The Annals of Statistics
1976, Vol. 4, No. 4, 685–711

ON ASYMPTOTICALLY OPTIMAL SEQUENTIAL BAYES INTERVAL ESTIMATION PROCEDURES[1]

BY LEON JAY GLESER AND SUDHAKAR KUNTE

Purdue University and University of Poona

A theory of sequential Bayes interval estimation procedures for a single parameter is developed for the case where the loss for using an interval I is a linear combination of the length of I, the indicator of noncoverage of I, and the number of observations taken. A class of stopping rules $\{t(c): c > 0\}$ is shown to be asymptotically pointwise optimal (A.P.O.) and asymptotically optimal (A.O.) for the confidence interval problem as the cost c per observation tends to 0. The results require generalization of Bickel and Yahav's (1968) general conditions for the existence of A.P.O. and A.O. stopping rules to the case where the terminal risk Y_n satisfies $f(n)Y_n \to V$ for $f(n)$ a regularly varying function on the integers.

1. Introduction and summary. There has been much recent interest in decision-theoretic approaches to fixed-sample confidence interval estimation [for example, Winkler (1972), Cohen and Strawderman (1973), and Joshi (1969)]. Fixed-width, fixed-confidence, sequential confidence interval estimation has also received attention in papers by Chow and Robbins (1965), Sen and Ghosh (1967), Paulson (1969), and Weiss and Wolfowitz (1972), among numerous others. Along with all fixed-sample designs, the fixed-sample confidence interval approach has the disadvantage of potentially inefficient use of data. On the other hand, the sequential fixed-width, fixed-confidence approach involves choices for the width and confidence of the intervals which may not correspond to any rational balancing of the merits of high confidence as opposed to narrow intervals.

In the present paper, we develop asymptotically optimal sequential Bayes interval estimation procedures in the case where the loss is a linear combination of the length of the interval, the indicator function for noncoverage, and the sample size. Apart from the cost of sampling, this loss function is the one considered by Joshi (1969) for fixed-sample inference on the mean of a normal distribution, and is a special case of the loss function of Cohen and Strawderman (1973). Our loss function has been used for sequential interval estimation of

Received January 1975; revised January 1976.

[1] Research sponsored in part by the Air Force Office of Scientific Research, Air Force Systems Command, USAF, under Grants AFOSR-73-2432 and AFOSR-72-2350 at Purdue University. The U.S. government is authorized to reproduce and distribute reprints for government purposes.

AMS 1970 *subject classifications.* Primary 62L12, 62L15, 62F15, 60G40; Secondary 62F10, 62F25, 60F99.

Key words and phrases. Bayes interval estimation procedures, sequential inference, stopping rules, asymptotically pointwise optimal, asymptotically optimal, moderate deviation theory, regularly varying function.

the positive mean of a folded normal distribution with known variance by Blumenthal (1970), and for sequential interval estimation of the mean of a normal distribution with both known and unknown variance by Kunte (1973). Kunte (1973) also gave some theoretical results on sequential Bayes interval estimators for a parameter of a general distribution. These results are completed and generalized in the present paper.

In Section 2, the fixed-sample Bayes interval estimation problem is solved in its entirety—including representations for the optimal interval and for the posterior Bayes risk Y_n based on n observations. In the sequential case, the Bayes terminal decision rule for a given stopping rule t is easily seen to be "when $t = n$, use the fixed-sample Bayes rule based on n observations," but discovery of the explicit form of the Bayes optimal stopping rule appears intractible, prompting us to turn to asymptotic methods.

We assume that the cost, c, per observation tends to 0 and adopt an approach parallel to that previously followed by Kiefer and Sacks (1963) and Bickel and Yahav (1967, 1968, 1969a) in the cases of sequential Bayesian rules for hypothesis testing and point estimation. That is, we determine a collection of stopping rules $\{t(c) : c > 0\}$ which is *asymptotically pointwise optimal* (A.P.O.) *and asymptotically optimal* (A.O.) for our Bayesian sequential interval estimation problem. Bickel and Yahav (1967, 1968) have given sufficient conditions for stopping rules to be A.P.O. and A.O. Their regularity conditions include the condition that for some $\beta > 0$,

$$n^{\beta} Y_n \to V\,, \quad \text{a.s.},$$

where V is a positive random variable, and Y_n is the posterior risk of terminal decision of the Bayes rule based on n observations. The results of Blumenthal (1970, Lemma 4.1) and Kunte (1973), however, suggest that in our interval estimation problem,

$$(n/\log n)^{\frac{1}{2}} Y_n \to V\,, \quad \text{a.s.}$$

Thus, in Sections 3 and 4, we extend the results of Bickel and Yahav to cover both this rate of convergence, and other similar rates of convergence. The context of these sections are abstract and, except for the main results (Theorems 3.1 and 4.1), can be skipped by readers only interested in the interval estimation problem.

In Sections 5 and 6, we return to the interval estimation problem, and determine a collection of stopping rules $\{t(c) : c > 0\}$ which is both A.P.O. and A.O. Although the conditions which we adopt for proving asymptotic pointwise optimality (Section 5) and asymptotic optimality (Section 6) closely parallel the conditions in Bickel and Yahav (1968, 1969b), and although we make use of some of their results, our large-sample analysis bears the same relationship to their work as the theory of moderate deviations [Rubin and Sethuraman (1965a, 1965b)] bears to the theory of large deviations.

2. Bayesian sequential interval estimation. Suppose that we observe independent, identically distributed random variables $X_1, X_2 \cdots$ whose common probability measure P_θ belongs to a family $\{P_\theta : \theta \in \Theta\}$ of measures defined on a measure space $(\mathscr{X}, \mathscr{B})$, indexed by an open subinterval Θ of the real line, and dominated by a σ-finite measure μ. Let $f(x \mid \theta) = dP_\theta/d\mu$ be the density function of P_θ with respect to μ, and let $\psi(\theta)$ be a prior density (with respect to Lebesgue measure) on Θ.

For our terminal action space, we take the class $\mathscr{I}$ of all subintervals (including point sets and the empty set) of Θ. For our loss function, we choose a linear combination of the length of our interval, the indicator of noncoverage of our interval, and the cost of sampling. That is, if n observations have been observed, the loss for choosing $I \in \mathscr{I}$ when θ is the true parameter is

$$L(\theta, I, n) = al(I) + b(1 - \delta_I(\theta)) + cn , \tag{2.1}$$

where a, b, c are finite positive constants, $l(I)$ is the length of I, and

$$\begin{aligned} \delta_I(\theta) &= 1 , \qquad \text{if} \quad \theta \in I \\ &= 0 , \qquad \text{otherwise.} \end{aligned}$$

Our decisions are pairs (I_t, t), where t is a stopping rule, and I_t is a terminal decision rule. It follows directly from the work of Arrow, Blackwell and Girshick (1949) that for every stopping rule t, the Bayes optimal terminal decision rule $I_t{}^*$, when $t = n$, is the fixed-sample Bayes estimation procedure based on n observations. This fixed-sample rule is determined as follows.

Let us first note that we could have started with a wider action space—namely, the class $\mathscr{C}$ of all Lebesgue measurable subsets of Θ—and a loss function

$$L(\theta, C, n) = al(C) + b(1 - \delta_C(\theta)) + cn ,$$

where $l(C)$ is the Lebesgue measure of C and $\delta_C(\theta) = 1$ if $\theta \in C$ and 0 otherwise, $C \in \mathscr{C}$. Let $\mathbf{X}_n = (X_1, X_2, \cdots, X_n)$ represent the vector of sample observations. A *regional estimation* procedure, $C(\mathbf{X}_n)$, is then a measurable assignment of regions $C \in \mathscr{C}$ to samples $\mathbf{X}_n$. There is at least one procedure with finite Bayes risk (namely, $C(\mathbf{X}_n) \equiv$ empty set, all $\mathbf{X}_n$). Hence, in finding the Bayes procedure $C^*(\mathbf{X}_n)$ against the prior ψ, we need consider only procedures with finite Bayes risk. For any such procedure, application of the Fubini–Tonelli theorem [cf. Royden (1968, Theorem 20)] yields

$$R(\psi, C(\mathbf{X}_n)) = \int_{\mathscr{X}^n} [\rho(\psi, C(\mathbf{X}_n)) + cn] \, dG_n(\mathbf{X}_n) , \tag{2.2}$$

where

$$\begin{aligned} \rho(\psi, C(\mathbf{X}_n)) &= al(C(\mathbf{X}_n)) + b[1 - \textstyle\int_{C(\mathbf{X}_n)} \psi(\theta \mid \mathbf{X}_n) \, d\theta] \\ &= b[1 + \textstyle\int_{C(\mathbf{X}_n)} (ab^{-1} - \psi(\theta \mid \mathbf{X}_n)) \, d\theta] \end{aligned} \tag{2.3}$$

is the posterior Bayes risk of terminal decision, $\psi(\theta \mid \mathbf{X}_n)$ is the posterior density of θ given $\mathbf{X}_n$, and $G_n(\mathbf{X}_n)$ is the marginal distribution function of $\mathbf{X}_n$. From

(2.2) and (2.3), a Bayes regional estimation procedure $C^*(\mathbf{X}_n)$ against ψ based on n observations is

$$C^*(\mathbf{X}_n) = \text{closure (in } \Theta) \quad \text{of} \quad \{\theta : \psi(\theta \mid \mathbf{X}_n) \geqq ab^{-1}\}, \tag{2.4}$$

since (2.4) clearly minimizes (2.3) and can be shown to be a measurable assignment of subsets of Θ to samples $\mathbf{X}_n$.

If $C^*(\mathbf{X}_n)$ is an interval $I^*(\mathbf{X}_n)$ for each $\mathbf{X}_n$, then $I^*(\mathbf{X}_n)$ provides us with our Bayes interval estimation procedure based on n observations. If $C^*(\mathbf{X}_n)$ is not always an interval, then our search for an *interval* estimation procedure can be justified only on the grounds of convenience, since any Bayes interval estimation procedure (i.e., a procedure which minimizes (2.2) over the subclass of procedures which assign only intervals in Θ to samples $\mathbf{X}_n$) will be improved upon by $C^*(\mathbf{X}_n)$. Hence, in the remainder of this paper we make the assumption:

A.O. The region $C^*(\mathbf{X}_n)$ defined by (2.4) is a closed (in Θ) subinterval $I^*(\mathbf{X}_n) \equiv [\alpha_{1n}^*(\mathbf{X}_n), \alpha_{2n}^*(\mathbf{X}_n)]$ of Θ for almost all $\mathbf{X}_n$, where "almost all" is defined by the measure corresponding to $G_n(\mathbf{X}_n)$.

We note that it is possible for $I^*(\mathbf{X}_n)$ to be empty (when $\psi(\theta \mid \mathbf{X}_n) < ab^{-1}$, all $\theta \in \Theta$) or a one-point set. In both of these cases, $\rho(\psi, I^*(\mathbf{X}_n)) = b$. In general,

$$\begin{aligned} \alpha_{1n}^*(\mathbf{X}_n) &= \inf \{\theta : \psi(\theta \mid \mathbf{X}_n) \geqq ab^{-1}\}, \\ \alpha_{2n}^*(\mathbf{X}_n) &= \sup \{\theta : \psi(\theta \mid \mathbf{X}_n) \geqq ab^{-1}\}, \end{aligned} \tag{2.5}$$

and

$$\begin{aligned} Y_n &\equiv \rho(\psi, I^*(\mathbf{X}_n)) \\ &= a(\alpha_{1n}^*(\mathbf{X}_n) - \alpha_{2n}^*(\mathbf{X}_n)) + b(1 - \textstyle\int_{\alpha_{1n}^*(\mathbf{X}_n)}^{\alpha_{2n}^*(\mathbf{X}_n)} \psi(\theta \mid \mathbf{X}_n)\, d\theta). \end{aligned} \tag{2.6}$$

EXAMPLE. (Blumenthal (1970), Kunte (1973)). If $X_1, X_2, \cdots, X_n$ are i.i.d. $N(\theta, 1)$, and θ has a normal prior distribution with mean μ and variance σ^2, then the posterior density $\psi(\theta \mid \mathbf{X}_n)$ is that of a normal distribution with mean

$$\hat{\theta}_n(\mathbf{X}_n) = (\textstyle\sum_{i=1}^n X_i + \sigma^{-2}\mu)/(n + \sigma^{-2})$$

and variance $n^{-1}(1 + \varepsilon_n)$, where $\varepsilon_n = -(n\sigma^2 + 1)^{-1}$. The Bayes interval estimation procedure is determined by the endpoints

$$\alpha_{in}^*(\mathbf{X}_n) = \hat{\theta}_n(\mathbf{X}_n) + (-1)^i n^{-\frac{1}{2}}(1 + \varepsilon_n)^{\frac{1}{2}} \log^{\frac{1}{2}} (nb^2/2\pi a^2(1 + \varepsilon_n)), \quad i = 1, 2. \tag{2.7}$$

For future use, we note that in this case

$$(n/\log n)^{\frac{1}{2}} Y_n \to 2a, \quad \text{a.s.}, \tag{2.8}$$

as can be seen from (2.6), (2.7) and the well-known inequality

$$\textstyle\int_x^\infty e^{-\frac{1}{2}t^2}\, dt = \int_{-\infty}^{-x} e^{-\frac{1}{2}t^2}\, dt \leqq x^{-1} e^{-\frac{1}{2}x^2}, \qquad x > 0. \tag{2.9}$$

Let us now return to the sequential Bayes interval estimation problem. We have seen that for any stopping rule t, the Bayes risk of terminal decision is minimized by choosing the interval $I^*(\mathbf{X}_n)$ defined in (2.5) whenever sampling terminates with $t = n$. The Bayes risk for any stopping rule t, using the optimal terminal decision $I^*(\mathbf{X}_t)$, is given by

$$(2.10) \qquad R(\psi, I^*(\mathbf{X}_t), t) = \sum_{n=0}^{\infty} \int_{\{t=n\}} [\rho(\psi, I^*(\mathbf{X}_n)) + cn]\, dG_n(\mathbf{X}_n) = E(Y_t + ct)\,,$$

where the expectation of $Y_t + ct$ is taken over the joint distribution of θ and $X_1, X_2, \cdots$. We must now seek a stopping rule t^* which minimizes (2.10), and then we will have found the Bayes sequential interval estimation procedure $(I^*(\mathbf{X}_{t^*}), t^*)$. This minimization problem appears to be intractible in general,[2] so we instead assume that the cost c of sampling is very small ($c \to 0$) and look for asymptotically pointwise optimal (A.P.O.) and asymptotically optimal (A.O.) stopping rules in the sense defined in Bickel and Yahav (1967, 1968) and in the next section (Section 3). To date, the most general set of sufficient conditions used to find A.P.O. and A.O. rules are those given by Bickel and Yahav (1968). Their sufficient conditions assume that

$$n^\beta Y_n \to V\,, \quad \text{a.s.},$$

for some $\beta > 0$, and some positive random variable V, whereas we have seen in our example (see equation (2.8)) that a function $(n/\log n)^{\frac{1}{2}}$ not of the form n^β is needed. Thus, in Sections 3 and 4, we generalize Bickel and Yahav's results to cover cases where $f(n)Y_n \to V$, a.s., where $f(x)$ is a member of a certain general class of functions containing x^β and $(x/\log x)^{\frac{1}{2}}$ as special cases. Because this generalization may have applications beyond the present problem, our results in Sections 3 and 4 are stated in terms of the general abstract framework introduced in Bickel and Yahav (1968). Once we have the necessary generalization of the results of Bickel and Yahav, we return in Sections 5 and 6 to derivation of the A.P.O. and A.O. stopping rules for our interval estimation problem.

3. A.P.O. stopping rules. Let $\{Y_n\}$ be a sequence of random variables defined on a probability space $(\Omega, \mathscr{F}, P)$, where Y_n is $\mathscr{F}_n$-measurable and $\mathscr{F}_1 \subset \mathscr{F}_2 \subset \cdots \subset \mathscr{F}$ is an increasing sequence of sub σ-fields. Let $K(x)$ be a strictly increasing positive function of $x \geqq 0$ satisfying

$$(3.1) \qquad \lim_{x\to\infty} K(x) = \infty\,.$$

Let T denote the class of all proper stopping variables t defined on $(\Omega, \mathscr{F}, P)$, where a proper stopping variable t is a natural-number-valued random variable for which the event $\{t = n\} \in \mathscr{F}_n$, $n = 1, 2, \cdots$, and $P(t < \infty) = 1$. Define

$$(3.2) \qquad X(n, c) = Y_n + cK(n)\,.$$

The goal is to choose $t \in T$ to minimize $E(X(t, c))$.

[2] In our example of estimating the mean of a normal distribution, however, the optimal sequential procedure is known, and has a fixed sample size.

Following Bickel and Yahav (1968), we say that a class $\{t(c): c > 0\}$ of stopping variables is *asymptotically pointwise optimal* (A.P.O.) if

$$P\left\{\lim_{c\to 0} \frac{X(t(c), c)}{\inf_{s\in T} X(s, c)} = 1\right\} = 1 , \tag{3.3}$$

and is *asymptotically optimal* (A.O.) if

$$\lim_{c\to 0}\left[\frac{E(X(t(c), c))}{\inf_{s\in T} E(X(s, c))}\right] = 1 . \tag{3.4}$$

We make the following assumptions:

B.0. $P\{Y_n > 0\} = 1$, all n, and $P\{\lim_{n\to\infty} Y_n = 0\} \doteq 1$.

B.1. There exists a strictly increasing, positive function $f(x)$ defined on $[0, \infty)$ and an almost surely positive random variable V such that

$$P\{\lim_{n\to\infty} f(n) Y_n = V\} = 1 .$$

B.2. For each $x > 0$ and $c > 0$, there exists an integer $N(x, c)$ which minimizes the function

$$h(x, c, n) = (f(n))^{-1}x + cK(n) . \tag{3.5}$$

Further, $N(x, c)$ may be taken as the first integer n such that $\Delta h(x, c, n) = h(x, c, n+1) - h(x, c, n) \geqq 0$.

B.3. The function

$$G(x) = \frac{K(x+1)(f(x+1) - f(x))}{f(x)(K(x+1) - K(x))} \tag{3.6}$$

is bounded, and

$$\lim_{x\to\infty} G(x) = M , \qquad 0 \leqq M < \infty . \tag{3.7}$$

B.4. Either $f(x)/f(x+1) \to 1$ or $K(x+1)/K(x) \to 1$ as $x \to \infty$.

Under these assumptions, we have the following result.

THEOREM 3.1. *For each* $c > 0$, *let*

$$t(c) = \textit{first} \quad n \geqq 1 \quad \textit{such that} \quad \left(1 - \frac{f(n)}{f(n+1)}\right) Y_n \leqq c\,\Delta K(n) , \tag{3.8}$$

where $\Delta K(n) = K(n+1) - K(n)$. *Then the class of stopping rules* $\{t(c): c > 0\}$ *is* A.P.O.

In order to prove Theorem 3.1, we need some preliminary results. Let

$$H(x) = G(x)/f(x+1)K(x+1) , \tag{3.9}$$

and note that

$$t(c) = \text{first} \quad n \geqq 1 \quad \text{such that} \quad H(n)f(n)Y_n \leqq c . \tag{3.10}$$

By using (3.6) to rewrite $H(x)$ solely in terms of $f(x)$ and $K(x)$, and using the fact that $f(x)$ and $K(x)$ are strictly increasing, we see that $H(x) > 0$ for all x. Also by (3.1), assumption B.3, and (3.9), we have

$$\lim_{x\to\infty} H(x) = 0 . \tag{3.11}$$

LEMMA 3.1. *For each $c > 0$, $t(c)$ is a proper stopping rule; that is $P\{t(c) < \infty\} = 1$. Also,*

$$P\{\lim_{c\to 0} t(c) = \infty\} = 1 , \tag{3.12}$$

$$P\{\lim_{c\to 0} f(t(c))Y_{t(c)} = V\} = 1 , \tag{3.13}$$

$$P\{\lim_{c\to 0} cK(t(c))f(t(c)) = MV\} = 1 . \tag{3.14}$$

PROOF. From assumption B.1 and (3.11),

$$P\{\lim_{n\to\infty} H(n)f(n)Y_n = 0\} = 1 . \tag{3.15}$$

Hence, since $c > 0$, the inequality defining $t(c)$ in (3.10) is a.s. satisfied by some $n < \infty$, thus proving that $t(c)$ is proper. On the other hand, $H(n)f(n)Y_n > 0$, a.s., for all $n \geqq 1$, as can be seen from assumptions B.0 and B.1 and the fact that $H(x) > 0$. It therefore follows from (3.10) and (3.15) that (3.12) holds. The result (3.13) now follows from (3.12) and assumption B.1. Finally, it follows from the definition (3.10) of $t \equiv t(c)$ that

$$\left(\frac{K(t)f(t)}{K(t+1)f(t+1)}\right) G(t)f(t)Y_t \leqq cK(t)f(t) \leqq G(t-1)f(t-1)Y_{t-1} . \tag{3.16}$$

Also, (3.12), (3.13) and assumption B.3 imply that

$$\lim_{c\to 0} G(t)f(t)Y_t = \lim_{c\to 0} G(t-1)f(t-1)Y_{t-1} = MV , \quad \text{a.s.} \tag{3.17}$$

When $M > 0$, we can use assumptions B.3 and B.4 to show that $\lim_{x\to\infty} [K(x)f(x)/K(x+1)f(x+1)] = 1$. Use of this result, (3.12), and (3.17) in (3.16) establishes (3.14). When $M = 0$, the right-hand inequality in (3.16), (3.17), and the fact that $cK(t)f(t) \geqq 0$ show that (3.14) holds. □

Let

$$n^*(c) = \text{first} \quad n \geqq 1 \quad \text{such that} \quad \Delta h(f(t(c))Y_{t(c)}, c, n) \geqq 0 . \tag{3.18}$$

It follows from assumptions B.0—B.2 that

$$h(f(t(c))Y_{t(c)}, c, n^*(c)) = \min_n h(f(t(c))Y_{t(c)}, c, n) . \tag{3.19}$$

Using the definition (3.9) of $H(x)$, we can also define $n^*(c)$ equivalently by

$$n^*(c) = \text{first} \quad n \geqq 1 \quad \text{such that} \quad H(n)f(t(c))Y_{t(c)} \leqq c . \tag{3.20}$$

By using arguments similar to those used to prove Lemma 3.1, we can show that

$$P(\lim_{c\to 0} n^*(c) = \infty) = P\{\lim_{c\to 0} f(n^*(c))Y_{n^*(c)} = V\} = 1 . \tag{3.21}$$

$$P\{\lim_{c\to 0} cK(n^*(c))f(n^*(c)) = MV\} = 1 . \tag{3.22}$$

We need to define one additional random variable. Let

(3.23) $$n_0(c) = \text{first} \quad n \geqq 1 \quad \text{satisfying} \quad X(n, c) = \min_k X(k, c) .$$

The a.s. existence of $n_0(c)$ follows from (3.1) and assumption B.0. Also

(3.24) $$P\{\lim_{c\to 0} n_0(c) = \infty\} = P\{\lim_{c\to 0} f(n_0(c)) Y_{n_0(c)} = V\} = 1 .$$

The proof of this result, which follows from assumptions B.0—B.2, is given in Gleser and Kunte (1973, Lemma 2.3), and will not be repeated here.

Comparing (3.10) and (3.20), we see that $n^*(c) \leqq t(c)$, a.s. Hence, since $f(x)$ is increasing in x, it follows that $f(n^*(c))/f(t(c)) \leqq 1$, a.s. Thus, letting $t \equiv t(c)$, $n^* \equiv n^*(c)$, and making use of (3.2) and (3.19),

$$\frac{X(t, c)}{\inf_n h(f(t)Y_t, c, n)} \leqq \frac{f(t)Y_t + cK(t)f(t)}{f(t)Y_t + cK(n^*)f(n^*)} .$$

Taking $\limsup_{c\to 0}$ on both sides of this inequality, we conclude from (3.13), (3.14) and (3.22) that

(3.25) $$P\left\{\limsup_{c\to 0} \frac{X(t(c), c)}{\inf_n h(f(t(c))Y_{t(c)}, c, n)} \leqq 1\right\} = 1 .$$

Similarly, with $t \equiv t(c)$, $n_0 \equiv n_0(c)$,

(3.26) $$\frac{h(f(t)Y_t, c, n_0)}{X(n_0, c)} = \frac{f(t)Y_t + cK(n_0)f(n_0)}{f(n_0)Y_{n_0} + cK(n_0)f(n_0)} .$$

Noting that $cK(n_0)f(n_0) \geqq 0$ and thus is bounded away from $-V$ for all $c > 0$, it follows from (3.13), (3.24) and (3.26) that

(3.27) $$P\left\{\lim_{c\to 0} \frac{h(f(t)Y_t, c, n_0)}{X(n_0, c)} = 1\right\} = 1 .$$

Since by Lemma 3.1, $t(c)$ is a proper stopping rule, we know that $X(t(c), c) \geqq \inf_{s\in T} X(s, c)$, a.s. Thus, to prove that (3.3) holds, and hence verify Theorem 3.1, it is enough to show that

(3.28) $$P\left\{\limsup_{c\to 0} \frac{X(t(c), c)}{X(n_0(c), c)} \leqq 1\right\} = 1 ,$$

since $X(n_0(c), c) \leqq \inf_{s\in T} X(s, c)$. However, letting $t \equiv t(c)$, $n_0 \equiv n_0(c)$,

$$\frac{X(t, c)}{X(n_0, c)} \leqq \frac{X(t, c)}{\inf_n h(f(t)Y_t, c, n)} \frac{h(f(t)Y_t, c, n_0)}{X(n_0, c)} ,$$

and (3.28) follows directly from (3.25) and (3.27). This completes the proof of Theorem 3.1.

REMARK I. Bickel and Yahav (1968, Theorem 2.1) prove our Theorem 3.1 for the case $f(x) = x^\beta$, $\beta > 0$. Our assumptions B.0—B.2 directly generalize the corresponding assumptions [(2.1)—(2.3), A.1, A.2] in their paper, but their assumption A.3 is less restrictive than our assumptions B.3 and B.4 [which in the case $f(x) = x^\beta$ are satisfied if $K(x+1)/x(K(x+1) - K(x)) \to M < \infty$, as $x \to \infty$]. On the other hand, there is a major gap in Bickel and Yahav's proof. In place of our equation (3.25), they claim that

$$\min_n h(f(t)Y_t, c, n) = X(t, c)\,, \quad t \equiv t(c)\,. \tag{3.29}$$

If $h(x, c, n)$ is strictly convex in n (as it is, for example, when $f(x) = x^\beta$, $K(x) = x$), then (3.29) can hold only when $t(c) = n^*(c)$, a.s. However, as we noted in the discussion leading to (3.25), the best that can be inferred from (3.10) and (3.20) is that $n^*(c) \leqq t(c)$, a.s. We remark that Bickel and Yahav have noted the gap in their proof (personal communication from Professor Bickel), and have suggested assumptions and a proof which, in the case $f(x) = x^\beta$, closely resemble those for our Theorem 3.1. Whether their original assumptions are sufficient to prove the theorem is still an open question.

REMARK II. In applications, assumption B.2 is likely to be the hardest to verify. It thus may be of some help to note that assumption B.2 holds if $1/f(x)$ and $K(x)$ are both convex on the integers.

4. A.O. stopping rules. In the special case where $K(x) = x$, we now consider sufficient conditions for the class $\{t(c): c > 0\}$ of stopping rules defined in Theorem 3.1 to be asymptotically optimal. Note that when $K(x) = x$,

$$G(x) = (x+1)\left(\frac{f(x+1)}{f(x)} - 1\right), \tag{4.1}$$

and assumption B.4 is trivially satisfied. We noted in the proof of Lemma 3.1 that when $\lim_{x\to\infty} G(x) = M > 0$, then $\lim_{x\to\infty} K(x+1)f(x+1)/K(x)f(x) = 1$. When $K(x) = x$, this last result implies that $\lim_{x\to\infty} f(x+1)/f(x) = 1$. Since we need this consequence of $M > 0$ in our proof of asymptotic optimality, we replace assumption B.3 by the following.

B.3′. The function $G(x)$ defined by (4.1) is bounded and

$$\lim_{x\to\infty} G(x) = M\,, \qquad 0 < M < \infty\,.$$

THEOREM 4.1. *When* $K(x) = x$ *and assumptions* B.0—B.2 *and* B.3′ *hold, the class* $\{t(c): c > 0\}$ *of stopping rules defined by*

$$t(c) = \textit{first} \quad n \geqq 1 \quad \textit{such that} \quad \left(1 - \frac{f(n)}{f(n+1)}\right) Y_n \leqq c \tag{4.2}$$

is A.P.O. *Further, if*

$$\sup_n E(f(n)Y_n) < \infty\,, \tag{4.3}$$

this class of stopping rules is A.O.

The class $\{t(c): c > 0\}$ defined by (4.2) is just the class of rules defined in Theorem 3.1 specialized to the case $K(x) = x$. Since $K(x) = x$ and assumptions B.0—B.2 and B.3′ imply assumptions B.0—B.4, the A.P.O. character of the stopping rules (4.2) follows from Theorem 3.1.

To prove that the class of rules (4.2) is A.O., we proceed in a series of lemmas. We begin by noting some consequences of assumption B.3′.

LEMMA 4.1. *If assumption* B.3′ *holds, then*

(a) $\lim_{x\to\infty} f(x+1)/f(x) = 1$,
(b) $\sum_{n=1}^{\infty} (1/nf(n)) < \infty$,
(c) *there exists an increasing positive function* $h(x)$ *such that for every* $x \in [0, \infty)$,

$$\liminf_{y\to\infty} \frac{f(xy)}{f(x)h(y)} \geqq 1$$

and

$$\sum_{n=1}^{\infty} (1/nh(n)) < \infty .$$

PROOF. We have already noted that assumption B.3′ implies (a). From (4.1), we see that

$$\text{(4.4)} \qquad \sum_{n=1}^{\infty} \frac{1}{nf(n)} = \sum_{n=1}^{\infty} \frac{1}{G(n-1)} \left(\frac{1}{f(n-1)} - \frac{1}{f(n)}\right) \leqq \frac{1}{\inf_{x\geqq 0} G(x)} \sum_{n=1}^{\infty} \left(\frac{1}{f(n-1)} - \frac{1}{f(n)}\right).$$

Since $f(x)$ is strictly increasing, we see from (4.1) that $G(x) > 0$, all x. This fact, the fact that $f(0) > 0$, and assumption B.3′ can now be used to show that the upper bound in (4.4) is finite, thus proving (b). Kunte and Gurjar (1973) have shown that (a) and (b) together imply (c). □

LEMMA 4.2. *Define*

$$\text{(4.5)} \qquad L(x) = x^{-M} f(x) ,$$

where $M = \lim_{x\to\infty} G(x)$. *If* $\{u_n\}$ *and* $\{v_n\}$ *are two sequences of integers for which* $u_n \leqq v_n$, *all* n, *and* $u_n \to \infty$ *as* $n \to \infty$, *and if there exists an integer* k *such that* $v_n \leqq ku_n$ *for all large enough* n, *then*

$$\text{(4.6)} \qquad \lim_{n\to\infty} \frac{L(v_n)}{L(u_n)} = 1 .$$

PROOF. From (4.1), it is easily demonstrated that for any integer j,

$$f(j) = f(0) \prod_{i=1}^{j} \left(1 + \frac{G(i-1)}{i}\right) = f(0) \frac{\Gamma(j+M+1)}{\Gamma(M+1)\Gamma(j+1)} \exp\left\{\sum_{i=1}^{j} i^{-1} \log\left(1 + \frac{G(i-1)-M}{i+M}\right)^{i}\right\},$$

so that if we let

$$W(i) = \frac{f(0)}{\Gamma(M+1)} \frac{\Gamma(i+M+1)}{\Gamma(i+1)i^{M}}, \qquad \varepsilon(i) = \log\left(1 + \frac{G(i-1)-M}{i+1}\right)^{i}.$$

we have

$$L(j) = W(j) \exp\left(\sum_{i=1}^{j} \frac{\varepsilon(i)}{i}\right).$$

Using the well-known fact that $\lim_{n\to\infty} \Gamma(n+M)/\Gamma(n)n^M = 1$, it follows that $\lim_{n\to\infty} W(n) = f(0)/\Gamma(M+1) \neq 0$. Using assumption B.3′, it can easily be established that

(4.7) $$\lim_{n\to\infty} \varepsilon(n) = 0 .$$

Now

(4.8) $$\frac{L(v_n)}{L(u_n)} = \frac{W(v_n)}{W(u_n)} \exp\left(\sum_{i=u_n+1}^{v_n} \frac{\varepsilon(i)}{i}\right).$$

Also, by the given, for large enough n,

(4.9) $$\left|\sum_{i=u_n+1}^{v_n} \frac{\varepsilon(i)}{i}\right| \leqq \sum_{i=u_n+1}^{ku_n} \frac{|\varepsilon(i)|}{i} ;$$

and (4.7) and the given fact that $u_n \to \infty$ as $n \to \infty$ can be used to show that the right-hand side of (4.9) tends to 0 as $n \to \infty$. Thus,

(4.10) $$\lim_{n\to\infty} \sum_{i=u_n+1}^{v_n} \frac{\varepsilon(i)}{i} = 0 .$$

Since we have already shown that $\lim_{n\to\infty} W(n) \neq 0$, (4.8) and (4.10) imply (4.6). □

Loosely speaking, Lemma 4.2 establishes that $f(x)$ is regularly varying on the integers, behaving almost like x^M for large values of x.

Let

(4.11) $$\gamma(c) = \text{first} \quad n \geqq 1 \quad \text{such that} \quad nf(n) \geqq c^{-1}M .$$

Since, from assumption B.1, $xf(x)$ strictly increases from 0 (at $x = 0$) to ∞, $\gamma(c)$ is well defined for all c, and $\lim_{c\to 0} \gamma(c) = \infty$. It therefore follows from assumption B.1 that

(4.12) $$P\{\lim_{c\to 0} f(\gamma(c)) Y_{\gamma(c)} = V\} = 1 .$$

From (4.11), we have

(4.13) $$M \frac{\gamma f(\gamma)}{(\gamma - 1)f(\gamma - 1)} \geqq c\gamma f(\gamma) \geqq M ,$$

where $\gamma = \gamma(c)$. Hence, since $\lim_{c\to 0} \gamma(c) = \infty$, it follows from Lemma 4.1 (a) and (4.13) that

(4.14) $$\lim_{c\to 0} c\gamma(c) f(\gamma(c)) = M .$$

LEMMA 4.3. *For the class* $\{t(c) : c > 0\}$ *of stopping rules defined by* (4.2), *we have*

(4.15) $$P\left\{\lim_{c\to 0} \frac{f(\gamma(c))}{f(t(c))} = V^{-M/(M+1)}\right\} = 1 .$$

PROOF. From (4.14) and (3.14) with $K(x) = x$, we see that (4.15) will hold if we can show that

(4.16) $$P\left\{\lim_{c\to 0} \frac{t}{\gamma} = V^{1/(M+1)}\right\} = 1 ,$$

where $\gamma = \gamma(c)$, $t = t(c)$. Let

$$u = u(V, c) = [V^{1/(M+1)}\gamma(c)] , \tag{4.17}$$

where

$$[x] = \text{largest integer} \leqq x .$$

It follows from the definition of u that with probability 1

$$\left| V^{1/(M+1)} - \frac{u}{\gamma} \right| \leqq \gamma^{-1} , \qquad \text{all } c > 0, \text{ all } V .$$

Thus, since $\lim_{c\to 0} \gamma(c) = \infty$, we have

$$P\left\{\lim_{c\to 0} \frac{u}{\gamma} = V^{1/(M+1)}\right\} = 1 . \tag{4.18}$$

It also follows from the definition of u that

$$\gamma \leqq u \leqq [V + 1]\gamma , \qquad \text{if } V \geqq 1 ,$$
$$u \leqq \gamma \leqq [V^{-1} + 1]u , \qquad \text{if } V < 1 ,$$

so that from Lemma 4.2 and the fact that $\lim_{c\to 0} \gamma(c) = \infty$, we have

$$P\left\{\lim_{c\to 0} \frac{L(u)}{L(\gamma)} = 1\right\} . \tag{4.19}$$

Now

$$cuf(u) = cu^{M+1}L(u) = c\gamma^{M+1}L(\gamma)\left(\frac{u}{\gamma}\right)^{M+1} \frac{L(u)}{L(\gamma)} ,$$

and thus from (4.14), (4.18) and (4.19),

$$P\{\lim_{c\to 0} cuf(u) = MV\} = 1 . \tag{4.20}$$

In turn, it follows from (3.14) with $K(x) = x$, and (4.20) that

$$P\left\{\lim_{c\to 0} \frac{tf(t)}{uf(u)} = 1\right\} = 1 . \tag{4.21}$$

However, since $f(x)$ is strictly increasing,

$$\left|\frac{t}{u} - 1\right| \leqq \left|\frac{tf(t)}{uf(u)} - 1\right| ,$$

and hence

$$P\left\{\lim_{c\to 0} \frac{t}{u} = 1\right\} = 1 . \tag{4.22}$$

Equation (4.16), and hence the result (4.15), now follows directly from (4.18) and (4.22). □

LEMMA 4.4.

$$\lim_{c\to 0} E(f(\gamma)ct) = ME(V^{1/(M+1)}) < \infty . \tag{4.23}$$

PROOF. Since $\gamma(c)$ is not random, it follows from (4.14) that (4.23) is equivalent to

$$\lim_{c\to 0} E\left(\frac{t}{\gamma}\right) = E(V^{1/(M+1)}) < \infty . \tag{4.24}$$

We will prove (4.24). Let $N = [m\gamma]$, and let

$$a_{c,m} = P\{t\gamma^{-1} > m\} . \tag{4.25}$$

Then from (4.2) and (4.25),

$$a_{c,m} = P\{t > N\} \leqq P\left\{f(N)Y_N > c\left\{\frac{1}{f(N)} - \frac{1}{f(N+1)}\right\}^{-1}\right\} .$$

Using Markov's inequality, (4.1), (4.3), and assumption B.3′, we have

$$a_{c,m} \leqq \frac{\alpha}{c}\left\{\frac{1}{f(N)} - \frac{1}{f(N+1)}\right\} \leqq \frac{\alpha\beta}{c(N+1)f(N+1)} , \tag{4.26}$$

where $\alpha = \sup_n E[f(n)Y_n]$, $\beta = \sup_n G(n) < \infty$. However, since $xf(x)$ is increasing in x and $N + 1 \geqq m\gamma$, we have from (4.11) and (4.26) that

$$a_{c,m} \leqq \frac{\alpha\beta}{cm\gamma f(m\gamma)} \leqq \frac{\alpha\beta}{Mm\{f(m\gamma)/f(\gamma)\}} . \tag{4.27}$$

It follows from Lemma 4.1 (c) that given ε, $0 < \varepsilon < 1$, there exists N_ε such that $m \geqq N_\varepsilon$ implies

$$\frac{f(mx)}{f(x)} \geqq (1 - \varepsilon)h(m) , \qquad \text{all } x ,$$

where $\sum_{m=1}^{\infty} (mh(m))^{-1} < \infty$. Therefore, for all $m \geqq N_\varepsilon$,

$$a_m \equiv \sup_c a_{c,m} \leqq \frac{\alpha\beta}{(1-\varepsilon)Mmh(m)} , \qquad \text{all } m \geqq N_\varepsilon . \tag{4.28}$$

It thus follows from (4.28) that $\sum_{m=1}^{\infty} a_m < \infty$. Since from (4.14) and (4.15) we have

$$P\{\lim_{c\to 0} t\gamma^{-1} = V^{1/(M+1)}\} = 1 , \tag{4.29}$$

Lemma 3.2 of Bickel and Yahav (1968) can now be used to show that (4.24), and thus (4.23) holds. □

The previous analysis was necessary to allow us to replace the random quantity $f(t(c))$ by the nonrandom quantity $f(\gamma(c))$ in our operation with expectations. The importance of this step is already apparent in the proof of Lemma 4.4, and is further apparent in the conclusion of our proof of Theorem 4.1, which now follows.

CONCLUSION OF PROOF OF THEOREM 4.1. From (3.4) and the fact that $t(c) \in T$, all $c > 0$ (Lemma 3.1), we see that the class $\{t(c) : c > 0\}$ is A.O. if

$$\limsup_{c\to 0}\left\{\frac{E(X(t(c), c))}{\inf_{s\in T} E(X(s, c))}\right\} \leqq 1 ,$$

or, equivalently, if for every $s \in T$ we can show that

$$\limsup_{c\to 0} \frac{E(X(t(c), c))}{E(X(s, c))} \leqq 1 . \tag{4.30}$$

We note that from (3.13), (3.14), with $K(x) = x$, and (4.15),

$$P\{\lim_{c\to 0} f(\gamma)X(t, c) = (1 + M)V^{1/(M+1)}\} = 1 , \tag{4.31}$$

and since the class $\{t(c) : c > 0\}$ is A.P.O., it thus follows that for every $s \in T$,

$$P\{\liminf_{c\to 0} f(\gamma)X(s, c) \geqq (1 + M)V^{1/(M+1)}\} = 1 .$$

Fatou's lemma then yields,

$$\liminf_{c\to 0} E(f(\gamma)X(s, c)) \geqq (1 + M)E(V^{1/(M+1)}) . \tag{4.32}$$

Since

$$\limsup_{c\to 0} \frac{E(X(t, c))}{E(X(s, c))} \leqq \frac{\limsup_{c\to 0} E(f(\gamma)X(t(c), c))}{\liminf_{c\to 0} E(f(\gamma)X(s, c))} ,$$

for every $s \in T$, we see from (4.32) that (4.30) will hold if we can show that

$$\lim_{c\to 0} E(f(\gamma)X(t, c)) = (1 + M)E(V^{1/(M+1)}) . \tag{4.33}$$

From (4.1) and (4.2),

$$Y_t \leqq ct \frac{(t + 1)f(t + 1)}{tf(t)} \frac{1}{G(t)} \leqq Kct ,$$

where (as explained in the proof of Lemma 4.1)

$$K \equiv \sup_x \frac{(x + 1)f(x + 1)}{xf(x)} (\inf_x G(x))^{-1} < \infty .$$

Thus,

$$f(\gamma)X(t, c) \leqq (1 + K)f(\gamma)ct . \tag{4.34}$$

Using a well-known generalization of the dominated convergence theorem [see Royden (1968, page 89)], (4.34), (4.16), Lemma 3.1, (4.23), and (4.31), we conclude that (4.33) holds. This in turn, as already explained, proves that $\{t(c): c > 0\}$ is A.O., completing the proof of Theorem 4.1. □

As we will see in Section 6, the condition (4.3) in Theorem 4.1 is not always easy to verify. However, some condition like (4.3) seems to be required, as shown by the following example.

EXAMPLE. Let the probability space be the interval [0, 1] under Lebesgue measure. Let $f(x)$ be any increasing positive function satisfying assumptions B.2 and B.3′ (where $K(x) = x$). For each $\omega \in [0, 1]$ and each positive integer n, let

$$\begin{aligned} f(n)Y_n(\omega) &= \left(\frac{1}{f(n)} - \frac{1}{f(n + 1)}\right)^{-1} V , && \text{if} \quad \omega \in \left[0, \frac{1}{n}\right) , \\ &= V , && \text{otherwise,} \end{aligned}$$

where $V > 0$ is any positive constant. It is easily shown that $f(x)$ and $Y_1, Y_2, \cdots$ satisfy assumptions B.0—B.2 and B.3′, but that $E(f(n)Y_n) \to \infty$ as $n \to \infty$, violating (4.3). It is also straightforward to show that if $0 < c < V$, then $P\{t(c) > n\} \geqq n^{-1}$, and thus $E(t(c)) = \infty$. Hence, for $0 < c < V$, $E(X(t(c), c)) = \infty$. However, the stopping rule s which a.s. takes exactly one observation has $E(X(s, c)) = E(Y_1 + c)$ bounded for all c, demonstrating that $\{t(c): c > 0\}$ cannot be A.O.

5. A.P.O. sequential Bayes interval estimation procedures. In Section 2, we saw that when assumption A.O. holds, the prior distribution has density $\psi(\theta)$, and the loss function is (2.1), then the Bayes optimal interval estimation procedure is to use the interval

$$I^*(\mathbf{X}_n) = [\alpha_{1n}^*(\mathbf{X}_n), \alpha_{2n}^*(\mathbf{X}_n)] ,$$

where the endpoints $\alpha_{in}^*(\mathbf{X}_n)$, $i = 1, 2$, are defined by (2.5). The posterior risk

$$Y_n = \rho(\psi, I^*(\mathbf{X}_n))$$

of this Bayes procedure is given by (2.6). For any stopping rule t, the Bayes optimal terminal decision rule is to use $I^*(\mathbf{X}_n)$ when $t = n$, incurring the Bayes risk

$$R(\psi, I^*(\mathbf{X}_t), t) = E(Y_t + ct) .$$

The example of estimation of the mean of a normal distribution $N(\theta, 1)$ under a normal prior for θ (see Section 2) suggests that if we can establish sufficient conditions for the posterior density $\psi(\theta \mid \mathbf{X}_n)$ obtained from $f(x \mid \theta)$ and $\psi(\theta)$ to have approximately the form of a normal density for large n, then we will be able to show that

$$\text{(5.1)} \qquad \lim_{n\to\infty} (n/\log n)^{\frac{1}{2}} Y_n = 2aK^{\frac{1}{2}}(\theta) , \qquad \text{a.s.} \quad (P_\theta) ,$$

where $K(\theta)$ is a certain function of the parameter θ, and a.s. (P_θ) refers to the conditional distribution of $X_1, X_2, \cdots$, given θ. [To verify assumption B.1 of Section 3 in this context, we need to show almost sure convergence with respect to the joint probability distribution of $X_1, X_2, \cdots$, and θ_0. However, since we are dealing with probabilities, use of (5.1) for all $\theta_0 \in \Theta$ and the dominated convergence theorem will establish the desired result.]

One set of conditions which is more than sufficient to make $\psi(\theta \mid \mathbf{X}_n)$ asymptotically of normal form can be obtained by taking $k = 1$, $r = 0$ in assumptions A2.2, A2.6, A2.7, A2.8 and A2.9 of Bickel and Yahav (1969b). In order to make the present paper as self-contained as possible, and because our notation is slightly different, we restate these conditions here.

A.1. The prior density $\psi(\theta)$ is continuous, positive and bounded on Θ.

A.2. The first two derivatives $\varphi^{(i)}(\theta, X) = d^i\varphi(\theta, X)/(d\theta)^i$, $i = 1, 2$, of $\varphi(\theta, X) = \log f(X \mid \theta)$ exist and are continuous in θ, a.s. (P_{θ_0}), for all $\theta_0 \in \Theta$.

A.3. For each $\theta \in \Theta$, there exists $\varepsilon(\theta) > 0$ such that

$$E_\theta(\sup\{\varphi^{(2)}(s, X): |s - \theta| < \varepsilon(\theta), s \in \Theta\}) < \infty ,$$

where the expectation, E_θ, is taken over X with respect to the density $f(x \mid \theta)$.

A.4. The function

$$A(\theta) = E_\theta(\varphi^{(2)}(\theta, X)) = -E_\theta(\varphi^{(1)}(\theta, X))^2$$

is a strictly negative function of $\theta \in \Theta$.

A.5. For all $\theta \in \Theta$ and all $\varepsilon > 0$,

$$E_\theta(\sup\{[\varphi(s, X) - \varphi(\theta, X)] : |s - \theta| \geqq \varepsilon, s \in \Theta\}) < 0 .$$

We note that since $\psi(\theta)$ is a density, and from assumptions A.1 and A.2, it follows that for all $\theta_0 \in \Theta$,

$$0 < \int_\Theta \psi(\theta) \prod_{i=1}^n f(X_i \mid \theta)\, d\theta < \infty , \qquad \text{a.s.} \quad (P_{\theta_0}) .$$

This result corresponds to assumption A2.5 of Bickel and Yahav (1969b).

Under our assumptions A.2—A.4, Bickel and Yahav (1969b, Lemma 2.1) show that for each $\theta_0 \in \Theta$, a strongly consistent estimator $\hat\theta_n = \hat\theta_n(\mathbf{X}_n, \theta_0)$ of θ_0 can be constructed by maximizing the likelihood over the subset $U_{\theta_0} = \{s : |s - \theta_0| \leqq 2^{-1}\varepsilon(\theta_0), s \in \Theta\}$ of Θ, where $\varepsilon(\theta_0)$ is defined by assumption A.3. They further show that there exists $N_1 = N(X_1, X_2, \cdots; \theta_0)$ such that $\hat\theta_n$ a.s. (P_{θ_0}) satisfies the likelihood equation

$$\sum_{i=1}^n \varphi^{(1)}(\hat\theta_n, X_i) = 0 \tag{5.2}$$

for all $n \geqq N_1$. They then consider the posterior density

$$\psi^*(z \mid \mathbf{X}_n) = n^{\frac{1}{2}}\psi(n^{-\frac{1}{2}}z + \hat\theta_n \mid \mathbf{X}_n) \tag{5.3}$$

of $z = n^{\frac{1}{2}}(\theta_0 - \hat\theta_n)$, and show [Bickel and Yahav (1969b, Theorem 2.2)] that $\psi^*(z \mid \mathbf{X}_n)$ is closely approximated for large n by the density, $N(z; 0, K(\theta_0))$, of a normal distribution with mean 0 and variance

$$K(\theta_0) = (-A(\theta_0))^{-1} . \tag{5.4}$$

Indeed, what they show is that

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} |\psi^*(z \mid \mathbf{X}_n) - N(z; 0, K(\theta_0))|\, dz = 0 , \qquad \text{a.s.} \quad (P_{\theta_0}) . \tag{5.5}$$

The strength of the approximation (5.5) gives us every reason to hope that under the assumptions A.0—A.5, the result (5.1) will hold.

Before stating the main result of this section, we digress briefly to take care of one minor technical problem—namely, that $(x/\log x)^{\frac{1}{2}}$ is not well-defined for $x < e$. However, for $\log x > 3^{\frac{1}{2}}$, $(x/\log x)^{\frac{1}{2}}$ is positive, strictly increasing, and strictly concave. Thus, it is clearly possible to define at least one function $f^*(x)$ which equals $(x/\log x)^{\frac{1}{2}}$ for $\log x > 3^{\frac{1}{2}}$ and which is positive, strictly increasing and strictly concave on $[0, \infty)$. By Remark II of Section 3, such an $f^*(x)$ will satisfy assumption B.2, and, since $f^*(x)$ equals $(x/\log x)^{\frac{1}{2}}$ for large x, will also satisfy assumption B.3′ with $M = \frac{1}{2}$. For asymptotic purposes, the beginning few values of $f^*(x)$ are unimportant. Therefore, if we can demonstrate the validity of equation (5.1), the following theorem will be an immediate consequence of the above discussion and Theorem 4.1.

THEOREM 5.1. *Under assumptions* A.0—A.5, *the class* $\{t(c): c > 0\}$ *of stopping rules defined by*

$$(5.6)\qquad t(c) = \text{first}\quad n \geqq 3\quad \text{such that}\quad \left(1 - \left(\frac{n \log (n+1)}{(n+1)\log n}\right)^{\frac{1}{2}}\right) Y_n \leqq c$$

is A.P.O.

As already remarked, Theorem 5.1 holds if we can establish (5.1). We first establish the asymptotic properties of $\alpha_{in}^* = \alpha_{in}^*(\mathbf{X}_n)$, and then use these properties to verify (5.1). To this end, fix arbitrary $\theta_0 \in \Theta$. Let

$$(5.7)\qquad \nu_n(z) = \exp\{\textstyle\sum_{i=1}^n (\varphi(n^{-\frac{1}{2}}z + \hat{\theta}_n, X_i) - \varphi(\hat{\theta}_n, X_i))\}\,,$$

and

$$(5.8)\qquad c_n = \textstyle\int_{-\infty}^{\infty} \nu_n(z)\psi(n^{-\frac{1}{2}}z + \hat{\theta}_n)\, dz\,,$$

where $\hat{\theta}_n = \hat{\theta}_n(\mathbf{X}_n, \theta_0)$. It is easily verified that

$$(5.9)\qquad \psi^*(z \mid \mathbf{X}_n) = (c_n)^{-1}\psi(n^{-\frac{1}{2}}z + \hat{\theta}_n)\nu_n(z)\,.$$

From equation (2.29) of Bickel and Yahav (1969b, page 263), assumption A.1, the strong consistency of $\hat{\theta}_n$, and the dominated convergence theorem, it can be shown that

$$(5.10)\qquad \lim_{n\to\infty} c_n = \psi(\theta_0)(2\pi K(\theta_0))^{\frac{1}{2}}\,,\qquad \text{a.s.}\quad (P_{\theta_0})\,.$$

Using (5.10), the strong consistency of $\hat{\theta}_n$, and the continuity of $\psi(s)$,

$$(5.11)\qquad \lim_{n\to\infty} \psi^*(0 \mid \mathbf{X}_n) = \lim_{n\to\infty} \frac{\psi(\hat{\theta}_n)}{c_n} = (2\pi K(\theta_0))^{-\frac{1}{2}} > 0\,,\qquad \text{a.s.}\quad (P_{\theta_0})\,;$$

while, on the other hand, it is not hard to show, using assumption A.5, that for all $\delta \neq 0$ such that $\theta_0 + \delta \in \Theta$, we have

$$(5.12)\qquad \begin{aligned} \lim_{n\to\infty} \psi(\theta_0 + \delta \mid \mathbf{X}_n) &= \lim_{n\to\infty} n^{-\frac{1}{2}}\psi(n^{\frac{1}{2}}(\theta_0 + \delta - \hat{\theta}_n) \mid \mathbf{X}_n) \\ &= 0\,,\qquad \text{a.s.}\quad (P_{\theta_0})\,. \end{aligned}$$

Let

$$(5.13)\qquad \beta_{in}^* = n^{\frac{1}{2}}(\alpha_{in}^* - \hat{\theta}_n)\,,\qquad i = 1, 2\,.$$

From (2.5) and (5.3), it is apparent that

$$(5.14)\qquad \begin{aligned} \beta_{1n}^* &= \inf\{z : \psi^*(z \mid \mathbf{X}_n) \geqq ab^{-1}n^{-\frac{1}{2}},\ n^{-\frac{1}{2}}z + \hat{\theta}_n \in \Theta\}\,, \\ \beta_{2n}^* &= \sup\{z : \psi^*(z \mid \mathbf{X}_n) \geqq ab^{-1}n^{-\frac{1}{2}},\ n^{-\frac{1}{2}}z + \hat{\theta}_n \in \Theta\}\,. \end{aligned}$$

LEMMA 5.1. *There exists* $N_2 = N(X_1, X_2, \cdots; \theta_0)$, *such that for all* $n \geqq N_2$, $\beta_{1n}^* < 0$ *and* $\beta_{2n}^* > 0$, *and*

$$(5.15)\qquad \psi^*(\beta_{in}^* \mid \mathbf{X}_n) = ab^{-1}n^{-\frac{1}{2}}\,,\qquad i = 1, 2\,,\quad \text{a.s.}\quad (P_{\theta_0})\,.$$

Further,

$$\lim_{n\to\infty} \beta_{1n}^* = -\infty\,, \qquad \lim_{n\to\infty} \beta_{2n}^* = \infty\,, \quad \text{a.s.} \quad (P_{\theta_0})\,. \tag{5.16}$$

PROOF. By (5.3) and assumptions A.1 and A.2, $\psi^*(z \mid \mathbf{X}_n)$ is continuous in z. Hence, the existence of N_2 follows from (5.11), (5.12) and (5.14).

To prove (5.16), we proceed by contraposition to (5.5). Taking β_{2n}^*, for example, we note that for every sequence of observations $X_1, X_2, \cdots$, for which $\beta_{2n}^* \nrightarrow \infty$, there exists a subsequence $\{n_k\}$, $\lim_{k\to\infty} n_k = \infty$, and Q, $0 < Q < \infty$, such that $\beta_{2n_k}^* < Q$, all k. Hence, from (5.14), we have $\psi^*(z \mid \mathbf{X}_{n_k}) < ab^{-1}n_k^{-\frac{1}{2}}$ for all $z \geqq Q$, and applying the dominated convergence theorem, we obtain

$$\lim_{k\to\infty} \int_Q^{Q+1} |\psi^*(z \mid \mathbf{X}_{n_k}) - N(z; 0, K(\theta_0))|\, dz = \int_Q^{Q+1} N(z; 0, K(\theta_0))\, dz > 0\,,$$

which contradicts (5.5). Similarly, $\beta_{1n}^* \nrightarrow -\infty$ leads to a contradiction with (5.5). □

LEMMA 5.2.

$$\lim_{n\to\infty} n^{-\frac{1}{2}}\beta_{in}^* = 0\,, \qquad i = 1, 2\,, \quad \text{a.s.} \quad (P_{\theta_0})\,. \tag{5.17}$$

PROOF. Let

$$q_n = K^{\frac{1}{2}}(\theta_0)\{\log n + \log (b^2/2\pi a^2 K(\theta_0)\}^{\frac{1}{2}}\,. \tag{5.18}$$

It is easily shown that $N(\pm q_n; 0, K(\theta_0)) = ab^{-1}n^{-\frac{1}{2}}$. Also, since the normal density is unimodal

$$N(z; 0, K(\theta_0)) \geqq ab^{-1}n^{-\frac{1}{2}} \qquad \text{if and only if} \quad z \in [-q_n, q_n]\,. \tag{5.19}$$

We prove (5.17) for $i = 2$; the case $i = 1$ follows in similar fashion. Thus, let

$$J_n = [\min (\beta_{2n}^*, q_n), \max (\beta_{2n}^*, q_n)]$$

and note that by (5.14) and (5.19),

$$\Delta_n(z) = |ab^{-1}n^{-\frac{1}{2}} - N(z; 0, K(\theta_0))| \leqq |\psi^*(z \mid \mathbf{X}_n) - N(z; 0, K(\theta_0))|$$

for all $z \in J_n$. Thus, from (5.5), it follows that

$$\lim_{n\to\infty} \int_{J_n} \Delta_n(z)\, dz = 0\,. \tag{5.20}$$

However,

$$n^{-\frac{1}{2}}|\beta_{2n}^* - q_n| = n^{-\frac{1}{2}} \int_{J_n} dz \leqq ba^{-1}\{\int_{J_n} \Delta_n(z)\, dz + \int_{J_n} N(z; 0, K(\theta_0))\, dz\}\,, \tag{5.21}$$

and since $q_n \to \infty$ [see (5.18)] and $\beta_{2n}^* \to \infty$, a.s. (P_{θ_0}), it follows that

$$\lim_{n\to\infty} \int_{J_n} N(z; 0, K(\theta_0))\, dz = 0\,, \qquad \text{a.s.} \quad (P_{\theta_0})\,. \tag{5.22}$$

Thus, from (5.20) — (5.22), we have

$$\lim_{n\to\infty} n^{-\frac{1}{2}}\, |\beta_{2n}^* - q_n| = 0\,, \qquad \text{a.s.} \quad (P_{\theta_0})\,.$$

Noting from (5.18) that $\lim_{n\to\infty} n^{-\frac{1}{2}} q_n = 0$ completes the proof of (5.17) for $i = 2$. □

As an immediate consequence of (5.13), the strong consistency of $\hat{\theta}_n$, and (5.17), we have

$$\lim_{n\to\infty} \alpha_{in}^* = \lim_{n\to\infty} \hat{\theta}_n = \theta_0 , \qquad i = 1, 2 , \quad \text{a.s.} \quad (P_{\theta_0}) . \tag{5.23}$$

Also, from (5.9), (5.13) and (5.15), we have

$$\nu_n(\beta_{in}^*) = \frac{c_n a}{b n^{\frac{1}{2}} \psi(\alpha_{in}^*)} , \tag{5.24}$$

while expansion of $\log \nu_n(\beta_{in}^*)$ in terms of $n^{-\frac{1}{2}}\beta_{in}^*$ by means of Taylor's formula, using (5.2), yields

$$\log \nu_n(\beta_{in}^*) = \frac{(\beta_{in}^*)^2}{2}\left\{\frac{1}{n} \sum_{i=1}^n \varphi^{(2)}(\tilde{\theta}_{in}, X_i)\right\} , \tag{5.25}$$

where $\tilde{\theta}_{in}$ lies between $\hat{\theta}_n$ and $\alpha_{in}^* = \hat{\theta}_n + n^{-\frac{1}{2}}\beta_{in}^*$.

Using assumptions A.2—A.4 and Lemma 2 of DeGroot (1970, Section 10.6), we see that $\varphi^{(2)}(s, X)$ is a supercontinuous function at all values $s \in \Theta$. It then follows from Theorem 1 of DeGroot (1970, Section 10.8) and (5.23) that

$$\lim_{n\to\infty} n^{-1} \sum_{i=1}^n \varphi^{(2)}(\tilde{\theta}_{in}, X_i) = A(\theta_0) , \qquad i = 1, 2 , \quad \text{a.s.} \quad (P_{\theta_0}) . \tag{5.26}$$

Noting that assumption A.1 and (5.23) imply that $\lim_{n\to\infty} \psi(\alpha_{in}^*) = \psi(\theta_0)$, a.s. (P_{θ_0}), $i = 1, 2$, we conclude from (5.10), (5.24), (5.25) and (5.26) that

$$\lim_{n\to\infty} \frac{(\beta_{in}^*)^2}{\log n} = K(\theta_0) , \qquad i = 1, 2 , \quad \text{a.s.} \quad (P_{\theta_0}) . \tag{5.27}$$

Thus, remembering from Lemma 5.1 that for large enough n, $\beta_{1n}^* < 0$ and $\beta_{2n}^* > 0$ a.s. (P_{θ_0}), we conclude that

$$\begin{aligned} \lim_{n\to\infty} (n/\log n)^{\frac{1}{2}}(\alpha_{2n}^* - \alpha_{1n}^*) &= \lim_{n\to\infty} (\log n)^{\frac{1}{2}}(\beta_{2n}^* - \beta_{1n}^*) \\ &= 2K^{\frac{1}{2}}(\theta_0) , \quad \text{a.s.} \quad (P_{\theta_0}) . \end{aligned} \tag{5.28}$$

Comparing (2.6) and (5.28), we see that (5.1) will be established, and Theorem 5.1 verified, if we can prove the following result.

LEMMA 5.3. *For all* $\theta_0 \in \Theta$,

$$\begin{aligned} &\lim_{n\to\infty} (n/\log n)^{\frac{1}{2}} \int_{\alpha_{2n}^*}^{\infty} \psi(\theta \mid \mathbf{X}_n)\, d\theta \\ &\qquad = \lim_{n\to\infty} (n/\log n)^{\frac{1}{2}} \int_{-\infty}^{\alpha_{1n}^*} \psi(\theta \mid \mathbf{X}_n)\, d\theta = 0 , \end{aligned} \tag{5.29}$$

or equivalently

$$\begin{aligned} &\lim_{n\to\infty} (n/\log n)^{\frac{1}{2}} \int_{\beta_{2n}^*}^{\infty} \psi^*(z \mid \mathbf{X}_n)\, dz \\ &\qquad = \lim_{n\to\infty} (n/\log n)^{\frac{1}{2}} \int_{-\infty}^{\beta_{1n}^*} \psi^*(z \mid \mathbf{X}_n)\, dz = 0 , \quad \text{a.s.} \quad (P_{\theta_0}) . \end{aligned} \tag{5.30}$$

PROOF. The equivalence of (5.29) and (5.30) is apparent from (5.3) and (5.13). We will demonstrate that

$$\lim_{n\to\infty} (n/\log n)^{\frac{1}{2}} \int_{\beta_{2n}^*}^{\infty} \psi^*(z \mid \mathbf{X}_n)\, dz = 0 , \quad \text{a.s.} \quad (P_{\theta_0}) . \tag{5.30a}$$

The other equality in (5.30) follows by a similar proof.

Bickel and Yahav (1969b, equation (2.40), with $k = 1$)[3] have shown that for each $\varepsilon > 0$, there exists $\delta(\varepsilon) > 0$ and $N_3 = N(X_1, X_2, \cdots, \theta_0, \delta(\varepsilon))$ such that for every $n \geqq N_3$ and all z, $|z| \leqq n^{\frac{1}{2}}\delta(\varepsilon)$,

$$\nu_n(z) \leqq \exp\{-1/2K(\theta_0, \varepsilon)\}\,, \qquad \text{a.s.} \quad (P_{\theta_0})\,, \tag{5.31}$$

where

$$K(\theta_0, \varepsilon) = (-A(\theta_0) - \varepsilon)^{-1}\,.$$

Note that by assumption A.4, we may choose $\varepsilon > 0$ as small as we wish and still have $K(\theta_0, \varepsilon) > 0$. Let $q_n(\varepsilon)$ be q_n of (5.18) with $K(\theta_0, \varepsilon)$ substituted for $K(\theta_0)$. It follows from (2.9), (5.9) and (5.31) that

$$\left(\frac{n}{\log n}\right)^{\frac{1}{2}} \int_{q_n(\varepsilon)}^{n^{\frac{1}{2}}\delta(\varepsilon)} \psi^*(z \mid \mathbf{X}_n)\, dz \leqq \left(\frac{n}{\log n}\right)^{\frac{1}{2}} \frac{L}{c_n q_n(\varepsilon)} \exp\left\{-\frac{1}{2}\frac{q_n^2(\varepsilon)}{K(\theta_0, \varepsilon)}\right\}\,,$$

where $L = \max\{\psi(s) : s \in \Theta\} < \infty$ by assumption A.1. Using the definition of $q_n(\varepsilon)$ and (5.10), we see that the right-hand side of this inequality goes to 0 a.s. (P_{θ_0}) as $n \to \infty$, and thus

$$\lim_{n\to\infty} \left(\frac{n}{\log n}\right)^{\frac{1}{2}} \int_{q_n(\varepsilon)}^{n^{\frac{1}{2}}\delta(\varepsilon)} \psi^*(z \mid \mathbf{X}_n)\, dz = 0\,, \qquad \text{a.s.} \quad (P_{\theta_0})\,. \tag{5.32}$$

Next, it follows from (5.9), (5.10) and equation (2.30) of Bickel and Yahav (1969b) that

$$\lim_{n\to\infty} \left(\frac{n}{\log n}\right)^{\frac{1}{2}} \int_{n^{\frac{1}{2}}\delta(\varepsilon)}^{\infty} \psi^*(z \mid \mathbf{X}_n)\, dz = 0\,, \qquad \text{a.s.} \quad (P_{\theta_0})\,. \tag{5.33}$$

Finally, it follows from (5.14) that

$$\int_{\beta_{2n}^*}^{q_n(\varepsilon)} \psi^*(z \mid \mathbf{X}_n)\, dz \leqq ab^{-1}n^{-\frac{1}{2}}|q_n(\varepsilon) - \beta_{2n}^*|\,,$$

and thus, using the fact that $\lim_{n\to\infty} (\log n)^{-\frac{1}{2}} q_n(\varepsilon) = K^{\frac{1}{2}}(\theta_0, \varepsilon)$ and (5.27), we have

$$\lim_{n\to\infty} \left(\frac{n}{\log n}\right)^{\frac{1}{2}} \int_{\beta_{2n}^*}^{q_n(\varepsilon)} \psi^*(z \mid \mathbf{X}_n)\, dz \leqq ab^{-1}(K^{\frac{1}{2}}(\theta_0, \varepsilon) - K^{\frac{1}{2}}(\theta_0))\,. \tag{5.34}$$

Since $\lim_{\varepsilon\to 0} K^{\frac{1}{2}}(\theta_0, \varepsilon) = K^{\frac{1}{2}}(\theta_0)$, and since, as we have remarked above, ε can be chosen arbitrarily small, it follows from (5.34) that

$$\lim_{n\to\infty} \left(\frac{n}{\log n}\right)^{\frac{1}{2}} \int_{\beta_{2n}^*}^{q_n(\varepsilon)} \psi^*(z \mid \mathbf{X}_n)\, dz = 0\,.$$

This last result, together with (5.32) and (5.33), establishes (5.30a). □

Lemma 5.3 completes the proof of (5.1), and thus establishes Theorem 5.1. Note from (5.28) and (5.29) that asymptotically, as $n \to \infty$, the posterior Bayes risk of the Bayes optimal terminal decision rule $I^*(\mathbf{X}_n)$ is dominated by the length of $I^*(\mathbf{X}_n)$. That is, the length $l(I^*(\mathbf{X}_n)) = O((\log n/n)^{\frac{1}{2}})$, while the posterior probability of noncoverage is $o((\log n/n)^{\frac{1}{2}})$, a.s. (P_{θ_0}).

[3] The "$\frac{1}{2}$" in the exponent of (5.31) is mistakenly omitted in equation (2.40) of Bickel and Yahav (1969b), as can be seen from their equation (2.36).

6. A.O. sequential Bayes interval estimation procedures. It follows from the discussion in Sections 2 and 5 and Theorem 4.1 that the class $\{t(c): c > 0\}$ of stopping rules defined by (5.6) is A.O. if we can show that

$$\sup_n \{(n/\log n)^{\frac{1}{2}} \int_{\mathscr{X}^n} \rho(\psi, I^*(\mathbf{X}_n))\, dG_n(\mathbf{X}_n)\} < \infty \,. \tag{6.1}$$

THEOREM 6.1. *If there exists any sequence* $\{I_n(\mathbf{X}_n)\}$ *of interval estimation procedures for which*

$$\sup_n \{(n/\log n)^{\frac{1}{2}} \int_{\mathscr{X}^n} \rho(\psi, I_n(\mathbf{X}_n))\, dG_n(\mathbf{X}_n)\} < \infty \,, \tag{6.2}$$

then (6.1) *holds, and* (*under assumptions* A.0—A.5) *the class of stopping rules* $\{t(c): c > 0\}$ *defined by* (5.6) *is* A.O.

PROOF. Since $I^*(\mathbf{X}_n)$ is Bayes for each n,

$$\int_{\mathscr{X}^n} \rho(\psi, I^*(\mathbf{X}_n))\, dG_n(\mathbf{X}_n) \leqq \int_{\mathscr{X}^n} \rho(\psi, I_n(\mathbf{X}_n))\, dG_n(\mathbf{X}_n) \,,$$

and (6.1) follows directly from (6.2). □

We exhibit a sequence $\{I_n(\mathbf{X}_n)\}$ satisfying (6.2) under the following assumptions.

C.1. There exists a function $g(\theta)$ on Θ, a positive integer k, and a function $v(X_1, X_2, \cdots, X_k)$ from $\mathscr{X}^k$ to the range $g(\Theta)$ of $g(\cdot)$ such that

$$E_\theta v(X_1, X_2, \cdots, X_k) = g(\theta) \,, \qquad \text{all } \theta \in \Theta \,, \tag{6.3}$$

and

$$E_\theta |v(X_1, X_2, \cdots, X_k)|^3 < \infty \,, \qquad \text{all } \theta \in \Theta \,. \tag{6.4}$$

C.2. Let

$$\sigma^2(\theta) = E_\theta |v(X_1, X_2, \cdots, X_k) - g(\theta)|^2 \,,$$
$$\gamma(\theta) = (\sigma^2(\theta))^{-\frac{3}{2}} E_\theta |v(X_1, X_2, \cdots, X_k) - g(\theta)|^3 \,.$$

Then

$$\int_\Theta \sigma(\theta)\psi(\theta)\, d\theta < \infty \qquad \textit{and} \qquad \int_\Theta \gamma(\theta)\psi(\theta)\, d\theta < \infty \,. \tag{6.5}$$

C.3. The function $g(\theta)$ has an inverse function $h: g(\Theta) \to \Theta$ which satisfies a uniform, first-order Lipschitz condition. That is, there exists $\varepsilon > 0$, $Q > 0$ for which

$$|a - b| < \varepsilon \,, \quad a, b \in g(\Theta) \Rightarrow |h(a) - h(b)| < Q|a - b| \,.$$

THEOREM 6.2. *Under assumptions* A.0—A.5, C.1—C.3, *the class of stopping rules* $\{t(c): c > 0\}$ *defined by* (5.6) *is* A.O.

PROOF. Define $m(n) = [nk^{-1}] =$ greatest integer $\leqq nk^{-1}$ and let

$$v_n(\mathbf{X}_n) = (m(n))^{-1} \sum_{i=1}^{m(n)} v(X_{(i-1)k+1}, \cdots, X_{ik}) \,, \tag{6.6}$$

$$s_n^2(\mathbf{X}_n) = (m(n))^{-1} \sum_{i=1}^{m(n)} (v(X_{(i-1)k+1}, \cdots, X_{ik}) - v_n(\mathbf{X}_n))^2 \,. \tag{6.7}$$

Let

$$r_n(\mathbf{X}_n) = Q(\log m(n)/m(n))^{\frac{1}{2}} s_n(\mathbf{X}_n)(1+\varepsilon)^{\frac{1}{2}} \tag{6.8}$$

where $\varepsilon > 0$ and $Q > 0$ are defined by assumption C.3, and let

$$u_n(\mathbf{X}_n) = h(v_n(\mathbf{X}_n)) . \tag{6.9}$$

Finally, let

$$I_n(\mathbf{X}_n) = [u_n(\mathbf{X}_n) - r_n(\mathbf{X}_n), u_n(\mathbf{X}_n) + r_n(\mathbf{X}_n)] . \tag{6.10}$$

By the Fubini–Tonelli theorem,

$$\int_{\mathscr{X}^n} \rho(\psi, I_n(\mathbf{X}_n))\, dG_n(\mathbf{X}_n) = \int_\Theta \{2aE_\theta(r_n(\mathbf{X}_n)) + bP_\theta(E_n)\}\psi(\theta)\, d\theta , \tag{6.11}$$

where

$$E_n = \{\mathbf{X}_n : |u_n(\mathbf{X}_n) - \theta| \geqq r_n(\mathbf{X}_n)\} ,$$

provided either side of (6.11) is finite. To show that the right-hand side is finite, note that $P_\theta(E_n) \leqq 1$, and that

$$\begin{aligned} E_\theta(r_n(\mathbf{X}_n)) &= Q(1+\varepsilon)^{\frac{1}{2}} \left(\frac{\log m(n)}{m(n)}\right)^{\frac{1}{2}} E_\theta s_n(\mathbf{X}_n) \\ &\leqq Q(1+\varepsilon)^{\frac{1}{2}} \left(\frac{\log m(n)}{m(n)}\right)^{\frac{1}{2}} \sigma(\theta) , \end{aligned} \tag{6.12}$$

since $E_\theta s_n(\mathbf{X}_n) \leqq (E_\theta s_n^2(\mathbf{X}_n))^{\frac{1}{2}} < \sigma(\theta)$. Since $P_\theta(E_n) \leqq 1$ and $\psi(\theta)$ is a probability density, we have from (6.12) and assumption C.2 that the right-hand side of (6.11) is finite.

We now attempt to bound $P_\theta(E_n)$. Let

$$\delta_n(\theta) = (\log m(n)/m(n))^{\frac{1}{2}}\sigma(\theta) , \tag{6.13}$$

and let

$$\begin{aligned} F_n &= \{\mathbf{X}_n : |u_n(\mathbf{X}_n) - \theta| > Q\delta_n(\theta)\} , \\ G_n &= \{\mathbf{X}_n : |v_n(\mathbf{X}_n) - g(\theta)| > \delta_n(\theta)\} , \\ H_n &= \{\mathbf{X}_n : r_n(\mathbf{X}_n) \leqq Q\delta_n(\theta)\} , \end{aligned}$$

and

$$D_n = \{\theta : \delta_n(\theta) \geqq \varepsilon\} .$$

Note that

$$\begin{aligned} \int_\Theta P_\theta(E_n)\, d\theta &= \int_\Theta (P_\theta\{E_n \cap H_n\} + P_\theta\{E_n \cap H_n^c\})\psi(\theta)\, d\theta \\ &\leqq \int_\Theta (P_\theta\{H_n\} + P_\theta\{F_n\})\psi(\theta)\, d\theta \\ &\leqq \int_\Theta P_\theta\{H_n\}\psi(\theta)\, d\theta + \int_{D_n} \psi(\theta)\, d\theta + \int_{\Theta\setminus D_n} P_\theta\{F_n\}\psi(\theta)\, d\theta . \end{aligned} \tag{6.14}$$

By Markov's inequality,

$$\int_{D_n} \psi(\theta)\, d\theta \leqq \frac{1}{\varepsilon} (\log m(n)/m(n))^{\frac{1}{2}} \int_\Theta \sigma(\theta)\psi(\theta)\, d\theta . \tag{6.15}$$

On the other hand, when $\theta \in \Theta\setminus D_n$, so that $\delta_n(\theta) < \varepsilon$, we can apply the contrapositive of assumption C.3 to show that

$$\begin{aligned} F_n &= (F_n \cap G_n) \cup (F_n \cap G_n^c) \\ &\subset G_n \cup \{\mathbf{X}_n : Q|v_n(\mathbf{X}_n) - g(\theta)| \leqq |u_n(\mathbf{X}_n) - \theta|\} \\ &\subset G_n \cup \{\mathbf{X}_n : |v_n(\mathbf{X}_n) - g(\theta)| \geqq \varepsilon\} \\ &\subset (G_n \cup G_n) = G_n . \end{aligned}$$

Hence,

$$\int_{\Theta \setminus D_n} P_\theta\{F_n\}\psi(\theta)\, d\theta \leqq \int_{\Theta \setminus D_n} P_\theta(G_n)\psi(\theta)\, d\theta . \tag{6.16}$$

Applying the well-known result of Berry and Esseen, we find that

$$P_\theta\{G_n\} \leqq 2 \int_{(\log m(n))^{\frac{1}{2}}}^{\infty} (2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2}t^2}\, dt + \frac{H\gamma(\theta)}{(m(n))^{\frac{1}{2}}} , \qquad \text{all } n \text{ and } \theta , \tag{6.17}$$

where $0 < H < \infty$ is a constant independent of θ or n. Therefore, from (2.9) and (6.17), we obtain

$$P_\theta\{G_n\} \leqq 2(2\pi m(n) \log m(n))^{-\frac{1}{2}} + H\gamma(\theta)(m(n))^{-\frac{1}{2}} , \qquad \text{all } n \text{ and } \theta . \tag{6.18}$$

Finally, for all θ, n,

$$\begin{aligned} P_\theta\{H_n\} &= P_\theta\{s_n(\mathbf{X}_n) \leqq (1 + \varepsilon)^{-\frac{1}{2}}\sigma(\theta)\} \\ &\leqq P_\theta \left\{ |s_n^2(\mathbf{X}_n) - \sigma^2(\theta)| > \left(\frac{\varepsilon}{1+\varepsilon}\right) \sigma^2(\theta) \right\} \\ &\leqq \frac{E_\theta|s_n^2(\mathbf{X}_n) - \sigma^2(\theta)|^{\frac{3}{2}}}{(\varepsilon/(1+\varepsilon))^{\frac{3}{2}}(\sigma(\theta))^{\frac{3}{2}}} \\ &\leqq \frac{\gamma(\theta)}{(\varepsilon/(1+\varepsilon))^{\frac{3}{2}}(m(n))^{\frac{1}{2}}} \left(8 + \frac{K}{m(n)}\right), \end{aligned} \tag{6.19}$$

where K is a constant independent of n and θ. The last inequality follows from the following lemma, which may be of independent interest.

LEMMA 6.3. *Let $U_1, U_2, \cdots, U_n$ be* i.i.d. *with mean μ, variance σ^2, and $\xi = E|U_i - \mu|^{2r}$, $1 \leqq r \leqq 2$. Then there exists a constant K, $0 < K < \infty$, such that*

$$E|s_n^2 - \sigma^2|^r \leqq n^{1-r}\xi[2^{2r} + Kn^{-1}] , \qquad \textit{all } n ,$$

where $s_n^2 = n^{-1} \sum_{i=1}^n (U_i - \bar{U})^2$ is the sample variance.

PROOF. Without loss of generality we can assume that $\mu = 0$. Let $Y_i = U_i^2 - \sigma^2$, and note that by the C_r inequality,

$$E|Y_i|^r \leqq 2^{r-1}(E|U_i|^{2r} + \sigma^{2r}) \leqq 2^r\xi ,$$

while

$$E|s_n^2 - \sigma^2|^r \leqq 2^{r-1}(E|n^{-1} \textstyle\sum_{i=1}^n Y_i|^r + E|\bar{U}|^{2r}) .$$

By Theorem 2 of Von Bahr and Esseen (1965),

$$E|n^{-1} \textstyle\sum_{i=1}^n Y_i|^r \leqq 2n^{1-r}E|Y_i|^r ,$$

while by Chung's inequality (see Bickel and Yahav (1968, page 451)) we have

$$E|\bar{U}|^{2r} \leqq (n^{-r})(K_r)\xi ,$$

where K_r is a finite constant depending only on r. Putting all of these inequalities together (with $K = 2^{r-1}K_r$), yields the desired result. □

It now follows from (6.11), (6.12), (6.14), (6.15), (6.16), (6.18), (6.19), assumption C.2, and the fact that

$$(n/\log n)(m(n)/\log m(n))^{-1} \leqq 2k ,$$

that (6.2) holds. This, by Theorem 6.1, completes the proof of Theorem 6.2. □

Perhaps the most restrictive condition in assumptions C.1—C.3 is C.3. If $g(\theta) = \theta$, then (of course) this condition is satisfied trivially, resulting in the following important corollary to Theorem 6.2.

COROLLARY 6.4. *If assumptions* C.1 *and* C.2 *are satisfied for* $g(\theta) = \theta$ [*in particular, if* θ *is estimable by a statistic with finite third absolute moment, all* θ], *and if assumptions* A.0—A.5 *hold, then the class of stopping rules* $\{t(c): c > 0\}$ *defined by* (5.6) *is* A.O.

Corollary 6.4 in particular covers the problem of estimating the mean in an exponential family, and many other estimation problems connected with the exponential family. It is not obvious, however, that either Corollary 6.4 or Theorem 6.2 covers the problem of estimating the natural parameter $\theta = \log[p/(1-p)]$ in the exponential family representation of the binomial distribution, since θ in this case is not estimable and the obvious choice $g(\theta) = p = e^\theta(1 + e^\theta)^{-1}$ to use in Theorem 6.2 does not have an inverse which obeys a uniform Lipschitz condition. In cases, however, in which

$$f(x \mid \theta) = e^{\theta T(x) - c(\theta)}\, d\mu(x) ,$$

where $\mu(x)$ is a σ-finite, nonnegative measure, and $c''(\theta) \geqq q > 0$, all $\theta \in \Theta$, then Theorem 6.2 is applicable with $g(\theta) = c'(\theta)$, $k = 1$, $v(X) = T(x)$. Here, $g(\theta)$ is strictly increasing since

$$c''(\theta) = \operatorname{Var}(T(X)) > 0 ,$$

and $g^{-1}(\theta)$ exists and satisfies a uniform Lipschitz condition since

$$\frac{d}{d\theta}\, g^{-1}(\theta) = \frac{1}{c''(\theta)} \leqq \frac{1}{q} < \infty .$$

Of course, $\psi(\theta)$ is still restricted by (6.5), but the major obstacles to application of Theorem 6.2 (namely, C.1 and C.3) have been overcome.

In cases where Theorem 6.2 does not seem to be applicable, an almost complete relaxation of C.3 combined with a stronger version of C.2 may sometimes help. Let $g(\theta)$ satisfy condition C.1, and define for each $\theta \in \Theta$, $\varepsilon > 0$,

$$Q_\varepsilon(\theta) = \sup_{|a-\theta|<\varepsilon;\, a \in \Theta} \frac{|a - \theta|}{|g(a) - g(\theta)|} .$$

Replace C.2 by the stronger requirement:

C.2′. Equation 6.5 holds; in addition there exists $\varepsilon^* > 0$ such that

$$\int_\Theta Q_{\varepsilon^*}(\theta)\sigma(\theta)\psi(\theta)\,d\theta < \infty\,. \tag{6.20}$$

Then, except for assuming that $g(\theta)$ has an inverse function, we can delete condition C.3, and repeat the proof of Theorem 6.2, replacing the constant Q in that proof everywhere by $Q_{\varepsilon^*}(\theta)$. This approach yields:

THEOREM 6.5. *Under* A.0—A.5, C.1, C.2′ *and under the assumption that* $g(\theta)$ *defined in* C.1 *is invertible, the rules defined by* (5.6) *are* A.O.

7. Generalizations and conclusions. The present paper has dealt only with interval estimation of a scalar parameter θ. This specialization has been partly for the sake of clarity and convenience, but mostly because the criteria which we have used to evaluate regional estimators in the scalar parameter case (length and coverage) have a variety of possible generalizations when vector parameters are considered. Indeed, although intervals are the natural region to use in the scalar case, the "natural" region to use in the vector case is far from clear. We could use spheres [Gleser (1965)], ellipsoids [Albert (1966)], rectangles [Callahan (1969)], or any other convex region. Nonconvex regions could also be considered, although these do not necessarily yield *interval* estimators for linear combinations $\mathbf{d}'\boldsymbol{\theta}$ of the parameter via the standard projection method (a desirable property in applications).

Given that we have chosen a desirable shape for our regions, what do we use to replace length in the loss function (2.1)? The Lebesgue measure (volume) of the region is one obvious choice, but an equally good choice [Gleser (1965), Callahan (1969)] is the diameter of the region, since this is the longest length of any derived interval estimator for a linear combination $\mathbf{d}'\boldsymbol{\theta}$.

Suppose we decide to use the volume, $l(C)$, of the region C. That is, suppose that our loss function is of the form

$$L(\theta, C, n) = al(C) + b(1 - \delta_C(\theta)) + cn\,. \tag{7.1}$$

The results of Section 2 are then easily generalized to show that the region

$$C^*(X_n) = \text{closure} \quad (\text{in } \Theta) \quad \text{of} \quad \left\{\boldsymbol{\theta} : \psi(\boldsymbol{\theta} \mid \mathbf{X}_n) \geqq \frac{a}{b}\right\} \tag{7.2}$$

is the Bayes-optimal terminal estimation procedure against the prior $\psi(\boldsymbol{\theta})$—provided, of course, that $C^*(X_n)$ has the shape which we desire. We can refer to the multivariate normal posterior distribution for guidance as to the proper choice of the function, $f_p(x)$, to be used to replace $(x/\log x)^{\frac{1}{2}}$ when $\boldsymbol{\theta}$ is $p \times 1$. It can thus be shown that $f_p(x) = (x/\log x)^{p/2}$. Under conditions similar to those in Bickel and Yahav (1969 b), and which straightforwardly generalize conditions A.1—A.5 of Section 5, it should then be possible to construct a class of A.P.O. stopping rules. Then by suitably generalizing conditions C.1—C.3 we can prove that this A.P.O. class of stopping rules is A.O. The sufficient conditions and details of these results are planned for a forthcoming paper.

Another generalization which is left for future work is the problem of estimating a scalar function $g(\boldsymbol{\theta})$ of a vector parameter $\boldsymbol{\theta}$. One special case of this problem—estimating the mean of a normal distribution with unknown variance under a conjugate prior—is treated in detail in Kunte's thesis (1973). When $g(\boldsymbol{\theta})$ has bounded first partial derivatives in the elements of $\boldsymbol{\theta}$, this problem can also be treated by the methods of Sections 5 and 6. Treatment of the case when the function $g(\boldsymbol{\theta})$ is vector-valued is planned for a forthcoming paper.

Finally, we have not treated questions of the robustness of our sequential procedures against changes in the prior distribution ψ (*asymptotic minimaxity*). This problem appears to be amenable to the kinds of arguments used in Bickel and Yahav (1968) and in Pessoa's thesis (1971).

What we have shown in the present paper is that there exists a class of asymptotically optimal (and asymptotically pointwise optimal) stopping rules for our Bayes sequential interval estimation problem. This class is clearly not unique, and much work remains to be done, both in determining A.O. stopping rules which have superior second-order asymptotic properties, and in comparing the resulting stopping rules to the sequential rules of the kind proposed for the more classical fixed width, fixed confidence statistical approaches to sequential confidence interval estimation (see Section 1).

8. Acknowledgments. The authors are indebted to Herman Rubin and David H. Root for helpful discussions, and to an associate editor and referee whose careful reading found several mistakes and missing references.

REFERENCES

[1] ALBERT, A. (1966). Fixed size confidence ellipsoids for linear regression parameters. *Ann. Math. Statist.* **37** 1602-1630.

[2] ARROW, K., BLACKWELL, D. and GIRSHICK, M. (1949). Bayes and minimax solution of sequential decision problems. *Econometrica* **17** 213-244.

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

[4] BICKEL, P. J. and YAHAV, J. A. (1968). Asymptotically optimal Bayes and minimax procedures in sequential estimation. *Ann. Math. Statist.* **39** 442-456.

[5] BICKEL, P. J. and YAHAV, J. A. (1969a). On an A.P.O. rule in sequential estimation with quadratic loss. *Ann. Math. Statist.* **40** 417-426.

[6] BICKEL, P. J. and YAHAV, J. A. (1969b). Some contributions to the asymptotic theory of Bayes solutions. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **11** 257-276.

[7] BLUMENTHAL, S. (1970). Interval estimation of the normal mean subject to restrictions, when the variance is known. *Naval. Res. Logist. Quart.* **17** 485-505.

[8] CALLAHAN, J. (1969). On some topics in sequential multiparameter estimation. Ph. D. dissertation, The Johns Hopkins Univ.

[9] CHOW, Y. S. and ROBBINS, H. (1965). On the asymptotic theory of fixed width sequential confidence intervals for the mean. *Ann. Math. Statist.* **36** 457-462.

[10] COHEN, A. and STRAWDERMAN, W. (1973). Admissible confidence interval and point estimation for translation or scale parameters. *Ann. Statist.* **1** 545-550.

[11] DE GROOT, M. H. (1970). *Optimal Statistical Decisions*. McGraw-Hill, New York.
[12] FELLER, WILLIAM (1966). *An Introduction to Probability Theory and Its Applications* **2**. Wiley, New York.
[13] GLESER, L. J. (1965). On the asymptotic theory of fixed-size sequential confidence bounds for linear regression parameters. *Ann. Math. Statist.* **36** 463–467.
[14] GLESER, L. J. (1969). On limiting distributions for sums of a random number of independent random vectors. *Ann. Math. Statist.* **40** 935–941.
[15] GLESER, L. J. and KUNTE, S. (1973). On asymptotically optimal and asymptotically pointwise optimal stopping rules. Mimeograph Series No. 341, Department of Statistics, Purdue Univ.
[16] JOSHI, V. M. (1969). Admissibility of the usual confidence sets for the mean of a univariate or bivariate normal population. *Ann. Math. Statist.* **40** 1042–1067.
[17] KIEFER, J. and SACKS, J. (1963). Asymptotically optimal sequential inference and design. *Ann. Math. Statist.* **34** 705–750.
[18] KUNTE, S. (1973). Asymptotically pointwise optimal and asymptotically optimal stopping rules for sequential Bayes confidence interval estimation. Mimeograph Series No. 328, Department of Statistics, Purdue Univ. (Ph. D. dissertation.)
[19] KUNTE, S. and GURJAR, R. V. (1973). Characterization theorem for a class of functions. Mimeograph Series No. 338, Department of Statistics, Purdue Univ.
[20] PAULSON, E. (1969). Sequential interval estimation for the mean of normal populations. *Ann. Math. Statist.* **40** 509–516.
[21] PESSOA, D. G. C. (1971). Asymptotically minimax fixed length confidence intervals. Ph.D. dissertation, Univ. of California, Berkeley.
[22] ROYDEN, H. L. (1968). *Real Analysis*, 2nd ed. MacMillan, London.
[23] RUBIN, H. and SETHURAMAN, J. (1965 a). Probabilities of moderate deviations. *Sankhyā Ser. A* **27** 325–346.
[24] RUBIN, H. and SETHURAMAN, J. (1965 b). Bayes risk efficiency. *Sankhyā Ser. A* **27** 347–356.
[25] SEN, P. K. and GHOSH, M. (1971). On bounded length sequential confidence intervals based on one-sample rank-order statistics. *Ann. Math. Statist.* **42** 189–203.
[26] VON BAHR, B. and ESSEEN, C. (1965). Inequalities for the rth absolute moment of a sum of random variables, $1 \leqq r \leqq 2$. *Ann. Math. Statist.* **36** 299–303.
[27] WEISS, L. and WOLFOWITZ, J. (1972). Optimal, fixed length, nonparametric sequential confidence limits for a translation parameter. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **24** 203–209.
[28] WINKLER, R. L. (1972). Decision-theoretic approach to interval estimation. *J. Amer. Statist. Assoc.* **67** 187–191.

DEPARTMENT OF STATISTICS
PURDUE UNIVERSITY
WEST LAFAYETTE, INDIANA 47907

DEPARTMENT OF MATHEMATICS AND STATISTICS
UNIVERSITY OF POONA
POONA 7, INDIA