The Annals of Mathematical Statistics
1968, Vol. 39, No. 3, 973–982

BAYES SEQUENTIAL DESIGN OF FRACTIONAL FACTORIAL EXPERIMENTS FOR THE ESTIMATION OF A SUBGROUP OF PRE-ASSIGNED PARAMETERS[1]

By S. Zacks

Kansas State University

0. Introduction. The objective of the present study is to investigate and find optimal sequential designs of fractional factorial experiments, for the purpose of estimating a pre-assigned subgroup of parameters. We are given a factorial system of $N = 2^m$ treatment combinations. Without loss of generality, assume that the pre-assigned subgroup of parameters consists of the first $S = 2^s (s < m)$ main effects and interactions. (The parameters are arranged according to the standard order, see Ehrenfeld and Zacks [2].) All the other $K = N - S$ parameters are considered as nuisance parameters. It is desired to estimate the subvector of S pre-assigned parameters. Randomized unbiased procedures were studies by Ehrenfeld and Zacks [5], [6], [13] in which $n (1 \leqq n \leqq M = 2^{m-s})$ blocks consisting of S treatment combinations are chosen. The design of these n fractional replicates is carried according to those procedures by a non-sequential manner. The question is whether one can improve and reduce total estimation risk by designing the fractional replicates sequentially, and after each stage of experimentation adjusting the appropriate estimator according to the information obtained concerning the parameters (pre-assigned as well as nuisance). It is well known that the best truncated sequential procedure for estimating the mean of a normal distribution with a known variance, $\mathfrak{N}(\alpha, 1)$ say, when the loss function is quadratic is a fixed sample procedure (see Hodges and Lehmann [8], and J. Wolfowitz [12]). In the present model we have observations following an S-variate normal distribution with a known covariance matrix $\sigma^2 I_S$. The mean vector is $A(\alpha, \beta)'$, where A is an $S \times N$ matrix; α is an S-dimensional vector to be estimated, and β is a K-dimensional vector of nuisance parameters. The question is whether under this model there exists a sequential procedure which is better than any fixed sample procedure. The present study is devoted to the solution of this problem. We search for a Bayes sequential design procedure, when the a-priori distribution of the parameters is assumed to be normal and the loss function is quadratic. As expected, the result is that the Bayes procedure for a prior normal distribution is of a fixed sample size. The optimal number of fractional replicates to perform is a function of the cost of experimentation and the prior dispersion matrix of the parameters. The main part of the investigation is then devoted to the problem of the best choice of n fractional replicates out of the $M = 2^{m-s}$ possible ones. It is proven that the Bayes procedure is to choose n different blocks

Received April 26, 1967; revised October 4, 1967.

[1] The research was partially supported by a National Science Foundation Grant GP-6129 to the Department of Statistics, Kansas State University.

(fractional replicates) of treatment combinations. The choice of the blocks of treatment combinations can be done in a non-randomized fashion, and is immaterial of which blocks are chosen. It is also proven that the common least squares estimators, with a non-randomized choice of fractional replicates are minimax.

1. Preliminaries. Let $\mathcal{E}_0, \mathcal{E}_1, \cdots, \mathcal{E}_{M-1}$ be $M = 2^{m-s}$ available experiments (fractional replicates). If experiment $\mathcal{E}_v$ is performed ($v = 0, 1, \cdots, M-1$), an $S = 2^s$ dimensional random vector Y_v is observed. The statistical model for Y_v is:

$$Y_v = C\alpha + H_v\beta + \epsilon_v \qquad (v = 0, \cdots, M-1) \tag{1.1}$$

where $\epsilon_v \sim N(0, \sigma^2 I_S)$ independently of v, α is an S-dimensional vector of the pre-assigned parameter; β is a $K = N - S (N = 2^m)$ dimensional vector of the nusiance parameters. $C \equiv C^{(S)}$ is an $S \times S$ Hadamard matrix. $H_v = (c_{v1}^{(M)}, \cdots, c_{v,M-1}^{(M)}) \otimes (C^{(S)})$, $(v = 0, \cdots, M-1)$ where $(1, c_{v1}^{(M)}, \cdots, c_{v,M-1}^{(M)})$ is the vth row vector of $C^{(M)}$, and the matrices $C^{(2k)} (k = 0, 1, \cdots)$ are generated by a Kronecker direct multiplication of $C^{(2)} = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$ with itself k times. $\otimes$ designates the Kronecker direct multiplication operator (see details in [13]). We are concerned with the problem of estimating α. The following lemma from normal regression theory will play a role in the sequel:

LEMMA 1. *If* $\mathcal{L}(X_1 \mid X_2) = \mathfrak{N}(AX_2, \Sigma_{11})$ and $\mathcal{L}(X_2) = \mathfrak{N}(\eta, \Sigma_{22})$, *where* $\mathcal{L}(Y)$ *designates the distribution law of* Y, *and* $\mathcal{L}(Y \mid Z)$ *designates the conditional distribution law of* Y, *given* Z, *then* $\mathcal{L}(X_2 \mid X_1) = \mathfrak{N}(E(X_2 \mid X_1), \Sigma_{22.1})$ *where:*

$$E(X_2 \mid X_1) = \eta + \Sigma_{22}A'(\Sigma_{11} + A\Sigma_{22}A')^{-1}(X_1 - A\eta), \tag{1.2}$$

and

$$\Sigma_{22.1} = \Sigma_{22} - \Sigma_{22}A'(\Sigma_{11} + A\Sigma_{22}A')^{-1}A\Sigma_{22}. \tag{1.3}$$

PROOF. Let $\mathcal{L}(X_1, X_2)$ denote the joint distribution law of X_1 and X_2. According to the hypotheses,

$$\mathcal{L}(X_1, X_2) = \mathfrak{N}\left((A\eta, \eta), \left[\begin{array}{c|c} \Sigma_{11}^* & \Sigma_{12} \\ \hline \Sigma_{21} & \Sigma_{22} \end{array}\right]\right),$$

where $\Sigma_{11}^* = \Sigma_{11} + A\Sigma_{22}A'$. By normal regression theory (see Anderson [1]),

$$\mathcal{L}(X_2 \mid X_1) = \mathfrak{N}(E(X_2 \mid X_1), \Sigma_{22.1}),$$

with:

$$E(X_2 \mid X_1) = \eta + \Sigma_{21}\Sigma_{11}^{*-1}(X_1 - A\eta), \tag{1.4}$$

and

$$\Sigma_{22.1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{*-1}\Sigma_{12}. \tag{1.5}$$

The Lemma is proven by substituting

$$\Sigma_{21} = E(X_2 - \eta)(X_1 - A\eta)' = \Sigma_{22}A', \tag{1.6}$$

and $\Sigma_{12} = \Sigma_{21}'$ in (1.4) and (1.5).

2. The sequential Bayes procedure for prior normal distributions.

2.1. *The decision framework.* In Section 2.2 we shall derive the Bayes sequential procedure for estimating α, assuming that the prior distribution law of the parameters $(\alpha, \beta)'$ is

$$\mathfrak{N}\left((\mathbf{0}, \mathbf{0})', \left[\begin{array}{c|c} \theta^2 I_S & 0 \\ \hline 0 & \tau^2 I_K \end{array}\right]\right), \qquad 0 < \theta^2, \tau^2 < \infty.$$

We shall consider a quadratic loss function of the form

$$L(\hat{\alpha}, \alpha) = (\hat{\alpha} - \alpha)'(\hat{\alpha} - \alpha), \tag{2.1}$$

where $\hat{\alpha}$ is an estimator of α. According to the above assumption concerning the prior distribution of $(\alpha, \beta)'$, the Bayes risk at the nth stage of experimentation is the minimal trace (over all choices of experiments $v_1, \cdots, v_n$) of the posterior dispersion matrix of α. In order to describe the Bayes procedure let $\mathfrak{Z}_j (j = 0, 1, \cdots, n)$ denote a set of j pairs $\mathfrak{Z}_j = \{(v_1, Y_{v_1}), \cdots, (v_j, Y_{v_j})\}$, which contains the whole relevant data of the first j experiments. $\mathfrak{Z}_0 = \phi$ designates the null set, and corresponds to the case of no experimentation. We denote by $E\{(\alpha, \beta)' \mid \mathfrak{Z}_j\}$ the posterior expectation of $(\alpha, \beta)'$, given $\mathfrak{Z}_j (j = 0, 1, \cdots, n)$. Given $\mathfrak{Z}_j$, the best estimator of α, after the jth experiment has been performed, is

$$\hat{\alpha}_j(\mathfrak{Z}_j) = E\{\alpha \mid \mathfrak{Z}_j\}, \qquad (j = 0, 1, \cdots, n). \tag{2.2}$$

There are two decisions to be made after the jth experiment:

(i) Whether to terminate experimentation or perform the $(j + 1)$st experiment $(j = 0, 1, \cdots, n - 1)$;

(ii) If the decision is to perform the $(j + 1)$st experiment, which experiment from $\mathcal{E}_0, \cdots, \mathcal{E}_{M-1}$ to choose?

Experimentation always terminates after the nth experiment, if it has not been terminated before.

The Bayes sequential truncated sampling rule is determined by the n functions $\rho_{n-j}(\mathfrak{Z}_j), j = 1, \cdots, n$; which are given by the following recursive definition:

$$\rho_0(\mathfrak{Z}_n) = E\{(\alpha - \hat{\alpha}_n(\mathfrak{Z}_n))'(\alpha - \hat{\alpha}_n(\mathfrak{Z}_n)) \mid \mathfrak{Z}_n\} \tag{2.3}$$

$$\rho_1(\mathfrak{Z}_{n-1}) = \min_{v=0,\cdots,M-1}\{\min[\rho_0(\mathfrak{Z}_{n-1}), 1 + E^{\mathfrak{Z}_{n-1}}\{\rho_0(\mathfrak{Z}_{n-1}, (v, Y_v))\}]\}$$

and, for every $k = 2, \cdots, n - 1$,

$$\rho_k(\mathfrak{Z}_{n-k}) = \min_{v=0,\cdots,M-1}\{\min[\rho_0(\mathfrak{Z}_{n-k}), 1 + E^{\mathfrak{Z}_{n-k}}\{\rho_{k-1}(\mathfrak{Z}_{n-k}, (v, Y_v))\}]\}. \tag{2.4}$$

In these equations we designate

$$E^{\mathfrak{Z}_{n-k}}\{\rho_{k-1}(\mathfrak{Z}_{n-k}, (v, Y_v))\} = \int \rho_{k-1}(\mathfrak{Z}_{n-k}, (v, y))\, dF(y \mid \mathfrak{Z}_{n-k}, v), \tag{2.5}$$

where $F(y \mid \mathfrak{Z}_j, v)$ is the expected distribution of Y_v, under the posterior distribution of (α, β) given $\mathfrak{Z}_j (j = 1, \cdots, n)$. The optimal (Bayes) stopping rule is to terminate at the least integer j, $1 \leqq j \leqq n$, such that

$$\rho_0(\mathfrak{Z}_j) \leqq \rho_{n-j}(\mathfrak{Z}_j). \tag{2.6}$$

In this formulation we assumed that the cost of an experiment is 1 risk unit (in \$).

The decision theoretic framework formulated here has been obtained by an appropriate adjustment of the well established truncated sequential Bayes procedure for the particular design and estimation problem under consideration. Proofs of the optimality of this sequential Bayes procedure can be found in various references. Among the well known are: Blackwell and Girshick [2]; Haggstrom [7], Paulson [9], Whittle [11], Wald [10].

2.2. *The Bayes sequential procedure in the normal case.* The normal posterior distribution of (α, β) given $\mathfrak{I}_j (j = 0, 1, \cdots, n)$ can be found according to Lemma 1 in the following manner. Consider the vector $(Y'_{v_1}, \cdots, Y'_{v_j})'$. Its distribution, given (α, β) and $(v_1, \cdots, v_j)$ is normal, with expectation

$$E\{(Y'_{v_1}, \cdots, Y'_{v_j})' \mid (\alpha, \beta), (v_1, \cdots, v_j)\} = \left[\begin{array}{c|c} C & H_{v_1} \\ \hline \vdots & \vdots \\ \hline C & H_{v_j} \end{array}\right] (\alpha, \beta)'. \tag{2.7}$$

The dispersion matrix of $(Y'_{v_1}, \cdots, Y'_{v_j})'$, given $(\alpha, \beta)'$ and $(v_1, \cdots, v_j)$, is

$$\mathfrak{D}\{(Y'_{v_1}, \cdots, Y'_{v_j})' \mid (\alpha, \beta)', (v_1, \cdots, v_j)\} = \sigma^2 I_j \otimes I_S. \tag{2.8}$$

Thus, the posterior distribution of $(\alpha, \beta)'$ given $\mathfrak{I}_j (j = 0, \cdots, n)$ is normal with mean:

$$\begin{aligned} E\{(\alpha, \beta)' \mid \mathfrak{I}_j\} = & \left[\begin{array}{c|c} \theta^2 I_S & 0 \\ \hline 0 & \tau^2 I_k \end{array}\right] \left(\begin{array}{c|c|c} C' & \cdots & C' \\ \hline H'_{v_1} & \cdots & H'_{v_j} \end{array}\right) \\ & \cdot \left\{\sigma^2 I_j \otimes I_S + \left[\begin{array}{c|c} \theta^2 C & \tau^2 H_{v_1} \\ \hline \vdots & \vdots \\ \hline \theta^2 C & \tau^2 H_{v_j} \end{array}\right] \left(\begin{array}{c|c|c} C' & \cdots & C' \\ \hline H'_{v_1} & \cdots & H'_{v_j} \end{array}\right)\right\}^{-1} \left[\begin{array}{c} Y_{v_1} \\ \vdots \\ Y_{v_j} \end{array}\right]. \end{aligned} \tag{2.9}$$

We notice that,

$$\begin{aligned} H_v H_{v'} &= K I_S, && \text{if } v = v' \\ &= -S I_S, && \text{if } v \neq v'. \end{aligned} \tag{2.10}$$

Thus,

$$\left[\begin{array}{c|c} \theta^2 C & \tau^2 H_{v_1} \\ \hline \vdots & \\ \hline \theta^2 C & \tau^2 H_{v_j} \end{array}\right] \left(\begin{array}{c|c|c} C' & \cdots & C' \\ \hline H'_{v_1} & \cdots & H'_{v_j} \end{array}\right) = B \otimes I_S, \tag{2.11}$$

where B is a $j \times j$ symmetric matrix whose diagonal elements are all equal to $(\theta^2 S + \tau^2 K)$, and its off-diagonal elements are

$$\begin{aligned} b(i, i') &= \theta^2 S + \tau^2 K, && \text{if } v_i = v_{i'} \\ &= \theta^2 S(1 - \gamma), && \text{if } v_i \neq v_{i'}, \quad i \neq i' = 1, \cdots, j, \end{aligned} \tag{2.12}$$

and $\gamma = \tau^2/\theta^2$.

The posterior dispersion matrix of $(\alpha, \beta)'$, given $\mathfrak{Z}_j$, is according to Lemma 1,

$$\mathfrak{D}(\mathfrak{Z}_j) = \begin{bmatrix} \theta^2 I_S & 0 \\ 0 & \tau^2 I_K \end{bmatrix} \begin{pmatrix} C' & \cdots & C' \\ H'_{v_1} & \cdots & H'_{v_j} \end{pmatrix}$$

$$\cdot \{[\sigma^2 I_j + B]^{-1} \otimes I_S\} \begin{bmatrix} C & H_{v_1} \\ \vdots & \vdots \\ C & H_{v_j} \end{bmatrix} \begin{bmatrix} \theta^2 I_S & 0 \\ 0 & \tau^2 I_K \end{bmatrix}. \tag{2.13}$$

Therefore, the risk function $\rho_0(\mathfrak{Z}_j)$ of the Bayes estimator of α is, when $1_j' = (1, 1, \cdots, 1)$:

$$\rho_0(\mathfrak{Z}_j) = \theta^2 S\{1 - \theta^2 S(1_j'(\sigma^2 I_j + B)^{-1} 1_j)\}. \tag{2.14}$$

We have shown the rather expected result that the risk function associated with the Bayes estimator of α does not depend on the observations. It depends only on the number of experiments performed, j, and the experiments chosen $(\mathcal{E}_{v_1}, \cdots, \mathcal{E}_{v_j})$. The conclusion is that the Bayes sequential procedure is a fixed sample procedure. (See Blackwell and Girshick [2], Theorem 9.3.3.)

2.3. *The optimal allocation of experiments.* Before we proceed with the derivation of a more explicit formula of the Bayes estimator we investigate the question of the optimal choice of experiments.

Suppose that the Bayes sequential procedure prescribes to perform n experiments $\mathcal{E}_{v_1}, \cdots, \mathcal{E}_{v_n}$. Following Degroot [4], we shall call the choice of a specific set $(v_1, \cdots, v_n)$ of n experiments an *allocation*. We denote by $A(n; k; n_1, \cdots, n_k)$ an allocation of n experiments according to which k different experiments are chosen. $\mathcal{E}_{v_1}$ is chosen n_1 times; $\mathcal{E}_{v_2}$, n_2 times, and $\mathcal{E}_{v_k}$, n_k times; $v_1 \neq \cdots \neq v_k$. An allocation A_1 is called better than A_2 if the posterior Bayes risk associated with A_1, say $\rho_0(\mathfrak{Z}_n^{(1)})$, is smaller (or equal) to that associated with A_2, $\rho_0(\mathfrak{Z}_n^{(2)})$. In case $\rho_0(\mathfrak{Z}_n^{(1)}) < \rho_0(\mathfrak{Z}_n^{(2)})$ A_1 is strictly better than A_2.

THEOREM 2.1. *For every $j = 1, 2, \cdots, n$, $A(j; j; 1, \cdots, 1)$ is better than $A(j; 1; j)$, and strictly better if $j \geqq 2$.*

PROOF. Consider the allocation $A(j; j; 1, 1, \cdots, 1)$. According to Lemma 1, (2.7) and (2.8), the associated dispersion matrix of the posterior distribution of $(\alpha, \beta)'$ is:

$$\mathfrak{D}(\mathfrak{Z}_j) = \begin{bmatrix} \theta^2 I_S & 0 \\ 0 & \tau^2 I_K \end{bmatrix} - \begin{pmatrix} \theta^2 C' & \cdots & \theta^2 C' \\ \tau^2 H'_{v_1} & \cdots & \tau^2 H'_{v_j} \end{pmatrix}$$

$$\cdot \{(\sigma^2 I_j + \tau^2 N I_j + \theta^2 S(1 - \gamma) J_j) \otimes I_S\}^{-1} \begin{pmatrix} \theta^2 C & \tau^2 H_{v_1} \\ \vdots & \vdots \\ \theta^2 C & \tau^2 H_{v_j} \end{pmatrix}, \tag{2.15}$$

where J_j is a $j \times j$ matrix all of whose elements are 1. Without loss of generality,

assume $\sigma^2 = 1$. It is easy to prove that,

$$
\begin{aligned}
&\{[(1 + \tau^2 N)I_j + \theta^2 S(1 - \gamma)J_j] \otimes I_s\}^{-1} \\
&= (1 + \tau^2 N)^{-1}[I_j + \theta^2 S(1 - \gamma)(1 + \tau^2 N)^{-1} J_j]^{-1} \otimes I_s \qquad (2.16) \\
&= (1 + \tau^2 N)^{-1}[I_j - \theta^2 S(1 - \gamma)(1 + \tau^2 N + j\theta^2 S(1 - \gamma))^{-1} J_j] \otimes I_s .
\end{aligned}
$$

Let $\Sigma_{11}(\mathfrak{z}_j)$ denote the dispersion matrix of α, given $\mathfrak{z}_j$. From (2.15) and (2.16) we find,

$$
\begin{aligned}
&\text{tr. } \Sigma_{11}(\mathfrak{z}_j) \\
&= \theta^2 S\{1 - \theta^2 S(1 + \tau^2 N)^{-1} 1_j' \\
&\quad \cdot [I_j - \theta^2 S(1 - \gamma)(1 + \tau^2 N + j\theta^2 S(1 - \gamma))^{-1} J_j] 1_j\} \qquad (2.17) \\
&= \theta^2 S \cdot (1 + \tau^2 K - (j - 1)S\tau^2)(1 + \tau^2 K - (j - 1)S\tau^2 + j\theta^2 S)^{-1}, \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad j = 1, 2, \cdots, n.
\end{aligned}
$$

Similarly, the dispersion matrix of the posterior distribution of $(\alpha, \beta)'$, associated with the allocation $A(j; 1; j), j = 1, \cdots, n$, is:

$$
\begin{aligned}
\Sigma(\mathfrak{z}_j) &= \begin{bmatrix} \theta^2 I_s & 0 \\ \hline 0 & \tau^2 I_K \end{bmatrix} - \left(\begin{bmatrix} \theta^2 C' \\ \hline \tau^2 H_v' \end{bmatrix} \otimes 1_j' \right) \qquad (2.18) \\
&\cdot \{(I_j - (\theta^2 S + \tau^2 K)(1 + j(\theta^2 S + \tau^2 K))^{-1} J_j) \otimes I_s\} 1_j \otimes (\theta^2 C \mid \tau^2 H_v).
\end{aligned}
$$

From this we obtain that the Bayes risk associated with $A(j; 1; j)$ is:

$$
\text{tr. } \Sigma_{11}(\mathfrak{z}_j) = \theta^2 S(1 + j\tau^2 K)(1 + j(\theta^2 S + \tau^2 K))^{-1}, \quad j = 1, \cdots, n. \qquad (2.19)
$$

Finally,

$$
\begin{aligned}
&(1 + \tau^2 K - (j - 1)S\tau^2)(1 + \tau^2 K - (j - 1)S\tau^2 + j\theta^2 S)^{-1} \qquad (2.20) \\
&\qquad \leqq (1 + j\tau^2 K)(1 + j(\theta^2 S + \tau^2 K))^{-1}, \qquad j = 1, \cdots, n,
\end{aligned}
$$

for all θ^2 and τ^2, with strict inequality for all $j \geqq 2$. This proves the theorem. We shall prove now that the allocation $A(n; n; 1, 1, \cdots, 1)$ is optimal. Observe that the combination of two allocations $A(j_1; 1; j_1)$ and $A(j_2; 1; j_2)$, where the experiments chosen for the two allocations are different, is the allocation $A(j_1 + j_2; 2; j_1, j_2)$.

THEOREM 2.2. *For every $j = 2, \cdots, n$, if $j_1 + j_2 = j$ the allocation $A(j_1 + j_2; j_1 + j_2; 1, \cdots, 1)$ is better than $A(j_1 + j_2; 2; j_1, j_2)$ and strictly better for every $j_1 \geqq 2$ and $j_2 \geqq 2$.*

PROOF. Consider the two allocations $A(j_1; 1, j_1)$ and $A(j_2; 1, j_2)$. Let $\{Y_{v_1}^{(1)}, \cdots, Y_{v_1}^{(j_1)}\}$ and $\{Y_{v_2}^{(1)}, \cdots, Y_{v_2}^{(j_2)}\}$, with $v_1 \neq v_2$, be the random vector associated with these allocations. Define

$$
Z_i = j_i^{-\frac{1}{2}} \sum_{\nu=1}^{j_i} Y_{v_i}^{(\nu)}, \qquad (i = 1, 2). \qquad (2.21)
$$

We derive now the form of the Bayes estimator of α for the allocations $A(j_i; 1; j_i)$,

$i = 1, 2$, and for the combined allocation $A(j_1 + j_2 ; 2; j_1 , j_2)$. Assuming $\sigma^2 = 1$ we have,

$$\mathfrak{L}(Z_i \mid (\alpha, \beta), v_i) = \mathfrak{N}(j_i^{\frac{1}{2}}(C\alpha + H_v\beta), I_S), \qquad i = 1, 2. \tag{2.22}$$

Let $\hat{\alpha}(Z_i ; (j_i ; 1; j_i)), i = 1, 2$, denote the Bayes estimator of α, under the allocation $A(j_i ; 1; j_i)$. Since

$$(j_i^{\frac{1}{2}}C \mid j_i^{\frac{1}{2}}H_{v_i}) \left[\begin{array}{c|c} \theta^2 I_S & 0 \\ \hline 0 & \tau^2 I_K \end{array}\right] \begin{pmatrix} j_i^{\frac{1}{2}}C' \\ \hline j_i^{\frac{1}{2}}H'_{v_i} \end{pmatrix} = j_i\,(\theta^2 S + \tau^2 K) I_S, \qquad i = 1, 2, \tag{2.23}$$

it is simple to verify that the Bayes estimator of α, from each of the single allocation, is:

$$\hat{\alpha}(Z_i ; (j_i ; 1; j_i)) = \theta^2 Z(1 + j_i(\theta^2 S + \tau^2 K))^{-1} C^{-1} \textstyle\sum_{v=1}^{j_i} Y_{v_i}^{(v)}, \; i = 1, 2. \tag{2.24}$$

We derive now the Bayes estimator of the combined allocation. Let

$$\zeta = \theta^2 S + \tau^2 K, \qquad \omega = \theta^2 S(1 - \gamma). \tag{2.25}$$

It is a straightforward matter to prove that

$$\begin{aligned} E\{(\alpha, \beta)' \mid Z_1, Z_2\} &= ((1 + j_1\zeta)(1 + j_2\zeta) - j_1 j_2 \omega^2)^{-1} \\ &\cdot \left[\begin{array}{c|c} \theta^2 SC^{-1} & \theta^2 SC^{-1} \\ \hline \tau^2 H_{v_1} & \tau^2 H'_{v_2} \end{array}\right] \left\{ \left[\begin{array}{c|c} 1 + j_2\zeta & -(j_1 j_2)^{\frac{1}{2}}\omega \\ \hline -(j_1 j_2)^{\frac{1}{2}}\omega & 1 + j_1\zeta \end{array}\right] \otimes I_S \right\} \begin{pmatrix} j_1^{\frac{1}{2}} Z_1 \\ j_1^{\frac{1}{2}} Z_2 \end{pmatrix}. \end{aligned} \tag{2.26}$$

From (2.24) and (2.26) we obtain that the Bayes estimator of α, for the combined allocation $A(j_1 + j_2 ; 2; j_1 , j_2)$ is:

$$\begin{aligned} &\hat{\alpha}(Z_1, Z_2 ; (j_1 + j_2 ; 2; j_1 , j_2)) \\ &\qquad = k_1(\zeta, \omega)\hat{\alpha}(Z_1 ; (j_1 ; 1; j_1)) + k_2(\zeta, \omega)\hat{\alpha}(Z_2 ; (j_2 ; 1; j_2)), \end{aligned} \tag{2.27}$$

where:

$$\begin{aligned} k_1(\zeta, \omega) &= ((1 + j_1\zeta)[1 + j_2\zeta - (j_1 j_2)^{\frac{1}{2}}\omega])((1 + j_1\zeta)(1 + j_2\zeta) - j_1 j_2 \omega^2)^{-1} \\ k_2(\zeta, \omega) &= ((1 + j_2\zeta)[1 + j_1\zeta - (j_1 j_2)^{\frac{1}{2}}\omega])((1 + j_1\zeta)(1 + j_2\zeta) - j_1 j_2 \omega^2)^{-1}. \end{aligned} \tag{2.28}$$

The Bayes risk of $\hat{\alpha}(Z; (j; 1; j))$ is given by (2.19). Since $\hat{\alpha}(Z_1 ; (j_1 ; 1; j_2))$ and $\hat{\alpha}(Z_2 ; (j_2 ; 1; j_2))$ are independent we obtain from (2.27) that the Bayes risk corresponding to the allocation $A(j + j; 2; j, j)$ is:

$$\begin{aligned} &\text{tr. } \mathfrak{V}(j_1 + j_2 ; 2; j_1 , j_2) \\ &\qquad = k_1^2(\zeta, \omega) \text{ tr. } \mathfrak{V}(j_1 ; 1; j_1) + k_2^2(\zeta, \omega) \text{ tr. } \mathfrak{V}(j_2 ; 1; j_2). \end{aligned} \tag{2.29}$$

Let $\hat{\alpha}(Z; (j; j; 1, \cdots, 1))$ denote the Bayes estimator of α for the allocation $A(j; j; 1, \cdots, 1)$. The Bayes risk of this estimator, tr. $\mathfrak{V}(j; j; 1, \cdots, 1)$ is given by (2.17). According to Theorem 2.1,

$$\text{tr. } \mathfrak{V}(j; j; 1, \cdots, 1) \leqq \text{tr. } \mathfrak{V}(j; 1; j), \qquad \text{with strict inequality for every } j \geqq 2.$$

Consider the estimator,

$$(2.30)\quad \tilde{\alpha}(Z_1, Z_2; (j_1 + j_2; j_1 + j_2; 1, \cdots, 1)) = k_1(\zeta, \omega)\hat{\alpha}(Z_1; (j_1; j_1; 1, \cdots, 1)) + k_2(\zeta, \omega)\hat{\alpha}(Z_2; (j_2; j_2; 1, \cdots, 1)).$$

The posterior risk of this estimator is

$$k_1^2(\zeta, \omega) \text{ tr. } \Psi(j_1; j_1; 1, \cdots, 1) + k_2^2(\zeta, \omega) \text{ tr. } \Psi(j_2; j_2; 1, \cdots, 1),$$

which is not greater than (2.29) for all j_1, j_2, and is strictly smaller than (2.29) for all $j_1 \geqq 2$ and $j_2 \geqq 2$. Finally, the (Bayes) posterior risk of the Bayes estimator $\hat{\alpha}(Z_1, Z_2; (j_1 + j_2; j_1 + j_2; 1, \cdots, 1))$ is minimal and therefore smaller or equal to $k_1^2(\zeta, \omega)$ tr. $\Psi(j_1; j_1; 1, \cdots, 1) + k_2^2(\zeta, \omega)$ tr. $\Psi(j_2; j_2; 1, \cdots, 1)$. This proves that the allocation $A(j_1 + j_2; j_1 + j_2; 1, \cdots, 1)$ is better than the allocation $A(j_1 + j_2; 2; j_1, j_2)$; it is strictly better whenever $j_1 \geqq 2$ and $j_2 \geqq 2$.

COROLLARY 2.3. *For every $n = 1, 2, \cdots, M$, the allocation $A(n; n; 1, \cdots, 1)$ is optimal with respect to the normal prior distribution.*

PROOF. Let $A(n; k; n_1, \cdots, n_k)$ be an arbitrary allocation, $1 \leqq k \leqq n$. This allocation can be obtained as a combination of k allocations over different experiments $A(n_i; 1; n_i)$, $i = 1, \cdots, k$. The Bayes estimator $\hat{\alpha}(Z_1, \cdots, Z_k; (n; k; n_1, \cdots, n_k))$ is, in analogy to (2.27), a linear combination of the corresponding Bayes estimators $\hat{\alpha}(Z_i; (n_i; 1; n_i))$, $i = 1, \cdots, k$. The same argument as in the proof of Theorem 2.2 yields that $A(n; n; 1, \cdots, 1)$ is a better allocation than $A(n; k; n_1, \cdots, n_k)$.

Corollary 2.3 gives a similar result to that of Theorem 2.8 of Degroot [4].

2.4. *The Bayes estimators.* We have proven that the Bayes sequential procedure for a prior normal distribution of $(\alpha, \beta)'$ consists of the allocation $A(n; n; 1, \cdots, 1)$, where n is the optimal number of experiments (fractional replicates). For this optimal allocation the Bayes estimator of α, at the nth stage, is, according to (1.2) and (2.16),

$$(2.31)\quad \begin{aligned}\hat{\alpha}_n(\mathfrak{I}_n) &= \theta^2 S(1 + \tau^2 N)^{-1}[1_n' \otimes C^{-1}] \\ &\cdot \{[I_n - \theta^2 S(1 - \gamma)(1 + \tau^2 N + n\theta^2 S(1 - \gamma))^{-1} J_n] \otimes I_s\} \\ &\cdot \begin{bmatrix} Y_{v_1} \\ \vdots \\ Y_{v_n} \end{bmatrix} = \theta^2 S(1 + \tau^2 N)^{-1}(1 - n\theta^2 S(1 - \gamma) \\ &\cdot (1 + \tau^2 N + n\theta^2 S(1 - \gamma))^{-1}) \sum_{i=1}^n C^{-1} Y_{v_i}.\end{aligned}$$

Accordingly,

$$(2.32)\quad \hat{\alpha}_n(\mathfrak{I}_n) = (\lambda + n - (n - 1)\gamma)^{-1} \sum_{i=1}^n C^{-1} Y_{v_i},$$

where $\lambda = (1 + \tau^2 K)/\theta^2 S$. The commonly used least squares estimator is

$$(2.33)\quad \tilde{\alpha}_n = n^{-1} \sum_{i=1}^n C^{-1} Y_{v_i}.$$

This estimator is obtained as a limit of the sequence of Bayes estimators $\hat{\alpha}_n(\mathfrak{I}_n)$ as $\theta^2 \to \infty$ and $\tau^2/\theta^2 \to 0$.

The posterior risk associated with the Bayes estimator $\hat{\alpha}_n(\mathfrak{I}_n)$ is, according to (2.17)

$$\rho_o(\mathfrak{I}_n) = \theta^2 S(\lambda - (n-1)\gamma)(n + \lambda - (n-1)\gamma)^{-1}. \tag{2.34}$$

We notice that the Bayes risk function $\rho_0(\mathfrak{I}_n)$ does not depend on the actual n experiments chosen $(v_1, \cdots, v_n)$. The only requirement is that $v_1 \neq v_2 \neq \cdots \neq v_n$.

According to (2.34), if n experiments are performed the reduction in the Bayes risk from the prior risk of $\theta^2 S$, is

$$\rho_0 - \rho_0(\mathfrak{I}_n) = n\theta^2 S(n + \lambda - (n-1)\gamma)^{-1}. \tag{2.35}$$

The number of experiments that should be performed can be determined by (2.35), which, acording to (2.6) should be in equivalent cost units larger than the cost of observing n blocks. Moreover, the size of the available budget determines whether $n < M$ or $n = M$ (full factorial experiment). In any event, the Bayes sequential procedure is actually a fixed sample size procedure; namely, a fractional replicate of size n/M.

We conclude with the following remark. As previously mentioned, the common least-squares estimators $\tilde{\alpha}_n$ (2.33) is obtained from the Bayes estimator by letting $\theta^2 \to \infty$, $\tau^2/\theta^2 \to 0$. The corresponding limit of the Bayes posterior risks is, for any fixed $0 < \tau^2 < \infty$,

$$\lim_{\theta^2 \to \infty} \rho_0(\mathfrak{F}_n) = n^{-1}(1 + \tau^2(N - n)), \qquad 0 < \tau^2 < \infty. \tag{2.36}$$

This limit approaches ∞ as $\tau^2 \to \infty$. The posterior risk of the least-squares estimator $\tilde{\alpha}_n$ is:

$$E_{\tau^2}\{n^{-2}(n + \textstyle\sum_{i=1}^{n} \beta' H'_{v_i} H_{v_i} \beta) \mid \mathfrak{F}_n\} = n^{-1}(1 + S^2(M-1)\tau^2). \tag{2.37}$$

Applying a result of Blyth [3], the least-squares estimator $\tilde{\alpha}_n$ is minimax (letting $\tau^2 \to \infty$). However, the minimax value is ∞, and obviously $\tilde{\alpha}_n$ is not a unique minimax. Moreover, from the Bayesian point of view, (2.36) and (2.37) show that for every $0 < \tau^2 < \infty$, $\tilde{\alpha}_n$ is inferior to the Bayes estmator $\hat{\alpha}_n(\mathfrak{F}_n)$.

REFERENCES

[1] Anderson, T. W. (1958). *Introduction to Multivariate Statistical Analysis.* Wiley, New York.

[2] Blackwell, D. and Girshick, M. A. (1958). *Theory of Games and Statistical Decisions.* Wiley, New York.

[3] Blyth, C. R. (1951). On minimax statistical decision procedures and their admissibility. *Ann. Math. Statist.* **22** 22–42.

[4] Degroot, M. (1966). Optimal allocation of observations. *Ann. Inst. Statist. Math.* **18** 13–28.

[5] Ehrenfeld, S. and Zacks, S. (1961). Randomization and factorial experiments. *Ann. Math. Statist.* **32** 270–297.

[6] Ehrenfeld, S. and Zacks, S. (1963). Optimal strategies in factorial experiments. *Ann. Math. Statist.* **34** 780–791.

[7] Haggstrom, G. (1966). Optimal stopping and experimental design. *Ann. Math. Stat.* **37** 7–29.

[8] HODGES, J. L. and LEHMANN, E. L. (1951). Some applications of the Cramer-Rao inequality. *Proc. Second Berkeley Symp. Math. Statist. Prob.* **I** 13–22. Univ. California Press.
[9] PAULSON, E. (1962). A sequential procedure for comparing several experimental categories with a standard or control. *Ann. Math. Statist.* **33** 438–443.
[10] WALD, A. (1950). *Statistical Decision Functions*. Wiley, New York.
[11] WHITTLE, P. (1964). Some general results in sequential analysis. *Biometrika* **51** 123–141.
[12] WOLFOWITZ, J. (1950). Minimax estimates of the mean of a normal distribution with known variance. *Ann. Math. Statist.* **21** 218–230.
[13] ZACKS, S. (1963). On a complete class of linear unbiased estimators for randomized factorial experiments. *Ann. Math. Statist.* **34** 769–779.