The Annals of Statistics
1973, Vol. 1, No. 5, 822–837

ASYMPTOTICALLY OPTIMAL BAYES SEQUENTIAL DESIGN OF EXPERIMENTS FOR ESTIMATION

BY VICTOR J. YOHAI

University of La Plata, Argentina

The purpose of this paper is to find asymptotically optimal Bayes sequential procedures for estimating a function $g(\theta_1, \theta_2, \cdots, \theta_k)$ when there are k experiments $E_1, E_2, \cdots, E_k$ and the performance of the experiment E_i conducts to the observation of a random variable whose distribution depends on the vector parameter θ_i. The term asymptotical refers here to the cost of experimentation tending to zero. The methods used are a generalization of those introduced by Bickel and Yahav.

1. Introduction. The problem of finding asymptotically optimal sequential procedures for testing in a design situation was treated first by Chernoff in [4] and then in a very general form by Kiefer and Sacks in [5].

Bickel and Yahav in [1] and [2], developed an asymptotic theory of sequential procedures for estimation when only one experiment is available. This theory is based on the concept of pointwise asymptotic optimality that they introduced.

The purpose of this paper is to apply the methods introduced by Bickel and Yahav in [1] and [2], to the problem of obtaining asymptotically optimal Bayes sequential procedures for estimation in a design situation. We consider the particular case when (i) there are a finite number of experiments $E_1, E_2, \cdots, E_k$; (ii) if the experiment E_i is used we observe a random variable whose distribution depends on a vector parameter θ_i; (iii) we want to estimate some function $g(\theta_1, \theta_2, \cdots, \theta_k)$.

In Section 2 we generalize Theorems 2.1 and 3.1 of [2] to cover the design situation. In Section 3 we present the problem of sequential procedures for estimation in a design situation and state the assumptions we shall use. In Section 4 we construct a sequential procedure that is asymptotically pointwise optimal as defined by Bickel and Yahav in [1] for the estimation of $g(\theta_1, \theta_2, \cdots, \theta_k)$ in the situation described in Section 3. In Section 5 we give additional conditions in order to prove that the procedure introduced in Section 4 is also asymptotically optimal as defined by Kiefer and Sacks in [5]. In Section 6 we give some indications of how the preceding results can be extended to the case in which the distributions corresponding to the different experiments depend on some common parameters. In Section 7 we give the proof of Theorem 4.1, used for proving that the design constructed in Section 4 is asymptotically pointwise optimal.

2. General results on asymptotic pointwise optimality and asymptotic optimality. Let

Received December 1971; revised January 1973.
AMS 1970 *subject classifications.* Primary 62; Secondary 45.
Key words and phrases. Sequential design, asymptotical optimality, Bayesian estimation.

$(\Omega, \mathscr{F}, P)$ be a probability space, D an arbitrary set that we call design space, and suppose that for each $d \in D$ we have a sequence of random variables $\{Y_{d,n}\}$, $n > 1$ where $Y_{d,n}$ is $\mathscr{F}_{d,n}$ measurable and $\mathscr{F}_{d,n} \subset \mathscr{F}_{d,n+1} \subset \mathscr{F}$ is an increasing sequence of σ-fields.

Define

$$X(n, d, c) = Y_{d,n} + nc .$$

$Y_{d,n}$ may be interpreted as the Bayes risk after n observations and c as the cost of each observation. Then $X(n, d, c)$ represents the total posterior expected cost when we stop after n observations using the design d and when the cost of each observation is c.

A stopping time for the design d will be a positive integer-valued random variable t such that $\{t = k\}$ belongs to $\mathscr{F}_{d,k}$.

A sequential procedure for our problem will be a pair (d, t) where d belongs to D and t is a stopping time for the design d.

Generalizing the concept of sequence of stopping times asymptotically pointwise optimal introduced by Bickel and Yahav in [1] we say that a sequence of sequential procedures $(d(c), t(c))$ is asymptotically pointwise optimal (A.P.O.) if for any other sequence of sequential procedures $(d'(c), t'(c))$ we have

$$\limsup_{c\to 0} X(t(c), d(c), c)/X(t'(c), d'(c), c) \leqq 1 \quad \text{a.s.}$$

The following theorem gives a general way to construct A.P.O. procedures.

Make the following assumptions:

A1. $\inf\{Y_{d,n} : d \in D\}$ is a random variable, $n \geqq 1$.

A2. $P(\inf\{Y_{d,n} : d \in D\} > 0) = 1$, $n \geqq 1$.

A3. There exists $\bar{d} \in D$, a finite random variable $V > 0$ and a real number $\beta > 0$ such that

(a) $$n^{\beta} Y_{\bar{d},n} \to V \quad \text{a.s.} \qquad \text{as } n \to \infty$$

(b) $$\liminf_{n\to\infty} n^{\beta} \inf_{d\in D} Y_{d,n} \geqq V \quad \text{a.s.}$$

THEOREM 2.1. *Let us define the stopping time $\bar{t}(c)$ by: "stop for the first n such that $Y_{\bar{d},n}(1 - (n/n+1)^{\beta}) \leqq c$" and $t'(c)$ by: "stop for the first n such that $Y_{\bar{d},n}\beta(1+n)^{-1} \leqq c$." Then under* A1, A2, *and* A3, *the sequence of procedures $(\bar{d}, \bar{t}(c))$ and $(\bar{d}, t'(c))$ are* A.P.O.

PROOF. Put $A(c) = (c^{\beta} V \beta)^{1/\beta+1} \beta^{-1}(1+\beta)$. It will be enough to prove the following three inequalities

$$\limsup_{c\to 0} X(\bar{t}(c), \bar{d}, c)/A(c) \leqq 1 \quad \text{a.s.} \tag{2.1}$$

$$\limsup_{c\to 0} X(t'(c), \bar{d}, c)/A(c) \leqq 1 \quad \text{a.s.} \tag{2.2}$$

$$\liminf_{c\to 0} \inf_n \inf_{d\in D} X(n, d, c)/A(c) \geqq 1 \quad \text{a.s.} \tag{2.3}$$

(2.1) and (2.2) are already proved in Theorem 2.1 of [3]. Now we prove (2.3). Put

$$X(n, c) = \inf_{d\in D} X(n, d, c)$$

and for every positive real u define

$$X^*(u, c) = u^{-\beta}V + uc\,.$$

Call $t(c)$ the n that minimizes $X(n, c)$. A3 (a) implies $\inf_{d \in D} Y_{d,n} \to 0$ as $n \to \infty$. Then using A2 we get

$$\lim_{c\to 0} t(c) = \infty \quad \text{a.s.} \tag{2.4}$$

Call $u^*(c)$ the value of u that minimizes $X^*(u, c)$. Differentiating we get

$$u^*(c) = (c/\beta V)^{-1/\beta+1}\,.$$

It may also be checked that

$$A(c) = X^*(u^*(c), c)\,.$$

Then we have

$$\begin{aligned}\inf_n X(n, c)/A(c) &= X(t(c), c)/X^*(u(c), c) \\ &\geqq X(t(c), c)/X^*(t(c), c) \\ &= \inf_{d\in D}(Y_{d,t(c)} + t(c)c)/(t^{-\beta}(c)V + t(c)c)\,.\end{aligned} \tag{2.5}$$

Using A2 (b), (2.4) and (2.5) it is easy to conclude that

$$\liminf_{c\to 0} \inf_n X(n, c)/A(c) \geqq 1 \quad \text{a.s.}$$ □

Following Kiefer and Sacks [5], we say that a sequence of procedures $(\bar{d}(c), \bar{t}(c))$ is asymptotically optimal (A.O.) if for any other sequence $(d(c), t(c))$ we have

$$\limsup_{c\to 0} E(X(\bar{t}(c), \bar{d}(c), c))/E(X(t(c), d(c), c)) \leqq 1\,.$$

The following theorem is a generalization of Theorem 3.1 of [3] for the design case.

THEOREM 2.2. *Under the same conditions of Theorem* 2.1, *and if*

$$\sup_n n^\beta E(Y_{\bar{d},n}) < \infty\,, \tag{2.6}$$

then the sequences of procedures $(\bar{d}(c), \bar{t}(c))$ *and* $(\bar{d}(c), t'(c))$ *are asymptotically optimal.*

PROOF. Completely similar to Theorem 3.1 of [2]. □

REMARK. In the Bayesian applications the probability P is determined by the conditional distributions given the parameter θ, P_θ, $\theta \in \Theta$ and the prior probability on Θ, γ. In this case in order to verify A3 it is enough to find $\Theta^* \subset \Theta$ such that

$$\lim_{n\to\infty} n^\beta Y_{\bar{d},n} = V \qquad \text{a.s.} \quad P_\theta\,, \qquad \forall \theta \in \Theta^*$$

and

$$\liminf_{n\to\infty} n^\beta \inf_{d\in D} Y_{d,n} \geqq V \qquad \text{a.s.} \quad P_\theta\,, \qquad \forall \theta \in \Theta^*$$

where $\gamma(\Theta^*) = 1$.

3. Bayes sequential designs for estimation. Consider a situation where there are k possible experiments, $E_1, E_2, \cdots, E_k$. We may perform each of them an infinite number of times. When we perform for the nth time the experiment E_i we observe the variable z_{in}. We assume all the z_{in}, $1 \leqq i \leqq k$, $1 \leqq n < \infty$ are independent random variables and for fixed i identically distributed. We also assume that z_{i1} has a density function $f_i(z_{i1}, \theta_i)$ with respect to some σ-finite measure μ on R^1. The parameter θ_i takes values on an open set of R^{p_i} that we denote by Θ_i. We denote by $\theta = (\theta_1, \cdots, \theta_k)$.

Let $\Omega = \times_{i=1}^{k} \times_{n=1}^{\infty} R_{in}$, where R_{in} are copies of the set of real numbers and $\mathscr{F}$ the corresponding product of Borel σ-fields. Call P_θ the measure induced by the z_{in}'s on $(\Omega, \mathscr{F})$.

We suppose a prior distribution of θ is given. This prior distribution has a density $\psi(\theta_1, \cdots, \theta_k)$ with respect to the Lebesgue measure on $\Theta = \Theta_1 \times \cdots \times \Theta_k$.

Our purpose will be to estimate a real-valued function $g(\theta_1, \theta_2, \cdots, \theta_k)$. In this paper we take for loss function the quadratic error. Then if the parameter is θ and our estimator of $g(\theta)$ is δ, the loss function will be

$$l(\theta, d) = (g(\theta) - \delta)^2 . \tag{3.1}$$

It is possible to generalize our results for a more general type of loss functions as is done in [2] for the case of one experiment, but for simplicity of exposition we shall consider only the loss function defined by (3.1).

A nonrandomized sequential design d will be a sequence $(d_1, d_2, \cdots, d_n, \cdots)$ where d_i is a Borel measurable function defined on R^{n-1} and taking values in $\{1, 2, \cdots, k\}$. In particular d_1 is a constant. We can interpret this design as indicating that if the first n observations take the values $x_1, x_2, \cdots, x_n$, then the $(n+1)$th experiment should be selected equal to E_i with $i = d_{n+1}(x_1, \cdots, x_n)$.

Suppose that using a design d the experiment E_i is chosen r_{in} times, $1 \leqq i \leqq k$, in the first n trials; then we put $z^{(n)} = z_{ij}$, $1 \leqq j \leqq r_{in}$, $1 \leqq i \leqq k$. Assume that

$$\int_\Theta (g(\theta))^2 \psi(\theta)\, d\theta < \infty . \tag{3.2}$$

Then if the loss function is given by (3.1), the Bayes estimate of $g(\theta)$ is the conditional expectation of $g(\theta)$ given $z^{(n)}$; we denote it by

$$E(g(\theta) \mid z^{(n)}) . \tag{3.3}$$

The Bayes risk $Y_{d,n}$ is given by the conditional variance of $g(\theta)$ given $z^{(n)}$. Then

$$Y_{d,n} = \operatorname{Var}(g(\theta) \mid z^{(n)}) . \tag{3.4}$$

Call $\mathscr{F}_{d,n}$ the σ-field generated by the first n variables observed when the design d is used. Then if we call c the cost of each experiment, we are in the situation described in Section 2 and we can look for a sequence of procedures A.P.O. and A.O. In order to use the results of Section 2 we need the following further assumptions.

B1.1. $\psi(\theta_1, \cdots, \theta_k)$ is positive, continuous and bounded on Θ.

B1.2. $\int_\Theta \|\theta\|^2 \psi(\theta)\, d\theta < \infty$.

B1.3. Put

$$\lambda_i(\theta_1, \cdots, \theta_{i-1}, \theta_{i+1}, \cdots, \theta_k) = \int ||\theta_i||^2 \psi(\theta_1, \cdots, \theta_k)\, d\theta_i ;$$

then $\lambda_i(\theta_1, \cdots, \theta_{i-1}, \theta_{i+1}, \cdots, \theta_k)$ is finite and continuous for all i.

B1.4. Properties B1.1, B1.2 and B1.3 hold for the marginal density of any subset $\{\theta_{i_1}, \cdots, \theta_{i_j}\}$.

Assumptions B1.3 and B1.4 are implied by B1.1 and B1.2 if $\theta_1, \theta_2, \cdots, \theta_k$ have independent prior distributions, i.e., if $\psi(\theta_1, \theta_2, \cdots, \theta_k) = \prod_{i=1}^k \psi_i(\theta_i)$.

B2.1. $\theta_i \neq \theta_i'$ implies $\mu\{f_i(z, \theta_i) \neq f_i(z, \theta_i')\} > 0$ for all i.

B2.2. $f_i(z, \theta_i)$ is continuous in θ_i for almost all z (a.e. μ) and all i.

B2.3. Put $\Phi_i(z, \theta_i) = \log f_i(z, \theta_i)$; then

$$E_{\theta_i}(|\Phi_i(z_{i1}, \theta_i)|) < \infty \qquad \text{for all} \quad i \quad \text{and} \quad \theta_i .$$

B2.4. For every i, there exists a sequence of compacts $K_{i,n}$ and a set $S_i \subset R^1$ such that (i) $\mu(S_i) = 1$, (ii) $K_{i,n} \uparrow \Theta_i$, (iii) if θ_{in} is a sequence in Θ_i with the property that for any positive integer j there exists another $n(j)$ such that $n > n(j)$ implies θ_{in} does not belong to $K_{i,j}$, then $f_i(z, \theta_{in}) \to 0$ as $n \to \infty$ for z in S_i.

B2.5. Put $\Phi_i^*(z, \theta_i, \rho) = \sup\{0, \sup\{\Phi_i(z, t) : ||t - \theta_i|| \leqq \rho\}\}$. Then for all θ_i there exists $\rho(\theta_i)$ such that $E_{\theta_i}(\Phi_i^*(z_{i1}, \theta_i, \rho(\theta_i))) < \infty$.

B2.6. Put $\varphi_i(z, n) = \sup\{0, \sup\{\Phi_i(z, \theta_i) : \theta_i \notin K_{i,n}\}\}$. Then for all θ_i there exists $n(\theta_i)$ such that $E_{\theta_i}(\varphi_i(z_{i1}, n(\theta_i))) < \infty$. The $K_{i,n}$'s are the same as in B2.4.

B2.7. Put $\theta_i = (\theta_{i,1}, \cdots, \theta_{i,p_i})$. Then $\partial^2\Phi_i(z, \theta_i)/\partial\theta_{i,r}\partial\theta_{i,s}$ is finite and continuous in θ_i for almost all z (a.e. μ) and all i, r and s.

B2.8. For every θ_i in Θ_i there exists $\varepsilon(\theta_i) > 0$ such that

$$E_{\theta_i}(\sup\{|\partial^2\Phi_i(z_{i1}, t)/\partial\theta_{i,r}\,\partial\theta_{i,s}| : ||t - \theta_i|| \leqq \varepsilon(\theta_i)\}) < \infty$$

for all i, r and s.

It is known that B2.7 and B2.8 imply

$$E_{\theta_i}(\partial\Phi_i(z_{i1}, \theta_i)/\partial\theta_{i,r}) = 0 \tag{3.5}$$

for all i and r.

Denote the covariance matrix of $(\partial\Phi_i(z_{i1}, \theta_i)/\partial\theta_{i,1}, \cdots, \partial\Phi_i(z_{i1}, \theta_i)/\partial\theta_{i,p_i})$ by $A_i(\theta_i)$. Set $A^*(\theta_i)$ the matrix whose r sth entry is $E_{\theta_i}(\partial^2\Phi_i(z_{i1}, \theta_i)/\partial\theta_{i,r}\,\partial\theta_{i,s})$. Then B2.7 and B2.8 imply

$$A_i(\theta_i) = -A_i^*(\theta_i) . \tag{3.6}$$

B2.9. $A_i(\theta_i)$ is positive definite for all i and θ_i.

B2.10. $A_i(\theta_i)$ is continuous.

B2.1—B2.6 are essentially the assumptions that Wald uses in [7] to prove the consistency of the maximum likelihood estimate. The only difference is that our assumptions B2.4 and B2.6 are more general than the corresponding ones in [7] in order to cover the case when Θ_i is open.

B2.7—B2.9 are regularity conditions that suffice for the asymptotic normality of the maximum likelihood estimator.

B3.1. $\partial g(\theta)/\partial\theta_{i,r}$ is continuous in θ for all i and r.

B3.2. Put $\text{grad}_i\, g(\theta) = (\partial g(\theta)/\partial\theta_{i,1}, \cdots, \partial g(\theta)/\partial\theta_{i,p_i})$. Then for every i and θ, $\text{grad}_i\, g(\theta) \neq 0$.

B3.3. $\sup\{\sum_{i=1}^k \|\text{grad}_i\, g(\theta)\| : \theta \in \Theta\} < \infty$.

Define the maximum likelihood estimator of θ_i, $\hat\theta_{in}(z_{i1}, \cdots, z_{in})$, as any function satisfying

$$\sum_{j=1}^n \Phi_i(z_{ij}, \hat\theta_{in}(z_{i1}, \cdots, z_{in})) = \max\{\sum_{j=1}^n \Phi_i(z_{ij}, \theta_i) : \theta_i \in \Theta_i\}.$$

Under B2.2 and B2.4 there exists a measurable version of $\hat\theta_{in}$. As in [7] it can be proved that under B2.1—B2.6, we have

$$\lim_{n\to\infty} \hat\theta_{in}(z_{i1}, \cdots, z_{in}) = \theta_{i0} \quad \text{a.s.} \quad P_{\theta_{i0}} \tag{3.7}$$

for any θ_{i0} in Θ_i.

4. A.P.O. Bayes sequential procedures for estimation. In this section we find a sequence of sequential procedures that is A.P.O. for the estimation of $g(\theta)$ in the situation described in Section 3.

We shall need the following definitions. Assume B2.9. Put

$$J_i(\theta) = (\text{grad}_i\, g(\theta))'(A_i(\theta_i))^{-1}\,\text{grad}_i\, g(\theta). \tag{4.1}$$

Set

$$Q = \{\lambda = (\lambda_1, \lambda_2, \cdots, \lambda_k) : \lambda_i \geqq 0, \sum_{i=1}^k \lambda_i = 1\}.$$

For $\theta = (\theta_1, \cdots, \theta_k)$ in Θ and $\lambda = (\lambda_1, \cdots, \lambda_k)$ in Q define

$$U(\theta, \lambda) = \sum_{i=1}^k (1/\lambda_i) J_i(\theta). \tag{4.2}$$

It is easy to see that $U(\theta, \lambda)$ with $\lambda = (\lambda_1, \cdots, \lambda_k)$ is the Rao–Cramér variance for estimating $g(\theta)$ when the frequency of the experiment E_i is equal to λ_i. It may be shown without difficulty that

$$V(\theta) = \inf_{\lambda\in Q} U(\theta, \lambda) = U(\theta, \bar\lambda(\theta)) \tag{4.3}$$

where $\bar\lambda(\theta) = (\bar\lambda_1(\theta), \cdots, \bar\lambda_k(\theta))$ with

$$\bar\lambda_i(\theta) = (J_i(\theta))^{\frac12} / \sum_{i=1}^k (J_i(\theta))^{\frac12}. \tag{4.4}$$

$\bar\lambda_i(\theta)$ gives the proportion in which the experiment E_i should be taken in order to obtain the minimum Rao–Cramér variance. Then it is reasonable to think than in a "good" design the asymptotic frequence of appearance of the experiment E_i should be $\bar\lambda_i(\theta)$. This is made rigorous by the following two theorems.

THEOREM 4.1. *Assume* B1.1—B1.4, B2.1—B2.9 *and* B3.1—B3.3. *Let* $Y_{d,n}$ *be as defined in* (3.4) *and* D *the set of all nonrandomized designs. Then*

(i) A1 *holds*;
(ii) A2 *holds*;

(iii) $\liminf_{n\to\infty} \inf_{d\in D} nY_{d,n} \geqq V(\theta_0)$ a.e. P_{θ_0}, *for every* θ_0 *in* Θ.

(iv) *Let* d_0 *be a nonrandomized design such that if* λ_{in} *is the frequency of the experiment* E_i *in the first n observations when the design* d_0 *is used, we have*

$$\lim_{n\to\infty} \lambda_{in} = \bar{\lambda}_i(\theta_0) \qquad \text{a.s.} \quad P_{\theta_0}\,. \tag{4.5}$$

Then

$$\lim_{n\to\infty} nY_{d_0,n} = V(\theta_0) \qquad \text{a.s.} \quad P_{\theta_0}\,. \tag{4.6}$$

The proof of this theorem is postponed until Section 7. Using Theorems 2.1 and 4.1 we obtain immediately Theorem 4.2.

THEOREM 4.2. *Under the same assumptions as in Theorem* 4.1, *if a design* d_0 *satisfies* (4.5), *then the sequence of procedures* $(d_0, t_0(c))$, *where* $t_0(c)$ *is the stopping time*: "*stop the first time* $Y_{d_0,n}/n + 1 \leqq c$," *is* A.P.O.

The rest of this section is devoted to constructing a design which satisfies (4.5). The basic idea for constructing this design is to choose the $(n + 1)$th experiment as if the true value of θ were some estimate $\tilde{\theta}$ based on the first n observations.

A similar idea was used first by Wald in [8] and then by Kiefer and Sacks in [5] in a design situation. They use a two-stage design: first take $n_i(c)$ observations of the experiment E_i in order to estimate θ, and then using the estimate of θ choose a complementary sample. We cannot use in our approach this type of two-stage design since it depends on c, and according to Theorem 2.1, the design d_0 should be independent of c.

Let $\tilde{\theta}_{in}$ $n \geqq 0$ $1 \leqq i \leqq k$ be any sequence of estimates of θ_i such that $\tilde{\theta}_{in}$ is in Θ_i and

$$\lim \tilde{\theta}_{in} = \theta_{i0} \qquad \text{a.s.} \quad P_{\theta_0}\,. \tag{4.7}$$

We define the design d_0 as follows. Let $\tilde{\theta}_0{}^* = (\tilde{\theta}_{10}, \cdots, \tilde{\theta}_{k0})$; then we select as the first experiment any E_{i_1} such that i_1 satisfies

$$\bar{\lambda}_{i_1}(\tilde{\theta}_0{}^*) = \max\{\lambda_i(\tilde{\theta}_0{}^*) : 1 \leqq i \leqq k\}\,. \tag{4.8}$$

This means that we take as the first experiment any one that according to our initial estimate of θ should appear with the greatest frequency.

Assume now that we have already selected the first n experiments and that the experiment E_i was chosen r_{in} times; then its frequency is $\lambda_{in} = r_{in}/n$. Call $\tilde{\theta}_n{}^* = (\tilde{\theta}_{1n}^*, \cdots, \tilde{\theta}_{kn}^*)$, where $\tilde{\theta}_{in}^* = \tilde{\theta}_{ir_{in}}^*$. Then $\tilde{\theta}_n{}^*$ is the estimate of θ based on the first n observations. The $(n + 1)$th will be any E_i such that i satisfies

$$\bar{\lambda}_i(\tilde{\theta}_n{}^*) - \lambda_{in} \geqq 0\,. \tag{4.9}$$

This means that the $(n + 1)$th experiment is chosen equal to any one whose frequency in the first n observations is smaller than the estimated optimal frequency.

We have the following theorem.

THEOREM 4.3. *Assume* B2.9, B2.10, B3.1, B3.2 *and* (4.8). *Then if* λ_{in} *is the frequency of the experiment* E_i *in the first n observations when the design* d_0 *constructed above is used, we have*

$$\lim_{n\to\infty} \lambda_{in} = \bar{\lambda}(\theta_0) \quad \text{a.s. } P_{\theta_0}, \qquad \forall \theta_0 \in \Theta . \tag{4.10}$$

PROOF. We start by proving that if $\tilde{\theta}_n^* \to \tilde{\theta}$ then $\lambda_{in} \to \bar{\lambda}_i(\tilde{\theta})$. Suppose that for i_0 $\lambda_{i_0 n} \nrightarrow \bar{\lambda}_{i_0}(\tilde{\theta})$. Then for some $\varepsilon > 0$, $|\lambda_{i_0 n} - \bar{\lambda}_{i_0}(\tilde{\theta})| \geqq \varepsilon$ holds infinitely many times. Since $\sum_{i=1}^k (\lambda_{in} - \bar{\lambda}_i(\tilde{\theta})) = 0$ it is easy to show that there exists i_1 such that $\lambda_{i_1 n} - \bar{\lambda}_{i_1}(\tilde{\theta}) \geqq \varepsilon/k = 2\varepsilon'$ holds infinitely many times. B2.10 and B3.1 imply that $\bar{\lambda}_{i_1}(\theta)$ is continuous. Then $\bar{\lambda}_{i_1}(\tilde{\theta}_n^*) \to \bar{\lambda}_{i_1}(\tilde{\theta})$ and we have

$$\lambda_{i_1 n} - \bar{\lambda}_{i_1}(\tilde{\theta}_n^*) \geqq \varepsilon' \qquad \text{for infinitely many } n. \tag{4.11}$$

It is easy to find n_0 such that

$$\lambda_{i_1 n} - \bar{\lambda}_{i_1}(\tilde{\theta}_n^*) \leqq \varepsilon'/2 \qquad \text{if } n \geqq n_0$$

and this clearly contradicts (4.11). Then $\lim_{n\to\infty} \lambda_{in} = \bar{\lambda}_i(\tilde{\theta})$, $1 \leqq i \leqq k$.

We show now that

$$\lim_{n\to\infty} \tilde{\theta}_n^* = \theta_0 \quad \text{a.s. } P_{\theta_0} . \tag{4.12}$$

From (4.7) we have that

$$\lim_{n\to\infty} \tilde{\theta}_n^* = \tilde{\theta}_0 \quad \text{a.s. } P_{\theta_0} \tag{4.13}$$

where $\tilde{\theta}_0 = (\tilde{\theta}_{10}, \cdots, \tilde{\theta}_{k0})$ with $\tilde{\theta}_{i0} = \theta_{i0}$ if the experiment E_i is taken infinitely many times and $\tilde{\theta}_{i0} = \tilde{\theta}_{in}$ if the experiment E_i is taken only n times. Then we have

$$\lim_{n\to\infty} \lambda_{in} = \bar{\lambda}_i(\tilde{\theta}_0) \quad \text{a.s. } P_{\theta_0}, \qquad 1 \leqq i \leqq k . \tag{4.14}$$

Using B2.9, B3.2, (4.1) and (4.4) we get that $\bar{\lambda}_i(\tilde{\theta}_0) > 0$, $1 \leqq i \leqq k$. Then from (4.14) we have that every experiment E_i is selected infinitely many times; then $\tilde{\theta}_{i0} = \theta_{i0}$, and from (4.13) we obtain (4.12). As we have proved above, (4.12) implies (4.10). □

THEOREM 4.4. *Assume* B1.1—B1.4, B2.1—B2.10 *and* B3.1—B3.3. *Then the sequence of procedures* $(d_0, t_0(c))$, *where* $t_0(c)$ *is the stopping time*: "*stop the first time* $Y_{d_0,n}/(n+1) \leqq c$," *is* A.P.O.

PROOF. Follows immediately from Theorems 4.2 and 4.3. □

5. A.O. Bayes sequential procedures for estimation. In this section we show that the A.P.O. sequence of sequential procedures $(d_0, t_0(c))$ is also under general conditions A.O.

THEOREM 5.1. *Assume* B1.1—B1.4, B2.1—B2.10 *and* B3.1—B3.3. *Suppose too that*

(i) *For each i there exists a sequence of estimators* $\bar{\theta}_{in}(z_{i1}, \cdots, z_{in})$ *such that*

$$\sup_n \int nE_{\theta_i}(||\bar{\theta}_{in}(z_{i1}, \cdots, z_{in}) - \theta_i||^2)\psi(\theta)\, d\theta < \infty . \tag{5.1}$$

(ii) *There exists* $\lambda > 0$ *such that*

$$\sup_{\theta,n} nP_\theta(\bigcup_{i=1}^k \{\lambda_{in} < \lambda\}) < \infty . \tag{5.2}$$

(λ_{in} is the frequency of the experiment E_i in the first n observations when the design d_0 is used.)

Then the sequence of procedures $(d_0, t_0(c))$ *is* A.O.

PROOF. According to Theorem 2.2 it suffices to prove that

$$\sup_n E(nY_{d_0,n}) < \infty . \tag{5.3}$$

Call $A_n = \bigcup_{i=1}^k \{\lambda_{in} < \lambda\}$. Then we have

$$nY_{d_0,n} = nY_{d_0,n}\chi_{A_n} + nY_{d_0,n}\chi_{A_n'} \tag{5.4}$$

where χ_A denotes the indicator function of the set A, and A' the complement of A.

Using B1.2, (5.2) and the definition of $Y_{d_0,n}$ we have

$$E(nY_{d_0,n}\chi_{A_n}) \leqq E(ng^2(\theta)\chi_{A_n}) = \int \psi(\theta)g^2(\theta)nP_\theta(A_n)\,d\theta < K_1 < \infty \tag{5.5}$$

where K_1 is a constant.

Using, B3.3., (5.1) and the definition of $Y_{d_0,n}$ we can write

$$\begin{aligned} E(nY_{d_0,n}\chi_{A_n'}) &\leqq nE(g(\bar\theta_{1[\lambda n]}, \cdots, \bar\theta_{k[\lambda n]}) - g(\theta_1, \cdots, \theta_k))^2 \\ &\leqq \frac{2}{\lambda}[\lambda n]K_2 \sum_{i=1}^k E(\|\bar\theta_{i[\lambda n]} - \theta_i\|^2) \leqq K_3 \end{aligned} \tag{5.6}$$

where $[s]$ denotes the largest integer smaller than or equal to s, and K_2 and K_3 are constants.

Then from (5.4), (5.5) and (5.6) we obtain (5.3). □

In order to find sufficient conditions for (i) see Theorems 4.3 and 4.4 of [2].

The following lemma gives a sufficient condition for (ii).

LEMMA 5.2. *If* $0 < a \leqq J_i(\theta) \leqq b < \infty$ *for all i and* θ, *then* (5.2) *is satisfied.*

PROOF. From the assumptions of the lemma we obtain

$$\bar\lambda_i(\theta) = (J_i(\theta))^{\frac12}/\sum_{i=1}^k (J_i(\theta))^{\frac12} \geqq a^{\frac12}/kb^{\frac12} = 2\lambda .$$

Then, from the construction of d_0 it follows that there exists a fixed positive integer n_0 such that for $n \geqq n_0$ we have $\lambda_{in} \geqq \lambda$ for all i. Then $P_\theta(\bigcup_{i=1}^k \{\lambda_{in} < \lambda\}) = 0$ for $n \geqq n_0$ and (5.2) follows. □

EXAMPLE 1. Suppose we have two Bernoulli populations with parameters p_1 and p_2, and we want to estimate $g(p_1, p_2) = p_1 - p_2$. Then

$$f_i(z, p_i) = p_i^z(1 - p_i)^{1-z} , \qquad z = 0, 1;\ 0 < p_i < 1;\ i = 1, 2 .$$

These densities satisfy Assumptions B2.1—B2.10, and $g(p_1, p_2)$ obviously satisfies Assumptions B3.1—B3.3. Suppose that the prior density of (p_1, p_2)

satisfies Assumptions B1.1—B1.4, and construct the design d_0 taking as $\hat{p}_{in} = \inf(1 - 1/n, \sup(\bar{p}_{in}, 1/n))$ where $\bar{p}_{in}$ is the sample mean. Then the sequence of sequential procedures $(d_0, t_0(c))$ is A.P.O.

Moreover, $f_i(z, p_i)$ satisfies the assumption (i) of Theorem 5.1, since $\operatorname{Var}(\bar{p}_{in}) = p_i(1 - p_i)/n$. In order to satisfy assumption (ii) of Theorem 5.1 a sufficient condition is to take as parameter space the set

$$\{(p_1, p_2) : a_1 < p_1 < b_1; a_2 < p_2 < b_2\}$$

where $0 < a_i < b_i < 1$.

This follows from the fact $J_i(p_i) = p_i(1 - p_i)$, and Lemma 5.2. Then if we construct the design d_0 using as $\hat{p}_{in} = \inf(b, \sup(\bar{p}_{in}, a))$, the sequence of sequential procedures $(d_0, t_0(c))$ is A.O.

EXAMPLE 2. Consider two normal populations with unknown means and variances, and suppose we want to estimate the difference of means, then

$$f_i(z, \mu_i, \sigma_i^2) = (2\pi\sigma_i^2)^{-\frac{1}{2}} \exp[-(z - \mu_i)^2/2\sigma^2]$$
$$i = 1, 2; \ -\infty < \mu_i < \infty; \ 0 < \sigma^2 < \infty$$

and $g(\mu_1, \sigma_1^2, \mu_2, \sigma_2^2) = \mu_2 - \mu_1$.

These densities satisfy Assumptions B2.1—B2.10 and g obviously satisfies B3.1—B3.3. Suppose that the prior density of $(\mu_1, \sigma_1^2, \mu_2, \sigma_2^2)$ satisfies B1.1—B1.4.

We have

$$J_i(p_i, \sigma_i^2) = \sigma_i^2 \tag{5.7}$$

and

$$\bar{\lambda}_i(\mu_i, \sigma_i^2) = \sigma_i/(\sigma_1 + \sigma_2)\ .$$

Then in order to construct the design d_0 we need only estimates of σ_i^2, $i = 1, 2$. We can use as the estimate of σ_i^2, $\hat{\sigma}_{in}^2 = \sum_{j=1}^{n} (z_{ij} - \bar{z}_{in})^2/n - 1$ for $n \geqq 2$ where $\bar{z}_{in} = \sum_{j=1}^{n} z_{ij}/n$.

Then according to Theorem 4.4, the sequence of procedures $(d_0, t_0(c))$ is A.P.O.

It is very easy to show that in this case assumption (i) of Theorem 5.1 may be replaced by the assumption that there exists a sequence of estimates $\hat{\mu}_{in}$ of μ_i such that

$$\sup_n \int_0^\infty \int_{-\infty}^\infty nE_{\mu_i,\sigma_i^2}(\hat{\mu}_{in} - \mu_i)^2 \psi_i(\mu_i, \sigma_i^2)\, d\mu_i\, d\sigma_i^2 < \infty \tag{5.8}$$

where $\psi_i(\mu_i, \sigma_i^2)$ is the marginal prior density of μ_i, σ_i^2.

We can take $\hat{\mu}_{in} = \bar{z}_{in}$; then

$$E_{\mu_i,\sigma_i^2}(\hat{\mu}_{in} - \mu_i) = \sigma_i^2/n$$

and (5.8) holds.

Then by Theorem 5.1, Lemma 5.2 and (5.7), in order that the sequence of procedures $(d_0, t_0(c))$ be A.O., it is sufficient that σ_i^2 be restricted to an interval (a_i, b_i) with $a_i > 0$.

6. Some extensions. The assumption that $\Theta = \Theta_1 \times \cdots \times \Theta_k$, together with the assumption that the prior distribution of θ is absolutely continuous with respect to the Lebesgue measure implies that the preceding formulation does not contemplate the case when the distributions corresponding to different experiments may depend on some unknown common parameters. This is the case, for example, when there are two experiments which lead to the observation of variables normally distributed with the same mean and different variance, and we want to estimate for example the common mean. Anyhow, the restriction $\Theta = \Theta_1 \times \cdots \times \Theta_k$ is not essential, but without it the proofs become very cumbersome, although they remain similar. In what follows we give without proof some extensions of the preceding results.

Consider the vector parameter $\theta = (\theta_1, \cdots, \theta_p)$ taking values in an open subset Θ of R^p. (Observe that in the setup of Section 3 the θ_i's are vectors while here they are real numbers.) Let M_i, $1 \leqq i \leqq k$ be subsets of $\{1, \cdots, p\}$ not necessarily disjoint with cardinal m_i, and put $\boldsymbol{\theta}_i = (\theta_h, h \in M_i)$. Suppose that the distribution of the variable observed when the experiment E_i is performed, $1 \leqq i \leqq k$, has a density $f_i(z, \boldsymbol{\theta}_i)$ with respect to a σ-finite measure μ.

As in the setup of Section 3 we have a prior density of θ, $\psi(\theta)$, with respect to the Lebesgue measure on Θ. We want to estimate a function $g(\theta^{(1)})$, where $\theta^{(1)} = (\theta_1, \theta_2, \cdots, \theta_r)$ $r \leqq p$, and where we have quadratic loss.

We need the following further assumptions

C1. $\psi(\theta)$ satisfies B2.1—B2.10.

C2. $f_i(z, \boldsymbol{\theta}_i)$ satisfies B2.1—B2.10.

C3. $\partial g(\theta^{(1)})/\partial\theta_i$ is continuous in $\theta^{(1)}$, different from 0 and bounded, $1 \leqq i \leqq r$.

Define by $A_i^{(1)}(\theta)$ the covariance matrix of the r dimensional vector whose j component is $\partial \lg f_i(z_{i1}, \boldsymbol{\theta}_i)/\partial\theta_j$ if j is in M_i and 0 if j is not in M_i, $1 \leqq j \leqq r$.

Put $S_h = \{i : h \in M_i\}$, $1 \leqq h \leqq p$. S_h is the set of indices corresponding to those experiments whose distribution depends on θ_h. Call

$$Q^* = \{\lambda = (\lambda_i, \cdots, \lambda_k) : \textstyle\sum_{i=1}^k \lambda_i = 1; \lambda_i \geqq 0, 1 \leqq i \leqq k;$$
$$\textstyle\sum_{i \in S_h} \lambda_i > 0, 1 \leqq h \leqq r\} .$$

Using B2.9 it is easy to show that if $\lambda = (\lambda_1, \cdots, \lambda_k)$ is in Q^* then $\sum_{i=1}^k \lambda_i A_i^{(1)}(\theta)$ is positive definite.

Put

$$\text{grad}\, g(\theta^{(1)}) = (\partial g(\theta^{(1)})/\partial\theta_1, \cdots, \partial g(\theta^{(1)})/\partial\theta_r) ;$$

then we define for θ in Θ and $\lambda = (\lambda_1, \cdots, \lambda_k)$ in Q,

$$U(\theta, \lambda) = (\text{grad}\, g(\theta^{(1)}))'(\textstyle\sum \lambda_i A_i^{(1)}(\theta))^{-1}\, \text{grad}\, g(\theta^{(1)})$$

and

$$V(\theta) = \inf_{\lambda \in Q^*} U(\theta, \lambda) .$$

Let $\bar{\lambda}(\theta)$ be a version which satisfies

$$V(\theta) = U(\theta, \bar{\lambda}(\theta)) .$$

Then Theorems 4.1 and 4.2 hold. Moreover, if there exists a continuous version of $\bar{\lambda}(\theta)$ we can construct a design d_0 similar to the one constructed in Section 4 such that the sequence of procedures $(d_0, t_0(c))$ is A.P.O. We can also give conditions similar to those of Section 5 which make this sequence A.O.

7. Appendix. In this Section we give the proof of Theorem 4.1. We start by studying the asymptotic behavior of the posterior distribution of θ after the first n observations, when these are chosen using a design of "simple" structure. In order to specify what we understand by "simple" we give the follow definition.

DEFINITION. A simple subdesign is a double sequence of nonnegative integers (h_{in}) $1 \leqq i \leqq k$, $1 \leqq n < \infty$ satisfying the following properties.

(i) For fixed i, h_{in} is non-decreasing.

(ii) $a(n) = \sum_{i=1}^k h_{in} \to \infty$ as $n \to \infty$.

h_{in} can be interpreted as the number of times that the experiment E_i is taken in the first $a(n)$ observations.

Fix a simple subdesign h_{in} $1 \leqq i \leqq k$, $1 \leqq n < \infty$. Call $J = \{i : \lim_{n\to\infty} h_{in} = \infty\}$ and J' its complement. Then there exists n_0 such that for i belonging to J' and $n \geqq n_0$ we have $h_{in} = h_{in_0}$. Assume for example $J = \{1, 2, \cdots, s\}$ $s \leqq k$. Put $\hat{\theta}_{in}^* = \hat{\theta}_{ih_{in}}$ and define $\theta_{in}^* = h_{in}^{\frac{1}{2}}(\theta_i - \hat{\theta}_{in}^*)$ for $1 \leqq i \leqq s$. Set $z^{(n)} = (z_{ij})$ $1 \leqq j \leqq h_{in}$, $1 \leqq i \leqq k$ and call $\psi_n{}^*(t_1, \cdots, t_k \mid z^{(n)})$ the density of the distribution of $(\theta_{1n}^*, \cdots, \theta_{sn}^*, \theta_{s+1}, \cdots, \theta_k)$ given $z^{(n)}$.

Put for $1 \leqq i \leqq s$

$$\nu_{in}(t_i, z^{(n)}) \exp[\textstyle\sum_{j=1}^{h_{in}} [\Phi_i(z_{ij}, \hat{\theta}_{in}^* + h_{in}^{-\frac{1}{2}} t_i) - \Phi_i(z_{ij}, \hat{\theta}_{in}^*)] \,.$$

Call $z^{(0)} = (z_{in})$ $1 \leqq n \leqq h_{in_0}$, $s + 1 \leqq i \leqq k$ and put

$$\nu_0(t_{s+1}, \cdots, t_k, z^{(0)}) = \exp[\textstyle\sum_{i=s+1}^k \sum_{j=1}^{h_{in_0}} \Phi_i(z_{ij}, t_i)] \,.$$

Put $t^{(1)} = (t_1, \cdots, t_s)$, $t^{(2)} = (t_{s+1}, \cdots, t_k)$ and $t = (t^{(1)}, t^{(2)})$. Set

$$\alpha(t, z^{(n)}) = [\psi(\hat{\theta}_{1n}^* + h_{1n}^{-\frac{1}{2}} t_1, \cdots, \hat{\theta}_{sn}^* + h_{sn}^{-\frac{1}{2}} t_s, t^{(2)}) \nu_0(t^{(2)}, z^{(0)}) \textstyle\prod_{i=1}^s \nu_i(t_i, z^{(n)})] \,.$$

Then we have for $n \geqq n_0$

$$\psi_n{}^*(t \mid z^{(n)}) = \alpha(t, z^{(n)}) / \textstyle\int \alpha(t, z^{(n)})\, dt \,. \tag{7.1}$$

Define $\psi_0{}^*(t^{(2)} \mid z^{(0)}, \theta_1, \cdots, \theta_s)$ the density function of $\theta^{(2)} = (\theta_{s+1}, \cdots, \theta_k)$ given $z^{(0)}$ and $\theta^{(1)} = (\theta_1, \cdots, \theta_s)$.

The following theorem is a generalization of Theorem 2.2. of [3]. It gives the asymptotic behavior of the posterior distribution of θ when a simple subdesign is used. The joint posterior distribution of θ_i, $1 \leqq i \leqq s$ behaves asymptotically up to the second moment as the distribution of independent multivariate normal random vectors with means $\hat{\theta}_{in}^*$ and covariance matrix $A_i(\theta_0)/n$.

THEOREM 7.1. *Assume* B1.1—B1.4 *and* B2.1—B2.9. *Then for any* $\theta_0 \in \Theta$, $\theta_0 = (\theta_{10}, \cdots, \theta_{k0})$, *there exists a* P_{θ_0} *null set* $F(\theta_0)$ *independent of the simple subdesign* (h_{in}) $1 \leqq i \leqq k$, $1 \leqq n < \infty$, *such that*

$$\lim_{n\to\infty} \textstyle\int \|t\|^q \,|\psi_n{}^*(t, z^{(n)}) - \psi_0{}^*(t^{(2)} \mid z^{(0)}, \theta_{10}, \cdots, \theta_{s0}) | \prod_{i=1}^s \varphi((A_i(\theta_{i0}))^{-1}, t_i)|\, dt = 0$$

for $0 \leqq q \leqq 2$ *and* $z = (z_{ij})$ $1 \leqq i \leqq k$, $1 \leqq j < \infty$ *not in* $F(\theta_0)$ *where* $\varphi(A, t)$ *denotes the density function corresponding to a multivariate normal distribution with mean* 0 *and covariance matrix* A.

PROOF. Using the same argument as that in Theorem 3.1 of [1] it will be enough to show that there exists a P_{θ_0} null set $F(\theta_0)$ such that

$$\lim_{n\to\infty} \int [(1 + \|t\|^2)\psi(\hat{\theta}_{1n}^* + h_{1n}^{-\frac{1}{2}}t_1, \cdots, \hat{\theta}_{sn}^* + h_{sn}^{-\frac{1}{2}}t_s, t^{(2)}) \times \nu_0(t^{(2)}, z^{(0)})|\prod_{i=1}^s \nu_{in}(t_i, z^{(n)}) - \prod_{i=1}^s \varphi_i^*(t_i)|]\, dt = 0 \tag{7.2}$$

for z not in $F(\theta_0)$, where $\varphi_i^*(t_i) = \exp[-\frac{1}{2}t_i' A_i(\theta_{i0})t_i]$.

B2.2 and B2.4 imply that there exists a μ-null set S, independent of the simple subdesign such that

$$\sup \{\nu_0(t^{(2)}, z^{(0)}) : t_{s+1} \in \Theta_{s+1}, \cdots, t_k \in \Theta_k\} < \infty \tag{7.3}$$

for z not in S.

Then in order to show (7.2), it is enough to find a P_{θ_0} null set $F(\theta_0) \supset S$ such that

$$\lim_{n\to\infty} \int [(1 + \|t^{(1)}\|^2)\bar{\psi}(\hat{\theta}_{1n}^* + h_{1n}^{-\frac{1}{2}}t_1, \cdots, \hat{\theta}_{sn}^* + h_{sn}^{-\frac{1}{2}}t_s) \times |\prod_{i=1}^s \nu_{in}(t_i, z^{(n)}) - \prod_{i=1}^s \varphi_i^*(t_i)|]\, dt^{(1)} = 0 \tag{7.4}$$

for z not in $F(\theta_0)$, and

$$\lim_{n\to\infty} \int \lambda(\hat{\theta}_{1n}^* + h_{1n}^{-\frac{1}{2}}t_1, \cdots, \hat{\theta}_{sn}^* + h_{sn}^{-\frac{1}{2}}t_s) \times |\prod_{i=1}^s \nu_{in}(t_i, z^{(n)}) - \prod_{i=1}^s \varphi_i^*(t_i)|\, dt^{(1)} = 0 \tag{7.5}$$

for z not in $F(\theta_0)$, where

$$\bar{\psi}(t_1, \cdots, t_s) = \int \psi(t^{(1)}, t^{(2)})\, dt^{(2)}$$

and

$$\lambda(t_1, \cdots, t_s) = \int \|t^{(2)}\|^2 \psi(t^{(1)}, t^{(2)})\, dt^{(2)} .$$

Then according to Assumptions B1.1—B1.4 we have that $\bar{\psi}(t_1, \cdots, t_s)$ is bounded and continuous and $\lambda(t_1, \cdots, t_s)$ is finite and continuous.

As in Theorem 2.2 of [3], for every $(z_{i1}, \cdots, z_{in}, \cdots)$ outside a $P_{\theta_{i0}}$ null set we can find $\delta_i > 0$ and $N_i(z_{i1}, \cdots, z_{in}, \cdots)$ such that we have

$$\log \nu_{in}(t_i, z^{(n)}) \leqq -t_i' B_i t_i \qquad \text{for} \quad \|t_i - \theta_{i0}\| \leqq 2\delta_i \tag{7.6}$$

where B_i are positive definite matrices that may depend on θ_{i0}.

As in Theorem 1 of [7] it may be proved that there exists $\varepsilon_i > 0$ such that

$$\limsup_{n\to\infty} \sup_{\|\theta_i - \theta_{i0}\| \geqq \delta_i} (1/n) \sum_{j=1}^n [\Phi_i(z_{ij}, \theta_i) - \Phi_i(z_{ij}, \theta_{i0})] \leqq -2\varepsilon_i \qquad \text{a.s.} \quad P_{\theta_0} . \tag{7.7}$$

Then using the same arguments as in Lemma 2.6 of [3] it may be proved that there exists $N_i^*(z_{i1}, \cdots, z_{in}, \cdots)$ such that for $n \geqq N_i^*(z_{i1}, \cdots, z_{in}, \cdots)$ we have

$$\sup_{\|\theta_i - \theta_{i0}\| \geqq \delta_i} \nu_{in}(t_i, z^{(n)}) \leqq \exp(-\varepsilon_i h_{in}) \qquad \text{a.s.} \quad P_{\theta_0} . \tag{7.8}$$

Using the strong law of large numbers we may prove

$$\text{(7.9)}\qquad \lim_{n\to\infty} \nu_{in}(t_i, z^{(n)}) = \varphi_i^*(t_i) \qquad \text{a.s.}\quad P_{\theta_0}, \qquad 1 \leqq i \leqq s .$$

Call $F(\theta_0)$ the union of S with the set where any one of (3.7), (7.6), (7.8) or (7.9) does not hold. Then $P_{\theta_0}(F(\theta_0)) = 0$.

For each subset M of $\{1, 2, \cdots, s\}$ call

$$A(M, n) = \{(t_1, \cdots, t_s) : t_i \in R^{p_i}, 1 \leqq i \leqq s, \|t_i\| \leqq \delta_i h_{in}^{\frac{1}{2}} \text{ for } i \in M, \|t_i\| > \delta_i h_{in}^{\frac{1}{2}} \text{ for } i \in M'\} .$$

Then in order to prove (7.4) and (7.5) it is enough to show that for any M we have

$$\text{(7.10)}\qquad \lim_{n\to\infty} \int_{A(M,n)} H_1(t^{(1)}, z^{(n)})\, dt^{(1)} = 0$$

for z not in $F(\theta_0)$, and

$$\text{(7.11)}\qquad \lim_{n\to\infty} \int_{A(M,n)} H_2(t^{(1)}, z^{(n)})\, dt^{(1)} = 0$$

for z not in $F(\theta_0)$, where $H_1(t^{(1)}, z^{(n)})$ and $H_2(t^{(1)}, z^{(n)})$ are the integrands of (7.4) and (7.5) respectively.

The proofs of (7.10) and (7.11) for the case in which $M \neq \{1, \cdots, s\}$ are straightforward but long. For details see Yohai [9].

Consider now $M = \{1, 2, \cdots, s\}$. Then using (7.6) it is easy to obtain integrable functions $\rho_i(t^{(1)})$ $i = 1, 2$, such that for z in $F(\theta_0)$ and $t^{(1)}$ in $A(M, n)$ we have

$$H_i(t^{(1)}, z^{(n)}) \leqq \rho_i(t^{(1)}) \qquad i = 1, 2 .$$

Then, since according to (7.9) we have $\lim_{n\to\infty} H_i(t^{(1)}, z^{(n)}) = 0$; $i = 1, 2$ for z not in $F(\theta_0)$ we get by dominated convergence, (7.10) and (7.11). □

The following two lemmas show the asymptotic behavior of the posterior risk when we use a simple subdesign in which each experiment is taken an infinite number of times.

Lemma 7.2. *Assume the same conditions as in Theorem* 7.1. *and* B3.1, B3.3. *Let* (h_{in}) $1 \leqq i \leqq k$, $1 \leqq n < \infty$ *be a simple subdesign such that*

$$\text{(7.12)}\qquad \lim_{n\to\infty} h_{in} = \infty \qquad \textit{for all } i,$$

and

$$\text{(7.13)}\qquad \lim_{n\to\infty} h_{in}/h_{i'n} = \mu_{ii'} , \qquad \textit{for all } i \textit{ and } i' .$$

($\mu_{ii'}$ *may be* ∞.) *Suppose* $\mu_{i_0 i} > \infty$ *for all* i. *Then for* z *not in* $F(\theta_0)$ *we have*

$$\lim_{n\to\infty} h_{i_0 n} \operatorname{Var}(g(\theta) \mid z^{(n)}) = \textstyle\sum_{i=1}^k \mu_{i_0 i} (\operatorname{grad}_i g(\theta_0))' (A_i(\theta_{i0}))^{-1} \operatorname{grad}_i g(\theta_0) .$$

$F(\theta_0)$ *is the same as in Theorem* 7.1.

Proof. It is enough to show that for z not in $F(\theta_0)$ we have

$$\text{(7.14)}\qquad \lim_{n\to\infty} h_{i_0 n} E^2(g(\theta) - g(\hat{\theta}_n^*) \mid z^{(n)}) = 0$$

and

$$(7.15)\qquad \lim_{n\to\infty} h_{i_0n} E((g(\theta) - g(\hat{\theta}_n^*))^2 \mid z^{(n)}) = \sum_{i=1}^k \mu_{i_0i}(\mathrm{grad}_i\, g(\theta_0))'(A_i(\theta_{i0}))^{-1}\, \mathrm{grad}_i\, g(\theta_0)\,.$$

(7.14) and (7.15) follows from Theorem 7.1 using a standard linear expansion for $g(\theta) - g(\theta_n^*)$. For details see Yohai [9]. □

LEMMA 7.3. *Assume the same conditions as in Lemma* 7.2. *Let* (h_{in}) $1 \leqq i \leqq k$, $1 \leqq n < \infty$ *be a simple subdesign satisfying* (7.12) *and* (7.13). *Put* $a(n) = \sum_{i=1}^k h_{in}$, *and* $\lim_{n\to\infty} h_{in}/a(n) = \lambda_i = 1/\sum_{j=1}^k \mu_{ji}$. *Then we have*

(i) *If* $\lambda_i > 0$ *for all i then*

$$\lim_{n\to\infty} a(n)\,\mathrm{Var}\,(g(\theta) \mid z^{(n)}) = \sum_{i=1}^k (1/\lambda_i)(\mathrm{grad}_i\, g(\theta_0))'(A_i(\theta_{i0}))^{-1}\, \mathrm{grad}_i\, g(\theta_0)$$

for z not in $F(\theta_0)$.

(ii) *Assume* B3.2 *too, then if* $\lambda_i = 0$ *for some i it follows that*

$$\lim_{n\to\infty} a(n)\,\mathrm{Var}\,(g(\theta) \mid z^{(n)}) = \infty \qquad \text{for } z \text{ not in } F(\theta_0)\,.$$

The following lemma shows that if we use a simple subdesign in which at least one of the experiments is taken a finite number of times, the posterior risk does not converge to 0.

LEMMA 7.4. *Assume the same conditions as in Lemma* 7.2 *and* B3.2. *Let* (h_{in}) $1 \leqq i \leqq k$, $1 \leqq n < \infty$ *be a simple subdesign such that* $\lim_{n\to\infty} h_{in} < \infty$ *for some i. Then for z not in* $F(\theta_0)$ *we have* $\lim_{n\to\infty} \mathrm{Var}\,(g(\theta) \mid z^{(n)}) > 0$.

PROOF. Suppose for example that $\lim_{n\to\infty} h_{in} = \infty$ for $i = 1, 2, \cdots, s$ and $\lim_{n\to\infty} h_{in} < \infty$ for $i = s+1, \cdots, k$. Then using Theorem 7.1 it is easy to show that

$$(7.16)\qquad \lim_{n\to\infty} \mathrm{Var}\,(g(\theta) \mid z^{(n)}) = \mathrm{Var}\,(g(\theta_{10}, \cdots, \theta_{s0}, \theta_{s+1}, \cdots, \theta_k) \mid z^{(0)}, \theta_1 = \theta_{10}, \cdots, \theta_s = \theta_{s0})$$

for z not in $F(\theta_0)$, where $z^{(0)}$ is defined as in Theorem 7.1.

Clearly B3.2 implies

$$\mathrm{Var}\,(g(\theta_{10}, \cdots, \theta_{s0}, \theta_{s+1}, \cdots, \theta_k) \mid z^{(0)}, \theta_1 = \theta_{10}, \cdots, \theta_s = \theta_{s0}) > 0\,. \qquad \square$$

PROOF OF THEOREM 3.1. From (3.4) we have

$$(7.17)\qquad \inf_{d\in D} Y_{d,n} = \text{minimum}\,\{\mathrm{Var}\,(g(\theta) \mid z_{ij}, 1 \leqq j \leqq s_{in}, 1 \leqq i \leqq k):\ s_{in} \geqq 0,\ \textstyle\sum_{i=1}^k s_{in} = n\}\,.$$

Then $\inf_{d\in D} Y_{d,n}$ is random variable and (i) holds.

In order to prove (ii), according to (7.17) it is enough to show that

$$P(\mathrm{Var}\,(g(\theta) \mid z_{ij}, 1 \leqq j \leqq s_{in}, 1 \leqq i \leqq k) > 0) = 1$$

for any set of nonnegative integers (s_{in}) $1 \leqq i \leqq k$ such that $\sum_{i=1}^k s_{in} = n$. But this follows from the fact that the posterior distribution of θ given z_{ij}, $1 \leqq j \leqq s_{in}$,

$1 \leqq i \leqq k$ is absolutely continuous with respect to the Lebesgue measure and the assumption B3.2.

In order to prove (iii) it suffices to prove that for any z not in $F(\theta_0)$ any strictly increasing sequence n_t has a subsequence n_t' such that

$$\liminf_{t\to\infty} \inf_{d\in D} n_t' Y_{d,n_t'} \geqq V(\theta_0) . \tag{7.18}$$

Fix z not in $F(\theta_0)$. Consider for each n the nonnegative integers (s_{in}^*) $1 \leqq j \leqq k$ such that minimize $\operatorname{Var}(g(\theta) \mid z_{ij}, 1 \leqq j \leqq s_{in}, 1 \leqq i \leqq k)$ subject to $\sum_{i=1}^k s_{in} = n$. Then clearly

$$\inf_{d\in D} nY_{d,n} = n \operatorname{Var}(g(\theta) \mid z_{ij}, 1 \leqq j \leqq s_{in}^*, 1 \leqq i \leqq k) . \tag{7.19}$$

It is easy to show that any strictly increasing sequence n_t has a subsequence n_t' such that if we put $h_{it} = s_{in_t'}^*$, then (h_{it}) $1 \leqq i \leqq k$, $1 \leqq t < \infty$ is a simple subdesign satisfying (7.13). Then if $h_{in} \to \infty$ as $n \to \infty$ for all i, from Lemma 7.3 and (7.19) we get (7.18). If $\lim_{n\to\infty} h_{in} < \infty$ for some i, using Lemma 7.4 we get

$$\lim_{n\to\infty} \inf_{d\in D} n_t' Y_{d,n_t'} = \infty \qquad \text{for } z \text{ not in } F(\theta_0) .$$

(iv) follows immediately from Lemma 7.3 and the fact that $P_{\theta_0}(F(\theta_0)) = 0$. □

Acknowledgment. The author is grateful to Mr. P. J. Bickel for having suggested the problem.

REFERENCES

[1] Bickel, P. J. and Yahav, J. A. (1967). Asymptotically pointwise optimal procedures in sequential analysis. *Proc. Fifth Berkeley Symp. Math. Statist. Prob.* Univ. of California Press, **1** 401–413.

[2] Bickel, P. J. and Yahav, J. A. (1968). Asymptotically optimal Bayes and minimax procedures in sequential estimation. *Ann. Math. Statist.* **39** 442–456.

[3] Bickel, P. J. and Yahav, J. A. (1969). Some contributions to the asymptotic theory of Bayes solutions. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **11** 257–276.

[4] Chernoff, H. (1959). Sequential design of experiments. *Ann. Math. Statist.* **30** 755–770.

[5] Kiefer, J. and Sacks, J. (1963). Asymptotically optimal sequential inference and design. *Ann. Math. Statist.* **34** 705–750.

[6] Loève, M. (1955). *Probability Theory.* Van Nostrand, Princeton.

[7] Wald, A. (1949). A note on the consistency of the maximum likelihood estimate. *Ann. Math. Statist.* **20** 595–601.

[8] Wald, A. (1951). Asymptotic minimax solution of sequential point estimation problems. *Proc. Second Berkeley Symp. Math. Statist Prob.* Univ. of California Press, 1–11.

[9] Yohai, V. J. (1971). Asymptotically optimal Bayes sequential design of experiments for estimation. Notas de Matemática No. 19, Univ. de La Plata.

Carlos Pellegrini 743 (8P 37)
Buenos Aires, Argentina