ON THE NONRANDOMIZED OPTIMALITY AND RANDOMIZED NONOPTIMALITY OF SYMMETRICAL DESIGNS[1]

By J. Kiefer[2]

Cornell University

0. Summary. Many commonly employed symmetrical designs such as Balanced Incomplete Block Designs (BIBD's), Latin Squares (LS's), Youden Squares (YS's), etc., are shown to have optimum properties among the class of *non-randomized*[1] designs (Section 3). This represents an extension of a property first proved by Wald for LS's in [1]; a similar property demonstrated by Ehrenfeld for LS's in [2] (as well as a third optimum property considered here) is shown to be an immediate consequence of the Wald property, and the Wald property is shown to be the more relevant when one considers optimality rigorously (Section 2). Surprisingly, all of these optimum properties fail to hold if *randomized*[1] designs are considered (Section 4); the results of Sections 2 and 3, as well as those appearing previously in the literature (as in [1], [2], [3]) must be interpreted in this sense. Generalizations of the BIBD's and YS's, for which analogous results hold, are introduced.

1. Introduction. Wald [1] stated an optimality criterion (called E-optimality in Section 2) for designs used in testing hypotheses in the setting of two-way soil heterogeneity where LS's are commonly employed, and succeeded in proving that a slightly different criterion (called D-optimality in Section 2) is satisfied by the LS design. Wald also stated that an analogous result holds for Graeco-Latin Squares and higher Latin Squares. This statement gives rise to speculation when one considers that, in a 3×3 Graeco-Latin Square (or, more generally, in an $n \times n$ square of order $n - 1$), there are no degrees of freedom for error: this implies that any test (e.g., of the hypothesis H_0 that there are no treatment effects) whose size (= supremum of the power function under H_0) is α, has a power function whose infimum over any of the contours usually considered ($\psi(\mu)/\sigma^2$ = constant, as discussed in the sequel) is $\leqq \alpha$. It is easy to construct a better design, i.e., one for which the infimum of the power function of some test over such a contour is > the size of the test; for example, for each of the two

Received July 8, 1957; revised January 22, 1958.

[1] One of the referees of this paper felt that the following remark on nomenclature should be included: Throughout this paper, the term *randomized design* is used in describing a statistical procedure which chooses according to a prescribed probability mechanism a member of a given class of ordinary designs, the chosen design being the one actually used; a precise definition is given in the text. The properties of such a procedure take into account the probabilities of the various possible choices. A *nonrandomized* design chooses one member of the given class with probability one. The customary usage of the phrase *randomized design* in the design of experiments can be viewed as a special case of the decision-theoretic usage employed here, but the reader is warned not to interpret the phrase in that narrower sense.

[2] Research sponsored by the Office of Naval Research.

factors, with probability $\frac{1}{2}$ use an ordinary LS design on the three levels of that factor holding the level of the other factor fixed.[3]

The phenomenon just described makes one wonder whether the optimality result for ordinary LS's also fails to hold if one permits comparison with randomized designs.[1] At the same time, the question arises whether an analogue of the limited optimality property of the LS (or Graeco-LS) design holds in a wide class of design settings for designs with suitable symmetry properties, and whether these designs fail to be optimum when compared with randomized designs.[1] This paper answers these questions affirmatively.

In Section 2A we define four optimality criteria (D-, E-, M-, and L-optimality) for designs (especially, for the normal case); Wald [1] and Ehrenfeld [2] proved D- and E-optimality, respectively, for the LS design. It is indicated why M-optimality, the strongest and least artificial of the four, seems very difficult to verify in most problems (although L-optimality, which is a local version of M-optimality, can sometimes be verified). At the same time, we list briefly for later reference the known results on the Analysis of Variance Test which are used in optimality considerations, and point out the incorrectness of tacitly assuming (as previous work in this area has done) that one should use that test, whatever design is chosen. In Section 2B we indicate by example why E-optimality seems, at least in the present state of knowledge indicated in 2A, the least satisfactory of the criteria considered; the connection of D-optimality with Isaacson's notion of type D tests [11] is examined. In Section 2C it is shown in a general setting where there is suitable symmetry that D-optimality implies E-optimality and L-optimality.

In Section 3A it is indicated why the treatment of LS's is much simpler than that of YS's, BIBD's, etc., and the general treatment of incomplete block designs

[3] It should be evident that the example of the 3×3 Graeco-Latin square, as well as the example discussed in the fourth paragraph below wherein two observations are taken, are of no *practical* importance; these simple examples are given to illustrate the general principles of Section 4. Those principles show that a precise study of certain optimality criteria for designs associated with familiar problems of testing hypotheses, can lead to the unexpected conclusion that certain intuitively unappealing randomized designs are superior to certain intuitively appealing nonrandomized symmetrical designs. The principles are less transparent (although applicable) in the context of applicationally meaningful problems such as those of Section 4, than in the simple examples; hence, the latter examples are discussed first. The present comments are included because two referees apparently read these simple examples as practical suggestions. In the same light, it is clear that the design δ in the fourth paragraph below, as well as its analogues in Section 4, is not suggested to the practical worker who wants *estimates* of all treatment effects; for these designs illustrate a non-optimality property of classical nonrandomized symmetrical designs in *hypothesis testing*, and a local property at that (see Section 5.4). In fact, the results of Section 4 are not even relevant for most estimation problems (see Section 5.2). To the practical worker who objects (as at least one has) to the conclusions of Section 4 on the grounds that one should not use a design which does not estimate all treatment effects, it should be pointed out that (1) the classical nonrandomized symmetrical design may still possibly possess certain *global* optimality properties (see Section 5.4), and (2) perhaps his problem is not really one of testing hypotheses.

of Bose [4] is briefly recalled; this treatment proves more useful in Section 3C than the more direct least squares approach used in [1] and [2] would be. In Section 3B several algebraic propositions (emphasizing the role of symmetry) are verified, which can be used to prove D- and E-optimality in important examples. Several such examples are considered in Section 3C, including generalizations of the BIBD's and the YS's.

Section 4 contains two theorems the consequences of which are that non-randomized symmetrical designs are not optimum if randomization is permitted. In Section 4B it is shown that, whether or not the variance is known, for α sufficiently small there is a randomized design whose power function is uniformly larger than that of the symmetrical design in some neighborhood of the hypotheses H_0 that all treatment effects are the same. This is slightly less transparent than the result of Section 4A, which gives an analogous result for *all* α when the above H_0 is replaced by the hypothesis that all treatment effects are equal to some specified value. The latter result can best be understood by considering the simplest example[3]: Suppose X_{ij} normal with unit variance and mean μ_i and that all X_{ij} are independent ($i, j = 1, 2$). Our problem is to select (before observation) exactly two of the X_{ij} and use them to test $\mu_1 = \mu_2 = 0$ against some class of alternatives. The symmetrical design d(say) selects X_{11} and X_{21} and uses the usual χ^2 test, and obviously has constant power $> \alpha$ on the contour $\mu_1^2 + \mu_2^2 = c > 0$, while either of the designs d_i ($i = 1, 2$), where d_i uses X_{i1} and X_{i2}, has α for the infimum of the power function on this contour. Let δ be the randomized design[1] obtained by using d_1 or d_2 with probability $\frac{1}{2}$ each. It is easily seen that, for μ_1 and μ_2 near 0, the power function of δ is $\alpha + c_1(\mu_1^2 + \mu_2^2)$ + terms of higher order, where $c_1 > 0$. Thus, on the contour $\mu_1^2 + \mu_2^2 = c > 0$ with c small, the power function of δ is almost constant and hence approximately equal to the value at $\mu_1 = \mu_2 = (c/2)^{\frac{1}{2}}$. Thus, in comparing d and δ near H_0, we may to a first approximation assume $\mu_1 = \mu_2$. But δ is clearly optimum for testing $\mu_1 = \mu_2 = 0$ *assuming* $\mu_1 = \mu_2$, while d (whose test is based on $X_1^2 + X_2^2$) is not. This explains why, for c small, δ has a power function greater than that of d.

Many of the results of this paper have counterparts for problems of point and interval estimation, for other distributions, etc. Such extensions and generalizations, as well as various other remarks, are stated in Section 5.

In design settings where no suitably symmetric design exists, it is often tedious algebraically to show that a design which is "closest to symmetrical" is optimum (if it *is* optimum: see the example of Section 2B), and we omit such considerations here. On the other hand, the conclusions of Section 4 have little to do with whether or not symmetrical designs are being considered.

Throughout this paper, except where explicitly stated to the contrary, Y will denote an N element column vector whose components Y_i are independent normal random variables with common variance σ^2 (it will be explicitly stated whenever σ^2 is assumed known; whether or not σ^2 is known has very little effect on our results); μ is an unknown m-vector, X_d is a known $N \times m$ matrix depending on an index d (the "design") and which will be described further below, and the expected value of Y when μ and σ^2 are the parameter values and when the design d is used is

$$E_{\mu,\sigma;d}Y = X_d\mu. \tag{1.1}$$

X_d is, within limits, subject to choice by the experimenter. (In many applications it is a matrix of zeros and ones.) We denote by Δ the set of choices of the index d which are available to the experimenter. A randomized design[1] δ is a probability measure on Δ (the latter will usually be finite in this paper, and measurability considerations will be trivial otherwise) which is used by selecting a d from Δ according to this measure and then using the selected d. We denote the class of available δ by Δ_R.

In many problems, one imposes an additional assumption of the form $\Gamma\mu = \gamma$, where Γ and γ are known $g \times m$ and $g \times 1$ matrices. Such an assumption can be absorbed into (1.1) and we suppose this to have been done, with no loss of generality.

A hypothesis H_0 will in this paper be of the form $R\mu = 0$, where R is a specified $r \times m$ matrix ($r \leqq m$) which we can take to be of rank r with no loss of generality. For simplicity, we can think of the class H_1 of alternatives as being all μ for which $R\mu \neq 0$. (For simplicity, we assume that σ^2 is either known exactly or else is known only to be positive, under both H_0 and H_1.) A hypothesis of the form $R\mu = \rho$ is easily reduced to the above form by letting p satisfy $Rp = \rho$ and replacing Y by $Y^* = Y - X_d p$ and μ by $\mu^* = \mu - p$ in (1.1).

We introduce some notation to be used in Section 2. We denote the $k \times k$ identity matrix by I_k. The transpose of a matrix A is written A'. It may or may not be that all r elements of $R\mu$ are estimable when a given design d is used. Suppose that there are s_d linearly independent linear combinations of the elements of $R\mu$ which have unbiased estimators when d is used, but not $s_d + 1$ such combinations. Then there is an $s_d \times r$ matrix Q_d such that there exist linear unbiased estimators of all components of $Q_d R\mu$ when design d is used; let t_d be the s_d-vector of such estimators with minimum variance ("best linear estimators" or b.l.e.'s), and let $\sigma^2 V_d$ be the convariance matrix of the components of t_d. When $s_d = r$, we may take Q_d to be the identity; for this choice of Q_d, we shall denote V_d by $\bar{V}_d$. Let b_d be the rank of X_d. Then there are b_d linearly independent combinations of the components of μ which are estimable when d is used. Of these, s_d of them can be taken to be the elements of $Q_d R\mu$; thus, there exists a $(b_d - s_d) \times m$ matrix J_d of rank $b_d - s_d$ whose rows are orthogonal to those of $Q_d R$ (i.e., $J_d' Q_d R = 0$) and such that all components of $J_d\mu$ have unbiased estimates when d is used. Let L_d be the $b_d \times m$ matrix whose first $b_d - s_d$ rows are J_d and whose last s_d rows are $Q_d R$. Let $\bar{S}_d$ be the usual best unbiased estimator of σ^2 (if it is unknown), so that $(N - b_d)\bar{S}_d/\sigma^2$ has the χ^2-distribution with $h_d = N - b_d$ degrees of freedom (it may be that $h_d = 0$ and there is no $\bar{S}_d$). For any test ϕ_d associated with d, let $\beta_{\phi_d}(\mu, \sigma^2)$ be the power function of ϕ_d (of course, β_{ϕ_d} actually depends on μ only through $L_d\mu$). For $0 < \alpha < 1$ we denote by

$H_d(\alpha)$ the class of all ϕ_d of size α, i.e., all ϕ_d for which

$$\beta_{\phi_d}(\mu, \sigma^2) \leqq \alpha \text{ whenever } R\mu = 0; \tag{1.2}$$

and by $H_d^*(\alpha)$, the class of similar tests of size α, i.e., those for which (1.2) holds with the inequality sign replaced by equality. Finally, let $F_{d,\alpha}$ denote the usual F-test of H_0 of size α with s_d and h_d degrees of freedom, based on $t_d' V_d^{-1} t_d / s_d \bar{S}_d$ (if σ^2 is known, this is replaced by the appropriate χ^2-test).

The symbol $g_{i,j}(\alpha)$ is used to denote the derivative at H_0 of the power function of the F-test of size α and i, j degrees of freedom, with respect to (a common choice of) the parameter on which it depends; specifically, if $r = m = i$, $N - r = j$, the matrices R, Q_d, and V_d are the identity, and the true values of μ and σ^2 are such that $\mu'\mu/\sigma^2 = \lambda$, then, as $\lambda \to 0$, the power function of $F_{d,\alpha}$ is

$$\alpha + g_{i,j}(\alpha)\lambda + O(\lambda^2). \tag{1.3}$$

The results of this paper can be stated in a very general setting involving invariance of Δ, of the restriction $R\mu = 0$, and of a generalization of the function ψ considered below, as well as of certain designs, under an appropriate group of permutations of the components of μ. However, in order to make our proofs (and, in particular, the role of symmetry) as transparent as possible, we will carry them out in two cases; the reader will not find it difficult to state our results more generally by making appropriate linear transformations, etc. The two cases (Δ and X_d being further specified in particular examples; the role of the function ψ which distinguishes contours on which the power function is examined, will be seen in Section 2A) are:

CASE I: $$\psi(\mu) = \sum_1^u \mu_i^2 \text{ and } R = R_{\mathrm{I}};$$

CASE II: $$\psi(\mu) = \sum_1^u (\mu_i - \bar{\mu})^2 \text{ and } R = R_{\mathrm{II}};$$

here we have written $\mu' = (\mu_1, \cdots, \mu_m)$, and $\bar{\mu} = \sum_1^u \mu_i/u$, while R_{I} is the $u \times u$ identity followed by $m - u$ columns of zeros (so $R_{\mathrm{I}}\mu = 0$ means $\mu_1 = \cdots = \mu_u = 0$), and R_{II} is a $(u - 1) \times u$ matrix P followed by $m - u$ columns of zeros, where P consists of the last $u - 1$ rows of a $u \times u$ orthogonal matrix $\bar{O}$ whose first row elements are all $1/\sqrt{u}$ (so $R_{\mathrm{II}}\mu = 0$ means $\mu_1 = \cdots = \mu_u$). The optimality results which hold in Case I are usually much more trivial to obtain than those of Case II, and Section 3B will therefore be mainly devoted to results applicable to the latter case, it being clear how to obtain the corresponding results in the former case.

2. Optimality criteria.

2A. *Preliminaries*. For a fixed design d, the test $F_{d,\alpha}$ is known to have several optimum properties, which we now list (there are obvious analogues when σ^2 is known):

(a) If $s_d = 1$ (and only then), among tests in $H_d(\alpha)$ which are unbiased (this implies that the tests are in $H_d^*(\alpha)$), $F_{d,\alpha}$ is uniformly most powerful (UMP). See [5] (a trivial completeness argument characterizing similar tests is all that is required to allow the $J_d\mu$ which is not present in [5] to be introduced, carrying through the argument there for each fixed value of the b.l.e. of $J_d\mu$).

(b) Among tests in $H_d(\alpha)$, $F_{d,\alpha}$ is UMP invariant (under the usual group of transformations when the problem is reduced to canonical form). See [5].

(c) (Wald's theorem) Among tests in $H_d^*(\alpha)$, for each $c > 0$, $\sigma^2 > 0$, and value of $J_d\mu$, the test $F_{d,\alpha}$ maximizes the Lebesgue integral of $\gamma_{\phi_d}(\nu, J_d\mu, \sigma^2)$ on the sphere $\nu'\nu = c$, where $\nu = G_dQ_dR\mu$ with G_d nonsingular $s_d \times s_d$ is such that the b.l.e.'s of the components of ν have σ^2 times the identity for their covariance matrix (i.e., ν is the vector of parameters about which H_0 is concerned in the canonical form of the problem), and where $\gamma_{\phi_d}(G_dQ_dR\mu, J_d\mu, \sigma^2) = \beta_{\phi_d}(\mu, \sigma^2)$. See [6] or [7] (the parenthetical remark at the end of (a) is relevant to [7] here).

(d) (Hsu's theorem, a consequence of (c)) Among tests in $H_d(\alpha)$ whose power function depends only on $\lambda_d = \mu'R'Q_d'V_d^{-1}Q_dR\mu/\sigma^2$ (this implies that the tests are in $H_d^*(\alpha)$), $F_{d,\alpha}$ is UMP. See [8].

(e) Among tests in $H_d(\alpha)$, $F_{d,\alpha}$ is minimax (over H_1) for a variety of weight functions, e.g., any nonnegative function of the λ_d of (d); in particular, $F_{d,\alpha}$ maximizes the minimum power on the contour $\lambda_d = c$ for each $c > 0$. See [9] or [10] (the result follows from (c) if we restrict consideration to $H_d^*(\alpha)$).

(f) (A special case of (e)) $F_{d,\alpha}$ is most stringent in $H_d(\alpha)$. See [9] or [10].

(g) (A consequence of (c)) $F_{d,\alpha}$ is of type D in $H_d(\alpha)$. (See [11] or Section 2B below for definition of type D, and Section 2B for a proof.)

It is to be noted that all the above criteria of optimality of the test $F_{d,\alpha}$ are relative to the design d. Thus, it is an error to assume (as has been done in previous papers on optimum designs) in a logical approach to optimum design problems that one should automatically use the test $F_{d,\alpha}$, whatever the chosen d, *when a reasonable criterion for optimality of a design, or of a test for a given design, may dictate the use of a test other than* $F_{d,\alpha}$. In fact, the example of Section 2B really illustrates that the use of $F_{d,\alpha}$ need not lead to an optimum design or test for many reasonable definitions of optimality; and the fact that it seems difficult (for many reasonable optimality criteria such as M-optimality, and for many common design problems) to characterize the appropriate test, is what makes it much harder than it has been thought to give a rigorous demonstration of the optimality of various common designs. We now list four optimality criteria for designs (there are many other obvious similar ones); the discussion of their meaning immediately follows the fourth definition.

M-optimality: For $c > 0$ and $0 < \alpha < 1$, a design d^* is said to be *$M_{\alpha,c}$-optimum in* Δ if, for some $\phi_{d^*}^*$ in $H_{d^*}(\alpha)$,

$$\inf_{\Gamma_c} \beta_{\phi^*_{d^*}}(\mu, \sigma^2) = \max_{d\epsilon\Delta} \sup_{\phi\epsilon H_d(\alpha)} \inf_{\Gamma_c} \beta_\phi(\mu, \sigma^2), \tag{2.1}$$

where Γ_c is the set of all μ, σ^2 for which $\psi(\mu)/\sigma^2 = c$.

L-optimality: A design is said to be *L_α-optimum in* Δ if, for some $\phi^*_{d^*}$ in $H_{d^*}(\alpha)$,

$$\lim_{c \to 0} [a_{\phi^*_{d^*}}(c) - \alpha]/[b(c) - \alpha] = 1, \tag{2.2}$$

where $a_{\phi^*_{d^*}}(c)$ and $b(c)$ are the expressions on the left and right sides of (2.1), respectively. A design is said to be *L-optimum in* Δ if it is L_α-optimum in Δ for $0 < \alpha < 1$.

D-optimality: A design d^* is said to be *D-optimum in* Δ if

$$\det \bar{V}_{d^*} = \min_{d \varepsilon \Delta'} \det \bar{V}_d, \tag{2.3}$$

where Δ' is the set of d in Δ for which $s_d = r$, and if $d^* \varepsilon \Delta'$.

E-optimality: A design d^* is said to be *E-optimum in* Δ if

$$\pi(\bar{V}_{d^*}) = \min_{d \varepsilon \Delta'} \pi(\bar{V}_d) \tag{2.4}$$

and if d^* is a member of Δ', where $\pi(\bar{V}_d)$ is the maximum eigenvalue of $\bar{V}_d$.

The above definitions will also be used with Δ replaced by Δ_R. In that case, for any δ, $\bar{V}_\delta^{-1}$ is defined to be the expected value under δ of $\bar{V}_d^{-1}$, the latter being replaced by the inverse of the covariance matrix of the b.l.e. of the estimable components of $R\mu$ (with zeros adjoined to this inverse in appropriate places to make it $r \times r$) if $s_d < r$; Δ_R' is then the set of δ for which $\bar{V}_\delta^{-1}$ is nonsingular. (This $\bar{V}_\delta^{-1}$ appears in computing certain β_{ϕ_δ} near H_0.)

D-optimality and E-optimality have been discussed in [1] and [2] and will also be discussed in Section 2B, where it will be seen that they have to do with local properties (near H_0) or optimum properties *assuming the use of* $F_{d,\alpha}$. Unfortunately, $M_{\alpha,c}$-optimality in Δ (or, better, $M_{\alpha,c}$-optimality in Δ simultaneously for all c) seems very difficult to verify, even in many simple problems, although it does not require much temerity to conjecture that it holds in such cases as those discussed in Section 2C. A similar remark applies to L-optimality (see, however, Lemma 2.2), a local version (near H_0) of M-optimality. The source of this difficulty in verifying M-optimality is illustrated by the example of Section 2B; it is simply that for fixed d the test which achieves the supremum over ϕ on the right side of (2.1) need not be $F_{d,\alpha}$ and is generally hard to compute (as is therefore the right side of (2.1)).

2B. *D- and E- optimality.* We begin by describing the meaning of E-optimality (which criterion is stated in [1] and is verified for the LS design in [2]). Suppose for fixed α, that we agreed to restrict ourselves to using $F_{d,\alpha}$, whatever d is chosen. The power function of $F_{d,\alpha}$ is then a strictly increasing function of λ_d (defined in Section 2A(d)). Now, in either Case I or II, for any $c > 0$, if we want a design d for which $F_{d,\alpha}$ maximizes the minimum power on the contour $\psi(\mu)/\sigma^2 = c$ (i.e., *which is $M_{\alpha,c}$-optimum in Δ under the additional restriction that we use $F_{d,\alpha}$*), we may restrict our attention to Δ' (since, for $s_d < r$, the infimum of $\beta_{F_{d,\alpha}}$ on the contour $\psi(\mu)/\sigma^2 = c$ is α; if Δ' is empty, there is no problem). $F_{d,\alpha}$ has the same number of numerator degrees of freedom for all d in Δ'; if also

b_d is the same for each d in Δ' (this is often the case in important examples such as those of Section 3C) so that the denominator degrees of freedom are the same for all $F_{d,\alpha}$, then a design which maximizes the minimum power on $\psi(\mu)/\sigma^2 = c$ *simultaneously for all* c is precisely one which maximizes the minimum of λ_d subject to $\psi(\mu)/\sigma^2 = c$. Since $\psi(\mu) = (R\mu)'(R\mu)$ in both Cases I and II, this means maximizing $\min_{\xi'\xi=1} \xi'\bar{V}_d^{-1}\xi = 1/\pi(\bar{V}_d)$. This is precisely the criterion of E-optimality.

One can cite many practical examples to illustrate that the restriction to using $F_{d,\alpha}$, which is imposed in order to make E-optimality meaningful, can have serious detrimental consequences. The simplest possible situation will suffice as an example: Suppose $N > 2$, $r = m = 2$, $R = R_I$, and Δ' to consist of two designs with

$$\bar{V}_{d_1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \bar{V}_{d_2} = \begin{pmatrix} 1+\epsilon & 0 \\ 0 & \epsilon \end{pmatrix},$$

where $\epsilon > 0$. Clearly, d_1 is E-optimum. Moreover, if d_1 is used, optimum property (e) above states that, for every c, $F_{d_1,\alpha}$ maximizes the minimum power on the contour $(\mu_1^2 + \mu_2^2)/\sigma^2 = c$ among all tests in $H_{d_1}(\alpha)$. However, if d_2 is used, $F_{d_2,\alpha}$ does not have this property. For example, if d_2 is used, let ϕ' be the test which with probability $(1+\epsilon)/(1+2\epsilon)$ uses the F-test (with 1 and N-2 degrees of freedom) of size α of the hypothesis $\mu_1 = 0$, and which with probability $\epsilon/(1+2\epsilon)$ uses the F-test of size α of the hypothesis $\mu_2 = 0$. The power function of ϕ' near $(\mu_1^2 + \mu_2^2)/\sigma^2 = 0$ is then

$$\alpha + g_{1,N-2}(\alpha)\,(\mu_1^2 + \mu_2^2)/(1+2\epsilon)\sigma^2 + o([\mu_1^2 + \mu_2^2]/\sigma^2),$$

while that of $F_{d_2,\alpha}$ is

$$\alpha + g_{2,N-2}(\alpha)\left(\frac{\mu_1^2}{1+\epsilon} + \frac{\mu_2^2}{\epsilon}\right)/\sigma^2 + o([\mu_1^2 + \mu_2^2]/\sigma^2).$$

The infimum of the expression multiplying $g_{2,N-2}(\alpha)$, taken on the contour $(\mu_1^2 + \mu_2^2)/\sigma^2 = c$, is $c/(1+\epsilon)$, compared with $c/(1+2\epsilon)$ for the coefficient of $g_{1,N-2}(\alpha)$; since $g_{1,N-2}(\alpha)/g_{2,N-2}(\alpha) \to 2$ as $\alpha \to 0$ (see Lemma 4.3 below) the assertion three sentences above regarding $F_{d_2,\alpha}$ is verified. Moreover, since the power function of $F_{d_1,\alpha}$

$$\alpha + g_{2,N-2}(\alpha)(\mu_1^2 + \mu_2^2)/\sigma^2 + o([\mu_1^2 + \mu_2^2]/\sigma^2),$$

we see similarly that, at least for α, ϵ, and c sufficiently small, d_1 is *not* $M_{\alpha,c}$-optimum or L_α-optimum, ϕ' being locally uniformly more powerful than $F_{d_1,\alpha}$; thus, the assertion of the first sentence of this paragraph regarding E-optimality is verified.

Of course, for any fixed α, ϵ, and c we have not asserted that the test ϕ' (considered above only for illustrative purposes) is $M_{\alpha,c}$-optimum. If one uses d_2, the power functions of ϕ', $F_{d_2,\alpha}$, etc., are not constant on $(\mu_1^2 + \mu_2^2)/\sigma^2 = c$ (the same is true of the test which minimizes the integral of the power function on that contour), and the computation of the supremum over ϕ on the right side of (2.1) does not seem easy (this will be discussed further in Section 5). Thus, the above example also illustrates why M-optimality (or L-optimality) seems so difficult to verify in many problems.

In order to see the meaning of D-optimality, we turn to the notion of a type D test as defined in [11] (we discuss the case where σ^2 is unknown, the other case being similar): For fixed d, let the function $\bar{\beta}_\phi(\eta, \tau, \sigma^2)$ be defined by $\bar{\beta}_\phi(Q_d R\mu, L_d\mu, \sigma^2) = \beta_\phi(\mu, \sigma^2)$ and let $\beta_\phi^i(\tau, \sigma^2)$ (resp., $\beta_\phi^{ij}(\tau, \sigma^2)$) be the derivative of $\bar{\beta}_\phi(\eta, \tau, \sigma^2)$ with respect to the ith (resp., ith and jth) component of η, evaluated at $\eta = 0$ (these derivatives always exist). A test ϕ in $H_d(\alpha)$ is said to be locally (near H_0) strictly unbiased if

(a) $\phi \in H_d^*(\alpha)$,

(b) $\beta_\phi^i(\tau, \sigma^2) = 0$ for all i, τ, and σ^2,

(c) the matrix $B_\phi(\tau, \sigma^2) = \| \beta_\phi^{ij}(\tau, \sigma^2) \|$ is positive definite for all τ and σ^2.

Clearly, (c) can be satisfied only if $d \in \Delta'$. Suppose then that $d \in \Delta'$ and that $Q_d =$ identity (we have mentioned the fact that we can make this choice of Q_d when $d \in \Delta'$). For any ϕ satisfying (a), (b), (c) just above, det $B_\phi(\tau, \sigma^2)$ is the Gaussian curvature of the surface given by the graph of $\bar{\beta}_\phi(\eta, \tau, \sigma^2)$ as a function of η for fixed τ, σ^2, at $\eta = 0$. A test ϕ is defined in [11] to be of type D if it maximizes this curvature for all τ and σ^2, among all locally strictly unbiased tests. This criterion of optimality, although a local one, has certain appealing features; for example, it is invariant under all one-to-one transformations of the parameter space which leave $\eta = 0$ fixed and which at $\eta = 0$ are twice differentiable with non-vanishing Jacobian [11]. Now, since without loss of generality we are taking $Q_d =$ identity, we can compare the behavior of the type D tests for various designs in Δ', assuming b_d to be the same for all d in Δ'. A design for which the Gaussian curvature at $\eta = 0$ of the test of maximum Gaussian curvature (for a given design) is a maximum (over all designs) is thus, if it exists, that d which maximizes $\max_{\phi_d}$ det $B_{\phi_d}(\tau, \sigma^2)$ simultaneously for all τ, σ^2. That such a design is precisely one which is D-optimum follows immediately from the following lemma[4] (there is an obvious analogue when σ^2 is known):

LEMMA 2.1. *For d in Δ' and $0 < \alpha < 1$, the test $F_{d,\alpha}$ is of type D.*

PROOF. $F_{d,\alpha}$ is clearly locally strictly unbiased. We again put $Q_d =$ identity, and a nonsingular linear transformation reduces the proof to the case where $G_d =$ identity (see Section 2A(c)), so that $\nu = \eta$. Wald's theorem can then be stated as

$$\int_{\eta'\eta=c} [\bar{\beta}_{F_{d,\alpha}}(\eta, \tau, \sigma^2) - \alpha] A(d\eta) \geqq \int_{\eta'\eta=c} [\bar{\beta}_\phi(\eta, \tau, \sigma^2) - \alpha] A(d\eta) \tag{2.5}$$

for every $c > 0$, $\sigma^2 > 0$, and ϕ in $H_d^*(\alpha)$, where $A(d\eta)$ is Lebesgue measure on the sphere $\eta'\eta = c$. Noting that

$$\int_{\eta'\eta=c} \eta_i \eta_j A(d\eta) = \begin{cases} K(c, r) & \text{if } i = j, \\ 0 & \text{if } i \neq j, \end{cases} \tag{2.6}$$

where η_i is the ith component of η and $K(c, r)$ is positive and depends only on c

[4] The author understands that Isaacson gave a longer, unpublished proof, earlier.

and r, we obtain from (2.5) by normalizing properly and letting $c \to 0$, for any ϕ satisfying conditions (a) and (b) above,

$$\sum_{i=1}^{r} \beta_{F_{d,\alpha}}^{ii}(\tau, \sigma^2) \geqq \sum_{i=1}^{r} \beta_{\phi}^{ii}(\tau, \sigma^2). \tag{2.7}$$

Since $B_{F_{d,\alpha}}(\tau, \sigma^2)$ is a constant times the identity in our reduction, using the inequality of the geometric and arithmetic means and the fact that the determinant of a positive-definite matrix is no greater than the product of its diagonal elements, we obtain (omitting some appearances of τ, σ^2),

$$\begin{aligned} \det B_{\phi}(\tau, \sigma^2) &\leqq \prod_{i=1}^{r} \beta_{\phi}^{ii}(\tau, \sigma^2) \leqq \left[\sum_{i=1}^{r} \beta_{\phi}^{ii}/r\right]^r \\ &\leqq \left[\sum_{i=1}^{r} \beta_{F_{d,\alpha}}^{ii}/r\right]^r = \prod_{i=1}^{r} \beta_{F_{d,\alpha}}^{ii} = \det B_{F_{d,\alpha}}(\tau, \sigma^2), \end{aligned} \tag{2.8}$$

which completes the proof.

To summarize, D-optimality and L-optimality, although local properties, seem more reasonable criteria than E-optimality, which is tied to the ad hoc assumption that $F_{d,\alpha}$ should always be used; M-optimality (and to a lesser extent L-optimality) seems difficult to verify in many examples.

2C. *Relationship among optimality criteria in symmetric cases.* For future reference we state the following simple result (which was alluded to in Section 0 in reference to the relation between [1] and [2]):

LEMMA 2.2. *Suppose b_d is constant for d in Δ'. If d^* is D-optimum and $\bar{V}_{d^*}$ is a multiple of the identity, then d^* is E-optimum and L-optimum.*

PROOF. E-optimality is obvious from the nature of $\bar{V}_{d^*}$. If d^* were not L-optimum, since $F_{d^*,\alpha}$ has property 2A(c), for some other design d' there would by (2.2) be an associated test $\phi_{d'}$ in $H_{d'}^*(\alpha)$ with

$$\inf_{\tau,\sigma^2} \det B_{\phi_{d'}}(\tau, \sigma^2) > \det B_{F_{d^*,\alpha}}(\tau, \sigma^2) \tag{2.9}$$

(the right side of (2.9) is constant); by Lemma 2.1, equation (2.9) is a fortiori true if $\phi_{d'}$ is replaced by $F_{d',\alpha}$; this yields the contradiction that $\det \bar{V}_{d'} < \det \bar{V}_{d^*}$.

In many examples of Case I where symmetrical designs exist, the condition on $\bar{V}_{d^*}$ in the hypothesis of Lemma 2.2 will be obvious. In Case II, as discussed in Section 3A, it is often convenient to write the normal equations in the form $C_d t_d^* = Z_d$, where C_d is a $u \times u$ matrix of rank $u - 1$, Z_d is a u-vector of linear forms in Y with covariance matrix C_d, and for any solution t_d^* of these equations one obtains the best linear estimator of any contrast $\sum_1^u c_i\mu_i$ with $\sum c_i = 0$ by forming $\sum c_i t_{di}^*$ where the t_{di}^* are the components of t_d^*. Clearly, Pt_d^* is the b.l.e. t_d of $R_{\text{II}}\,\mu$. Hence, if every diagonal element of C_d has the same value and if all off-diagonal elements have the same value, the fact that the first row of the orthogonal matrix $\bar{O}$ defined in Case II of Section 1 is constant immediately yields the fact (see Section 3A) that $\bar{V}_d^{-1} = PC_d P'$ is a multiple of the identity, so that

Lemma 2.2 may be applicable in such cases. For future reference, we state this simple computation (put $a + (u - 1)c = 0$) in

LEMMA 2.3. *If U is a $u \times u$ matrix with diagonal elements a and off-diagonal elements c, then*

$$\bar{O}U\bar{O}' = \begin{pmatrix} a + (u-1)c & 0 \\ 0 & (a-c)I_{u-1} \end{pmatrix}. \tag{2.10}$$

We remark that the form of R_{II} (associated with $\bar{O}$) used here makes computations and proofs simpler and emphasizes more the role of symmetry (e.g., as it appears in the form of $\bar{V}_d^{-1}$ just noted, when C_d has appropriate symmetry), than would be the case if R_{II} were replaced by a matrix obtained by adjoining a column of 1's and $m - u$ columns of 0's to I_{u-1}, as in [1] and [2].

3. Optimality of symmetrical designs.

3A. *Preliminaries.* The results of this section will be proved for the case where σ^2 is unknown, the other case being handled similarly. The setting of two-way heterogeneity where the LS design is employed is much easier to analyze (and thereby obtain an optimality proof) than other settings considered in Section 3B such as those where the YS and BIBD are used (and the remarks at the end of Section 2 indicate how this analysis can be made even simpler than in [1] and [2]). The reason for this is that in this setting where the LS is used, whether μ is considered to have $3u$ components (u each for row, column, and treatment effects in the $u \times u$ case) or $3u - 2$ components (to make $X_d'X_d$ nonsingular when $s_d = b_d = u - 1$), $X_d'X_d$ becomes particularly simple, having large blocks of 1's (each row and column occur together once, etc.) or multiples of an identity (rows by rows, etc.) in the former case, and large blocks of 0's (especially if $\bar{O}$ is used in reducing X_d) and multiples of an identity, in the latter. Other design situations yield more complicated forms of $X_d'X_d$. Therefore, although the examples of Section 3C could be analyzed in a manner analogous to that used for the LS in [1] and [2], it appears algebraically simpler to use the incomplete block design analysis of Bose [4], to which end we now briefly outline the notation. Of course, we are concerned here with the more difficult Case II, which includes most of the important examples.

The form of the Z_d and C_d mentioned in Section 2C depends on the design setting and, in particular, in this section, on whether we are in a setting of one-way or two-way heterogeneity of (for example) soil (since all block sizes will be the same in our example of the former, it could be considered as a special case of the latter under further restrictions on μ). We shall first state the pertinent results which apply in both of these settings, and then specify the particular forms (see [4] for details). The $u \times u$ symmetric matrix C_d has row (or column) sums equal to zero, and the sum of the components of the u-vector Z_d is zero. The covariance matrix of Z_d is $\sigma^2 C_d$ and the expected value of Z_d is $C_d\mu^{(u)}$, where $\mu^{(u)}$ is the vector of the first u components of μ. We may assume $d \varepsilon \Delta'$, which means the design d is connected and that C_d has rank $u - 1$. If t_d^* satisfies $C_d t_d^* =$

Z_d and P is the $(u - 1) \times u$ matrix defined in Case II in Section 1, then $t_d = Pt_d^*$ is the vector if b.l.e.'s of $R_{II}\mu$; the last $u - 1$ rows of the equation $\bar{O}C_d\bar{O}'\bar{O}t_d^* = \bar{O}Z_d$ are thus $PC_dP't_d = PZ_d$ (the first row and column of $\bar{O}C_d\bar{O}'$ are zero), so that $t_d = (PC_dP')^{-1}PZ_d$ (the inverse may be taken for d in Δ') and thus the components of t_d have covariance matrix $(PC_dP')^{-1}$.

In the one-way heterogeneity setting we have u treatments, to be planted in b blocks; in our example, each block will contain the same number k of plots, one "planting" to be allowed per plot. The component of Y corresponding to an appearance of treatment i in block j has expected value $\mu_i + b_j$; thus, $m = u + b$, with $\mu_{u+j} = b_j$. Let n_{dij} be the number of appearances of treatment i in block j. We do *not* restrict n_{dij} to be 0 or 1, as is often done. Thus, D consists of those d for which $'X_d$ is any matrix of 0's and 1's for which each row contains exactly one 1 among the first m elements and one 1 among the last b elements and for which the last b columns each contain k one's; of course, $N = bk$. Let $r_{di} = \sum_j n_{dij}$ = number of replications of treatment i, let T_{di} = sum of all components of Y corresponding to treatment i, and let B_{dj} = sum of all components of Y arising from block j. The ith component Z_{di} of Z_d ("adjusted yield of treatment i") is $Z_{di} = T_i - \sum_j n_{ij} B_j/k$, and the (i, j)th component c_{dij} of C_d is

$$c_{dij} = \delta_{ij}r_{di} - \lambda_{dij}/k, \tag{3.1}$$

where δ_{ij} is the Kronecker delta and $\lambda_{dij} = \sum_s n_{dis}n_{djs}$.

In the setting of two-way heterogeneity, we have u treatments and a $k_1 \times k_2$ array of plots, and the expected value of a component of Y corresponding to treatment i in row j and column h is $\mu_i + b_j^{(1)} + b_h^{(2)}$; thus, $m = u + k_1 + k_2$ with $b_j^{(1)} = \mu_{m+j}$ and $b_h^{(2)} = \mu_{m+k_1+h}$. Let $n_{dij}^{(1)}$ (resp., $n_{dih}^{(2)}$) be the number of times treatment i appears in row j (resp., column h), and let T_{di} be as before and $B_{dj}^{(1)}$ (resp., $B_{dh}^{(2)}$) be the sum corresponding to the jth row (resp., hth column). r_{di} is as above, while $\lambda_{dij}^{(q)} = \sum_s n_{dis}^{(q)}n_{djs}^{(q)}$ for $q = 1, 2$. In this case $Z_{di} = T_{di} - \sum_j n_{dij}^{(1)} B_{dj}^{(1)}/k_2 - \sum_h n_{dih}^{(2)} B_{dh}^{(2)}/k_1 + r_{di}\sum_s T_{ds}/k_1k_2$ and

$$c_{dij} = \delta_{ij}r_{di} - \frac{\lambda_{dij}^{(1)}}{k_2} - \frac{\lambda_{dij}^{(2)}}{k_1} + \frac{r_{di}r_{dj}}{k_1k_2}. \tag{3.2}$$

Many other design settings can be treated similarly; the above two will be used in the examples of Section 3C to illustrate our methods of proving optimality.

3B. *Algebraic results*. We now demonstrate the algebraic results used in proving optimality in the examples of Section 3C and which will be useful in other examples of Case II. The results proved here are meant to apply elsewhere than in the settings of Section 3A. We suppose in the present Section 3B that we are given a class $\{K_d, d \in \Delta'\}$ of $u \times u$ symmetric nonnegative definite matrices of rank $u - 1$ with row and column sums zero and define $W_d = PK_dP'$ (in our applications, $W_d = \bar{V}_d^{-1}$). The elements of $\bar{O}$, K_d, and W_d will be denoted by $\bar{o}_{ij}$,

k_{dij}, and w_{dij}, respectively. In Lemma 3.2 we consider an orthogonal matrix $\tilde{O} = \| o_{ij} \|$, not necessarily $\bar{O}$, and a diagonal matrix $D = \| d_{ij} \|$.

Our first lemma merely translates into terms of K_d the obvious fact that, if W_{d^*} has equal eigenvalues and if the sum of the eigenvalues (= trace) of W_d is a maximum for $d = d^*$, then the product of eigenvalues (= determinant) of W_d is a maximum for $d = d^*$.

LEMMA 3.1. *If all diagonal elements of K_{d^*} are equal and all off-diagonal elements of K_{d^*} are equal and $\sum_i k_{dii}$ is a maximum for $d = d^*$, then* det W_d *is a maximum for $d = d^*$.*

PROOF. Since $\bar{o}_{ij} = 1/\sqrt{u}$ and $\sum_{i,j} k_{dij} = 0$, the upper left-hand element of $\bar{O}K_d\bar{O}'$ is zero. Since the traces of $\bar{O}K_d\bar{O}'$ and K_d are equal, we conclude that the traces of K_d and W_d are equal, so that the trace of W_d is a maximum for $d = d^*$. The result now follows from Lemma 2.3 (follow the steps of (2.8) with W_d for B_ϕ and W_{d^*} for $B_{F_{d,\alpha}}$).

We shall actually prove in Theorems 3.1 and 3.2 that the trace of the matrix PC_dP' is a maximum and that all eigenvalues are equal when d is a BBD or GYS, so that Lemma 3.1 is relevant. However, there are settings in which the next three lemmas are more useful for proving D- or E-optimality directly when the hypothesis of Lemma 3.1 is difficult to verify or is false.

LEMMA 3.2. *For $u > 1$ if $\tilde{O}$ is orthogonal $u \times u$, D is diagonal $u \times u$, K is symmetric nonnegative definite $u \times u$ with row and column sums zero, and $\tilde{O}D\tilde{O}' = K$, then*

$$\left(\frac{u-1}{u}\right)^u \left(\prod_{i=1}^{u-1} d_{ii}\right)^{u/(u-1)} \leqq \prod_{i=1}^{u} k_{ii}. \tag{3.3}$$

PROOF. We assume $d_{uu} = 0 < d_{ii}$ for $i < u$, or the result is trivial. Since, then,

$$0 = \sum_{i=1}^{u}\sum_{j=1}^{u} k_{ij} = \sum_{i=1}^{u}\sum_{j=1}^{u}\sum_{s=1}^{u-1} o_{is}\,o_{js}\,d_{ss} = \sum_{s=1}^{u-1} d_{ss}\left(\sum_{i=1}^{u} o_{is}\right)^2, \tag{3.4}$$

we conclude that the first $u - 1$ columns of $\tilde{O}$ are orthogonal to the vector of ones. Hence, $o_{ju} = 1/\sqrt{u}$ (or its negative, which is treated in the same way).

Let the coordinates of a point ϵ in $u(u - 1)$-dimensional Euclidean space be denoted by ϵ_{ij} ($i = 1, \cdots, u; j = 1, \cdots, u - 1$), and let B be the set of points ϵ in this space for which all $\epsilon_{ij} \geqq 0$, for which $\sum_j \epsilon_{ij} = (u - 1)/u$ for all i, and for which $\sum_i \epsilon_{ij} = 1$ for all j. We shall prove below that ϵ in B implies

$$\prod_{i=1}^{u}\left(\sum_{j=1}^{u-1} \epsilon_{ij}\,d_{jj}\right) \geqq \left(\frac{u-1}{u}\right)^u \left(\prod_{i=1}^{u-1} d_{ii}\right)^{u/u-1}; \tag{3.5}$$

since the left side of (3.5) with $\epsilon_{ij} = o_{ij}^2$ gives the right side of (3.3) and since the restrictions on the ϵ_{ij} in B must be satisfied by the o_{ij}^2 (the orthogonality restrictions on the o_{ij} are omitted in defining B), (3.5) implies (3.3).

Call the left side of (3.5) $f(\epsilon)$. It is easy to verify that $-\log f(\epsilon)$ is convex in ϵ on $u(u - 1)$-space, and hence on B. Moreover, B is a convex body in $u(u - 1)$

space, and any extreme point of B is either

$$(3.6)\qquad \begin{bmatrix} \epsilon_{11} & \cdots & \epsilon_{1,u-1} \\ \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot \\ \epsilon_{u1} & \cdots & \epsilon_{u,u-1} \end{bmatrix} = \begin{bmatrix} \frac{u-1}{u} I_{u-1} \\ \frac{1}{u} \cdots\cdots\cdots \frac{1}{u} \end{bmatrix}$$

or is obtained by permuting the rows of the matrix on the right side of (3.6). Since a convex function on a convex set attains its maximum at an extreme point, we conclude that the minimum of f is attained at one of these extreme points. But f has the same value at any of these extreme points, namely,

$$(3.7)\qquad \min_B f(\epsilon) = \left(\sum_{i=1}^{u-1} d_{ii}/u\right) \prod_{i=1}^{u-1} \left(\frac{u-1}{u} d_{ii}\right).$$

Thus, it remains only to prove that the right side of (3.7) is no less than the right side of (3.5), i.e., that

$$(3.8)\qquad \prod_{i=1}^{u-1} d_{ii}^{1/(u-1)} \leqq \sum_{i=1}^{u-1} d_{ii}/(u-1);$$

but (3.8) is merely the well-known inequality between the geometric and arithmetic means.

The form of Lemma 3.2 which is useful in many applications is the following:

LEMMA 3.3. *If* $\prod_1^u k_{dii}$ *is a maximum for* $d = d^*$ *and if* K_{d^*} *has all diagonal elements equal and all off-diagonal elements equal, then* det W_d *is a maximum for* $d = d^*$.

PROOF: We use Lemma 3.2 with the product on the left side of (3.3) going from 2 to u, in order to conform to previous notation. In this form, with $\tilde{O} = \bar{O}$, it follows from Lemma 2.3 that the left and right sides of (3.3) are equal for $K = K_{d^*}$. Hence, from Lemma 3.2, $\prod_i w_{dii}$ is a maximum for $d = d^*$. Since $\prod_i w_{dii} \geqq$ det W_d with equality for the diagonal matrix W_{d^*}, the proof is complete.

The following lemma could be used in the case of the YS, and in more complicated problems where D-optimality is hard to prove or false, to prove E-optimality directly (i.e., without the use of Lemma 2.2):

LEMMA 3.4. *For* $u > 1$, *if* $m(W_d)$ *is the minimum eigenvalue of* W_d, *then*

$$(3.9)\qquad m(W_d) \leqq \frac{u}{u-1} \min_i k_{dii};$$

if all diagonal elements of K_d *are equal and all off-diagonal elements are equal, equality holds in* (3.9).

PROOF. Let δ_i be a u-vector with ith element one and all other elements zero. Let $\xi_i = P\delta_i$. Clearly, $\sqrt{u/(u-1)}\,\xi_i$ has unit length. Hence,

$$(3.10)\qquad \begin{aligned} k_{dii} &= \delta_i' K_d \delta_i = (\bar{O}\delta_i)'(\bar{O}K_d\bar{O}')(\bar{O}\delta_i) \\ &= \xi_i' W_d \xi_i \geqq \frac{u-1}{u} \min_{a'a=1} a' W_d a = \frac{u-1}{u} m(W_d), \end{aligned}$$

which proves (3.9); the result on equality follows from Lemma 2.3.

The results for Case I analogous to those proved for Case II in this subsection are trivial (since in Case I the analogue of K_d will be nonsingular and K_{d^*} will be a multiple of the identity), and will be omitted.

3C. *Examples.*[5] (1). *Optimality of* BIBD's. In the setting of one-way heterogeneity described in Section 3A (with $u > 1$), suppose b, u, and k to be such that there exists a design d^* for which all n_{d^*ij} are k/u if k/u is an integer and are either of the two integers closest to k/u otherwise, for which all r_{d^*i} are equal, and for which all λ_{d^*ij} are equal for $i \neq j$. Such a design is called a BIBD if $k < u$, but we do not impose this last restriction here, and therefore call such a design a Balanced Block Design (BBD). (For example, if $b = 2$, $u = 2$, $k = 3$, such a d^* is that for which $n_{d^*11} = n_{d^*22} = 1$ and $n_{d^*12} = n_{d^*21} = 2$.) Our result is:

THEOREM 3.1. *If a* BBD d^* *exists, it is D-optimum, E-optimum, and L-optimum.*

PROOF. From (3.1) we have

$$\sum_{i=1}^{u} c_{dii} = N - \textstyle\sum_i \sum_s n_{dis}^2/k; \tag{3.11}$$

since $\sum_i \sum_s n_{dis} = N$, it is clear that (3.11) is a maximum for $d = d^*$. The result now follows from Lemma 3.1 and Lemma 2.2.

(2). *Optimality of* YS's. In the setting of two-way heterogeneity described in Section 3A (with $u > 1$), suppose k_1, k_2, and u to be such that there exists a design d^* for which all r_{d^*i} are equal, for which all $\lambda_{d^*ij}^{(1)}$ are equal for $i \neq j$, for which all $\lambda_{d^*ij}^{(2)}$ are equal for $i \neq j$, and for which all $n_{d^*ij}^{(q)}$ are equal to k_q/u if k_q/u is an integer and are either of the two integers closest to k_q/u otherwise ($q = 1, 2$). Thus, d^* is a BBD when either the rows or the columns are considered to be the blocks. Such a design d^* is usually called a YS if $k_1 < u$ (and k_2/u is an integer); we do not impose this condition, and shall hence call such a design d^* a Generalized Youden Square (GYS). (For example, if $u = 2$, $k_1 = 4$, $k_2 = 3$, such a design d^* is easily constructed.) If $k_1 = k_2 = u$, such a d^* is of course a LS. Our result is:

THEOREM 3.2. *If k_1/u or k_2/u is an integer and if a* GYS d^* *exists, then* d^* *is D-optimum, E-optimum, and L-optimum.*

PROOF. We shall show that $\sum_i c_{dii}$ is a maximum for $d = d^*$; Lemma 3.1 then yields the desired result. In this proof only we write $[x]$ = greatest integer $\leqq x$. Let r be an integer. Subject to the restrictions that $\sum_1^k m_j = r$ and that all m_j are integers, the expression $\sum_1^k m_j^2$ is minimized by taking $k - r + k[r/k]$ of the m_i to be $[r/k]$ and $r - k[r/k]$ of them to be $[r/k] + 1$, the corresponding minimum of $\sum m_j^2$ being $r + (2r - k)[r/k] - k[r/k]^2 = h(r, k)$ (say). We may assume

---

[5] The Editor has informed the author that E-optimality of the BIBD's (as a subclass of the BBD's) has been proved independently by V. L. Mote, and that the minimization of the average variance (see numbered paragraph 2 of Section 5) and of the generalized variance (i.e., the attainment of D-optimality) achieved by the BIBD's and YS's (a subclass of the GYS's) has been proved independently by A. M. Kshirsagar; both of these authors prove their results under the restriction and that the n_{dij} and $n_{dij}^{(q)}$ are all 0 or 1. Under this restriction, these special cases of the results of this paper are a consequence of the following line of argument: the trace of C_d is the same for all d, and the results follow at once from the symmetry of the BIBD and YS.

k_1/u is an integer. From (3.2) we have, for any d,

$$(3.12)\quad k_1k_2(k_1k_2 - \sum_i c_{dii}) \geqq \sum_i \{k_2h(r_{di}, k_2) - r_{di}^2\} + \sum_i k_1h(r_{di}, k_1),$$

with equality in the case of a GYS. The theorem will be proved if we show that each of the two sums on the right side of (3.12) attains its minimum for $d = d^*$. Now, $h(r, k) \geqq r^2/k$, since the latter is the minimum of $\sum m_j^2$ subject to $\sum m_j = r$ without the restriction that the m_j be integers. Hence, the first sum on the right side of (3.12) is at least zero. Moreover, this lower bound is achieved by the first sum on the right in (3.12) when $d = d^*$, since $r_{d^*i}/k_2 = k_1/u$ is an integer. It remains to consider the last sum of (3.12). We shall show that, subject to $\sum_1^m z_i = c$, the expression

$$(3.13)\qquad q(z_1, \cdots, z_m) = \sum_{i=1}^{m} \{(2z_1 - 1)[z_i] - [z_i]^2\}$$

is a minimum when all z_i are equal; putting $z_i = [r_{di}/k_1]$, we see that this will yield the desired conclusion regarding the last sum of (3.12). The proof regarding (3.13) is by induction: assuming the conclusion to be true of $m = M$, in proving the case $m = M + 1$ we may put $z_1 = \cdots = z_M = s$ and $z_{M+1} = c - Ms$ in (3.13). The resulting expression is continuous in s and, except on a discrete set, has a derivative with respect to s which is equal to $2M([s] - [c - Ms])$. The latter is $\leqq 0$ if $s < c/(M + 1)$ and is $\geqq 0$ if $s > c/(M + 1)$, so that $s = c/(M + 1)$ yields a minimum. This completes the proof of Theorem 3.2.

We remark that, without the assumption that k_1/u or k_2/u is an integer, the above proof fails and Lemma 3.3 also fails to be applicable generally. To see this, consider the case $k_1 = k_2 = 6$, $u = 4$. A GYS d^* exists here, e.g., that one whose successive rows are (134324), (412233), (241342), (124123), (313412), (321441). We obtain $c_{d^*ii} = 25/4$ for all i. Let d' be the design whose rows are (133442), (213344), (421334), (442133), (344213), and (334421). Then $c_{d'11} = c_{d'22} = 5$, $c_{d'33} = c_{d'44} = 8$, $c_{d'12} = -1$, $c_{d'34} = -4$, and all other $c_{d'ij} = -2$. Thus, we obtain $\sum_i c_{d'ii} = 26 > 25 = \sum_i c_{d^*ii}$ and even $\prod_i c_{d'ii} = 1600 > (25/4)^4 = \prod_i c_{d^*ii}$. However, det $\bar{V}_{d'}^{-1} = 576 < (25/3)^3 =$ det $\bar{V}_{d^*}^{-1}$. Thus, between the designs d^* and d', the former is D-optimum, although Lemmas 3.1 and 3.3 cannot be used to prove it. Lemma 3.4 could still have been used to prove the E-optimality of d^* directly.

(3) *Other examples*. Many other design settings can be analyzed in a manner differing only slightly from the above examples and we mention but a few. One can treat similarly problems where the test concerns the b_j and $b_j^{(q)}$ of Section 3A. Problems involving Graeco-Latin Squares or higher Latin Squares, with or without replications, admit similar treatments. Higher-dimensional analogues (more than two directions of heterogeneity) can also be considered in a like fashion, as can complete or partial factorial arrangements. Many of the Case I analogues, such as the analogue of the BIBD treatment which assumes the b_j to be known, are trivial.

Other problems such as those for which E-optimality is considered in [2] (e.g., Hotelling's weighing problem and certain problems in the analysis of covariance) could be considered regarding D- and L-optimality by similar methods.

The treatment of some problems is in part parallel, but entails other considerations in addition to symmetry; such a problem is to test whether a regression function $\sum_{j=1}^{m} \mu_j f_j(x)$ is actually such that $\mu_1 = \cdots = \mu_r = 0$, where the f_j are given and N x's must be chosen from a given region of some space. (Many problems in the analysis of covariance involve similar considerations.) D- and E-optimality are also relevant in estimation problems (see Section 5.2).

The consideration of some of these other examples will appear elsewhere, in a paper by J. Wolfowitz and the author.

4. Nonoptimality of symmetrical nonrandomized designs among randomized designs.[1]

4A. Case I. We consider here the simplest general setting of Case I, namely, the extension of the example of Section 1 to more observations N and more treatments u. Other examples, such as the Case I analogues of the examples of Section 4B, have parallel analyses, and we omit them. We shall carry out the treatment when σ^2 is unknown, the treatment when σ^2 is known being similar. The underlying probabilistic property (of the normal distribution) which is relevant here will now be stated in a lemma. Let U/σ^2 have a non-central χ^2 distribution with N_1 degrees of freedom (d.f) and non-central parameter $\lambda = EU/\sigma^2 - N_1$, and let V/σ^2 have the central χ^2 distribution with N_2 d.f., with U and V independent. Let $P_{N_1,N_2}(\lambda; \alpha)$ denote the power function of the F-test of size α for testing $\lambda = 0$ based on N_2U/N_1V, and, as in (1.3), let $g_{N_1,N_2}(\alpha)$ denote the derivative of this power function with respect to λ at $\lambda = 0$.

Lemma 4.1. *If* $N_1 \leqq N_1'$ *and* $N_1 + N_2 \geqq N_1' + N_2'$ *with at least one of these a strict inequality, then* $P_{N_1,N_2}(\lambda; \alpha) > P_{N_1',N_2'}(\lambda; \alpha)$ *for* $\lambda > 0$ *and* $0 < \alpha < 1$, *and* $g_{N_1,N_2}(\alpha) > g_{N_1',N_2'}(\alpha)$ *for* $0 < \alpha < 1$.

Proof. Let U/σ^2 have a χ^2 distribution with parameter λ and N_1 d.f., and let V_1/σ^2, V_2/σ^2, and V_3/σ^2 have central χ^2 distributions with N_2', $N_1' - N_1$, and $N_1 + N_2 - N_1' - N_2'$ d.f., respectively (if any of the d.f.'s is 0, so is the corresponding V_i). U, V_1, V_2, V_3 are independent. For testing the hypothesis $\lambda = 0$ against alternatives $\lambda > 0$ based on U, V_1, V_2, V_3, it is easy to prove that the F-test based on $N_2U/N_1(V_1 + V_2 + V_3)$ is UMP unbiased of size α and is of type A, and is the unique (up to sets of measure zero) test with each of these properties; in particular, this is true in comparison with the F-test based on $N_2'(U + V_2)/N_1'V_1$, which proves the lemma.

The above lemma indicates both that the numerator d.f. should be as small as possible without affecting λ, which is also true when σ^2 is known, and also that for fixed $N_1 + N_2$, decreasing N_1 helps even more if σ^2 is *unknown*, since N_2 is increased (compare (4.5) and (4.7) below).

We now consider the following problem: Y_{ij} are independent and normally distributed random variables with unknown mean μ_i ($j = 1, \cdots, n_i$; $i = 1,$

$\cdots, u)$ and variance σ^2 (we use a convenient notation for the example, rather than that introduced in Section 1). The problem is to test $H_0: \mu_1 = \mu_2 = \cdots = \mu_u = 0$, and a design d in Δ is a specification of nonnegative integers n_i whose sum is N. For any such d, we denote by $M(d)$ the set of i for which $n_i > 0$; by $k(d)$, the number of integers in $M(d)$; by τd, the design associated with the values $n_i = n^*_{\tau(i)}$ when d is associated with the values $n_i = n_i^*$, where τ is any element of the symmetric group S_u on u symbols; by δ_d, the design in Δ_R which assigns probability $1/u!$ to each τd for τ in S_u; by $f_{d,\alpha}$ the test associated with δ_d which is obtained by using the appropriate F-test of size α with whatever τd is chosen by δ_d. We shall also use the symbol $a_\phi(c)$ of (2.2), with $\psi(\mu) = \sum_1^u \mu_i^2$, and shall denote by a'_ϕ its derivative with respect to c at $c = 0$. We shall also use the symbols $g_{ij}(\alpha)$ introduced in (1.3). Our result, which implies that the "symmetrical" design associated with $k(d) = u$ and all n_i equal (or as nearly so as possible) is not L_α-optimum in Δ_R, and that the δ_d associated with the d for which $n_1 = N$ (this δ_d chooses each i with probability $1/u$ and takes all Y_{ij} with the chosen i) is locally best among the δ_d, is the following:

THEOREM 4.1. *For every d, α, and c,*

$$a_{F_{d,\alpha}}(c) \leqq a_{f_{d,\alpha}}(c); \tag{4.1}$$

$a'_{f_{d,\alpha}}$ is strictly decreasing in $k(d)$, and the same is true of $a_{f_{d,\alpha}}(c)$ for all c in some neighborhood of $c = 0$.

PROOF. (4.1) is trivial, and we proceed to the rest of the proof. The numerator $t'_d V_d^{-1} t_d$ of $F_{d,\alpha}$ is of course

$$U_d = \sum_{i \in M(d)} n_i \left(\sum_{j=1}^{n_i} Y_{ij}/n_i \right)^2,$$

and U_d/σ^2 has a χ^2 distribution with $k(d)$ d.f. and non-central parameter

$$\sum_{i \in M(d)} n_i \mu_i^2/\sigma^2.$$

The denominator of $F_{d,\alpha}$ has $N - k(d)$ d.f. Write $\lambda = \sum_1^u \mu_i^2/\sigma^2$. From (1.3) we have, as $\lambda \to 0$,

$$\begin{aligned} \beta_{f_{d,\alpha}}(\mu, \sigma^2) &= \sum_{\tau \in S_u} \beta_{F_{\tau d,\alpha}}(\mu, \sigma^2)/u! \\ &= \sum_{\tau \in S_u} [\alpha + g_{k(d),N-k(d)}(\alpha) \sum_{\tau \in S_u} n_{\tau(i)} \mu_i^2/\sigma^2 + 0(\lambda^2)]/u! \\ &= \alpha + g_{k(d),N-k(d)}(\alpha) \sum_i (\sum_\tau n_{\tau(i)}/u!) \mu_i^2/\sigma^2 + 0(\lambda^2) \\ &= \alpha + \frac{N}{u} g_{k(d),N-k(d)}(\alpha)\lambda + 0(\lambda^2). \end{aligned} \tag{4.2}$$

The desired conclusion now follows from Lemma 4.1.

Existing tables and charts of the power functions of the F-test and χ^2-test are presented in such forms (in terms of $\sqrt{\lambda/(k(d) + 1)}$, usually in inverted form and with wide spacing of arguments) as to make accurate comparisons of the

$\beta_{f_{d,\alpha}}$ difficult. This difficulty is made the worse by the fact that $\beta_{f_{d,\alpha}}$ is not (with an obvious exception) constant on the contour λ = constant, making it somewhat of a task to obtain $a_{f_{d,\alpha}}(c)$. It is not true, as might be supposed, that this minimum power on the contour λ = constant is always attained for a μ with all components equal, or else is always attained for a μ with all components except one equal to zero. To see this, consider the problem of Section 1 ($N = u = 2$, σ^2 known). Let C_α be the value such that, if Y is a normal random variable with 0 mean and unit variance, then $P\{ \mid Y \mid > C_\alpha\} = \alpha$. A direct computation of the power function of δ near $\lambda \equiv \mu_1^2 + \mu_2^2 = 0$ yields

$$\begin{aligned}\beta_\delta(\mu) &= \alpha + \frac{C_\alpha \exp(-C_\alpha^2/2)}{2\sqrt{2\pi}} \\ &\quad \cdot \{2(\mu_1^2 + \mu_2^2) + (C_\alpha^2 - 3)(\mu_1^4 + \mu_2^4)/3 + O(\lambda^3)\}.\end{aligned} \tag{4.3}$$

Hence, when c is sufficiently small, the minimum of $\beta_\delta(\mu)$ on the contour $\lambda = c$, neglecting the term $O(\lambda^3)$, is located at $\mu_1 = \sqrt{c}$, $\mu_2 = 0$ (or $\mu_2 = \sqrt{c}$, $\mu_1 = 0$) if $C_\alpha \leqq \sqrt{3}$ and at $\mu_1 = \mu_2 = \sqrt{c/2}$ if $C_\alpha \geqq \sqrt{3}$. When we include terms of higher order in μ, it is no longer even evident that the minimum must be attained at one of these two values of μ.

We see from (4.3) that $g_{1,\infty}(\alpha) = (2\pi)^{-\frac{1}{2}}C_\alpha \exp(-C_\alpha^2/2)$ and it is not hard to show that $g_{2,\infty}(\alpha) = -\alpha (\log \alpha)/2$ (see [12], equation (6.27), where λ is our $\lambda/2$). Thus, a comparison of $a'_{f_{d,\alpha}}$ for $k(d) = 1$ and 2 is given in this example by the following table:

α	$g_{1,\infty}(\alpha)$	$g_{2,\infty}(\alpha)$
.01	.037	.023
.05	.114	.075
.10	.175	.115
.20	.225	.161
.30	.242	.181
.50	.214	.173
.90	.050	.047

The following lemma shows that, as $\alpha \to 0$, the ratio of the second to third column above goes to 2 and, more generally, that $g_{i,\infty}(\alpha)/g_{j,\infty}(\alpha) \to j/i$ (this gives a comparison of the various δ_d for general N and u and for various $k(d)$ when σ^2 is known, as $\alpha \to 0$; see Lemma 4.3 for the case when σ^2 is unknown):

LEMMA 4.2. *As* $\alpha \to 0$,

$$g_{j,\infty}(\alpha) = -[1 + o(1)]\alpha(\log \alpha)/j. \tag{4.5}$$

PROOF. Fix j. Let k_α be such that if Y is a random variable with central χ^2 distribution with j d.f., then $P\{Y > k_\alpha\} = \alpha$. Let f_λ be the χ^2 density function with j d.f. and non-central parameter λ. A simple calculation shows that $df_\lambda(u)/d\lambda$ at $\lambda = 0$ is $f_0(u)[(u/2j) - 1/2]$. Hence, as $k_\alpha \to \infty$,

$$g_{j,\infty}(\alpha) = \int_{k_\alpha}^\infty f_0(u)[(u/2j) - 1/2]\,du = 1 + o(1))f_0(k_\alpha)k_\alpha/j, \tag{4.6}$$

by partial integration. On the other hand, an integration by parts shows that

$\alpha = 2f_0(k_\alpha)[1 + o(1)]$ as $k_\alpha \to \infty$, and hence that $k_\alpha = -2[1 + o(1)] \log \alpha$. This completes the proof.

4B. CASE II. We again treat the case where σ^2 is unknown, the other case being handled similarly (mainly, use Lemma 4.2 for Lemma 4.3). We first prove two simple lemmas.

LEMMA 4.3. *As* $\alpha \to 0$,

$$g_{ji}(\alpha) = i\alpha/2j + o(\alpha). \tag{4.7}$$

(This does not contradict (4.5), since j is fixed in (4.7).)

PROOF: Fix j and i. Let h_α be such that if Y has a central F-distribution with j and i d.f., then $P\{Y > h_\alpha\} = \alpha$. Let G_λ be the F density function with j and i d.f. and non-central parameter λ. From [12], equation (6.29) (with λ there replaced by our $\lambda/2$), it is easy to compute that $dG_\lambda(u)/d\lambda$ at $\lambda = 0$ is $G_0(u)$ $[(j + i)\, u/j(1 + u) - 1]/2$. Hence, as $k_\alpha \to \infty$,

$$g_{ji}(\alpha) = \frac{1}{2}\int_{k_\alpha}^{\infty} G_0(u)\left[\frac{i}{j} - \frac{j+i}{j} \cdot \frac{1}{1+u}\right] du = i\alpha/2j + o(\alpha). \tag{4.8}$$

In the next lemma, we use the following notation: n_i $(i = 1, \cdots, u)$ are again nonnegative integers with sum N. S_u is the symmetric group on u symbols and, for τ in S_u, $\bar{\mu}(\tau) = N^{-1} \sum_i n_{\tau(i)}\mu_i$; finally, $\bar{\mu} = u^{-1} \sum_i \mu_i$.

LEMMA 4.4. *For all* $u > 1$, μ, *and* N,

$$\sum_{\tau \varepsilon S_u} \sum_i n_{\tau(i)}(\mu_i - \bar{\mu}(\tau))^2 = u(u-2)![N - N^{-1} \sum_i n_i^2] \sum_i (\mu_i - \bar{\mu})^2. \tag{4.9}$$

PROOF. Since

$$\sum_{\tau \varepsilon S_u} n_{\tau(i)}^2 = (u-1)! \sum_i n_i^2 \tag{4.10}$$

and, for $i \neq j$,

$$\sum_{\tau \varepsilon S_u} n_{\tau(i)} n_{\tau(j)} = (u-2)! \sum_{i \neq j} n_i n_j = (u-2)![N^2 - \sum_i n_i^2], \tag{4.11}$$

we have

$$\begin{aligned} N^2 \sum_{\tau \varepsilon S_u} \bar{\mu}(\tau)^2 &= \sum_{i,j} \mu_i \mu_j \sum_{\tau \varepsilon S_u} n_{\tau(i)} n_{\tau(j)} \\ &= (u-1)! \sum_i n_i^2 \sum_j \mu_j^2 \\ &\quad + (u-2)![N^2 - \sum_i n_i^2][u^2\bar{\mu}^2 - \sum_j \mu_j^2]. \end{aligned} \tag{4.12}$$

Also,

$$\sum_{\tau \varepsilon S_u} \sum_i n_{\tau(i)} \mu_i^2 = \sum_i \mu_i^2 \sum_\tau n_{\tau(i)} = (u-1)! N \sum_i \mu_i^2. \tag{4.13}$$

Equations (4.12) and (4.13), together with

$$\sum_\tau \sum_i n_{\tau(i)}(\mu_i - \bar{\mu}(\tau))^2 = \sum_\tau \sum_i n_{\tau(i)}\mu_i^2 - N \sum_\tau \bar{\mu}(\tau)^2, \tag{4.14}$$

give (4.9).

The maximum for fixed $k(d)$ of the factor in square brackets on the right side of (4.9) will of course be nondecreasing in $k(d)$. It is the factor $g_{k(d)-1,h_d(\alpha)}$ which will increase rapidly enough as $k(d)$ is decreased to more than make up for the decrease in this term in brackets.

We are now ready to give our nonoptimality result in several illustrative examples of Case II, including those of Section 3C(1) and 3C(2). In all of these examples we ignore the divisibility properties; considerations when the design does not "divide up" properly (e.g., when $k(d)$ does not divide N in Example (1) below) are messier and their consideration does not help in the understanding of the phenomenon we are illustrating; thus, we shall assume whatever divisibility properties of N are needed to make our examples simple.

(1). *One-way analysis of variance.* In our first and simplest example, the setup is that of Section 4A, except that we now are testing $\mu_1 = \cdots = \mu_u$, and the appropriate F-tests are changed accordingly. Our result has the same implication as that stated just above Theorem 4.1, except that it now holds only when α is sufficiently small, and the optimum δ chooses each pair (i, j) $(i \neq j)$ with equal probability and sets $n_i = n_j = N/2$.

THEOREM 4.21. *For every d, α, and c,* (4.1) *holds; for fixed $k(d)$, $a'_{f_{d,\alpha}}$ is strictly decreasing in $\sum_i n_i^2$, attaining its maximum for $n_1 = \cdots = n_{k(d)} = N/k(d)$; for this choice of the n_i and for all α in some neighborhood of* 0, *$a'_{f_{d,\alpha}}$ is strictly decreasing in $k(d)$ for $k(d) > 1$; the results just stated for $a'_{f_{d,\alpha}}$ hold also for $a_{f_{d,\alpha}}(c)$ for all c in some neighborhood of* 0.

PROOF. From Lemma 4.4 and an argument like that of (4.2), we have, setting $\lambda = \sum_i(\mu_i - \bar{\mu})^2/\sigma^2$,

$$\beta_{f_{d,\alpha}}(\mu, \sigma^2) = \alpha + g_{k(d)-1,N-k(d)}(\alpha)(u - 1)^{-1}(N - N^{-1} \sum_i n_i^2)\lambda + O(\lambda^2). \tag{4.15}$$

When $n_1 = \cdots = n_{k(d)} = N/k(d)$, the ratio of values of $a'_{f_{d,\alpha}}$ corresponding to two values k and k' of $k(d)$ with $1 < k < k'$ is thus

$$\frac{g_{k-1,N-k}(\alpha)(1 - 1/k)}{g_{k'-1,N-k'}(\alpha)(1 - 1/k')}; \tag{4.16}$$

as $\alpha \to 0$, by Lemma 4.3, this ratio approaches

$$\frac{(N - k)/k}{(N - k')/k'} > 1, \tag{4.17}$$

completing the proof.

For a numerical example, suppose $N = 6$, $u = 3$, with σ^2 known. Comparing the δ_d's for which $k = 2$ and $k' = 3$, we see that $(1 - 1/k)/(1 - 1/k') = \frac{3}{4}$; thus, the ratio of the two $a'_{f_{d,\alpha}}$ in this example is $\frac{3}{4}$ times the ratio of second to third column in the table above Lemma 4.2. For $\alpha < .3$, then, the design with $k(d) = 2$ is locally better than that with $k(d) = 3$, in this example.

(2). *Several-way analysis of variance.* With or without interactions, the considerations are very similar to those of Example (1), and we omit them.

(3). *One-way heterogeneity.* In the setting described in Section 3A, suppose for fixed b, k, and u that BBD's exist for two possible choices u_1 and u_2 of the "number of treatments" to be tested, say for u_1 and u_2 with $1 < u_1 < u_2 \leqq u$. Let $d_i(i = 1, 2)$ be the design which uses the BBD with parameters b, k, and u_i to test the hypothesis $\mu_1 = \cdots = \mu_{u_i}$, and let δ_{d_i} be the corresponding randomized design which replaces the subscripts $1, \cdots, u_i$ here by $\tau(1), \cdots, \tau(u_i)$ with probability $1/u!$ for each τ (or, which is the same thing, which chooses each of the possible subsets of u_i treatments with equal probability). Otherwise, we use the same notation as in Example (1) of this section.

For any design setting, the parameter of the non-central χ^2 variable $t_d' V_d^{-1} t_d/\sigma^2$ is $(Q_d R\mu)' V_d^{-1}(Q_d R\mu)$, and by Lemma 2.3 and equation (3.1) this reduces in the case of a BBD d^* with parameters b, k, and u to

$$[r_{d^*1} - (\lambda_{d^*11} - \lambda_{d^*12})/k] \sum_i (\mu_i - \bar{\mu})^2/\sigma^2. \tag{4.18}$$

For the sake of arithmetical simplicity only, suppose that k/u_i is either an integer or is < 1 (the phenomenon to be studied persists without this assumption). Then, for $d^* = d_i$, the term in square brackets in (4.18) is easily computed to be

$$f(u_i) = \begin{cases} b(k-1)/(u_i - 1) & \text{if } k/u_i \leqq 1, \\ bk/u_i & \text{if } k/u_i \geqq 1. \end{cases} \tag{4.19}$$

Using now the counterpart of (4.18) for the designs d_i and the fact that, for $n_1 = \cdots = n_{u_q} = 1$ and all other $n_j = 0$, (4.9) becomes

$$\sum_{\tau \varepsilon S_u} \sum_i n_{\tau(i)}(\mu_i - \bar{\mu}(\tau))^2/u! = (u-1)^{-1}(u_q - 1) \sum_{i=1} (\mu_i - \bar{\mu})^2, \tag{4.20}$$

we obtain, corresponding to (4.16),

$$\frac{a'_{f d_1,\alpha}}{a'_{f d_2,\alpha}} = \frac{g_{u_1-1, bk-u_1-b+1}(\alpha)(u_1 - 1)f(u_1)}{g_{u_2-1, bk-u_2-b+1}(\alpha)(u_2 - 1)f(u_2)}. \tag{4.21}$$

By Lemma 4.3, as $\alpha \to 0$ this ratio approaches

$$\frac{(bk - u_1 - b + 1)f(u_1)}{(bk - u_2 - b + 1)f(u_2)}. \tag{4.22}$$

It is trivial to verify that $(bk - u - b + 1)f(u)$ is strictly decreasing in u for $u > 1$, so that the expression of (4.22) is > 1. Thus, we have proved

THEOREM 4.22. *For fixed b, k, u and all α in some neighborhood of 0, $a'_{fd_i,\alpha}$ is strictly decreasing in u_i for $i > 1$; the same is true for $a_{fd_i,\alpha}(c)$ for all c in some neighborhood of* 0.

This result implies that, if k is even, the locally best δ_{d_i} is that which chooses each pair of treatments with equal probability and assigns each of the two chosen treatments to $k/2$ of the plots in *every* block.

(4). *Two-way heterogeneity.* Using (3.2) in place of (3.1), the analogue of Theorem 4.22 can be proved for the YS design by an argument very similar to that of Example (3) just above, and which we therefore omit. One can even give an example of the lack of optimality of the YS in Δ_R without resorting to this analysis: for the case $k_1 = 2$, $k_2 = 3$, $u = 3$, the usual YS gives no d.f. to error, while the design which chooses two treatments at random and assigns each treatment to three plots, at least once in each row and column (full symmetry is impossible here) is uniformly more powerful for all α and all alternatives.

(5) *Other examples.* Examples like those mentioned in Section 3C (3) can be considered similarly, with analogous results. In particular, a trivial example in the case of a higher LS has already been mentioned in the first paragraph of Section 1.

5. Remarks and extensions. We list a few of the variants of the examples considered in this paper for which similar results hold, and make a few comments on questions which arise in connection with the paper, some of which present unanswered research problems.

1. A few of the other problems to which modifications of our method apply have been mentioned in Section 3C, and some of these will be considered elsewhere. Some such results hold under various non-normal probability laws (the point of the results of Section 4 is not merely that they hold for *many* models, but that they hold for the simplest, classical, normal model). Of course, a design which is optimum for one model may fail to be optimum for another, and vice versa; in particular, the results are obviously sensitive to change in the function ψ (even to changes to other quadratic forms and for a fixed d, as indicated in Section 2). Optimality criteria can be altered in other ways; e.g., one can consider $M_{\alpha,c,\sigma}$-optimality, in imitation of 2A(c). The extent of completeness of nonoptimality results like those on the higher LS design (first paragraph of Section 1) and YS design (Section 4B(4)) obviously depends on whether or not σ^2 is known. The results for Model II and certain mixed models of the analysis of variance differ considerably from those for the model considered herein, since the dependence of the power function on the design (and on the test, for a fixed design) is so different; however, similar methods can be used there.

2. Besides changing the model, one can also change the decision space. From the examples cited just above regarding higher LS and YS designs, it is clear that *nonoptimality* results for some classical symmetrical designs hold for many decision problems. For normal and certain nonparametric point estimation problems, the discussion of [2] and [3] indicates why Section 3 yields *optimality* results (these actually hold for many weight functions other than squared error). Another typical estimation result is contained in the fact that the designs d^* of Theorems 3.1 and 3.2 maximize the trace of V_d^{-1} and that V_{d^*} is a multiple of the identity; from these it follows at once that *average variance of* t_d (= trace of $\sigma^2 V_d/(u - 1)$) *is a minimum for* d^*. However, the results of Section 4 are meaningless for many common weight functions, since V_δ is not the covariance matrix of b.l.e.'s. Similar results hold for some interval estimation problems; for estimating $\psi(\mu)/\sigma^2$ (e.g., in "multiple comparison" problems), Section 4 is now sometimes relevant. Multiple classification and ranking problems can be treated in like

manner. Of course, a D-optimum design minimizes the approxiate *generalized variance* in point estimation problems.

3. As we have mentioned, nonoptimality results like those of Section 4 do not depend on the nonrandomized design being symmetrical. Much more difficult is the problem of characterizing optimum designs in the sense of Section 3 when there is no appropriate symmetry. (Even the considerations of Sections 3B(2) and 4B(3 and 4) become messier without the restrictions on k_i/u and k/u; it would be nice if neat proofs could be given in such cases.) It seems often to be true that a design which is "closest to being symmetrical" in an appropriate sense (e.g., note the dependence on $\sum n_i^2$ in Theorem 4.21) is optimum, but the algebra involved in proving this can be tedious. Problems like that cited in the next to last paragraph of Section 3C(3) can be similarly unwieldy under heteroscedasticity. In connection with a general symmetry-invariance approach like that mentioned below (1.3), we note that appropriate symmetry of X_d is useful as a partial sufficient condition for some optimality results, but that appropriate symmetry of $X_d'X_d$ is what is really relevant (for the functions ψ we have considered).

4. We have mentioned in Section 2 some of the difficulties present in verifying M- (or sometimes L-) optimality. If b_d is not a constant for d in Δ', or if randomized designs are considered, this difficulty is increased by the nonconstancy of the d.f. for $\bar{S}_d$, etc. (We have not considered here a thorough investigation of the optimality properties of the procedures δ_d of Section 4). The difficulty encountered in connection with M-optimality in the nonconstancy of the power functions of competing tests on appropriate contours also manifests itself when one tries to find a *most stringent* design (the "envelope power function" being obtained by taking the supremum of β_ϕ over all ϕ in $H_d(\alpha)$ and all d in Δ or Δ_R). The method of invariance used to prove 2A(f) cannot even supply a start here, and the method of [6] or [7] used to prove 2A(c) yields no analogue here where d is not fixed. Thus, even in such a simple example as that of Section 2B, the stringency problem seems extremely difficult.

It is interesting to note that the δ_d of Section 4 lack a "consistency" property if $k(d) < r$, in that $a_{f_{d,\alpha}}(c)$ does not approach 1 as $c \to \infty$ (in fact, it is easy to see that the μ for which one component of $R\mu$ is $\sigma\sqrt{c}$ and all others are 0 is asymptotically worst on the contour $\psi(u)/\sigma^2 = c$ as $c \to \infty$, giving power approaching $[k(d) + (r - k(d))\alpha]/r$). Nevertheless, the question remains open as to whether any of these δ_d, or some other design and associated test which lacks this consistency property, is nevertheless most stringent.

The reader will not find it difficult in considerations like those of Section 3B to supply the details which show, in some problems, that the D-optimum (or L- or E-optimum) design is unique. When uniqueness is not present (e.g., for some α and ϵ, both designs in Section 2B will be L-optimum), questions of global admissibility arise. A related problem is to look not at a fixed contour or family of contours in the manner of Section 2, but rather to characterize complete classes of designs in the manner of [3]; in such considerations, especially for problems of testing hypotheses, Section 4 shows that results like those of [3] must be altered if Δ_R is considered rather than Δ.

Finally, we may remark that, for a fixed d, the problem of characterizing an L_α-optimum test is unsolved; the generalized Neyman-Pearson Lemma does not seem to yield explicit results easily, although it is not difficult to show that an L_α-optimum test is obtained by replacing the numerator of the F-test by some other quadratic form.

REFERENCES

[1] A. WALD, "On the efficient design of statistical investigations," *Ann. Math. Stat,* Vol. 14 (1943), pp. 134–140.

[2] S. EHRENFELD, "On the efficiency of experimental designs," *Ann. Math. Stat.*, Vol. 26 (1953), pp. 247–255.

[3] S. EHRENFELD, "Complete class theorems in experimental design," *Third Berkeley Symposium on Probability and Statistics,* Vol. 1, University of California Press, 1946.

[4] R. C. BOSE, "Least Squares Aspects of Analysis of Variance" (mimeographed notes), Institute of Statistics, North Carolina.

[5] E. L. LEHMANN, "Notes on Testing Hypotheses" (mimeographed), Associated Student Store, University of California, Berkeley.

[6] A. WALD, "On the power function of the analysis of variance test," *Ann. Math. Stat.*, Vol. 13 (1942), pp. 434–439, and Vol. 15 (1944), pp. 330–333.

[7] J. WOLFOWITZ, "The power of the classical tests associated with the normal distribution," *Ann. Math. Stat.*, Vol. 20 (1949), pp. 540–551.

[8] P. L. HSU, "Analysis of variance from the power function standpoint," *Biometrika,* Vol. 32 (1941), p. 62.

[9] G. A. HUNT AND C. STEIN, "Invariant tests," unpublished.

[10] J. KIEFER, "Invariance, sequential minimax estimation, and continuous time processes," *Ann. Math. Stat.*, Vol. 28 (1957).

[11] S. ISAACSON, "On the theory of unbiased tests of simple statistical hypotheses specifying the values of two or more parameters," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 217–234.

[12] H. B. MANN, *Analysis and Design of Experiments,* Dover, New York, 1949.