim}{x}}^*)^t}^*$. It is clear that if we substitute G^* for G in Lemma 3.2 and Lemma 3.3, the results still hold (note that the definition of W is *not* changed). Thus we have, for $d\in \underset{\sim}{D}^*$ and $\underset{\sim}{c}\in \underset{\sim}{C}_G^*$,

$$(23)\quad Er(\underset{\sim}{\phi}_{G^*}(d), \underset{\sim}{c}(\underset{\sim}{\phi}_{G^*}(d));s) = \|\underset{\sim}{c}(d)\underset{\sim}{x}^*\|^2 + \text{trace } \underset{\sim}{c}(d)\Big(\overline{(\underset{\sim}{x}-\bar{\underset{\sim}{x}}^*)(\underset{\sim}{x}-\bar{\underset{\sim}{x}}^*)^t}^* + \bar{\underset{\sim}{M}}^*\Big)\, \underset{\sim}{c}(d)^t = \|\underset{\sim}{c}(d)\bar{\underset{\sim}{x}}^*\|^2 + \text{trace } \underset{\sim}{c}(d)\,\overline{(\underset{\sim}{x}-\bar{\underset{\sim}{x}}^*)(\underset{\sim}{x}-\bar{\underset{\sim}{x}}^*)^t + \underset{\sim}{M}}^*\, \underset{\sim}{c}(d)^t$$

Write $\underset{\sim}{V} = (\underset{\sim}{x}-\bar{\underset{\sim}{x}}^*)(\underset{\sim}{x}-\bar{\underset{\sim}{x}}^*)^t + \underset{\sim}{M}$. We claim that for any $\underset{\sim}{v}\in W$,

$$(24)\qquad \underset{\sim}{v}^t \bar{\underset{\sim}{x}}^* = 0$$

and

$$(25)\qquad \underset{\sim}{v}^t \bar{\underset{\sim}{V}}^* \underset{\sim}{v} = \lambda\cdot \underset{\sim}{v}^t\underset{\sim}{v} \quad \text{for a constant } \lambda \text{ which depends on } \underset{\sim}{V}.$$

First, note that if $\underset{\sim}{v}\in W$, then $\underset{\sim\sim}{\pi v}\in W$ for all $\pi\in G_t$. Second, since it is clear that $G^* = G_t\cdot G$ and $g^* = g_t\cdot g$ (where g_t is the order of G_t), we have $\bar{\underset{\sim}{x}}^* = \frac{1}{g_t}\sum_{\pi\in G_t}\pi(\bar{\underset{\sim}{x}})$ and $\bar{\underset{\sim}{V}}^* = \frac{1}{g_t}\sum_{\pi\in G_t}\pi(\bar{\underset{\sim}{V}})$. Then (24) follows easily from Lemma 3.3.4 since

$$\underset{\sim}{v}^t \bar{\underset{\sim}{x}}^* = \frac{1}{g_t} \sum_{\pi \in G_t} \underset{\sim}{v}^t \underset{\sim}{\pi}\, \bar{\underset{\sim}{x}} = \frac{1}{g_t} \sum_{\pi \in G_t} (\underset{\sim}{\pi}^{-1}\underset{\sim}{v})^t \bar{\underset{\sim}{x}} = \frac{1}{g_t} \cdot \sum_{\pi \in G_t} 0 = 0 .$$

Similarly, by (12) we have

$$\underset{\sim}{v}^t \bar{\underset{\sim}{V}}^* \underset{\sim}{v} = \frac{1}{g_t} \sum_{\pi \in G_t} \underset{\sim}{v}^t \underset{\sim}{\pi}\, \bar{\underset{\sim}{V}}\, \underset{\sim}{\pi}^t \underset{\sim}{v} = \frac{1}{g_t} \sum_{\pi \in G_t} (\underset{\sim}{\pi}^t \underset{\sim}{v})^t \bar{\underset{\sim}{V}} (\underset{\sim}{\pi}^t \underset{\sim}{v})$$

$$= \frac{1}{g_t} \sum_{\pi \in G_t} \sum_{b=1}^{B} c_b \operatorname{trace}(\underset{\sim}{V}_{[b]}) \, \| (\underset{\sim}{\pi}^t \underset{\sim}{v})_{[b]} \|^2$$

$$= \frac{1}{g_t} \sum_{b=1}^{B} c_b \cdot \operatorname{trace}(\underset{\sim}{V}_{[b]}) \cdot \sum_{\pi \in G_t} \| (\underset{\sim}{\pi}^{-1} \underset{\sim}{v})_{[b]} \|^2$$

$$= \sum_{b=1}^{B} c_b \operatorname{trace}(\underset{\sim}{V}_{[b]}) \left(\frac{1}{g_t} \sum_{\pi \in G_t} \| \underset{\sim}{v}_{[\pi^{-1}(b)]} \|^2 \right.$$

$$= \sum_{b=1}^{B} c_b \operatorname{trace}(\underset{\sim}{V}_{[b]}) \cdot \left(\frac{1}{B} \| \underset{\sim}{v} \|^2 \right)$$

$$= \left(\frac{1}{B} \sum_{b=1}^{B} c_b \operatorname{trace}(\underset{\sim}{V}_{[b]}) \right) \cdot \| \underset{\sim}{v} \|^2 \equiv \lambda \cdot \| \underset{\sim}{v} \|^2$$

where the sixth equality is due to the transitivity of G_t'. Thus (25) holds. Now from (23), (24) and (25) we conclude that for $d \in \underset{\sim}{D}^*$ and $\underset{\sim}{c} \in \underset{\sim}{C}_G^*$,

(26) $\quad Er(\phi_{G^*}(d), \underset{\sim}{c}(\underset{\sim}{\phi}_{G^*}(d)); s) = \lambda(s) \operatorname{trace} \underset{\sim}{c}(d) \underset{\sim}{c}(d)^t$, where $\lambda(s)$ is some positive number.

The rest of the proof will be similar to the proof of Theorem 3.1. Replacing $\underset{\sim}{C}$ with $\underset{\sim}{C}^o$ and $\underset{\sim}{C}_{G^*}$ with $\underset{\sim}{C}^o_{G^*}$ in (8), we have

$$\min_{(\underset{\sim}{\mu},\underset{\sim}{c})\in \underset{\sim}{D}\times\underset{\sim}{C}^{0}} \max_{s\in S} Er(\underset{\sim}{\mu},\underset{\sim}{c}(\underset{\sim}{\mu});s) = \max_{s\in S} \min_{(d,s)\in \underset{\sim}{D}\times\underset{\sim}{C}_{G}^{0}} Er(\underset{\sim}{\phi}_{G^*}(d),\underset{\sim}{c}(\underset{\sim}{\phi}_{G^*}(d);s)$$

$$\geqslant \max_{s\in S} \min_{(d,s)\in \underset{\sim}{D}^{*}\times\underset{\sim}{C}_{G}^{*}} Er(\underset{\sim}{\phi}_{G}(d),\underset{\sim}{c}(\underset{\sim}{\phi}_{G^*}(d));s) \geqslant \max_{s\in S} \min_{(d,s)\in \underset{\sim}{D}^{*}\times\underset{\sim}{C}_{G}^{*}}$$

$$\lambda(s)\ \text{trace}\ \underset{\sim}{c}(d)\underset{\sim}{c}(d)^{t} = \left(\max_{s\in S}\lambda(s)\right)\cdot \text{trace}\ \underset{\sim}{c}^{0}(d^{0})\underset{\sim}{c}^{0}(d^{0})^{t}$$

$$= \max_{s\in S} Er(\underset{\sim}{\phi}_{G^*}(d^{0}),\underset{\sim}{c}^{0}(\underset{\sim}{\phi}_{G^*}(d^{0}));s)\ .$$

The last equality is by (26). The proof is thus complete. □

Acknowledgement. I am indebted to Professor Kiefer for his guidance and many helpful suggestions on this work. I am also indebted to Professor Ching-Shui Cheng for many useful comments.

REFERENCES

[1] Burnside,W. (1911). Theory of Group Finite Order, 2nd ed., Cambridge University Press, London.

[2] Cheng,C.S. (1978a). Optimality of certain asymmetrical experimental designs. Ann. Statist. 6, 1239-1261.

[3] Cheng,C.S. (1978b). Optimal designs for the elimination of multi-way heterogeneity. Ann. Statist. 6, 1262-1272.

[4] Cheng,C.S. (1980). Optimal designs for the elimination of multi-way heterogeneity. II. IMS Bull. 80t-57.

[5] Cheng,C.S. and Li,K.C. (1981). Identifying the covariance structures equivalent to those with homoscedastic uncorrelated errors for the estimation and design problems. Manuscript.

[6] Hardy, Littlewood and Polyà (1934). Inequalities. Cambridge University Press, London.

[7] Huber,P.J. (1975). Robustness and designs, in A Survey of Statistical Design and Linear Models, (J.N.Srivastava, ed.), 287-301. North Holland, Amsterdam.

[8] Kiefer,J. (1958). On the non-randomized optimality and randomized non-optimality of symmetrical designs. Ann. Math. Statis. 29, 675-699.

[9] Kiefer,J. (1975). Constructions and optimality of generalized Youden designs, in A Survey of Statistical Designs and Linear Models (J.N.Srivastava, ed), 333-353. North Holland, Amsterdam.

[10] Wu,C.F. (1980). On the robustness and efficiency of some randomized designs. Manuscript.

#81-44

The Annals of Statistics
1983, Vol. 11, No. 1, 225–239

MINIMAXITY FOR RANDOMIZED DESIGNS: SOME GENERAL RESULTS*

BY KER-CHAU LI

Purdue University

In many design settings where model violations are present, a "stochastic" minimaxity for many standard randomization procedures is demonstrated. This result requires no special analytic properties of the loss function and estimators. Next, under the squared loss and with the restriction to the use of linear estimators, a recipe for finding a randomized strategy is given. As a special case, randomizing an A-optimal design in the standard manner and using the least squares estimates yields a minimax strategy in most cases. These results generalize some aspects of Wu (1981).

1. Introduction. The role of randomization in the design of experiments has been discussed in numerous papers (see the references given in Wu, 1981). As it was summarized by Wu, the most popular of the arguments favoring the use of randomization are the following: it provides a solid basis for statistical inference; it ensures impartiality; it is a source of robustness against model inadequacies. Most of the literature has been addressed to the first and the second arguments.

While the third argument on the model robustness aspect of randomization has already been well accepted, Wu (1981) seems to be the first work devoted to giving it a formal definition and rigorous justification. For some basic design setups in comparative experiments where T treatments are to be assigned to N experiment units, Wu argued that since the experimenter's information about the model is never perfect, there is always the possibility that the "true" model deviates from the assumed model. Thus if G is the collection of all possible "true" models, he defined the concept of model-robustness with respect to G in terms of minimizing the maximum possible mean squared error of the corresponding best linear unbiased estimator (for the assumed model) over G. For the use of the model-robustness notion in other contexts, see, for example, Box and Draper (1959) and Huber (1975). Some randomized designs, including the balanced completely randomized design (coined by Wu), the randomized complete block design and the randomized Latin square design, were shown to be model-robust with respect to any G which possesses an appropriate invariance property in each setting. Furthermore, Wu compared some randomized designs in terms of maximum squared bias. In this paper, we shall discuss only the minimax results. Basically we adopt Wu's general framework on the model-violation consideration; i.e., G will be invariant in an appropriate sense. But we shall extend the results to quite general design settings after a careful study of Wu's ideas.

This paper is composed of two parts. The first part (Section 2) discusses the minimaxity of some commonly-used randomization procedures. The results obtained here are the generalization of Wu's Proposition 1, Theorem 4, and the minimaxity for the randomized complete block design and BIB designs. In these results, the competing class of designs was restricted to those having the same treatment replication numbers of each block (the block design case) and for each row and each column (Latin squares case). We shall make the same restriction in this section. For example, in the block design case, the block-

Received June 1981; revised June 1982.

[1] Research sponsored in part by the National Science Foundation under Grant No. MCS-78-25301. Revision supported by National Science Foundation under grant No. MCS-8200631.

AMS 1980 *subject classification.* Primary 62K99; secondary, 62A05, 62C20.

Key words and phrases. A-optimality, doubly transitive groups, H-uniform randomization, randomization procedures, randomized strategies, robustness.

treatment incidence matrix is fixed (but arbitrary). (The optimum choice of a block-treatment incidence matrix will be considered in Section 3). However, it is easier to convey these results in terms of a rigorously-defined notion of randomization procedures. Our definition is based on the observation that when applying any randomization procedure (in the usual sense, for example, the complete randomization) to a non-randomized design, the possible realized designs are those with the same treatment replication numbers as those for the original design. Thus by comparing different randomization procedures only, we avoid the more complicated problem of choosing an optimal non-randomized design to be randomized. Roughly speaking, if the class of all possible true models is invariant under a group H, then the randomization procedure (H-uniform randomization, to be defined later) generated by the uniform probability measure on H is minimax in the sense that for any fixed non-randomized design d and estimator δ, the maximum risk of applying a randomization procedure to d and using the corresponding permuted version of the estimator δ is minimized when H-uniform randomization is used. We do not require any special analytic property of δ; thus non-linear estimators (which are sometimes proposed for guarding against distribution violations) are allowed. The loss function could be arbitrary, athough some invariant properties should be satisfied to ensure that we are not estimating any feature of the nuisance parameters such as unit effects, etc.

An important observation leading to our broadened results is that to prove Wu's Proposition 1, no explicit expression of the risk functions is needed. The basic idea is the standard concept that "suitable invariance" implies "minimaxity", due to the Hunt-Stein Theorem (see Lehmann, 1971). However, the result of Blackwell and Girshick (1954), which shows the minimaxity of simple random sampling, is more relevant. This is because unlike the cases where the Hunt-Stein Theorem usually applies, what our group actually transforms are the nuisance parameters (unit effects, etc.), *not* the parameters of interest! Furthermore, there is one special feature about the manner of evaluating the randomized decision rules which makes our results different from any earlier results. Recall that in the standard decision theory, after defining the risk for a non-randomized rule, the risk for a randomized rule is defined to be the mean of the risks of its possible realized rules. However, it is quite obvious that instead of the means, several other location measures such as medians, quantiles, etc., may also be used to assess a randomized rule provided that the possible mathematical difficulties can be removed. In other words, ideally we should compare the randomized rules according to the stochastic orderings of their random (due to randomization) risks. Our minimaxity results in Section 2 are established under such considerations. Therefore, they provide a very sound basis for using randomization procedures in guarding against model-violations.

The second part of this paper concerns the choice of a randomized design under the same model-violation considerations as in Section 2 but *without any restriction to the competing class*. However, we do require that the estimators be linear (but not necessarily the least squares) and the loss function be the squared one. Furthermore, we evaluate the performance of a randomized rule by defining its risk in the standard way; i.e., by considering the mean risk only. These restrictions seem to be unavoidable for obtaining useful results since the usual work on optimal experimental designs (which assumed no model-violations) is based on these assumptions. Among the three commonly-used design criteria, (A, D, and E criteria), our results are most closely related to the A-criterion. In the block design (the two-way heterogeneity design, respectively) settings, we show that the randomized strategy (i.e., design and estimator) which first chooses an A-optimal design and then randomizes it in the standard way, i.e., randomizes completely the blocks and the units within each block (rows and columns, respectively), and uses the usual least squares estimators, is minimax.

These results extend Wu's Theorem 1 and Theorem 3 which justified *randomization* as well as *balance* from the model-robustness viewpoint for the no-blocking setup. We also justify the use of least squares estimators in the appropriate randomization procedures. In using A-optimal designs, we assume (by the loss function) that all treatments are of equal interest. Recently, there have been considerable research interests on designs for comparing

test treatments to a control treatment, Bechhofer and Tamhane (1981). In such cases, the loss function should reflect the relative importance of the control and the test treatments. In general, if the loss function is of the form $(L\alpha - \mathbf{a})'(L\alpha - \mathbf{a})$ where L is a $p \times T$ matrix and $\mathbf{a}$ is $p \times 1$ vector estimating $L\alpha$, which was referred to as a linear criterion or L-criterion (see Kiefer, 1974, Karlin and Studden, 1966, or Fedorov, 1972), then our results show that to obtain the minimax randomized strategy one only has to first search for the corresponding optimal designs for the "assumed" (or ideal) model. After finding an optimal one, then we should randomize it in the standard way and use the standard least squares estimates. Having seen such results for the one-way and two-way settings, one is easily led to the conclusion that similar results should hold for the k-way settings. Unfortunately, this is true only when the complicated model which assumes the existence of all higher order (up to k–1) interactions among the units is considered. For the usual additivity model, the minimax randomized design may depend on the actual form of the class G of possible models if $k \geq 3$.

Furthermore, our counter-example shows that for certain invariant class G, randomizing the most symmetric design may sometimes be inferior to randomizing a less symmetric design. This example illustrates the need for rigorous justification in applying randomization in various settings. Like elsewhere, a careless application of Hunt-Stein's idea or any related concepts may incur misleading conclusions. But the crucial issue involved here is not the compactness of the transformation group on the class G (permutation groups are always finite and hence compact). The issue is how the group works. Usually, the transformation group used here can be naturally decomposed into some basic subgroups. (For example, in two-way heterogeneity settings, the transformation group involved is the product of a row permutation group and a column permutation group). The relation between the orbits of these subgroups and the block effects, row, column effects, or higher order interaction effects turns out to be a very important consideration in obtaining the results. (Such a consideration was implicit in Cheng and Li, 1980). If the orbits of the subgroups correspond to block or interaction effects then our minimax results hold. (For example, the orbits of the row permutation group correspond to the column effects and the orbits of the column permutation group correspond to the row effects). However, for the $k \geq 3$ way settings, the orbits of each subgroup which permute the levels of one factor will correspond to (k–1)-interaction effects among factors which are not assumed in the usual additivity models. This explains why we need a complicated model to ensure minimaxity. Also, as a simple consequence, we obtain other randomization procedures which are generated by groups of very small orders and are of the same efficiency as the commonly-used ones when the squared loss function is assumed. Section 4 is devoted to the proofs.

2. Minimax randomization procedures under general loss functions. Suppose T treatments are to be assigned to N experimental units. A (non-randomized) design is a function d from $\{1, \cdots, N\}$ to $\{1, \cdots, T\}$ with the uth unit receiving treatment $d(u)$. Let D be the class of all designs. In this paper, instead of defining a randomized design as a probability measure on D, we shall conveniently treat it as a *random element* with the nonrandomized designs as possible realizations. Denote the ith treatment effect by α_i and let $\alpha = (\alpha_1, \cdots, \alpha_T)'$. We now first present a simple example to illustrate the general results we shall obtain. This example was already considered by Wu.

EXAMPLE 1. *No blocking.* Suppose the yield (or response) y_u of the uth unit satisfies the following additivity assumption:

$$y_u = \alpha_{d(u)} + g_u + \varepsilon_u, \quad u = 1, \cdots, N, \tag{2.1}$$

where g_u is the uth unit effect and ε_u is the random error with mean 0. In the ideal case we would assume $g_u = 0$ (or a constant) and the random errors are homogeneous and uncorrelated. But this certainly is not a good situation for justifying the use of randomization. In fact, there is always the possibility that the "true" model deviates from the ideal

one. Let G be the set of all possible $\mathbf{g} = (g_1, \cdots, g_n)'$. Let $\mathscr{E}$ be the set of all possible probability measures of $\boldsymbol{\varepsilon} = (\varepsilon_1, \cdots, \varepsilon_n)'$. To reflect the vagueness of the experimenter's knowledge, we require that G and $\mathscr{E}$ are invariant under the group H of all permutations on $\{1, \cdots, N\}$; i.e., $\mathbf{g} \in G \Rightarrow \pi\mathbf{g} \in G$ and $\xi \in \mathscr{E} \Rightarrow \pi\xi \in \mathscr{E}$ for all $\pi \in H$, where $\pi\mathbf{g} = (g_{\pi^{-1}(1)}, \cdots, g_{\pi^{-1}(N)})'$ and $\pi(\xi)$ is the probability measure of $\pi\boldsymbol{\varepsilon}$ when ξ is the probability measure of $\boldsymbol{\varepsilon}$. We denote the triple $(\boldsymbol{\alpha}, \mathbf{g}, \xi)$ by s, let S be the set of all possible s, and write $\pi s = (\boldsymbol{\alpha}, \pi\mathbf{g}, \pi\xi)$. Thus we have

$$s \in S \Rightarrow \pi s \in S \text{ for all } \pi \in H. \tag{2.2}$$

Now we shall define the concept of a randomization procedure rigorously. Recall that when applying any randomization procedure (e.g., the complete randomization) to a given non-randomized design d, the possible realized designs will have the same replication numbers as those for the original design d. Since it is clear to see that $H(d) = \{\pi d \mid \pi \in H\}$ is the class of designs with the same replication numbers as d, we define a randomization procedure to be a *a function $\boldsymbol{\Phi}$ on D such that $\boldsymbol{\Phi}(d)$ is a random element with possible realizations in $H(d)$ for any $d \in D$*. In particular, the complete randomization is a function which maps $d \in D$ to $\mathbf{h}d$, where $\mathbf{h}$ *is the random permutation generated by the uniform distribution on H*. Thus we denote the complete randomization by $\mathbf{h}$. We shall demonstrate a minimax property for $\mathbf{h}$, after discussing the problem of choice of estimators and loss functions and the problem of evaluating a randomized strategy.

The loss function $\mathscr{L}$ considered in this section does not need any special analytic property. We only require $\mathscr{L}$ to be invariant in the following sense:

(2.3) $\mathscr{L}(\pi(s), a) = \mathscr{L}(s, a)$ for any $\pi \in H$, $s \in S$ and any a in the action space $\mathscr{A}$.

This invariance requirement amounts to claiming that what we estimate depends only on $\boldsymbol{\alpha}$ and in no way on $\mathbf{g}$ or ξ. For instance, we may take $\mathscr{L}(s, a) = (L\boldsymbol{\alpha} - a)'(L\boldsymbol{\alpha} - a)$ where L is a $p \times T$ matrix and $a \in \mathscr{A} = R^p$.

The choice of estimators should also be invariant under H in the following sense. Suppose for design d, an estimator δ, which is a function mapping $\mathbf{y} = (y_1, \cdots, y_N)'$ to an element in $\mathscr{A}$, is used. Then we require that for design πd, the estimator $\pi\delta(\mathbf{y}) = \delta(\pi^{-1}(\mathbf{y}))$ should also be used. This is a reasonable restriction, similar to that imposed by Blackwell and Girshick (1954) in justifying simple random sampling, because when there is no model-violation, the distribution of $\pi^{-1}(\mathbf{y})$ under design πd is the same as the distribution of $\mathbf{y}$ under design d. Thus when a randomization procedure ϕ is applied to design d for which estimator δ is used, the realized strategy (i.e., design and estimator) is determined and will be denoted by $\phi(d, \delta)$.

Now we discuss the problem of evaluating a randomized strategy. As usual, the risk of a non-randomized design and estimator (d, δ) under $s \in S$ is defined by $r(d, \delta; s) = E\,\mathscr{L}(s, \delta(\mathbf{y}))$. But we do not assess a randomized strategy merely by its expected risks. Instead, we consider the class $\mathscr{F}$ of real functions f on the class of all probability measures of R such that $f(p\mu_1 + (1-p)\mu_2) \leq \max\{f(\mu_1), f(\mu_2)\}$ for any $0 \leq p \leq 1$ and any probability measures μ_1 and μ_2 on R. We also write $f(X) = f(\mu)$ if X is a random variable with the probability measure μ. This broad class $\mathscr{F}$ includes the mean, median or any quantile (all given a convenient definition if not unique) functionals of random variables. The following result is what we shall prove:

> Under (2.1)–(2.3), the complete randomization $\mathbf{h}$ is minimax in the sense that it minimizes $\max_{s \in S}(r(\phi(d, \delta); s))$ over all randomization procedures ϕ, for any $f \in \mathscr{F}$ and any design d and estimator δ.

The above statement follows from Theorem 2.1 below. We call this a "stochastic" minimax property for $\mathbf{h}$ for the following reasons.

For any $t \in R$ and any random variable X, define $f_t(X) = P(X > t)$. Observe that for two random variable X and Y, X is usually said to be stochastically at least as large as Y if $f_t(X) \geq f_t(Y)$ for any $t \in R$. Also, it is clear that $\mathscr{F}$ contains any f_t. Therefore the minimax result preserves the genuine spirit of stochastic ordering. When taking f to be the mean

functional, our "stochastic" minimaxity result is then reduced to a form with the standard sense of risks for randomized strategies. This standard sense of minimaxity was already explored by Wu with the use of the squared loss that reflects the experimenter's equal interests among all treatment effects (i.e., A-criterion) and the use of the least squares estimates. But Wu further justified balance for treatment replication numbers. We shall take up the same task in Section 3 for more complicated designs.

Now, we shall generalize the above notions and results to other design settings such as block design or higher-way heterogeneity settings. This extension requires only an abstraction of the framework of Example 1.

Let S be a set of possible "true states" of nature, and H be a permutation group on $\{1, \cdots, N\}$. Assume that for any $\pi \in H$ and $s \in S$, πs is well-defined and (2.2) holds. (Implicitly, π will transform only the nuisance parameter part of s such as unit effects, etc.) Replace the model assumption (2.1) by the following:

(2.4) for $\pi \in H$, $d \in D$ and $s \in S$, the yield $\mathbf{y}$ under πs and design πd has a probability measure equal to that of $\pi \mathbf{z}$ where $\mathbf{z}$ is the yield under s and design d.

The loss function should satisfy (2.3). The definition of randomization procedures is the same as before and we take Φ_H to be the class of all randomization procedures. We also require the choice of estimators to be invariant under H. The randomization procedure $\mathbf{h}$ will be referred to as the H-uniform randomization. We have the following.

THEOREM 2.1. *Under* (2.2)–(2.4), *the H-uniform randomization* $\mathbf{h}$ *is minimax in the sense that it minimizes* $\max_{s\in S} f(r(\phi(d, \delta); s))$ *over all randomization procedures* $\phi \in \Phi_H$ *for any* $f \in \mathcal{F}$, $d \in D$ *and estimator* δ.

The proof of this theorem will be given in Section 4. We now present two examples to illustrate the application. These examples were already treated by Wu, but our theorem strengthens his results.

EXAMPLE 2. *Block design setup.* Suppose the $N = \sum_{b=1}^{B} N_b$ units are arranged into b blocks with sizes $N_1, \cdots, N_B$ respectively. Consider the model:

$$y_u = \alpha_{d(u)} + \beta_b + g_u + \varepsilon_u, \quad u = 1, \cdots, N, \tag{2.5}$$

where d is the design, g_u is the uth unit effect, ε_u is the random error, β_b is the bth block effect with unit u in block b. Take H to be the group of all permutations within blocks. The class G of all possible $\mathbf{g}$ and the class $\mathcal{E}$ of all possible error distributions ξ are assumed to be invariant under H. Now take $s = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{g}, \xi)$ where $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_B)'$. Observe that $\pi s = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \pi\mathbf{g}, \pi\xi)$; (2.2) and (2.4) are satisfied; (2.3) amounts to claiming that what we estimate depends only on $\boldsymbol{\alpha}$ or $\boldsymbol{\beta}$ but not on $\mathbf{g}$ or ξ; for any design d, $H(d)$ is the class of designs possessing the same treatment replication numbers for each block as those of d; $\mathbf{h}$ is the procedure of complete randomization within blocks. Applying Theorem 2.1, we obtain a stochastic minimax property for $\mathbf{h}$.

When the block sizes are equal, i.e., $N_1 = \cdots = N_B$, we may consider a larger group H generated by all permutations within blocks and all block permutations. Observe that if $\pi = \pi_1 \cdot \pi_2$ where π_1 is a permutation within blocks and π_2 is a block permutation, then $\pi s = (\boldsymbol{\alpha}, \pi_2\boldsymbol{\beta}, \pi g, \pi\xi)$; (2.3) claims that what we are interested in depends only on $\boldsymbol{\alpha}$ and not on $\boldsymbol{\beta}$, $\mathbf{g}$, or ξ; the H-uniform randomization $\mathbf{h}$ is the procedure of completely randomizing the blocks and the units within each block. Thus if the model is invariant under H, then $\mathbf{h}$ will be a minimax randomization procedure.

EXAMPLE 3. *Two-way heterogeneity setup.* The $N = \ell_1\ell_2$ units are now arranged in an $\ell_1 \times \ell_2$ array. Suppose the model is

$$y_{ij} = \alpha_{d(i,j)} + \beta_i + r_j + g_{ij} + \varepsilon_{ij}, \; i = 1, \cdots, \ell_1, j = 1, \cdots, \ell_2, \tag{2.6}$$

where β_i is the ith row effect and r_j is the jth column effect, g_{ij} is the (i, j)th unit effect and ε_{ij} is the random error. Take H to be the group generated by all row permutations and column permutations. Let $s = (\alpha, \beta, \mathbf{r}, \mathbf{g}, \xi)$ and $\pi s = (\alpha, \pi_1\beta, \pi_2\mathbf{r}, \pi\mathbf{g}, \pi\xi)$ where $\pi = \pi_1 \cdot \pi_2$ and π_1, π_2 are row and column permutations respectively. The H-uniform randomization $\mathbf{h}$ is the procedure of completely randomizing the rows and the columns. Thus our theorem gives a minimaxity for this procedure. It is interesting to note that under the squared loss $\sum_{t,s}\{\alpha_t - \alpha_s - (\hat{\alpha}_t - \hat{\alpha}_s)\}^2$ (A-criterion) and by using the least squares estimates $\{\hat{\alpha}_t\}$, Wu showed that randomly permuting the rows (or the columns) of a Latin square is of the same efficiency as that of permuting both the rows and the columns and treatment numbers. This greatly simplifies the Fisher-Yates "recipe" of randomization procedure for Latin squares. However, if not for the Latin square with the A-criterion, the mean functional, and the least squares estimator, permuting the rows (or the columns) only will typically be inferior to permuting both rows and columns. Thus, the latter is necessary in general cases.

3. Minimax randomized designs under the squared loss. In Section 2, we justify several commonly-used randomization procedures from the viewpoint of robustness against model-violations. However, to which designs should these procedures be actually applied and what estimators should be used are still not solved. By the knowledge of the classical optimal designs, we would expect the solution to be dependent on the criterion used. Thus to successfully attack these problems, in this section we shall merely focus on the situation where the loss function is a squared one and the mean functional is used to assess randomized strategies. Moreover, we shall only consider the linear estimators (but not necessarily the least squares ones). These restrictions are necessary because we need explicit expressions for the risks of randomized strategies. Confining to these and under the H-invariant considerations of Section 2, we shall obtain some general results which reduce the problem of finding a minimax strategy to the classical problem of finding an optimal design under an ideal model. For the cases where the classical optimal design theory has provided a solution, say d^0, our results then provide a minimax strategy $\mathbf{h}(d^0, \delta^0)$, where $\mathbf{h}$ is the H-uniform randomization procedure and δ^0 is the (weighted) least squares estimators; in other words, H-uniformly randomizing a classical optimal design (for suitable criterion) and using the (weighted) least squares estimators is a minimax randomized strategy when the class of possible true models is H-invariant. Specifically, we consider the following setting.

Suppose T treatments are to be assigned to the $N = \sum_{b=1}^{B} N_b$ units which are classified into B blocks, where N_b is the bth block size. Within block b, the $N_b = \prod_{i=1}^{n(b)} \ell_i^{(b)}$ units are arranged according to $n(b)$ factors so that when $n(b) \geq 2$ they form an $n(b)$-dimensional hyper-rectangle of size $\ell_1^{(b)} \times \cdots \times \ell_{n(b)}^{(b)}$, where $\ell_i^{(b)}$ is the number of levels of the ith factor in block b, and when $n(b) = 1$, the $N_b = \ell_1^{(b)}$ units are assumed to be of the same level. To avoid trivialities, we assume $\ell_i^{(b)} \geq 2$. The uth unit, when it falls in block b, is now labeled by $(i_1^{(b)}, \cdots, i_j^{(b)}, \cdots, i_{n(b)}^{(b)})$, where $1 \leq i_j^{(b)} \leq \ell_j^{(b)}$. Assume that

$$y_u = \alpha_{d(u)} + \beta_b + \sum_{j=1}^{n(b)} \beta^{(b)}_{(i_1^{(b)}, \cdots, i_j^{(b)}, \cdots, i_{n(b)}^{(b)})} + g_u + \varepsilon_u, \tag{3.1}$$

where β_b is the bth block effect and

$$\beta^{(b)}_{(i_1^{(b)}, \cdots, i_j^{(b)}, \cdots i_{n(b)}^{(b)})}$$

is the interaction effect of all but the jth factor in block b at levels $i_1^{(b)}$, $\cdots$, $i_{j-1}^{(b)}$, $i_{j+1}^{(b)}$, $\cdots$, $i_{n(b)}^{(b)}$ respectively. Note that all the lower level interactions of factors in the same block are implicit in this model. The set G of all possible $\mathbf{g}$ and the set $\mathcal{E}$ of all possible error distributions are assumed to be invariant under a group H to be specified below.

For any b and j such that $1 \leq b \leq B$ and $1 \leq j \leq n(b)$, take $H_j^{(b)}$ to be a doubly transitive group on $\{1, \cdots, \ell_j^{(b)}\}$; i.e., for any $k_1, k_2, k_3, k_4 \in \{1, \cdots, \ell_j^{(b)}\}$ such that $k_1 \neq k_2$ and $k_3 \neq k_4$, there exists some $\pi' \in G_j^{(b)}$ such that $\pi'(k_1) = k_3$ and $\pi'(k_2) = k_4$. For any $\pi' \in H_j^{(b)}$, define a permutation π on $\{1, \cdots, N\}$ by letting $\pi(u) = u$ for $u \notin$ block b, and $\pi(u) =$

$(i_1^{(b)}, \cdots, i_{j-1}^{(b)}, \pi'(i_j^{(b)}), i_{j+1}^{(b)}, \cdots, i_{n(b)}^{(b)})$ when $u = (i_1^{(b)}, \cdots, i_{n(b)}^{(b)})$. Let H_j^b be the group of all such π derived from $\pi' \in H_j^{(b)}$. It is clear that an element in H_j^b permutes the levels of factor j in block b and any two H_j^b commute. Let H be the group generated by all H_j^b, i.e.,

$$H = \prod_{b=1}^{B} \prod_{j=1}^{n(b)} H_j^b. \tag{3.2}$$

Note that (3.1) is reduced to (2.1) ((2.5); (2.6), respectively) when $B = 1$ and $n(1) = 1$ ($n(b) = 1$ for $1 \le b \le B$; $B = 1$ and $n(1) = 2$, respectively). It is also clear that the H's considered in these examples are of the form (3.2).

We are interested in estimating p contrasts among the treatment effects and want to use a squared loss function and linear estimators only. More explicitly, take $s = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{g}, \xi)$ where $\boldsymbol{\beta}$ is the vector of block and interaction effects, and define

$$\mathscr{L}(s, \mathbf{a}) = (L\boldsymbol{\alpha} - \mathbf{a})'(L\boldsymbol{\alpha} - \mathbf{a}), \tag{3.3}$$

where L is a $p \times T$ matrix with zero row sums, and $\mathbf{a} \in \mathscr{A} = R^p$. Also denote a linear estimator by a $p \times N$ matrix δ and a randomized strategy by $(\mathbf{d}, \boldsymbol{\delta})$; and recall the definition of the risk function r. We shall find a minimax randomized strategy that achieves

$$\min_{(\mathbf{d},\boldsymbol{\delta})} \max_{s \in S} Er(\mathbf{d}, \boldsymbol{\delta}; s). \tag{3.4}$$

To make risks finite, we assume that $\mathscr{E}$ contains only the probability measure ξ with finite second moments. Since only linear estimators are considered, the risks depends on the covariance matrix $\mathbf{V}$ of ξ. Thus hereafter we replace ξ by $\mathbf{V}$ and let $\mathscr{V}$ be the set of all possible $\mathbf{V}$'s.

The minimax solution of (3.4) is related to the following classical optimal design problem. Set $g_u = 0$ in (3.1). Assume that ε_u's are uncorrelated, with means 0 and known (up to a constant) variances σ_b^2 where b is the label of the block containing unit u. Under such an ideal model and the squared loss (3.3), it is clear that no randomization is necessary. Also, the best linear unbiased estimator (b.l.u.e.) is the weighted least squares one. Using the b.l.u.e., we reduce the problem to finding a design which minimizes

$$\operatorname{trace}[(L, \mathbf{0})\{X' \operatorname{diag}(\sigma_b^{-2})X\}^{-}(L, \mathbf{0})'],$$

where X is the usual design matrix, $\operatorname{diag}(\sigma_b^{-2})$ is the inverse of the covariance matrix of ε_u's, $\mathbf{0}$ is the zero matrix, and A^- denotes any generalized inverse of A. These were called the L- or linear criteria in the optimal design literature. Denote any optimum design under this criterion by d^0. For the case where the σ_b^2's are equal and L is chosen so that (3.3) becomes $\mathscr{L}(s, \alpha) = \sum_{i,j=1} (\alpha_i - \alpha_j - a_i + a_j)^2$, the linear criterion is often called the A-criterion. A-optimal designs have been found in many settings; for example, the balanced block designs (Kiefer, 1958), some group divisible designs in the block design settings (Cheng, 1978a), Generalized Youden designs (G. Y. D.) (Kiefer, 1975). When there exists a control treatment (say, the first treatment is a control), one may want to use an L for which the loss function (3.3) becomes $\mathscr{L}(\mathbf{s}, \mathbf{a}) = \sum_{i=2}^{T} (\alpha_i - \alpha_1 - a_i + a_1)^2$. Several balanced treatment incomplete block designs are found to be optimal under this criterion (Notz, 1981).

The following is a recipe for finding a minimax solution of (3.4): (i) suitably define the σ_b's by some feature of G and $\mathscr{V}$; (ii) find an L-optimal design d^0; (iii) H-uniformly randomize d^0 and use the b.l.u.e. δ^0.

In short, $\mathbf{h}(d^0, \delta^0)$ is minimax. To define σ_b we need some notation. The cardinal number of a set (or a group) A is denoted by $\#A$. For each $b \in \{1, \cdots, B\}$, let $\Lambda^b = \{\gamma \mid \gamma \subset \{1, \cdots, n(b)\}\}$ and define $H_\gamma^b = \prod_{j \in \gamma} H_j^b$ for $\gamma \in \Lambda^b$. For $\mathbf{g} \in R^N$ define $\mathbf{g}^\gamma = (-1)^k \sum_{\pi \in H_\gamma^b} \pi\mathbf{g}/\#H_\gamma^b$, where $k = \#\gamma$. Now let $\mathbf{g}_{[b]}$ be the projection of $\mathbf{g} \in R^N$ on block b, .e., the uth coordinate of $\mathbf{g}_{[b]}$ equals that of $\mathbf{g}$ or 0, depending on whether u falls in block b or not. Then define $\tilde{\mathbf{g}}_{[b]} = \sum_{\gamma \in \Lambda^b} \mathbf{g}_{[b]}^\gamma$. For a $N \times N$ matrix $\mathbf{V}$ with the uth column vector $\mathbf{v}_u$, let $\mathbf{V}_{[b]}^*$ be the $N \times N$ matrix with the uth column vector $\tilde{\mathbf{v}}_{u[b]}$ and define $\tilde{\mathbf{V}}_{[b]} = ((\mathbf{V}_{[b]}^*)')_{[b]}^*$. Finally, define

$$\sigma_b^2 = c_b \cdot \max_{(\mathbf{g},\mathbf{V}) \in G \times \mathscr{V}} \{\|\tilde{\mathbf{g}}_{[b]}\|^2 + \operatorname{trace} \tilde{\mathbf{V}}_{[b]}\}, \qquad b = 1, \cdots, B, \tag{3.5}$$

where $\|\cdot\|$ is the Euclidean norm and

$$c_b = \frac{1}{N_b} \cdot \sum_{\gamma \in \Lambda^b} \prod_{j \in \gamma} (\ell_j^{(b)} - 1)^{-1}$$

(the product over an empty set is 1).

THEOREM 3.1. *Suppose* (3.1)–(3.5) *hold. If there exists a* $(\mathbf{g}^0, \mathbf{V}^0) \in G \times \mathscr{V}$ *which simultaneously achieves all the maxima of* (3.5), *then* $\mathbf{h}(d^0, \delta^0)$ *is a randomized strategy which achieves* (3.4).

Now we apply this theorem to the settings of the three examples of Section 3. For the settings of Examples 1 and 3, we only have one block; i.e., $B = 1$. This greatly simplifies the matter. We do not have to verify the existence of $(\mathbf{g}^0, \mathbf{V}^0)$ and the b.l.u.e. is the least squares estimate. Thus in Example 1, a minimax randomized strategy is to completely randomize an L-optimal design and to use the least squares estimator. In particular, when the A-criterion is assumed, the balanced completely randomized design together with the least squares estimator is a minimax strategy. This slightly strengthens Wu's Theorem 1 and Theorem 3 which justified balance as well as randomization but the use of the least squares estimator was assumed. For Example 3, completely randomizing the rows and the columns of a G.Y.D. (whenever existent) together with the use of the least squares estimator is a minimax strategy. In general, when k-way setting is assumed and all higher order interactions are present so that (3.1) holds with $B = 1$, the minimax randomized strategy can be found in a similar way. However, for $k \geq 3$ if the setting does not include the interaction effects, then Theorem 3.1 does not apply. This is demonstrated in the following example.

EXAMPLE 4. Suppose 8 experimental units are classified by 3 factors. Each factor has 2 levels (high and low). Thus each unit can be labeled by (i, j, k) where $i, j, k = 1$ or 2. Suppose there are only two treatments. Instead of (3.1), we consider the following simpler additivity assumption:

$$\mathbf{y}_u = \alpha_{d(u)} + \beta_{1i} + \beta_{2j} + \beta_{3k} + g_u + \varepsilon_u, \tag{3.6}$$

where β_{1i} is the first factor's ith level effect, and β_{2j}, β_{3k} are defined similarly. This model is valid when the interaction effects are known to be negligible. The class G of all possible unit effects and the class $\mathscr{V}$ of all possible covariance matrices for the random errors are again assumed to be invariant under the group H of all the permutations of the factor levels. Since (3.6) is not of the form (3.1), Theorem 3.1 does not apply. In fact, for the A-criterion, the associated classical design problem (i.e., finding an A-optimal design for model (3.6) when $g_u = 0$ and the ε_u's are uncorrelated with a common variance) has the following two solutions d_1 and d_2:

d_1: $\begin{array}{|c|c|}\hline A & B \\ \hline B & A \\ \hline\end{array}$ $\begin{array}{|c|c|}\hline B & A \\ \hline A & B \\ \hline\end{array}$ $\qquad$ d_2: $\begin{array}{|c|c|}\hline A & B \\ \hline B & A \\ \hline\end{array}$ $\begin{array}{|c|c|}\hline A & B \\ \hline B & A \\ \hline\end{array}$

where A and B denote the treatment labels; for each design, units in the first and the second squares are those with the third factor at low level and high level, respectively. If the conclusion of Theorem 3.1 were true, then we would expect that randomly permuting the factor levels for d_1 and d_2 (and using least squares estimators) would yield the same maximum risks for any H-invariant G and $\mathscr{V}$ because they should be both minimax. However, the following two special G's disprove this assertion:

(i) $G_1 = \left\{ \left(\begin{array}{|c|c|}\hline 1 & 0 \\ \hline 0 & 1 \\ \hline\end{array}, \begin{array}{|c|c|}\hline 1 & 0 \\ \hline 0 & 1 \\ \hline\end{array} \right), \left(\begin{array}{|c|c|}\hline 0 & 1 \\ \hline 1 & 0 \\ \hline\end{array}, \begin{array}{|c|c|}\hline 0 & 1 \\ \hline 1 & 0 \\ \hline\end{array} \right) \right\}$

(ii) $G_2 = \left\{ \left(\begin{array}{|c|c|}\hline 1 & 0 \\ \hline 0 & 1 \\ \hline\end{array}, \begin{array}{|c|c|}\hline 0 & 1 \\ \hline 1 & 0 \\ \hline\end{array} \right), \left(\begin{array}{|c|c|}\hline 0 & 1 \\ \hline 1 & 0 \\ \hline\end{array}, \begin{array}{|c|c|}\hline 1 & 0 \\ \hline 0 & 1 \\ \hline\end{array} \right) \right\}$

Both G_1 and G_2 are H-invariant. To simplify the computation, assume that the random errors vanish. Consider the case $G = G_1$. It is easy to find that the maximum mean squared error for the strategy of H-uniformly randomizing d_1(respectively d_2) and using the least squares estimator is equal to 0 (respectively 1). Thus d_1 performs better than d_2 when $G = G_1$. However, similar observation leads to the opposite conclusion (i.e., d_2 is better than d_1) when $G = G_2$. Intuitively, d_1 should always be more desirable than d_2 since it possesses better symmetry properties. However, the above discussion shows that this intuition may sometimes be misleading. This also demonstrates the importance of Wu's rigorous treatment on the justification of the role of randomization from the model-robustness viewpoint although it seems to have been well accepted.

For the k-way ($k \geq 3$) settings without interaction effects, Cheng (1978b, 1980) showed that the Youden hyperrectangles (Y.H.R.) are A-optimal. (The two designs d_1, d_2 in Example 4 are both Y.H.R.'s). However, due to the above consideration, H-uniformly randomizing a Y.H.R. does not necessarily provide a minimax strategy. Moreover, the actual minimax strategy may depend on the actual form of G. This then creates many difficulties in finding a solution and we have no satisfactory answer yet.

We turn to the block design settings of Example 2. The H here contains all permutations within each block. It is clear that

$$\tilde{\mathbf{g}}_{[b]} = \mathbf{g}_{[b]} - \frac{1}{N_b} \cdot \left(\sum_{u \in \text{block } b} g_u \right) \mathbf{1}_{[b]}$$

where $\mathbf{1}$ is the vector of ones, $c_b = 1/(N_b - 1)$, and with $\mathbf{V} = (v_{uu'}) \in \mathscr{V}$,

$$\text{trace } \tilde{\mathbf{V}}_{[b]} = \sum_{u \in \text{block } b} v_{uu} - \frac{1}{N_b} \sum_{u,u' \in \text{block } b} v_{uu'}.$$

Thus for any specified G, it is not hard to actually compute σ_b^2. But the conclusion of Theorem 3.1 may or may not be true, depending on whether the σ_b^2's are achieved by a common $(\mathbf{g}^0, \mathbf{V}^0)$ or not. If we take, for instance,

$$G = G_1 = \{\mathbf{g} : \|\mathbf{g}\| \leq k\},\ G = G_2 = \{\mathbf{g} : |g_u| \leq k \text{ for each } u\},$$

or

$$G = G_3 = \{\mathbf{g} \mid \text{for each } b,\ \textstyle\sum_{u \in \text{block } b} g_u = 0 \quad \text{and} \quad |g_u| \leq k_b \text{ for } u \in \text{block } b\},$$

where k and the k_b's are constants, then the σ_b^2's can be achieved by a $\mathbf{g}^0$. Thus a minimax solution can be found by the general recipe given before. In particular, consider the case where block sizes are equal, $\mathscr{V}$ contains only the identity matrix, and A-criterion is desired. If the σ_b^2's are equal (for instance $G = G_1$, $G = G_2$, or $G = G_3$ with equal k_b's), then randomizing the units within each block of an A-optimal block design (e.g. a B.B.D.) is minimax. Note that in such cases we do not need to randomize the blocks. However, if we take, for instance,

$$G = G_4 = \cup_{b=1}^{B} \{\mathbf{g} \mid |g_u| \leq k \text{ for } u \in \text{block } b \text{ and } g_u = 0 \text{ elsewhere}\},$$

then for different blocks their σ_b^2's are achieved by different $\mathbf{g}$'s. For such a case, only randomizing the units within each block is not minimax. The common sense suggests that one should randomize both the blocks and the units within each block. But to justify this, we need to consider a larger transformation group H^0 which contains both the original group H and a group H_c that permutes the blocks. This enlarged group H^0 cannot be represented as the form of (3.2) because H_c and H do not commute. Therefore we need a different theorem to handle this case. The following development is mostly motivated by this consideration.

Suppose that $N_1 = N_2 = \cdots = N_B$, $n(1) = n(2) = \cdots = n(B) = n$, and $\ell_j^{(1)} = \ell_j^{(2)} = \cdots = \ell_j^{(B)} = \ell_j$, $j = 1, 2, \cdots, n$. Take a transitive group H_c' on $\{1, \cdots, B\}$; i.e., for any $k_1, k_2 \in \{1, \cdots, B\}$, there exists some $\pi' \in H_c'$ such that $\pi'(k_1) = k_2$. For any $\pi' \in H_c'$, define a permutation π on $\{1, \cdots, N\}$ by $\pi(i_1^{(b)}, \cdots, i_n^{(b)}) = (i_1^{(\pi'(b))}, \cdots, i_n^{(\pi'(b))})$; clearly, π is a block

permutation. Let H_c be the group of all such π's. Now define H^0 to be a group generated by the H of (3.2) and H_c. Denote the H^0-uniform randomization by $\mathbf{h}^0$.

THEOREM 3.2. *Suppose* (3.1) *and* (3.3) *hold. If* G *and* $\mathscr{V}$ *are invariant under* H^0, *then* $\mathbf{h}^0(d^0, \delta^0)$ *achieves* (3.4) *where* d^0 *is an* L*-optimal design (defined before) with* $\sigma_1^2 = \cdots = \sigma_B^2$, *and* δ^0 *is the least squares estimator.*

Applying this theorem to the block design setups of Example 2 with equal block sizes, we see that applying the standard randomization procedure to an A-optimal design and using the least squares estimator is minimax. Moreover, it suggests a simpler randomization procedure; i.e., instead of completely randomly permuting the blocks we may just randomly rotate the blocks. This is because we may take H_c' to be a cyclic group. This simplified procedure is not only easier to implement but also enjoys at least as many robustness properties as the standard one when the squared loss is concerned. To see this, we simply observe that if a $(G, \mathscr{V})$ is invariant under the group of all block permutations and the permutations within each block, then it is also invariant under a smaller group that contains only a cyclic group of block permutations and any permutations within each block. Similar argument also applies to the cases covered by Theorem 3.1 and we conclude that randomization procedure generated by a doubly transitive group is as good as the complete randomization. But in general a doubly transitive group with a simple form and a small order is not easy to obtain. For some particularly simple cases (e.g., a product of two cyclic groups may sometimes be doubly transitive), see Burnside (1911).

4. Proofs. To save space, some of the proofs are only outlined. For details, see Li (1981).

PROOF OF THEOREM 2.1. It is clear that we need only to show that for any $f \in \mathscr{F}$ and $s_0 \in S$,

$$\min_\phi \max_{\pi \in H} f(r(\phi(d, \delta); \pi s_0)) = \max_{\pi \in H} f(r(\mathbf{h}(d, \delta); \pi s_0)). \tag{4.1}$$

Observe that for any $\pi \in H$, $r(\pi d, \pi\delta; \pi s_0) = r(d, \delta; s_0)$; this is due to the invariance properties of the model (2.4), the estimator and the loss function (2.3). Define a real function l on H by $l(\pi) = r(\pi d, \pi\delta; s_0)$. Then,

$$r(\phi(d, \delta); \pi s_0) = r(\pi^{-1}\phi(d, \delta); s_0) = l(\pi^{-1}\phi).$$

We may write (4.1) as $\min_\phi \max_{\pi \in H} f(l(\pi\phi)) = \max_{\pi \in H} f(l(\pi \mathbf{h}))$; equivalently,

$$\min_{\mu \in \mathscr{M}} \max_{\pi \in H} f(l(\pi\mu)) = \max_{\pi \in H} f(l(\pi\mu_0)), \tag{4.2}$$

where μ_0 is the uniform distribution on H, $\mathscr{M}$ is the class of all probability measures on H, and $l(\pi\mu)$ is the distribution of $l(\pi\phi))$ when ϕ has probability measure μ.

Obviously, for any μ,

$$\frac{1}{\#H}\sum_{\pi \in H} \pi\mu = \mu_0 \quad \text{and} \quad \frac{1}{\#H}\sum_{\pi \in H} l(\pi\mu) = l(\mu_0).$$

Hence for any $\mu \in \mathscr{M}$ and $f \in \mathscr{F}$, $f(l(\mu_0)) \leq \max_{\pi \in H} f(l(\pi\mu))$ by the definition of $\mathscr{F}$.

Therefore (4.2) holds since $\pi\mu_0 = \mu_0$ for any $\pi \in H$. □

PROOF OF THEOREM 3.1. To proceed, a sequence of lemmas will be presented first.

LEMMA 4.1. *Under* (3.1)–(3.3), *we have*

$$\min_{(\mathbf{d}, \boldsymbol{\delta})} \max_{s \in S} Er(\mathbf{d}, \boldsymbol{\delta}; s) \geq \max_{s \in S} \min_{(d, \delta)} Er(\mathbf{h}(d, \delta); s). \tag{4.3}$$

Roughly speaking, this lemma suggests that in order to find a randomized strategy achieving (3.4) we may first choose a suitable non-randomized strategy (d, δ); then we

apply the H-uniform randomization on the design d, and use the estimator accordingly. We now proceed to evaluate $Er(\mathbf{h}(d, \delta); s)$. Write $\mathbf{Y} = (y_1, \cdots, y_N)'$ and $\boldsymbol{\alpha}_d = (\alpha_{d(1)}, \cdots, \alpha_{d(N)})'$; let $\mathscr{B}$ be $E\mathbf{Y}$ under $\boldsymbol{\alpha} = 0$ and $\mathbf{g} = 0$ ($\mathscr{B}$ depends only on $\boldsymbol{\beta}$); for any $\mathbf{g} \in R^N$, define $\bar{\mathbf{g}} = \sum_{\pi \in H} \pi\mathbf{g}/\#H$; for any symmetric matrix $\mathbf{V}$, define

$$\bar{\mathbf{V}} = \frac{1}{\#H} \sum_{\pi \in H} \pi(\mathbf{V}) \text{ where } \pi(\mathbf{V}) = (v_{\pi^{-1}(i)\pi^{-1}(j)}) \text{ for } \mathbf{V} = (v_{ij}).$$

LEMMA 4.2. *For any* (d, δ),

$$Er(\mathbf{h}(d, \delta); s) = \| L\boldsymbol{\alpha} - \delta(\boldsymbol{\alpha}_d + \mathscr{B} + \bar{\mathbf{g}}) \|^2 + \text{trace}\delta(\overline{(\mathbf{g} - \bar{\mathbf{g}})(\mathbf{g} - \bar{\mathbf{g}})'} + \bar{\mathbf{V}})\delta',$$

where $s = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{g}, \mathbf{V})$.

To avoid having an infinite maximum risk, it is necessary for the estimator to satisfy an unbiasedness condition when $\mathbf{g} = 0$. This will be made explicit by the notion of orbits. The orbit of an element u under a group K is the set $\{\pi(u) \mid \pi \in K\}$ and is denoted by $K(u)$. For any subset $A \subset \{1, \cdots, N\}$, define $\mathbf{1}_A$ to be the vector in R^N with the ith coordinate equal to 1 or 0 depending on whether $i \in A$ or not. Thus for $n(b) \geq 2$, $1 \leq j \leq n(b)$ and $u = (i_1, \cdots, i_{n(b)})$, $\mathbf{1}_{H^b_{j(u)}}$ is the expectation of $\mathbf{Y}$ when $\beta^{(b)}_{(i_1, \cdots, \hat{i}_j, \cdots, i_{n(b)})} = 1$ and all other parameter values are 0. Similarly, for $n(b) = 1$, $\mathbf{1}_{H^b_1(u)}$ is the expectation of $\mathbf{Y}$ when $\beta_b = 1$ and all other parameter values are 0. Let

$$W = \{\mathbf{a} \mid \mathbf{a} \in R^N, \mathbf{a}'\mathbf{1}_{H^b_{j(u)}} = 0, u \in \text{block } b, \quad b = 1, \cdots, B \text{ and } j = 1, \cdots, n(b)\}.$$

LEMMA 4.3. *A necessary condition for* $\max_{s \in S} Er(\mathbf{h}(d, \delta); s)$ *to be finite is that each column vector of* δ' *belongs to* W *and* $\delta\boldsymbol{\alpha}_d = L\boldsymbol{\alpha}$ *for all* $\boldsymbol{\alpha} \in R^T$.

Let U be the class of any non-randomized strategies (d, δ) such that δ satisfies the necessary condition in Lemma 4.3. Without loss of generality, we may restrict the randomized strategies to have supports on U. The following lemma will be used to simplify $Er(\mathbf{h}(d, \delta), s)$. Recall some notations from Section 3.

LEMMA 4.4. *For any* $\mathbf{a} \in W$, $\mathbf{g} \in R^N$, *and any symmetric* $N \times N$ *matrix* $\mathbf{V}$, *we have*

$$\mathbf{a}'\bar{\mathbf{g}} = 0, \tag{4.4}$$

$$\mathbf{a}'\overline{(\mathbf{g} - \bar{\mathbf{g}})(\mathbf{g} - \bar{\mathbf{g}})'}\mathbf{a} = \mathbf{a}'\overline{\mathbf{g}\mathbf{g}'}\mathbf{a}, \tag{4.5}$$

$$\mathbf{a}'\overline{\mathbf{g}\mathbf{g}'}\,\mathbf{a} = \sum_{b=1}^{B} c_b \| \tilde{\mathbf{g}}_{[b]} \|^2 \cdot \| \mathbf{a}_{[b]} \|^2, \tag{4.6}$$

and

$$\mathbf{a}'\bar{\mathbf{V}}\mathbf{a} = \sum_{b=1}^{B} c_b \text{ trace } \tilde{\mathbf{V}}_{[b]} \cdot \| \mathbf{a}_{[b]} \|^2. \tag{4.7}$$

We now combine these lemmas to complete the proof of Theorem 3.1.

First, by Lemmas 4.2, 4.3 and 4.4, we obtain

$$Er(\mathbf{h}(d, \delta); s) = \sum_{b=1}^{B} c_b(\| \tilde{\mathbf{g}}_{[b]} \|^2 + \text{trace } \tilde{\mathbf{V}}_{[b]}) \cdot \text{trace } \delta_{[b]}(\delta_{[b]})' \tag{4.8}$$

where $\delta_{[b]}$ is a $p \times N_b$ matrix with each row vector equal to the component of the corresponding row vector of δ on the block b.

Next, by Lemma 4.1, (4.8) and the assumptions of Theorem 3.1, we have

$$\begin{aligned}
\min_{(\mathbf{d}, \delta)} \max_{s \in S} Er(\mathbf{d}, \delta; s) &\geq \max_{s \in S} \min_{(d, \delta)} Er(\mathbf{h}(d, \delta); s) \\
&\geq \max_{s \in S} \min_{(d, \delta) \in U} \sum_{b=1}^{B} c_b(\| \tilde{\mathbf{g}}_{[b]} \|^2 + \text{trace } \tilde{\mathbf{V}}_{[b]}) \cdot \text{trace}(\delta_{[b]}\delta'_{[b]}) \\
&\geq \min_{(d, \delta) \in U} \sum_{b=1}^{B} c_b(\| \tilde{\mathbf{g}}^0_{[b]} \|^2 + \text{trace } \tilde{\mathbf{V}}^0_{[b]}) \cdot \text{trace}(\delta_{[b]}\delta'_{[b]}) \\
&= \min_{(d, \delta) \in U} \sum_{b=1}^{B} \sigma_b^2 \text{ trace}(\delta_{[b]}\delta'_{[b]}).
\end{aligned}$$

By the definition of (d^0, δ^0), we see that the above minimum is achieved by (d^0, δ^0). On the other hand, for any $s \in S$,

$$Er(\mathbf{h}(d^0, \delta^0); s) \leq \sum_{b=1}^{B} \sigma_b^2 \operatorname{trace}(\delta_{[b]}^0 \delta_{[b]}^{0\prime}).$$

Therefore, $\max_{s\in S} Er(\mathbf{h}(d^0, \delta^0); s) = \min_{(\mathbf{d},\delta)} \max_{s\in S} Er(\mathbf{d}, \boldsymbol{\delta}; s)$ and Theorem 3.1 is proved.

It remains to establish Lemmas 4.1–4.4.

PROOF OF LEMMA 4.1. For any $s^* \in S$ and any $(\mathbf{d}, \boldsymbol{\delta})$, if $\mathbf{d}$ has probability measure μ,

$$\begin{aligned}\max_{s\in S} Er(\mathbf{d}, \boldsymbol{\delta}; s) &\geq \max_{\pi\in H} Er(\mathbf{d}, \boldsymbol{\delta}; \pi^{-1}s^*) \geq \frac{1}{\#H} \sum_{\pi\in H} Er(\mathbf{d}, \boldsymbol{\delta}; \pi^{-1}s^*) \\ &= \sum_{(d,\delta)} \mu(d) \cdot \sum_{\pi\in H} r(\pi d, \pi\delta; s^*)/\#H \\ &= \sum_{(d,\delta)} \mu(d) \cdot Er(\mathbf{h}(d, \delta); s^*) \geq \min_{(d,\delta)} Er(\mathbf{h}(d, \delta); s^*).\end{aligned}$$

It follows that $\min_{(d,\delta)} \max_{s\in S} Er(\mathbf{d}, \boldsymbol{\delta}; s) \geq \min_{(d,\delta)} Er(\mathbf{h}(d, \delta); s^*)$ for any $s^* \in S$. Hence the lemma holds. □

PROOF OF LEMMA 4.2. By (3.1), (3.3), we have

$$\begin{aligned}Er(\mathbf{h}(d, \delta); s) &= Er(d, \delta; \mathbf{h}^{-1}s) \\ &= E[\| L\boldsymbol{\alpha} - \delta \cdot (\boldsymbol{\alpha}_d + \mathscr{B} + \mathbf{h}^{-1}\mathbf{g}) \|^2 + \operatorname{trace}\{\delta \mathbf{h}^{-1}(\mathbf{V})\delta'\}] \\ &= \| L\boldsymbol{\alpha} - \delta \cdot (\boldsymbol{\alpha}_d + \mathscr{B} + E\mathbf{h}^{-1}\mathbf{g}) \|^2 + \operatorname{trace}\{\delta \operatorname{Cov}(\mathbf{h}^{-1}\mathbf{g})\delta'\} + \operatorname{trace}(\delta\bar{\mathbf{V}}\delta') \\ &= \| L\boldsymbol{\alpha} - \delta \cdot (\boldsymbol{\alpha}_d + \mathscr{B} + \bar{\mathbf{g}}) \|^2 + \operatorname{trace}\{\delta(\mathbf{g} - \bar{\mathbf{g}})(\mathbf{g} - \bar{\mathbf{g}})'\delta'\} + \operatorname{trace}(\delta\bar{\mathbf{V}}\delta'). \square\end{aligned}$$

PROOF OF LEMMA 4.3. In view of Lemma 4.2, the proof becomes straightforward. □

PROOF OF LEMMA 4.4. Observe that

$$\pi\bar{\mathbf{g}} = \bar{\mathbf{g}} \quad \text{for any} \quad \pi \in H. \tag{4.9}$$

Take π to be any permutation in H_j^b and we see that $\bar{\mathbf{g}}$ is a constant for any coordinate in $H_j^b(u)$ where $u \in$ block b. By the definition of W, we see that (4.4) holds.

Next, to verify (4.5), it suffices to show $\mathbf{a}'\overline{\bar{\mathbf{g}}\mathbf{g}'}\mathbf{a} = 0 = \mathbf{a}'\overline{\bar{\mathbf{g}}\bar{\mathbf{g}}'}\mathbf{a}$. Now,

$$\mathbf{a}'\overline{\bar{\mathbf{g}}\mathbf{g}'}\mathbf{a} = \frac{1}{\#H} \sum_{\pi\in H} \mathbf{a}'\pi\bar{\mathbf{g}}(\pi\mathbf{g})'\mathbf{a} = \frac{1}{\#H} \sum_{\pi\in H} \mathbf{a}'\bar{\mathbf{g}}(\pi\mathbf{g})'\mathbf{a} = 0,$$

where the second equality holds by (4.9) and the last equality is due to (4.4). A similar argument completes the proof of (4.5).

Next, we need some lemmas to prove (4.6) and (4.7).

LEMMA 4.5. *For any* $\mathbf{a} \in W$, $\mathbf{a}'\overline{\mathbf{g}\mathbf{g}'}\mathbf{a} = \sum_{b=1}^{B} \mathbf{a}_{[b]}'\overline{\mathbf{g}_{[b]}\mathbf{g}_{[b]}'}\mathbf{a}_{[b]}$.

This lemma suggests that we may assume $B = 1$ in proving (4.6) without any loss of generality. Thus we delete the block label b from all the notations hereafter (e.g., $H_i = H_j^b$, $\tilde{\mathbf{g}} = \tilde{\mathbf{g}}_{[b]}$, $n(b) = n$, $H_\gamma^b = H_\gamma$, etc.)

LEMMA 4.6. *For* $\gamma \neq \emptyset$, $(\tilde{\mathbf{g}})^\gamma = 0$.

LEMMA 4.7. *For any* $\mathbf{a} \in W$, $\mathbf{a}'\overline{\mathbf{g}\mathbf{g}'}\mathbf{a} = \mathbf{a}'\overline{\tilde{\mathbf{g}}\tilde{\mathbf{g}}'}\mathbf{a}$.

In view of Lemma 4.6 and Lemma 4.7, to establish (4.6), we may assume $\mathbf{g}^\gamma = 0$ for $\gamma \neq \emptyset$ and show that $\mathbf{a}'\overline{\mathbf{g}\mathbf{g}'}\mathbf{a} = c \| \mathbf{g} \|^2 \| \mathbf{a} \|^2$ where $c = \sum_\gamma \prod_{j\in\gamma} (\ell_j - 1)^{-1}/N$. Thus, fixing u

$\in [1, \cdots, N]$, denoting the uth column of $\mathbf{gg}'$ by $\mathbf{w} = (w_1, \cdots, w_N)'$ and writing $\mathbf{a} = (a_1, \cdots, a_N)'$, it suffices to show that

$$\mathbf{w}'\mathbf{a} = c \| \mathbf{g} \|^2 a_u. \tag{4.10}$$

To proceed, define $U_\gamma = H_\gamma(u) - \cup_{j\in\gamma} H_{\gamma-\{j\}}(u)$ for $\gamma \neq \emptyset$. Evidently,

$$\mathbf{1}_{U_\gamma \cup \{j\}} = \sum_{u'\in U_j} \mathbf{1}_{H_j(u')} - \mathbf{1}_{U_\gamma}, \tag{4.11}$$

and

$$U_\gamma \cap U_{\gamma'} = \emptyset \quad \text{for any} \quad \gamma \neq \gamma'. \tag{4.12}$$

The following lemma is crucial to get (4.10).

LEMMA 4.8. *For any u', $u'' \in U_\gamma$, we have $w_{u'} = w_{u''}$. This constant equals $(-1)^k N^{-1} \| \mathbf{g} \|^2 \prod_{j\in\gamma} (\ell_j - 1)^{-1}$, where $k = \#\gamma$. In particular, we have $w_u = N^{-1} \| \mathbf{g} \|^2$.*

Write $\lambda_\gamma = (-1)^k N^{-1} \| \mathbf{g} \|^2 \prod_{j\in\gamma} (\ell_j - 1)^{-1}$. By (4.12) and Lemma 4.8, we get $\mathbf{w}'\mathbf{a} = \sum_{\gamma\neq\emptyset} \lambda_\gamma \mathbf{1}'_{U_\gamma}\mathbf{a} + N^{-1} \| \mathbf{g} \|^2 a_u$. Hence it suffices to show that $\mathbf{1}'_{U_\gamma}\mathbf{a} = (-1)^k a_u$, where $k = \#\gamma$. This will be proved by mathematical induction. When $\gamma = \{j\}$, by the definition of W, $0 = \mathbf{1}'_{H_j(u)}\mathbf{a} = \mathbf{1}'_{U_\gamma}\mathbf{a} + \alpha a_u$. Thus our assertion is true for $k = 1$. Suppose it is true for $\#\gamma = k$. For $j \notin \gamma$, by (4.11) and the induction hypothesis we have

$$\mathbf{1}'_{U_\gamma\cup\{j\}}\mathbf{a} = \sum_{u'\in U_\gamma} \mathbf{1}'_{H_j(u')}\mathbf{a} - \mathbf{1}'_{U_\gamma}\mathbf{a} = 0 - \mathbf{1}'_{U_\gamma}\mathbf{a} = (-1)^{k+1} a_u.$$

Therefore (4.10) is established and so is (4.6).

Turning now to (4.7), let $\mathbf{V}^{1/2}$ be any square root of $\mathbf{V}$. Denote the uth column vector of $\mathbf{V}^{1/2}$ by $\mathbf{e}_u$. By (4.6), we have

$$\mathbf{a}'\tilde{\mathbf{V}}\mathbf{a} = \mathbf{a}' \overline{\sum_{u=1}^N \mathbf{e}_u\mathbf{e}'_u}\mathbf{a} = \sum_{u=1}^N \mathbf{a}'\overline{\mathbf{e}_u\mathbf{e}'_u}\mathbf{a} = \sum_{b=1}^B \sum_{u=1}^N c_b \| \tilde{\mathbf{e}}_{u[b]} \|^2 \cdot \mathbf{a}'_{[b]}\mathbf{a}_{[b]}.$$

Thus, it suffices to show that trace $\tilde{\mathbf{V}}_{[b]} = \sum_{u=1}^N \| \tilde{\mathbf{e}}_{u[b]} \|^2$. Let A_b be the $N \times N$ matrix such that $A_b\mathbf{g} = \tilde{\mathbf{g}}_{[b]}$ for any $\mathbf{g}$. Now we have

$$\begin{aligned}\sum_{u=1}^N \| \tilde{\mathbf{e}}_{u[b]} \|^2 &= \text{trace} \sum_{u=1}^N \tilde{\mathbf{e}}_{u[b]}\tilde{\mathbf{e}}'_{u[b]} = \text{trace} \sum_{u=1}^N A_b\mathbf{e}_u\mathbf{e}'_uA'_b = \text{trace } A_b\mathbf{V}A'_b \\ &= \text{trace } \mathbf{V}^*_{[b]}A'_b = \text{trace } A_b(\mathbf{V}^*_{[b]})' = \text{trace } \tilde{\mathbf{V}}_{[b]}.\end{aligned}$$

Hence (4.7) is established. The proof of Lemma 4.4 is complete. □

PROOF OF LEMMA 4.5. An argument similar to that used in proving (4.4) leads to the conclusion that for any u, u' in different blocks b, b', the (u, u')th cell of $\mathbf{gg}'$ equals the $(u, \pi(u'))$th cell, for any $\pi \in H_j^{b'}$, $j = 1, \cdots, n(b')$. In view of this and by the definition of W, Lemma 4.5 follows. □

PROOF OF LEMMA 4.6. This can be verified by using mathematical induction and observing the following two facts:

$$(\mathbf{g}_1 + \mathbf{g}_2)^\gamma = \mathbf{g}_1^\gamma + \mathbf{g}_2^\gamma \quad \text{for any} \quad \mathbf{g}_1, \mathbf{g}_2 \in R^N; \tag{4.13}$$

$$\begin{aligned}(\mathbf{g}^\gamma)^{\{j\}} &= \mathbf{g}^{\gamma\cup\{j\}} \quad \text{if} \quad j \notin \gamma, \\ &= -\mathbf{g}^\gamma \quad \text{if} \quad j \in \gamma. \square\end{aligned} \tag{4.14}$$

PROOF OF LEMMA 4.7. Write $\tilde{\mathbf{g}} = \mathbf{g} + \sum_{\gamma\neq\emptyset} \mathbf{g}^\gamma$ and compare both sides of the equation to be established in Lemma 4.7. It suffices to show that $\mathbf{a}'\pi\mathbf{g}^\gamma = 0$ for any $\pi \in H$, any $\gamma \neq \emptyset$ and any $\mathbf{a} \in W$. Since W is invariant under H, we may only show $\mathbf{a}'\mathbf{g}^\gamma = 0$ for any $\mathbf{a} \in W$. Now, this assertion can be verified by using mathematical induction, (4.14), and the definition of W. □

PROOF OF LEMMA 4.8. Since each H_j is doubly transitive, it is clear that for any u' and $u'' \in U_{\{j\}}$ (which implies $u \neq u'$ and $u \neq u''$) there exists a π such that $\pi u = u$, $\pi u' = u''$. Now, since $\mathbf{gg}' = \pi \mathbf{gg}' \pi$, we get $w_{u'} = w_{u''}$ by comparing the (u, u')th cells of these two matrices. Thus the first statement is proved for $\#\gamma = 1$. For general γ, the proof is similar. We now compute the constant for each U_γ by mathematical induction.

First, the diagonal elements of $\mathbf{gg}'$ are the same due to the transitivity of the group $\prod_{j=1}^{n} H_j$. This constant is easily verified to be $N^{-1} \| \mathbf{g} \|^2$, by considering the trace of $\mathbf{gg}'$. Hence we have shown $w_u = N^{-1} \| \mathbf{g} \|^2$.

Next, let $\mathbf{z} = (z_1, \cdots, z_n)' = \mathbf{hg}$. Clearly, $E\mathbf{zz}' = \mathbf{gg}'$ and $w_{u'} = E z_u z_{u'}$. By the assumption that $\mathbf{g}^{(j)} = 0$, it follows that $\sum_{\pi \in H_j} z_{\pi(u)} = 0$. Thus $-z_u^2 = \sum_{u' \in U_{(j)}} z_u z_{u'}$. Taking the expectations on both sides, we obtain $-w_u = \sum_{u' \in U_{(j)}} w_{u'} = (\ell_j - 1) w_{u'}$. Hence $w_{u'} = -(\ell_j - 1)^{-1} \cdot N^{-1} \cdot \| \mathbf{g} \|^2$, as desired.

Next, suppose that our lemma is true for some γ. We shall find the desired constant for $\gamma \cup \{j\}$ where $j \notin \gamma$. For any $u' \in U_{\gamma \cup \{j\}}$, there exist some $\pi_1 \in H_\gamma$ and some $\pi_2 \in H_j$ such that $\pi_1(u) \in U_\gamma$ and $\pi_2 \pi_1(u) = u'$. Now, $\mathbf{g}^{(j)} = 0$ implies $\sum_{\pi \in H_j} z_{\pi \pi_1(u)} = 0$ and thus we get $z_u z_{\pi_1(u)} = -\sum_{\pi \in H_j - \{\mathcal{I}\}} z_u z_{\pi \pi_1(u)}$, where $\mathcal{I}$ is the identity permutation. Observe that $\pi \pi_1(u) \in \cup_{\gamma \cup \{j\}}$ for any $\pi \in H_j - \{\mathcal{I}\}$ and take the expectations on both sides of the last equality. It follows that $w_{\pi_1(u)} = -(\ell_j - 1) w_{u'}$. Hence

$$w_{u'} = -(\ell_j - 1)^{-1} w_{\pi_1(u)} = (-1)^{k+1} N^{-1} \| \mathbf{g} \|^2 \prod_{i \in \gamma \cup \{j\}} (\ell_i - 1)^{-1},$$

where the last equality is due to the induction hypothesis. The proof for Lemma 4.8 is now complete. □

PROOF OF THEOREM 3.2. Define

$$\mathbf{g}^0 = \frac{1}{\#H^0} \sum_{\pi \in H^0} \pi \mathbf{g} \quad \text{and} \quad \mathbf{V}^0 = \frac{1}{\#H^0} \sum_{\pi \in H^0} \pi(\mathbf{V}).$$

By arguments similar to those in the proof of Lemmas 4.1–4.3, we see that

$$Er(\mathbf{h}^0 (d, \delta); s) = \| \delta \mathbf{g}^0 \|^2 + \operatorname{trace}[\delta \{(\mathbf{g} - \mathbf{g}^0)(\mathbf{g} - \mathbf{g}^0)' + \mathbf{V}\}^0 \delta'], \tag{4.15}$$

for (d, δ) such that the maximum risk is finite. Write $\mathbf{M} = (\mathbf{g} - \mathbf{g}^0)(\mathbf{g} - \mathbf{g}^0)' + \mathbf{V}$. We claim that for any $\mathbf{a} \in W$,

$$\mathbf{a}' \mathbf{g}^0 = 0 \tag{4.16}$$

and

$$\mathbf{a}' \mathbf{M}^0 \mathbf{a} = \lambda \| \mathbf{a} \|^2 \tag{4.17}$$

for a constant λ depending on $\mathbf{M}$ but not on $\mathbf{a}$.

First, observe that

$$\mathbf{g}^0 = \frac{1}{\#H_c} \sum_{\pi \in H_c} \pi \bar{\mathbf{g}}.$$

Then (4.16) follows easily from (4.4). Next,

$$\begin{aligned}
\mathbf{a}' \mathbf{M}^0 \mathbf{a} &= \frac{1}{\#H_c} \sum_{\pi \in H_c} \mathbf{a}' \pi \bar{\mathbf{M}} \pi^{-1} \mathbf{a} \qquad \text{(where } \pi \text{ is treated as a matrix)} \\
&= \frac{1}{\#H_c} \sum_{\pi \in H_c} \sum_{b=1}^{B} c_b \operatorname{trace} \tilde{\mathbf{M}}_{[b]} \cdot \| (\pi^{-1} \mathbf{a})_{[b]} \|^2 \qquad \text{(by (4.7))} \\
&= \sum_{b=1}^{B} c_b \operatorname{trace} \tilde{\mathbf{M}}_{[b]} \cdot \left(\frac{1}{\#H_c} \sum_{\pi \in H_c} \| \mathbf{a}_{[\pi^{-1}(b)]} \|^2 \right) \\
&= \sum_{b=1}^{B} c_b \operatorname{trace} \tilde{\mathbf{M}}_{[b]} \cdot (B^{-1} \| \mathbf{a} \|^2) \\
&= (B^{-1} \sum_{b=1}^{B} c_b \operatorname{trace} \tilde{\mathbf{M}}_{[b]}) \cdot \| \mathbf{a} \|^2 \equiv \lambda \cdot \| \mathbf{a} \|^2,
\end{aligned}$$

where the fourth equality is due to the transitivity of H_c'. Thus (4.17) holds. From (4.15)–(4.17), we get

$$Er(\mathbf{h}^0(d, \delta); s) = \lambda(s)\text{trace}(\delta\delta'),$$

where $\lambda(s)$ is some positive constant depending on s. The rest of the proof is similar to that of Theorem 3.1.

Acknowledgments. This paper is based on parts of my Ph.D. thesis. I am indebted to my advisor, Professor Jack Kiefer, for guidance, encouragement, and many valuable suggestions. I sincerely thank an associate editor, a referee, Professor C. S. Cheng and Professor C. F. Wu for providing me with numerous useful comments which greatly improved the presentation and quality of this paper. Finally, this paper is dedicated to the memory of Professor Jack Kiefer, a great teacher, advisor, and friend.

REFERENCES

BLACKWELL, D. and GIRSHICK, M. A. (1954). *Theory of Games and Statistical Decisions.* Wiley, New York.

BECHHOFER, R. E. and TAMHANE, A. C. (1981). Incomplete block designs for comparing treatments with a control: general theory. *Technometrics* **23** 45–58.

BOX, G. E. P. and DRAPER, N. R. (1959). A basis for the selection of a response surface design. *J. Amer. Statist. Assoc.* **54** 622–654.

BURNSIDE, W. (1911). *Theory of Group with Finite Order*, 2nd ed., Cambridge University Press, London.

CHENG, C. S. (1978a). Optimality of certain asymmetrical experimental designs. *Ann. Statist.* **6** 1239–1261.

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

CHENG, C. S. (1980). Optimal designs for the elimination of multi-way heterogeneity. II. *IMS Bull.* 80t–57.

CHENG, C. S. and LI, K. C. (1981). Identifying the covariance structures equivalent to those with homoscedastic uncorrelated errors for the estimation and design problems. Unpublished manuscript.

FEDOROV, V. V. (1972). *Theory of Optimal Experiments.* Academic, New York.

HUBER, P. J. (1975). Robustness and designs, in *A Survey of Statistical Design and Linear Models*, (J. N. Srivastava, ed.), 287–301. North Holland, Amsterdam.

KARLIN, S. and STUDDEN, W. J. (1966). Optimal experimental designs. *Ann. Math. Statist.* **37** 783–815.

KIEFER, J. (1958). On the nonrandomized optimality and randomized non-optimality of symmetrical designs. *Ann. Math. Statist.* **29** 675–699.

KIEFER, J. (1974). General equivalence theory for optimum designs (approximation theory). *Ann. Statist.* **2** 849–879.

KIEFER, J. (1975). Constructions and optimality of generalized Youden designs, in *A Survey of Statistical Designs and Linear Models* (J. N. Srivastava, ed). 333–353. North Holland, Amsterdam.

LEHMANN, E. (1971). *Testing Statistical Hypotheses.* Wiley, New York.

LI, K. C. (1981). Minimax property for randomized designs—a unified approach. Mimeograph Series #81-44, Department of Statistics, Purdue University.

NOTZ, W. (1981). Optimal incomplete block designs for comparing treatments to a control. Mimeograph Series #81-34, Department of Statistics, Purdue University.

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

DEPARTMENT OF STATISTICS
PURDUE UNIVERSITY
WEST LAFAYETTE, INDIANA 47907