OPTIMUM AND MINIMAX EXACT TREATMENT DESIGNS FOR ONE-DIMENSIONAL AUTOREGRESSIVE ERROR PROCESSES[1]

By J. Kiefer and H. P. Wynn

University of California, Berkeley and Imperial College, London

A theory is developed following work by Williams (1952) and Kiefer (1960) for exact treatment designs in one dimension in which the errors are a stationary process. It is shown that the designs which achieve the minimax value of any of a wide class of functionals on the information matrix for estimation of treatment differences have a special property. If the process is autoregressive of order p then a random piece of the design of length $p + 1$ exhibits uncorrelated treatment values. Such designs can be formed using full length cyclic error-correcting codes of a suitable order. A new technique is developed for classifying the ergodic combinatorial structure of exact designs of arbitrary or infinite length. It is shown that all designs are, to pth order, generated by a finite number of sequences with finite length. The classification is given explicitly up to order 3. The method is used to find asymptotically optimum designs for different processes. It is also shown that the designs can be achieved to within an arbitrarily good approximation as the realization of an ergodic Markov chain of sufficiently high order.

1. Introduction. The problem of laying treatments out in a line, or in time, so as to obtain good estimation of between-treatment contrasts when the error process is autocorrelated was first systematically studied by Cox (1951) and Williams (1952). One of the present authors, Kiefer (1960), extended the results of Williams and proved a number of optimality results, some of which were conjectures from these earlier papers. In particular, first and second order autoregressive processes were studied in some detail. The present paper extends these results using a flexible formulation in which a general minimax property can be established for certain special designs. Briefly, it is sufficient that in every "window" of certain fixed length, every vector of treatments occurs once (or the same number of times). We shall show that such designs can be formed using full-length cyclic error correcting codes. These are an example of a pseudo-random sequence used in communication theory.

Most of the related work in experimental design theory has been on two- and higher-dimensional designs, often with added row, column, or block effects. We mention particularly a recent paper by the authors, Kiefer and Wynn (1981). The approach taken in that paper was somewhat different. The analysis was based on the ordinary least squares estimator and restricted to a nearest neighbor model equivalent to a first order moving average model. The models were more complex, involving row, column and treatment effects. Here we carry out the full

431

best linear unbiased (BLU) estimation but confine ourselves to an in-depth study of the one-dimensional case. We should particularly like to acknowledge the work of Martin (1977, 1979), and papers by Berenblutt and Webb (1974) and Duby et al. (1977). A representative paper on sampling for continuous processes in which there is moderate literature is Bickel and Herzberg (1979).

To start with, we formulate the problem on the circle rather than the line. The extension to the line can be carried out straightforwardly but is best treated in an asymptotic way and is left until Sections 6 and 7.

Consider N units labeled by the letter t $(t = 1, \cdots, N)$ situated in order around a circle. To unit t attach a random variable Y_t $(t = 1, \cdots, N)$. Let the Y_t have a joint distribution having the autocorrelation structure

(1) $$Y_t + \rho_1 Y_{t-1} + \cdots + \rho_{N-1} Y_{t-N+1} = \varepsilon_t \quad (t = 1, \cdots, N)$$

where because of the circulant nature of the set-up we identify unit $-i$ with unit $N - i$ $(i = 1, \cdots, N)$. The "innovation" vector $\varepsilon = (\varepsilon_1, \cdots, \varepsilon_N)^T$ has mean and covariance matrix $E(\varepsilon) = 0$, $\mathrm{Cov}(\varepsilon) = \sigma^2 I_{N \times N}$. Throughout this paper we shall assume that the ρ_r $(r = 1, \cdots, N-1)$ are known. (Lemma 1 and the end of Section 7 consider the ρ_r as variable. From a practical point of view, this might mean estimating the ρ_r from the data and using those estimates to "adapt" the BLU estimate for fixed ρ_r). At this stage we make no restriction on the ρ_r, but later, for a corresponding process on the line instead of on the circle, a stationarity condition is required. Whether σ^2 is known is immaterial. The model is said to be pth *order* (or less) if $\rho_{p+1} = \rho_{p+2} = \cdots = \rho_{N-1} = 0$. Usually, we shall treat the order $p(0 \le \rho \le N - 1)$ as fixed and note here that it will play a crucial role in the construction of designs. Write $Y = (Y_1, \cdots, Y_N)^T$ and take X_p to denote the $N \times N$ circulant matrix

$$X_p = \begin{bmatrix} 1 & \rho_{N-1} & \rho_{N-2} & \cdot & \cdot & \cdot & \rho_1 \\ \rho_1 & 1 & \rho_{N-1} & \cdot & \cdot & \cdot & \cdot \\ \rho_2 & \rho_1 & 1 & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & 1 & \rho_{N-1} \\ \rho_{N-1} & \cdot & \cdot & \cdot & \cdot & \rho_1 & 1 \end{bmatrix}$$

with $\rho_i = 0$ for $i > p$. Then assuming the model is pth order (1) can be written

(2) $$X_p Y = \varepsilon.$$

Now assume that there are k treatments labeled $1, \cdots, k$, one applied to each unit. With treatment i, associate an unknown constant α_i $(i = 1, \cdots, k)$. Let $[t]$ denote the label of the treatment applied to unit t. The actual observation on unit t is then assumed to be $W_t = Y_t + \alpha_{[t]}$. Denote by U the $N \times k$ matrix describing the allocation of treatments so that its entries are

$$U_{ti} = 1 \quad \text{if } [t] = i$$

$$= 0 \quad \text{otherwise.}$$

Thus, U has exactly one unity in each row, the other entries being zero. If

$W = (W_1, \cdots, W_N)^T$, the full treatment model becomes

(3) $W = Y + U\alpha,$

where $\alpha = (\alpha_1, \cdots, \alpha_k)^T$.

We shall be interested in considering the BLU estimates of contrasts among the α_i based on the observations W. Following Kiefer and Wynn (1981), we consider a vector θ of standardized treatment contrasts

$$\theta = (I_k - k^{-1}J_k)\alpha$$

where I_k is the $k \times k$ identity matrix and J_k the $k \times k$ matrix of ones. Write $P = I_k - k^{-1}J_k$ and notice that this is the usual projection operator associated with obtaining the treatment effects in one-way analysis of variance. Multiplying (3) through by X_p and using (2) we obtain

$$X_p W = X_p U\alpha + \varepsilon.$$

This is now in standard regression form and the covariance matrix for $\hat{\alpha}$, the BLU estimate of α, is

$$\mathrm{Cov}(\hat{\alpha}) = \sigma^2(U^T X_p^T X_p U)^{-1},$$

where we have assumed that $\mathrm{Rank}(X_p U) = k$. Define $A = U^T X_p^T X_p U$. Thus the BLU estimate of θ, $\hat{\theta}$, has covariance matrix

$$\mathrm{Cov}(\hat{\theta}) = \sigma^2 P A^{-1} P.$$

It is more tractable to work with

$$C = PAP$$

(ignoring σ^2). However, C is not in general an inverse of $PA^{-1}P$. Writing B^- to denote the Moore-Penrose g-inverse of a matrix B, we have (Zyskind, 1967, Theorem 2)

(4) $\bar{C} \overset{\text{def}}{=} (PA^{-1}P)^- = C - D$

where D is a nonnegative definite matrix which is zero when P and A (or P and A^{-1}) commute. Notice that both C and $\bar{C}$ have row and column sums equal to zero.

2. Optimum design. The design is defined to be the allocation of treatments to units and enters through $[t]$ or the matrix U. It is labeled by d. Where we are concerned to denote the dependence of the information matrix on d we write A_d or C_d. However, the vector of ρ_i values $\rho = (\rho_1, \cdots, \rho_p)^T$ (assuming a pth order model) also affects C through X_p so occasionally we may write $C_d(\rho)$ or $A_d(\rho)$.

We shall restrict slightly the definition of universally optimum designs of Kiefer (1958, 1975) and give a slighty different version of the key result (Kiefer, 1975, Proposition 1). This version is required to cope with the matrix D in (4).

Let $\mathscr{C}$ be the class of all nonnegative definite matrices with row and column sums zero. A design (here) is called universally optimum among the class of all

designs if it minimizes $\Phi(\bar{C})$ for every real valued function Φ on $\mathscr{L}$ which is (i) convex, (ii) invariant under permutations of coordinates (rows and columns simultaneously), (iii) nonincreasing in the usual (Loewner) sense: $G - H$ non-negative definite $\Rightarrow \Phi(G) \leq \Phi(H)$. Notice that we assume a fixed value of ρ.

We now show that a sufficient condition for a design d^* to be universally optimum is that

(a) A_{d^*} is a multiple of $aI_k + bJ_k$ for constants a, b.

(b) $\mathrm{trace}(C_{d^*}) = \mathrm{trace}(PA_{d^*}P) = \max_d \mathrm{trace}(C_d)$.

In proving this, we use Proposition 1 of Kiefer (1975) on the matrices C, which we may do since our condition (iii) is stronger than Kiefer's. From (a) and (b) we thus conclude that $\Phi(C_d)$ is minimized by d^*. By (4) and (iii) we have $\Phi(\bar{C}_d) \geq \Phi(C_d)$ with equality when $D = 0$. But (a) implies that A_{d^*} commutes with P and hence that $D = 0$ for d^*. This completes the proof.

Condition (a) is summarized by saying A_{d^*} is completely symmetric (CS). For individual values of ρ, the problem of finding designs d^* is fairly difficult, particularly because of the complex nature of the class of designs. We give a "large N" solution of the problem in Section 7. We also discuss, at the end of that section, "complete classes" of designs in terms of the function (18) of the ρ_r, which may be thought of as analogous to the risk function of a procedure (here, of a design, in terms of its π_r's).

An alternative to finding optimum designs for given ρ, or a complete class, is to take a minimax approach. A design d^* is said to be universally minimax if

$$\max_\rho \Phi(C_{d^*}) = \min_d \max_\rho \Phi(C_d),$$

for all Φ satisfying (i), (ii) and (iii) above. The maximum is taken over all vectors ρ in R^P.

It will now be shown that if d^* is completely symmetric and satisfies the corresponding maximin condition on its *trace* then it is minimax. We note from the form of X_ρ that $\mathrm{trace}(C_d(\rho))$ is a nonnegative quadratic in ρ, so that $\min_\rho C_d(\rho)$ is always attained.

LEMMA 1. *Suppose d^* is a design for which $A_{d^*}(\rho)$ is completely symmetric for all ρ and*

(5) $$\min_\rho \mathrm{trace}(C_{d^*}(\rho)) = \max_d \min_\rho \mathrm{trace}(C_d(\rho)).$$

Then d^* is universally minimax.

PROOF. Let ρ^* be the value of ρ achieving $\min_\rho \mathrm{trace}(C_{d^*}(\rho))$. Since $C_{d^*}(\rho)$ is completely symmetric for all ρ, all these $C_{d^*}(\rho)$'s (as ρ varies) are multiples of $C_{d^*}(\rho^*)$ and hence, by (iii),

(6) $$\Phi(C_{d^*}(\rho^*)) = \max_\rho \Phi(C_{d^*}(\rho)), \quad \text{for all } \Phi.$$

Now take any other d. By (5) there is a value of ρ, say ρ', such that

$$\mathrm{trace}\, C_d(\rho') \leq \mathrm{trace}(C_{d^*}(\rho^*)).$$

Using the argument following (a) and (b) in Section 1 applied to the *pair* of

matrices $C_d(\rho')$ and $C_{d^*}(\rho^*)$ we have

$$\Phi(C_d(\rho')) \geq \Phi(C_{d^*}(\rho^*))$$

for all Φ. Thus

(7) $$\max_\rho \Phi(C_d(\rho)) \geq \Phi(C_{d^*}(\rho^*)).$$

Combining (7) with (6) gives the result. □

The lemma helps considerably in finding minimax designs. The method is first to get the trace maximin and then hope that we can find a design which in addition makes C_d completely symmetric.

3. The structure of designs on the circle. ' Consider the treatments arranged on their units around the circle. Consider a "window" of length $p + 1$ placed on the circle so that when it is in position t the locations $1, \cdots, p + 1$ in the window show the adjacent units $t - p, t - p + 1, \cdots, t - 1, t$, respectively. The treatments on these units will be $[t - p]$, $[t - p + 1]$, $\cdots$, $[t - 1]$, $[t]$, respectively. Now suppose that the window, instead of being in a fixed position, is placed at random with probability $1/N$ attached to each of the N possible positions. The vector of treatments in the random window is now a random vector $T_1, \cdots, T_{p+1}$ of treatment labels. Note that it is stationary under shifts; that is, $T_{i_1}, \cdots, T_{i_r}$ has the same joint distribution as $T_{i_1+\tau}, \cdots, T_{i_r+\tau}$ for $1 \leq i_1 < \cdots < i_r \leq p + 1$ and $0 < \tau \leq p + 1 - i_r$ (all integers).

Define $T_1^{(i)}, \cdots, T_{p+1}^{(i)}$ by

$$T_r^{(i)} = 1 \quad \text{if} \quad T_r = i \ (i = 1, \cdots, k; r = 1, \cdots, p + 1),$$

$$= 0 \text{ otherwise}$$

so that $T_r^{(i)}$ is the indicator for the ith treatment in the rth location in the window.

From this *randomized* version of the design we can recapture certain useful design quantities. Thus define

$$n_i = \#\{t \mid [t] = i\} \quad (i = 1, \cdots, k),$$

that is, the number of times treatment i occurs. We put $m_i = n_i/N$ and call the design treatment-balanced if $n_1 = n_2 = \cdots = n_k = N/k$. Define, for $r = 1, \cdots, p$,

$$\pi_r^{(ij)} = N^{-1}\#\{t \mid [t] = i, [t - r] = j, 1 \leq t \leq N\} \quad (i, j = 1, \cdots, k).$$

This is the proportion of times treatment j is r steps "behind" treatment i. (Here, $t = 1, \cdots, N$ but $t - r$ is identified, when ≤ 0, with $N + t - r$.) When $i = j$ define $\pi_r^{(i)} = \pi_r^{(ii)}$. Of considerable importance will be

$$\pi_r = \sum_{i=1}^k \pi_r^{(i)} \quad (r = 1, \cdots, p),$$

which is the proportion of times the same treatment occurs r units apart, regardless of treatment label. We may connect the $T_r^{(i)}$ variables with these design

quantities: since $T_r^{(i)} T_r^{(i')} = \delta_{ii'} T_r^{(i)}$ where $\delta_{ii'} =$ Kronecker symbol,

$$E(T_r^{(i)}) = m_i, \quad E(T_r^{(i)} T_r^{(i')}) = \delta_{ii'} m_i,$$

and, whether or not $i = i'$,

$$E(T_j^{(i)} T_{j+r}^{(i')}) = \pi_r^{(ii')} \quad \text{for } r > 0.$$

We may also express the matrix C in terms of these quantities. If we consider the individual contributions to the entries A_{ij} of A_d from different positions of the window, we can obtain these entries as expectations with respect to the window process $\{T_j^{(i)}\}$. Thus, writing $\rho_0 = 1$, whether or not $i = i'$ we have

(8)
$$N^{-1} A_{ii'} = E\{(\sum_{r=0}^{p} \rho_r T_{p+1-r}^{(i)})(\sum_{r=0}^{p} \rho_r T_{p+1-r}^{(i')})\}$$

$$= \delta_{ii'} m_i (\sum_{r=0}^{p} \rho_r^2) + 2 \sum_{r=1}^{p} \pi_r^{(ii')} \sum_{s=0}^{p-r} \rho_s \rho_{s+r}.$$

Since $\sum_i T_j^{(i)} = 1$, we have $N^{-1} \sum_{i=1}^{k} \sum_{i'=1}^{k} A_{ii'} = (\sum_{r=0}^{p} \rho_r)^2$. Consequently,

(9)
$$N^{-1}\text{trace}(C) = N^{-1}\text{trace}(PAP) = N^{-1}\text{trace } A - (Nk)^{-1} \sum_{i=1}^{k} \sum_{i'=1}^{k} A_{ii'}$$

$$= \sum_{r=0}^{p} \rho_r^2 - k^{-1}(\sum_{r=0}^{p} \rho_r)^2 + 2 \sum_{r=1}^{p} \pi_r \sum_{s=0}^{p-r} \rho_s \rho_{s+r}.$$

4. Minimax designs on the circle. We are now in a position to translate the conditions of minimaxity in Lemma 1 into conditions on the design. First consider the condition on the trace, (5). Write the covariance matrix of $T^{(i)} = (T_{p+1}^{(i)}, \cdots, T_1^{(i)})^T$ in the form

$$\Gamma^{(i)} = \begin{bmatrix} \tilde{\Gamma}^{(i)} & \gamma^{(i)} \\ \gamma^{(i)T} & m_i(1 - m_i) \end{bmatrix}.$$

We also write Γ, $\tilde{\Gamma}$, γ for the sum over i of the corresponding quantities. Also write $\bar{\rho}^T = (1, \rho^T) = (1, \rho_1, \cdots, \rho_p)$ and $\bar{T}^{(i)} = \bar{\rho}^T T^{(i)}$. In place of (8) we can also write

$$N^{-1} A_{ii} = E(\bar{T}^{(i)})^2 = E(\bar{\rho}^T T^{(i)})^2$$

$$= \bar{\rho}^T [\Gamma^{(i)} + (ET^{(i)})(ET^{(i)})^T] \bar{\rho}$$

$$= \bar{\rho}^T [\Gamma^{(i)} + m_i^2 J_{p+1}] \bar{\rho}.$$

From this and the first line of (9), with e the p-vector $(1, 1, \cdots, 1)^T$ and $S = \sum_i (m_i - k^{-1})^2$, we obtain

$$N^{-1}\text{trace } C = \bar{\rho}^T [\Gamma + S J_{p+1}] \bar{\rho}$$

$$= 1 - k^{-1} + 2\rho^T [\gamma + Se] + \rho^T [\tilde{\Gamma} + S J_p] \rho.$$

When $h = [\gamma + Se]$ is in the column space of $H = [\tilde{\Gamma} + S J_p]$, the minimum of $2\rho^T h + \rho^T H \rho$ is $-h^T H^- h$, attained at $\rho = -H^- h$ (not uniquely if H is singular). Thus,

(10) $$N^{-1} \min_\rho \text{trace}(C(\rho)) = 1 - k^{-1} - [\gamma + Se]^T [\tilde{\Gamma} + S J_p]^- [\gamma + Se].$$

This is clearly maximized when $S = 0$ and $\gamma = 0$. The other case gives the same values.

Since $\Gamma^{(i)}$ and hence Γ are circulant, $\gamma = 0$ implies that all of the off-diagonal elements of Γ are zero. Thus max min trace(C) is achieved when

(11)
$$m_1 = m_2 = \cdots = m_k = k^{-1}(\text{treatment balance}),$$
$$\Gamma \equiv \sum_i \Gamma^{(i)} = (1 - k^{-1})I_{p+1}.$$

Then it is easily seen that (11) is equivalent to $m_1 = \cdots = m_k = k^{-1}$ and $\pi_r = k^{-1}$ $(r = 1, \cdots, p)$. Now (11) says that the average covariance matrix of the separate indicators for the separate treatments in the window process is a nonzero multiple of the identity matrix. It is, of course, sufficient for (11) that

(12)
$$\Gamma^{(i)} = k^{-1}(1 - k^{-1})I_{p+1} \text{ ,}$$

$(i = 1, \cdots, k)$ so that each of the indicator processes (considered by itself) is uncorrelated. We shall return to this in a moment.

The condition in Lemma 1 that $A_{d^*}(\rho)$ be completely symmetric for all ρ is more restrictive. Assume that $m_1 = \cdots = m_k = k^{-1}$ already holds. Then it is easy to see, looking at (8), that all $A_{ii'}$ equal and A_{ii} all equal forces

(13)
$$\pi_r^{(i_1, i_1')} = \pi_r^{(i_2, i_2')} \,(r = 1, \cdots, p)$$

for all $i_1 \neq i_1'$ and $i_2 \neq i_2'$, as well as

(14)
$$\pi_r^{(i)} = \pi_r^{(i')}$$

for all $i \neq i'$ $(r = 1, \cdots, p)$. This says that any pair of different treatments (i, i') $(i \neq i')$ is r units apart the same number of times as any other pair of different treatments and any pair of identical treatments is r units apart the same number of times as any other pair of identical treatments. We call a design with properties (13) and (14) *completely balanced*. Actually, since $m_i = \pi_r^{(i)} + \sum_{j \neq i} \pi_r^{(ij)}$, complete balance implies treatment balance, although it does not by itself imply $\pi_r = k^{-1}$ $(r = 1, \cdots, p)$. Furthermore (11) together with complete balance implies (12). Thus also $\pi_r^{(i)} = 1/k^2$ $(r = 1, \cdots, p)$. We summarize these results as follows.

THEOREM 1. *A sufficient condition for a design d^* to be universally minimax is that*
(1) *it is completely balanced,*
(2) $\pi_r = 1/k$ $(r = 1, p)$.

This theorem generalizes to arbitrary order Theorems 3.1.2 and 3.1.4 of Kiefer (1960).

In the next section we give a method of constructing designs having the properties in Theorem 1.

5. Cyclic codes as minimax designs. It is simple to show that complete balance and the "window uncorrelated property" ((1) and (2) of Theorem 1) can be obtained by allowing each of the k^{p+1} possible $(p + 1)$-tuples of treatment

numbers to occupy our window of length $p + 1$ exactly once, as it moves around the circle. For example, let $k = 3$, $p = 2$ and label the treatment 0, 1, 2 rather than 1, 2, 3. Take the design with 27 units and treatments laid out as follows:

(15)   001101021222100220201211120.

The circle is completed by making the last "0" a neighbor of the first. As we move a window of length 3 along the sequence every triple i, j, k $(i, j, k = 0, 1, 2)$ occurs just once (200 and 000 cover the join of the two ends). Properties (1) and (2) of Theorem 1 are easily verified. Such a sequence in fact has a stronger property, namely, the random window exhibits not just uncorrelated entries but an *independent* 3-state process. A sequence with this property is called pseudo-random in the communications theory literature. The sequence was generated in the following way. Let $V_t = [t]$, the treatment label (now 0, 1 or 2) of the tth unit. Then the sequence was generated by the recurrence relation

(16) $$V_t = V_{t-1} + 2V_{t-2} + 2V_{t-3}$$

where all integers are identified with their residues mod 3. To start with, take 001. Continuing up to the end of the first cycle excludes 000 so one "dummy" zero must be added at the end before joining up to form the circle. This is a special case of a full length cyclic error-correcting code. There is a huge literature on the subject and several excellent books (see, for example, Berlekamp, 1968, MacWilliams and Sloane, 1977); and, we confine ourselves to a brief summary.

Let $k = q^m$ where q is prime. Consider a general recurrence relationship

(17) $$V_t = \sum_{i=1}^{p} \alpha_i V_{t-1}$$

in which all treatment values V_t and coefficients α_i are identified with members of the Galois field $GF(q^m)$ of order q^m. In order for the recurrence (17) to generate a full length cyclic code with the required property that every $(p + 1)$-tuple of elements appears just once (except for $00 \cdots 0$), the α_i must be of a special form.

Consider the extended Galois field consisting of all polynomials of degree p whose coefficients lie in $GF(q^m)$. This is itself a field if every polynomial of higher order than p is identified with its residue mod some fixed polynomial of degree $p + 1$ which is irreducible over $GF(q^m)$. The polynomial field (which can be identified with $GF(k^p)$) has nonzero elements which form a cyclic group under multiplication. An irreducible polynomial $h(x)$ of degree $p + 1$ over $GF(q^m)$ is called *primitive* if all the elements of the cyclic group are generated as powers of a solution x in $GF(k^p)$ of $h(x) = 0$. The element x is said to be a primitive element of $GF(k^p)$. A primitive polynomial always exists.

A necessary and sufficient condition that the recurrence (17) generate a full length code with the required property is that

$$h(x) = x^k - \alpha_1 x^{k-1} - \alpha_2 x^{k-2} - \cdots - \alpha_p$$

is a primitive polynomial over $GF(q^m)$. Thus in our example (16), the polynomial is $x^3 - x^2 - 2x - 2$ $(\equiv x^3 + 2x^2 + x + 1)$, which is primitive over $GF(3)$. The polynomial is not unique; $x^3 + 2x + 1$ would do equally well, producing a different sequence.

Lists of primitive polynomials are available in many books on algebra, coding theory and combinatorial theory. In the communications literature, the designs are called m-sequences or pseudo-random sequences (see MacWilliams and Sloane (1977), Chapter 14). Techniques are available for generating pseudo-random sequences for values of k other than prime powers (see Shedd and Sarware, 1979).

We have shown, then, that it is possible to find minimax designs exactly for certain values of N and k.

6. Asymptotic designs for the line. There is no obvious way of extending the results on the circle to results on the line in an exact fashion for *finite* N. For example, we may form a design with $N = 29$ by adding 00 to the end of (15) so that triples 200 and 000 are included. However, since the first two observations are included, the conditions for minimaxity are upset. Moreover, for fixed N there can be at most a finite number of different designs and so the set of possible design quantities m_i, π_r, $\pi_r^{(i)}$, etc. is finite. This means that some values of these quantities which are useful from an optimization point of view may not be achievable. We have seen in Section 5, for example, that we have to choose the value of N carefully to get an exactly minimax design. Following Kiefer (1960) we therefore allow N to tend to infinity and discuss allowable asymptotic values of the design quantities.

First we shall redefine the quantities to emphasize the dependence on N. The designs will be on the line although the cyclic nature of some designs will play an important role.

Let d_N be a design on the line; that is, $d_N = (V_1, \cdots, V_N)$ where, as before, $V_t = [t]$ is one of the integers $1, \cdots, k$ ($t = 1, \cdots, N$). Let $\mathscr{D}_N$ be the set of all d_N so that $\#(\mathscr{D}_N) = k^N$. For any $d_N \in \mathscr{D}_N$ let

$$\pi_{r,N}^{(i)} = (N - p)^{-1} \#\{t \mid V_t = V_{t-r} = i, \quad p + 1 \le t \le N\}$$

and

$$\pi_{r,N} = \sum_{i=1}^k \pi_{r,N}^{(i)}.$$

If there is a sequence of designs $\{d_N \in \mathscr{D}_N, N \in \mathscr{N}\}$ where $\mathscr{N}$ is an unbounded set of natural numbers, for which, for $1 \le r \le k$,

$$\pi_{r,N} \to \text{a limit } \pi_r \text{ as } N \to \infty, \quad N \in \mathscr{N},$$

we say that $\pi_1, \cdots, \pi_p$ or simply the vector $\pi = (\pi_1, \cdots, \pi_p)$ is *asymptotically achievable* (aa). Throughout this section the π_r will be (aa) values.

The rest of this section is devoted to the difficult problem of finding the set of asymptotically achievable π_r values ($1 \le r \le p$). We do not need to use the individual $\pi_{r,N}^{(i)}$ because a symmeterization procedure to be described shortly will yield sequences for which $\pi_{r,N}^{(i)} \to (\pi_r/k)$ ($N \to \infty$) ($i = 1, \cdots, k; r = 1, \cdots, p$). Let $\Pi_p \equiv \Pi_p(k)$ be the set of all (aa) vectors π for given k. The development is based on a connection between the (aa) π-vectors and the covariance of stationary discrete time processes. The technique is discussed in some detail in Kiefer and Wynn (1983) for the binary ($k = 2$) case.

Consider the sequence for a given i, $\{Z_t\}_0^\infty$:

$$Z_t = 1 \quad \text{if} \quad V_t = i$$

$$= -1, \quad \text{otherwise.}$$

Then if

$$C_{r,N} = \frac{1}{N-p} \sum_{t=p+1}^N Z_t Z_{t-r},$$

we have $\pi_{r,N} = \frac{1}{2}(1 + C_{r,N})$. Now $C_{r,N}$ is essentially the sample covariance function of the $\{Z_t\}$ values. If $\pi_{r,N} \to \pi_r$ $(r = 1, \cdots p)$ then $C_{r,N} \to 2\pi_r - 1$. If the $\{Z_t\}$ values are generated as a realization of a stationary ergodic zero mean binary process then $C_{r,N} \to C_r$ with probability one where the C_r is the covariance function of the process. If $(1/N) \sum Z_t \to 0$ the set of π_r values $(r = 1, \cdots, p)$ is contained in the set $\frac{1}{2}(C_r + 1)$ values allowable as the covariance of a zero mean stationary binary process. Work by Shepp (1967) and Masry (1972) show that the two sets are the same and that the zero mean condition is irrelevant. We follow the results of Hobby and Ylvisaker (1964) to characterize stationary k state processes in terms of the distribution induced in a window of $p + 1$ neighbouring time points. This theory can be considered as the correct way to "unwrap" the structure of designs on a circle discussed in Section 3 onto the real line (discrete time).

Let $\{Z_t\}_0^\infty$ be a stationary k state process. Consider a window of $p + 2$ neighbouring time points, say $t = 0, 1, \cdots, p + 1$. In this window $\{Z_t\}$ induces a distribution $Z_0, \cdots, Z_{p+1}$. The stationarity property says that we can express the marginal joint distribution of $Z_1, \cdots, Z_p$ in two ways, summing over Z_0 or Z_{p+1}.

$$\text{Prob}(Z_1 = i_1, \cdots, Z_p = i_p)$$

$$= \sum_{i_0=1}^k \text{Prob}(Z_0 = i_0, Z_1 = i, \cdots, Z_p = i_p)$$

$$= \sum_{i_{p+1}=1}^k \text{Prob}(Z_1 = i_1, \cdots, Z_{p+1} = i_{p+1}).$$

The equations are necessary conditions on the quantities $\text{Prob}(Z_0 = i_0, \cdots, Z_p = i_p)$ which characterise the distribution induced on the $p + 1$ window. In fact the conditions are also sufficient. Moreover they lead to a very simple characterisation of the *extreme* distributions, analogous to the characterisation of exchangeable distributions. The results are also found in unpublished notes of the late Walter Weissblum (Bell Telephone Laboratories.)

THEOREM (Hobby-Ylvisaker-Weissblum). *The set of all extreme distributions induced in a $p + 1$ window by a stationary k state discrete process are all those for periodic processes which do not repeat a p-tuple within a cycle (together with a uniform random phase shift).*

To explain the theorem, consider the case $p = 2$, $k = 2$. We may list all the strings which do repeat a neighbouring pair of integers (0, 1). They are (up to

permutation of the labels 0, 1):

sequences	π-vector
$\overline{0}$	(1, 1)
$\overline{01}$	(0, 1)
$\overline{001}$	(⅓, ⅓)
$\overline{0011}$	(½, 0)

Here the "bar" notation means we repeat the sequence infinitely often. The periodic processes obtained by giving each periodic sequence a uniform random phase shift is a stationary process. It is clear that no extreme cycle can have period greater than k^p.

The (aa) π_r values are merely the convex hull of the π-vectors from each of the periodic sequences. Every extreme π-vector derives from an extreme distribution but not vice-versa (see the example above). We summarize the result as

COROLLARY.   *The set* $\Pi_p(k)$ *is the convex hull of* π-vectors generated by a finite number of cyclic designs.

The convexity of $\Pi_p(k)$ can also be established by working directly with infinite design sequences. Let π and π' be two vectors in Π_p. Let $\{d_N, N \in \mathcal{N}\}$ and $\{d'_N, N \in \mathcal{N}'\}$ be their defining sequences. Let $\alpha = r/s$ be rational, $0 \leq \alpha \leq 1$, with r and s integers. Take d_N with $N \in \mathcal{N}$ and d'_N, with $N' \in \mathcal{N}'$. Form a design $d''_{N''}$ with $N'' = Nr + N'(s - r)$ formed by writing $N'(s - r)$ copies of d_N followed by Nr copies of $d'_{N'}$. Let $N, N' \to \infty$, $N \in \mathcal{N}$ and $N' \in \mathcal{N}'$. Then it can be shown with a little analysis that the sequence $\{d''_N, N \in \mathcal{N}''\}$, where $\mathcal{N}'' = \{N''\}$, gives the (aa) value $(1 - \alpha)\pi + \alpha\pi'$. The extension to irrational α is again simple. This is an alternative to the ingenious method of Shepp (1967) for "mixing" together two infinite sequences which preserves the π_r (C_r) values for all $1 \leq r < \infty$.

Given any sequence (design) with a given π_r vector, it is always possible to find a symmetric version for which $\lim \pi_{r,N}^{(i)} = \pi_r/k$ $(i = 1, \cdots, k)$. This is done by mixing, using our method or the method of Shepp, all the sequences obtained by permuting the treatment labels.

We prove one more general result on Π_p before analyzing the structure of Π_1, Π_2 and Π_3. The form of Π_p depends not just on p but also on k, the number of treatments. However, for large k this dependence ceases.

LEMMA 2.   *For all* $k \geq p + 1$ *the* $\Pi_p(k)$ *are identical.*

PROOF.   Let π be a point in $\Pi_p(k)$ for $k > p + 1$. Let it be generated by $\{d_N, N \in \mathcal{N}\}$. Consider a typical $d_N = \{V_1, V_2, \cdots, V_N\}$. We shall convert d_N to d'_N which has all $V_t \leq p + 1$ without changing its $\pi_{r,N}$ values. If no $V_t > p + 1$, we are done. Consider the first unit t for which $V_t > p + 1$. Select a treatment number j with $1 \leq j \leq p + 1$ but $j \neq V_{t-s}$, $s = 1, \cdots, p$. (Ignore any values $t - s < 1$.) This is obviously possible. Now interchange the treatment label j with the

treatment label V_t both at unit t and also wherever they occur after t. This does not effect $\pi_{r,N}$. Performing this operation repeatedly on this and on all other d_N, $N \in \mathcal{N}$, we create a sequence $\{d_N', N \in \mathcal{N}\}$ generating π but which by definition is now a point of $\Pi_p(k)$ with $k = p + 1$. The reverse inclusion is trivial. □

We state Π_p for $p = 1, 2$ and different values of k.

$$\Pi_1 = \{\pi_1 \mid 0 \leq \pi_1 \leq 1\} \qquad \text{for } k \geq 2,$$

$$= \{1\} \qquad \text{for } k = 1.$$

$$\Pi_2 = \{(\pi_1, \pi_2) \mid 0 \leq \pi_2 \leq 1, 0 \leq 2\pi_1 \leq 1 + \pi_2\} \quad \text{for } k \geq 3,$$

$$= \{(\pi_1, \pi_2) \mid \pi_2 \leq 1, 1 - \pi_2 \leq 2\pi_1 \leq 1 + \pi_2\} \quad \text{for } k = 2,$$

$$= \{(1, 1)\} \qquad \text{for } k = 1.$$

Finding Π_1 is trivial. Kiefer (1960) essentially discovered the Π_2 region.

Extensive computations have yielded $\Pi_3(k)$ for $k = 2, 3$ and ≥ 4. The case $k = 2$ (binary) is covered by Martins de Carvalho and Clark (1983) up to $p = 5$. The case $p = 3$, $k = 2$ was known to Shepp (private communication). The essential method is to search among the extreme distributions for the extreme π-vectors by testing "new" vectors to see if they lie in the convex hull of the "old" ones.

Consider now for $p = 3$ the following designs $d^1, \cdots, d^{10}$ written in cyclic notation and using capital letters. The π-vectors follow each design.

$$d^1: \quad \overline{A} \qquad (1, 1, 1)$$

$$d^2: \quad \overline{AB} \qquad (0, 1, 0)$$

$$d^3: \quad \overline{AAB} \qquad (1/3, 1/3, 1)$$

$$d^4: \quad \overline{AABB} \qquad (1/2, 0, 1/2)$$

$$d^5: \quad \overline{AAABBB} \qquad (2/3, 1/3, 0)$$

$$d^6: \quad \overline{ABC} \qquad (0, 0, 1)$$

$$d^7: \quad \overline{AABBCC} \qquad (1/2, 0, 0)$$

$$d^8: \quad \overline{ACBC} \qquad (0, 1/2, 0)$$

$$d^9: \quad \overline{ABBCAABCC} \qquad (1/3, 0, 0)$$

$$d^{10}: \quad \overline{ABCD} \qquad (0, 0, 0).$$

These generate the following ten half-spaces $H_1, \cdots, H_{10}$. The designs whose

π-vectors lie on the planes (sometimes more than 3) are given in brackets.

$$H_1, \quad \pi_1 \geq 0 \qquad\qquad (d^2, d^6, d^8, d^{10})$$

$$H_2, \quad \pi_2 \geq 0 \qquad\qquad (d^4, d^6, d^7, d^9, d^{10})$$

$$H_3, \quad \pi_3 \geq 0 \qquad\qquad (d^2, d^5, d^7, d^8, d^9, d^{10})$$

$$H_4, \quad \pi_1 - \pi_2 - \pi_3 + 1 \geq 0 \qquad (d^1, d^2, d^3, d^6)$$

$$H_5, \quad -\pi_1 - \pi_2 + \pi_3 + 1 \geq 0 \qquad (d^1, d^2, d^5)$$

$$H_6, \quad -2\pi_1 + \pi_2 + 1 \geq 0 \qquad (d^1, d^4, d^5, d^7)$$

$$H_7, \quad -\pi_1 + \pi_2 - \pi_3 + 1 \geq 0 \qquad (d^1, d^3, d^4, d^6)$$

$$H_8, \quad 3\pi_1 + 2\pi_2 + \pi_3 - 1 \geq 0 \qquad (d^6, d^8, d^9)$$

$$H_9, \quad \pi_1 + \pi_2 + \pi_3 - 1 \geq 0 \qquad (d^2, d^4, d^5)$$

$$H_{10}, \quad 2\pi_1 + \pi_2 - 1 \geq 0 \qquad (d^2, d^3, d^4).$$

$k \geq 4$. The extreme points are given by d^1, d^2, d^4, d^5, d^6, d^7 and d^{10}, and the determining half-spaces are H_1, H_2, H_3, H_4, H_5, H_6 and H_7. See Figure 1.

$k = 3$. The point d^{10}, $ABCD$, $(0, 0, 0)$ is no longer available, and the extreme points are those for $k > 3$ with d^{10} deleted and d^8 and d^9 adjoined. The half-spaces are those of the previous case with the addition of H_8. Each of the eight determining planes is seen to be generated by 3 or 4 of the extreme points. See Figure 2.

$k = 2$. The points d^6, d^7, d^8, d^9 of the case $k = 3$ (and, of course, d^{10} of $k \geq 4$) are no longer available. The remaining four points and d^3 are now extreme: d^1, d^2, d^3, d^4, d^5. Since the π-vector for $AAB = d^3$ is on the segment from the π-vectors for $A = d^1$ to $ABC = d^6$, we still obtain the planes determining H_4 and H_7, and similarly H_5 and H_6. But the planes associated with H_1, H_2, H_3 and H_8 are no longer determined by these extreme points and are lost. Two new half-spaces are adjoined and we have Π_3 determined by H_4, H_5, H_6, H_7, H_9 and H_{10}. See Figure 3.

We complete this section with a brief discussion of random allocation. It is obvious that complete random allocation independently at each time point (Bernoulli) gives with probability one $\pi_r = 1/k$ $(r = 1, \cdots, p)$. Alternatively we may use a length of a pseudo random code as explained in Section 5 ignoring edge effects.

Random allocation of treatments to units has been widely advocated and is probably the most frequently used procedure, expecially in clinical trials (see for example Zelen, 1974). Under classical assumptions completely random allocation
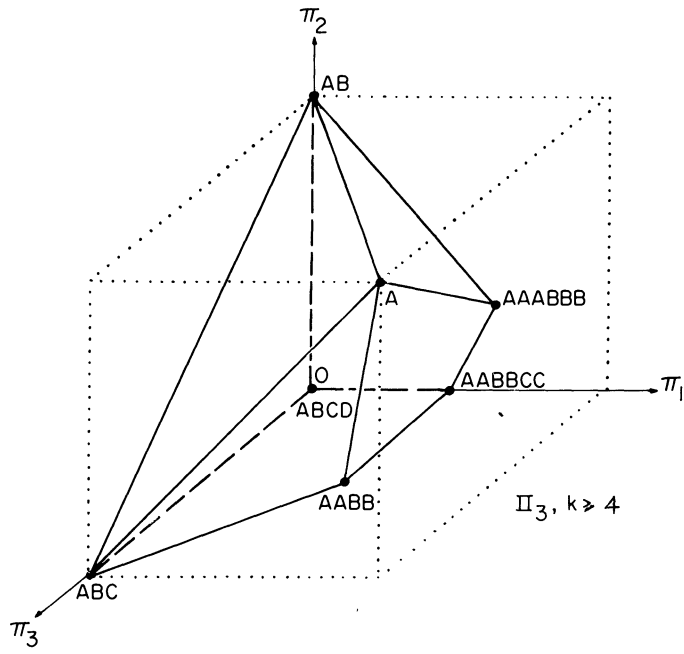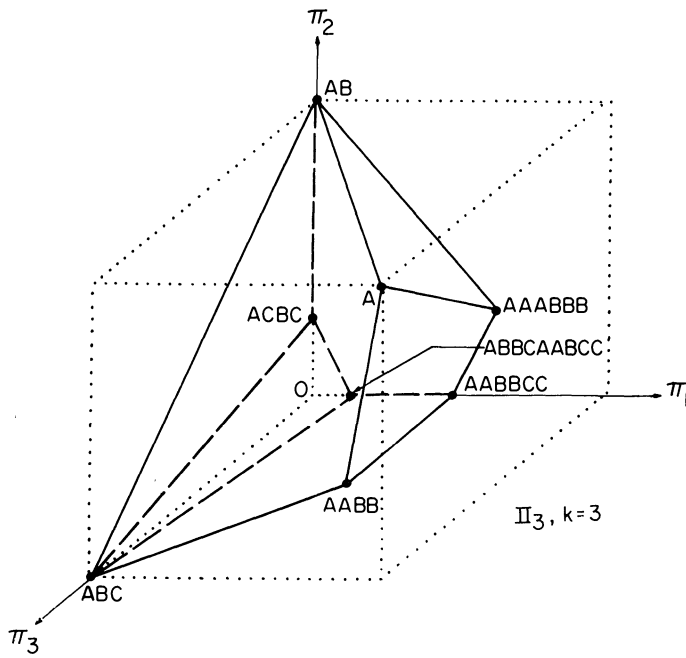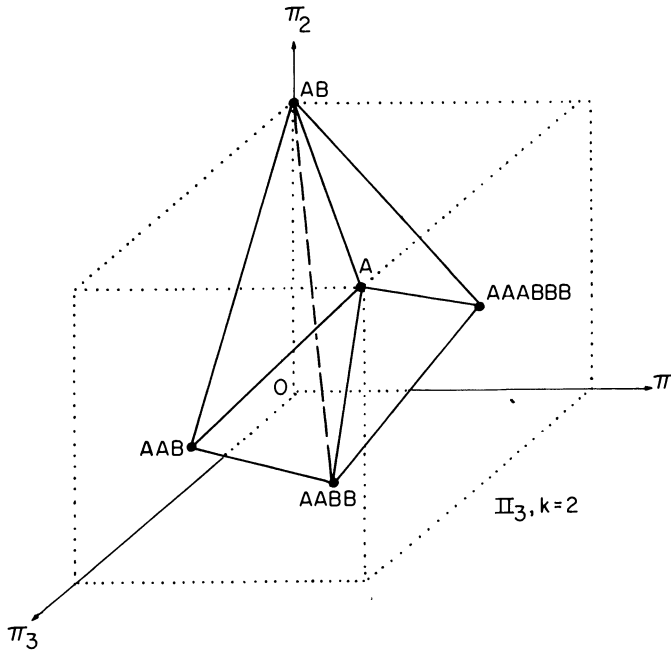
FIG. 1

FIG. 2

FIG. 3.

has been shown to be minimax by Blackwell and Hodges (1957) in the case of two treatments. They refer to the designs as "truncated binomal." Work by Stigler (1969), Efron (1971), and more recently, Wei (1977, 1978) advocates "adaptive" or Markovian methods of allocation. Complete randomization or a more restricted version certainly seems attractive if only to guard against deliberate or unwitting treatment/unit partiality. See also recent work by Atkinson (1982), Smith (1983).

Suppose that, indeed, the treatments are allocated according to a Markov chain of particular order, say L. We may call such a procedure purely random if the chain considered as a first order Markov chain on the states consisting of all L^k of the L-tuples is irreducible and aperiodic. Applying the Ergodic Theorem, we can say that the chain will, with probability one, give a realization which, considered as an infinite design $\{V_1, V_2, \cdots\}$, generates an $(a\ a)$ π-vector. Moreover, for the same chain every such π-vector is the same. The Markov chain thus uniquely determines the π-vector and, of course, the vector is easily obtained from the limiting distribution of the chain, over the L-tuples. We shall be interested in the extent to which the converse is the case: Can we with probability one obtain a *given* π-vector using the sample path of a Markov chain of sufficiently high (but finite) order?

Consider such a sample path. Then with probability zero it will be an infinite cyclic design. However, more strongly, since it gives positive limiting probability to every L-tuple (being nontransient), it must give it likewise to every π-vector

corresponding to the $\pi_{r,L}$ of an L-tuple. This argument shows that *no* extreme point of a Π_p can be obtained in this way. In fact, no purely random Markovian sampling scheme can be optimum in the sense of Section 7. As a simple illustration let $L = 2$, $p = 1$ and suppose the second-order chain attaches limiting probabilities α_1, α_2, α_3 and α_4 to the pairs 11, 12, 21 and 22 respectively. Then with probability one any sample path gives a π-vector (now a single value) $0 < \pi_1 = \alpha_1 + \alpha_4 < 1$, which is never extreme. The details are straightforward in the general case.

A cyclic design is made stationary by giving it a random shift. However, such a random process is not an irreducible Markov aperiodic chain of any order, but a purely deterministic process. (See Doob (1953) for details of these distinctions.) In sampling terminology we have a systematic sample following a random start. We can obtain the same process by a periodic chain with a random initial distribution but this rather destroys our idea of a purely random sampling scheme.

Despite these careful distinctions between deterministic and nondeterministic schemes it is possible to obtain an irreducible aperiodic Markovian scheme whose π-vector is within ε of any given π-vector. We write $\| \ \|$ for the L_∞-norm in what follows. The proof is omitted.

THEOREM 3.   *Let π in Π_p be an asymptotically achievable vector generated by a sequence $\{d_N, N \in \mathcal{N}\}$. Then given $\varepsilon > 0$, there is an irreducible Markov aperiodic chain of finite order such that with probability one its sample path generates a π' in Π_p with $\| \pi - \pi' \| < \varepsilon$.*

7. Optimum designs on the line.   Now that the class of π-vectors has been defined for infinite designs (or technically *sequences* of designs) we are in a better position to carry out the optimization and minimax procedures of Section 2. To get asymptotic results as $N \to \infty$, $N \in \mathcal{N}$ we need to do two things: (1) use the limiting interpretation of the π_r from Section 6; and (2) choose $\rho = (\rho_1, \cdots, \rho_p)$ which makes the process $\{Y_t\}$ satisfying

$$Y_t + \rho_1 Y_{t-1} + \cdots + \rho_p Y_{t-p} = \varepsilon_t \quad (t = 1, 2, \cdots)$$

stationary in the usual wide sense. Here $\{\varepsilon_t\}$ is an uncorrelated sequence with finite variance σ^2. The details of the asymptotic development, which we omit, allow us to write "$+ O(N^{-1})$, $N \in \mathcal{N}$" after the key expression (8) and (9). Identifying d with the design sequence $\{d_N, N \in \mathcal{N}\}$ we define C_d as $\lim(N^{-1}C_{d_N})$ as $N \to \infty$, N in $\mathcal{N}$, when the limit exists. All the results go through and it will be sufficient for C_d to have the form satisfying (a) and (b) of Section 2 for optimality and (1) and (2) of Theorem 1 for minimaxity, the π_r now being interpreted as limiting π_r values. Both cases require complete balance and we turn to the idea of asymptotic symmetrization.

If we have a design sequence $\{d_N, N \in \mathcal{N}\}$ with given π_r values (say from a minimax or optimization argument) it is an easy matter to construct a sequence of $\{\tilde{d}_N, N \in \tilde{\mathcal{N}}\}$ with the same π_r values but which is asymptotically completely symmetric. Let $d_N^s(s = 1, \cdots, k!)$ be all designs of length N constructed from d_N by permutation of treatment labels. Let $\tilde{N} = k!N$ and write all the d_N^s in sequence

so that $\tilde{d}_{\tilde{N}} = d_N^1 d_N^2 \cdots d_N^{k!}$. Then clearly letting $\mathscr{N}$ be $\{\tilde{N}\}$ we obtain a sequence with the required property. Note that it is sometimes possible to symmetrize using less than $k!$ permutations. Indeed the design may already be completely balanced.

The advantage of using codes for $k = q^m$ in Section 5 to construct minimax designs is that N is fairly small. Repetition of the code (with the adjoined dummy "O" treatment) will give an asymptotically minimax design. This seems an excellent procedure where it is possible. However, since the point $(1/k, \cdots, 1/k)$ always lies in Π_p, as can be easily seen, we can *always* construct asymptotically minimax designs by first finding a $\pi = (1/k, \cdots, 1/k)$ by suitable mixing of cyclic extreme point designs and then symmetrizing. Stopping at any $N \in \mathscr{N}$ then gives an almost (to within $O(N^{-1})$) minimax design. The details are left to the reader.

We want to concentrate now on optimization. Let $\rho = (\rho_1, \cdots \rho_p)$ be fixed. We have to maximize trace(C_d) (in the limit). But this maximum is given by a design with (1) $m_1 = \cdots = m_k$ and (2) maximum value of

(18) $$\sum_{r=1}^p \pi_r \sum_{s=0}^{p-r} \rho_s \rho_{s+r} \quad (\rho_0 = 1).$$

Point (1) we can ignore because symmetrization will yield it automatically. Thus we are left with a *linear programming* problem to maximize (18) over Π_p. This can be carried out either by using a standard algorithm on the relevant half-space constraints, or, as the extreme points (vertices) are listed (not usually the case in LP), by just running through them.

For designs on the circle, we may select any value for the vector ρ and if we are lucky enough to have N just a multiple of the length of the correct extreme symmetric cyclic design, then that design or a repetition of it will be optimum. For designs on the line we need ρ to satisfy the stationarity condition that the roots of

$$x^p + \rho_1 x^{p-1} + \cdots + \rho_p = 0$$

lie inside the unit circle. Then with that value of ρ we (1) find the optimum π-vector and the corresponding design, (2) repeat it a reasonable number of times (to make the edge effect small), (3) symmetrize it if it is not already symmetric.

As a simple example take $p = 3$ and $\rho = (-0.4, -0.3, -0.1)$. Negative ρ-values correspond to positive autocorrelation. We suspect that a reasonable amount of separation between occurences of the same treatment will be optimum. Expression (18) becomes

$$-0.25\pi_1 - 0.26\pi_2 - 0.1\pi_3.$$

When $k \geq 4$ this is maximized uniquely in Π_3 at d_{10}, $\overline{ABCD}$ when $\pi = (0, 0, 0)$. Thus with symmetrization taking all permutations of 1234, we obtain for large N an approximately optimum design 12341234 $\cdots$ 21342134 $\cdots$, etc. When $k = 3$ this design is not available. The unique optimum design is d^9, $\overline{ABBCAABCC}$. Symmetrization yields 122311233 $\cdots$ 211322133 $\cdots$ etc. When $k = 2$ the unique optimum design is d_4, $AABB$, given by 11221122 $\cdots$ without further symmetrization. The full computer implementation of the ideas of this paper is still being

carried out. The steps are (1) choice of autocorrelation, possibly after preliminary estimation, (2) choice of the number of treatments and sample size N, (3) elucidation of the extreme points and associated region for given p and k, (4) choice of optimum (cyclic) extreme design, and (5) packing of the sample length N with a suitably symmetrized version of the extreme designs.

As a final exercise we show that sometimes an optimum design can be found for large values of p, despite the fact that Π_p has not yet been fully worked out for large p. Consider, then, the first order moving average model on the circle given by

$$Y_t = \varepsilon_t - \lambda \varepsilon_{t-1} \quad (t = 1, \cdots, N)$$

where $|\lambda| < 1$ and the ε_t are uncorrelated with variance σ^2, as before. Rewriting this in an auto-regressive formulation,

$$\rho = \frac{1}{1 - (-\lambda)^N} (1, -\lambda, \lambda^2, \cdots, (-\lambda)^{N-1}).$$

Then (18) becomes

(19) $$[1 - (-\lambda)^N]^{-1} \sum_{r=1}^{N-1} \pi_r \sum_{s=0}^{N-r} (-\lambda)^{2s+r}.$$

An informal analysis runs as follows. Letting $N \to \infty$ expression (19) becomes

$$-\alpha_1 \pi_1 + \alpha_2 \pi_2 - \alpha_3 \pi_3 + \cdots, \quad \lambda > 0,$$

$$\alpha_1 \pi_1 + \alpha_2 \pi_2 + \alpha_3 \pi_3 + \cdots, \quad \lambda < 0,$$

where $\alpha_i \geq 0$, $i = 1, 2, \cdots$. Thus since this must be maximized and $0 \leq \pi_r \leq 1$ for all r, when $\lambda > 0$, a π-vector $(0, 1, 0, 1 \cdots)$ would be (asymptotically) optimum. Fortunately such a π-vector always lies in Π_p, namely, that arising from AB (repeated). Thus an approximately optimum design for the first order MA model is to take $\overline{AB}$ symmetrized, that is, $121212 \cdots 131313 \cdots 232323 \cdots$, etc. This design is intuitively appealing for a model only having a 1-step local autocorrelation.

Similarly, when $\lambda < 0$, the point A, with π-vector $(1, 1, 1, \cdots)$ yields an approximately optimum design $111 \cdots 11222 \cdots 22 \cdots kkk \cdots kk$.

We have discussed thus far in this section maximization of (18) for a particular vector ρ. One can also compare approximately balanced designs in terms of the integral of (18) wrt a positive measure on ρ's (approximately Bayes designs). Related is the notion of asymptotic admissibility of design sequences in terms of the behavior of (18) on the space of ρ's, of concern when ρ is unknown. Since (18) is linear in π, every (approximately balanced) design is seen to be approximately admissible, for large N. In practice when ρ is unknown and N is large, one would presumably replace the BLUE (MVUE) by an adaptive estimator that estimates ρ.

REFERENCES

ATKINSON, A. C. (1982). Optimum biased coin designs for sequential clinical trials with prognostic factors. *Biometrika* **69** 61–67.

BERENBLUTT, I. I. and WEBB, G. I. (1974). Experimental designs in the presence of autocorrelated errors. *Biometrika* **61** 427–437.

BERLEKAMP, E. R. (1968). *Algebraic Coding Theory.* McGraw-Hill, New York.

BICKEL, P. J. and HERZBERG, A. M. (1979). Robustness of designs against autocorrelation in time. I. Asymptotic theory for location and linear regression. *Ann. Statist.* **7** 77–95.

BLACKWELL, D. H. and HODGES, J. L. (1957). Design for the control of selection bias. *Ann. Math. Statist.* **28** 449–460.

COX, D. R. (1951). Some systematic experimental designs. *Biometrika* **38** 312–323.

DOOB, J. L. (1953). *Stochastic Processes.* Wiley, New York.

DUBY, C., GUYON, X. and PRUM, B. (1977). The precision of different experimental designs for a random field. *Biometrika* **64** 59–66.

EFRON, B. (1971). Forcing a sequential experiment to be balanced. *Biometrika* **58** 403–417.

HOBBY, C. and YLVISAKER, N. D. (1964). Some structure theorems for stationary probability measures on finite state spaces. *Ann. Math. Statist.* **35** 550–556.

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

KIEFER, J. (1960). Optimum experimental designs. *V*, with applications to systematic and rotatable designs. *Proc. Fourth Berk. Symp. Math. Statist. Probab.* **2** 381–405.

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

KIEFER, J. and WYNN H. P. (1981). Optimum balanced block and latin square designs for correlated observations. *Ann. Statist.* **9** 737–757.

KIEFER, J. and WYNN, H. P. (1983). Autocorrelation-robust design of experiments. *Scientific Inference and Data Analysis and Robustness.* Academic, New York.

MACWILLIAMS, F. J. and SLOANE, N. J. A. (1977). *The Theory of Error-Correcting Codes.* North-Holland, Amsterdam.

MARTIN, R. J. (1977). Spatial models with applications in sampling and experimental design. Ph.D. thesis, London School of Economics, University of London.

MARTIN, R. J. (1979). A subclass of lattice processes applied to a problem in planar sampling. *Biometrika* **66** 209–217.

MARTINS DE CARVALHO, J. L. and CLARK, J. M. C. (1983). Characterizing the autocorrelations of binary sequences. *IEEE Trans. Inform. Theory* **IT-29** 502–508.

MASRY, E. (1972). On covariance functions of unit processes. *Siam J. Appl. Math.* **23** 28–33.

SHEDD, D. A. and SARWATE, D. V. (1979). Construction of sequences with good correlation properties. *IEEE Trans. Inform. Theory.* **IT-25** 94–97.

SHEPP, L. A. (1967). Covariances of unit processes. *Proc. Working Conf. on Stochastic Process.* 205–218. Santa Barbara, Calif.

SMITH, R. L. (1983). Sequential treatment allocations using biased coin designs. *J. Roy. Statist. Soc. B.*, to appear.

STIGLER, S. M. (1969). The use of random allocation for control of selection bias. *Biometrika* **56** 553–560.

WEI, L. J. (1977). A class of designs for sequential clinical trials. *J. Amer. Statist. Assoc.* **72** 382–386.

WEI, L. J. (1978). The adaptive biased coin design for sequential experiments. *Ann. Statist.* **6** 92–100.

WILLIAMS, R. M. (1952). Experimental designs for serially correlated observations. *Biometrika* **39** 151–167.

ZELEN, M. (1974). The randomization and stratification of patients to clinical trials. *J. Chronic Diseases* **27** 365–375.

ZYSKIND, G. (1967). On canonical forms, non-negative covariance matrices and best and simple least squares linear estimators in linear models. *Ann. Math. Statist.* **38** 1092–1109.

DEPARTMENT OF STATISTICS   DEPARTMENT OF MATHEMATICS
EVANS HALL   IMPERIAL COLLEGE
UNIVERSITY OF CALIFORNIA   180 QUEENS GATE
BERKELEY, CALIFORNIA 94720   LONDON SW7 2BZ
                                                ENGLAND