The Annals of Mathematical Statistics
1968, Vol. 39, No. 2, 442–456

ASYMPTOTICALLY OPTIMAL BAYES AND MINIMAX PROCEDURES IN SEQUENTIAL ESTIMATION

By Peter J. Bickel[1] and Joseph A. Yahav[2]

University of California, Berkeley and University of Tel Aviv

1. Introduction and summary. In [4] we introduced a general method for obtaining asymptotically pointwise optimal procedures in sequential analysis when the cost of observation is constant. The validity of this method in both estimation and testing was established in [4] for Koopman-Darmois families, and in [5] for the general case.

Section 2 of this paper generalizes Theorem 2.1 of [4] to cover essentially the case of estimation with variable cost of observation.

In Section 3 we show that in estimation problems, under a very weak condition, for constant cost of observation, the asymptotically pointwise optimal rules we propose are optimal in the sense of Kiefer and Sacks [9]. The condition given is further investigated in the context of Bayesian sequential estimation in Section 4 and is shown to be satisfied if reasonable estimates based on the method of moments exist. In Section 5 we consider the robustness of our rules under a change of prior. The main result of this section is given by Theorem 5.1. Finally Theorem 5.2 deals with a generalization of Wald's [12] theory of asymptotically minimax rules and an application of that theory to the Bayesian model.

2. A general theorem on asymptotic pointwise optimality. We use the notation of [4]. $\{Y_n\}$, $n \geqq 1$, is a sequence of random variables defined on a probability space $(\Omega, \mathfrak{F}, P)$ where Y_n is $\mathfrak{F}_n$ measurable and $\mathfrak{F}_n \subset \mathfrak{F}_{n+1} \cdots \subset \mathfrak{F}$ is an increasing sequence of σ fields.

We assume,

$$P[Y_n > 0] = 1, \tag{2.1}$$

$$P[Y_n \to 0] = 1. \tag{2.2}$$

Let $K(n)$, $n \geqq 1$, be a sequence of positive constants such that,

$$K(n) \uparrow \infty \tag{2.3}$$

strictly as $n \to \infty$.

Y_n in the statistical model represents the Bayes stopping risk and $K(n)c$ the cost of n observations. c is a general cost parameter which we permit to tend to 0. The most interesting special case studied in [4] is, of course, $K(n) = n$. Define,

$$X(n, c) = Y_n + K(n)c. \tag{2.4}$$

Received 26 January 1967; revised 28 August 1967.

[1] Supported in part by Grant GP-5059 of the National Science Foundation.

[2] Supported in part by Grant GP-5705 of the National Science Foundation.

We are interested in finding a sequence of stopping variables $\{t(c)\}$ such that the $t(c)$ are asymptotically pointwise optimal (APO), that is, such that,

$$X(t(c), c)[\inf_{s \varepsilon T} X(s, c)]^{-1} \to 1 \quad \text{a.s.} \quad \text{as } c \to 0 \tag{2.5}$$

where T is the set of all stopping variables defined on the sequence $\{\mathfrak{F}_n\}$.

As usual, a stopping variable t is a natural number valued random variable such that the event $[t = n] \varepsilon \mathfrak{F}_n$.

We now specialize our model as follows: Suppose

A.1 $n^\beta Y_n \to V$ a.s, where $\beta > 0$ and $P[0 < V < \infty] = 1$.

A.2 For any $c > 0$, $x > 0$ there exists $N(x, c)$ which minimizes $h(x, c, n) = xn^{-\beta} + cK(n)$ and $N(x, c)$ may be taken to be the first n such that $\Delta(h(x, c, n)) \geqq 0$, where Δ is the one step difference operator.

A.3 $n^{-(\beta+1)} = o(\Delta(K(n)))$

THEOREM 2.1. *Under* (2.1), (2.2), (2.3), (2.4) *and* A.1 A.2 A.3 *the stopping variable* $\bar{t}(c)$: *stop for the first n such that*

$$Y_n(1 - (n/(n + 1))^\beta) \leqq c[K(n + 1) - K(n)] \tag{2.6}$$

is APO.

PROOF. From A.1 and A.3 we see that,

$$P[\bar{t}(c) < \infty] = 1 \quad \text{for all} \quad c > 0. \tag{2.7}$$

By A.1 $Y_n(1 - (n/(n + 1))^\beta$ is asymptotically equivalent to $Vn^{-(\beta+1)}$ and so by A.3 (2.7) holds. By A.2 we have:

$$X(\bar{t}(c), c) = \min_n h(\bar{t}^\beta(c) Y_{\bar{t}(c)}, c, n). \tag{2.8}$$

Let $n_0(c)$ be the first n for which

$$X(n_0(c), c) = \min_n X(n, c). \tag{2.9}$$

The a.s. existence of $n_0(c)$ follows from (2.2) and (2.3). By (2.8) and (2.9) we get

$$X(n_0(c), c) \leqq X(\bar{t}(c), c) \leqq \bar{t}^\beta(c) Y_{\bar{t}(c)}/n_0^\beta(c) + cK(n_0(c)). \tag{2.10}$$

Relations (2.1), (2.3) and (2.9) imply that as $c \to \infty$

$$n_0(c) \uparrow \infty \quad \text{a.s.} \tag{2.11}$$

and similarly (2.1), (2.3) and (2.8) imply that

$$\bar{t}(c) \uparrow \infty. \tag{2.12}$$

By (2.21) we get $\bar{t}^\beta(c)\ Y_{\bar{t}(c)} \to V$ a.s. and by (2.11) $n_o(c) Y_{n_o(c)} \to V$ a.s. Remark that for any three sequences of real numbers x_n, x_n', a_n such that $x_n \to x$, $x_n' \to x$ with $x \neq 0$ we have $(x_n + a_n)/(x_n' + a_n) \to 1$. Upon dividing the left-hand side of (2.10) by the right-hand side and using the preceding remarks, we see that,

$$\limsup_{c\to 0} X(\bar{t}(c), c)/X(n_0(c), c) = 1. \tag{2.13}$$

Of course, (2.13) implies the theorem.

For convenience we state as a corollary,

COROLLARY 2.2. *Under the assumptions of Theorem* 2.1,

$$\limsup_{c\to 0} \{X(\bar{t}(c), c)\} [\inf_n h(V, c, n)]^{-1} = 1. \tag{2.14}$$

PROOF. By A.1 given $\epsilon > 0$ there exists $N(\epsilon)$ possibly depending on the sample sequence such that

$$(1-\epsilon)n^{-\beta}V \leqq Y_n \leqq (1+\epsilon)n^{-\beta}V \quad \text{a.s.} \tag{2.15}$$

for $n > N(\epsilon)$. Then,

$$(1-\epsilon)\inf_{n\geqq N(\epsilon)}\{Vn^{-\beta} + cK(n)\} \leqq \inf_{n\geqq N(\epsilon)} X(n, c) \leqq (1+\epsilon)\inf_{n\geqq N(\epsilon)}\{Vn^{-\beta} + cK(n)\}. \tag{2.16}$$

Clearly a.s., whatever be $\epsilon > 0$, one may find c sufficiently small, possibly depending on the sample sequence, such that,

$$\inf_{n>N(\epsilon)} \{Vn^{-\beta} + cK(n)\} = \inf_n h(V, c, n) \tag{2.17}$$

and

$$\inf_{n>N(\epsilon)} X(n, c) = X(n_0(c), c). \tag{2.18}$$

By (2.16), (2.17) and (2.18) the corollary follows.

If we put $K(n) = n$ we can easily compute from (2.14)

$$X(t(c), c)c^{-\beta(\beta+1)^{-1}} \to (1+\beta^{-1})(V\beta)^{(\beta+1)^{-1}} \quad \text{a.s.} \tag{2.19}$$

Moreover by arguments similar to those used in [4] in this instance, if $s(c)/\bar{t}(c) \to 1$ a.s. then $s(c)$ is also APO.

In particular if we take the natural approximation to $t(c)$ and consider $t'(c)$ given by,

"Stop as soon as $Y_n\beta(n+1)^{-1} \leqq c$"

we can easily conclude that $t'(c)$ is APO.

We note that changes in $K(n)$ can radically affect the relative importance of the "stopping risk" $Y_{\bar{t}}$ and the cost of observation $K(\bar{t}(c))c$.

Thus in [4] it is shown, for $\beta = 1$, $K(n) = n$ that

$$Y_{\bar{t}}(c) \sim [cV]^{\frac{1}{2}}, \tag{2.20}$$

$$c\bar{t}(c) \sim Y_{\bar{t}(c)} \tag{2.21}$$

a.s. P.

More generally, by a careful examination of the argument of Theorem 2.1 or arguments similar to those employed in [4] we can show if $K(n) = n$, and β is arbitrary, positive

$$Y_{\bar{t}}(c) \sim (c^{\beta}V\beta)^{(\beta+1)^{-1}}\beta^{-1}, \tag{2.22}$$

$$c\bar{t}(c) \sim \beta Y_{\bar{t}}(c). \tag{2.23}$$

On the other hand if for instance, $K(n) = \log n$ or equivalently $K(n+1) - K(n) = n^{-1}$,

$$Y_{\bar{t}}(c) = o_p(c\bar{t}(c))$$

and if $K(n) = n^2$

$$c\bar{t}(c) = o_p(Y_{\bar{t}(c)}).$$

3. Asymptotic optimality. The easy but useful theorem of this section gives a simple criterion for the rules $\bar{t}$, t' to be asymptotically optimal rather than merely pointwise optimal for the special case $K(n) = n$. Following Kiefer and Sacks [9] we say $t(c)$ is an asymptotically optimal (sequence of) stopping rule(s) if,

$$\limsup_{c\to 0} E(X(t(c), c)) [\inf \{E(X(s, c)) : s \in S\}]^{-1} \leqq 1 \tag{3.1}$$

where S is the set of all stopping rules.

We have,

THEOREM 3.1. *If the sequence* $\{Y_n\}$ *obeys the conditions of Theorem* 2.1 *with* $K(n) = n$ *and*

$$\sup n^{\beta} E(Y_n) < \infty, \tag{3.2}$$

then the rules $\bar{t}(c)$, $t'(c)$ *are asymptotically optimal.*

The proof hinges on an elementary lemma.

LEMMA 3.2. *Suppose* $\{R_n\}$ *is a sequence of random variables on some probability space* $(\Omega, \mathfrak{F}, P)$ *tending in law to some random variable* R.

Let $a_{mn} = P\{|R_n| > m\}$,
and $\sup_n a_{mn} = a_m$. *Then if,*

$$\sum a_m < \infty, \tag{3.3}$$

$E(R_n)$ *and* $E(R)$ *are finite and,*

$$E(R_n) \to E(R). \tag{3.4}$$

PROOF.

$$\begin{aligned} \int_{[R_n > m]} |R_n|\, dP &\leqq \textstyle\sum_{k=m}^{\infty} (k+1) P[k < |R_n| \leqq k+1] \\ &= (m+1) P[|R_n| > m] + \textstyle\sum_{k=m}^{\infty} P[|R_n| > k] \\ &\leqq (m+1) a_m + \textstyle\sum_{k=m}^{\infty} a_k . \end{aligned} \tag{3.5}$$

If (3.3) holds since $a_m \downarrow$ by a classical lemma of Abel's, $m a_m \to 0$. (3.3) thus implies uniform integrability of the $\{R_n\}$ sequence and the lemma is proved.

PROOF (of Theorem 3.1). From Theorem 2.1 and (2.14) we can conclude that for any sequence of stopping rules $s(c)$,

$$\liminf_c X(s(c), c)[c^{-\beta/(\beta+1)}] \geqq (1 + \beta^{-1})(V\beta)^{(\beta+1)^{-1}} \quad \text{a.s.} \tag{3.6}$$

and hence by Fatou's theorem

$$\liminf_c E(X(s(c), c))\, c^{-\beta/(\beta+1)} \geqq (1 + \beta^{-1})\beta^{(\beta+1)^{-1}} E(V^{(\beta+1)^{-1}}). \tag{3.7}$$

Since by definition,

$$(3.8) \qquad Y_{t'(c)} \leqq c\beta^{-1}t'(c)$$

by (3.7) to prove the theorem it suffices to show

$$(3.9) \qquad c^{(\beta+1)^{-1}}E(t'(c)) \to \beta^{(\beta+1)^{-1}}E(V^{(\beta+1)^{-1}}).$$

By (2.23), (2.24) and the lemma it's enough to show,

$$(3.10) \qquad \sum_m a_m < \infty$$

where $a_m = \sup\{P[c^{(\beta+1)^{-1}}t'(c) > m] : 0 < c < 1\}$. Now,

$$(3.11) \qquad P[c^{-(\beta+1)^{-1}}t'(c) > m] \leqq P[Y_{[c^{-(\beta+1)^{-1}}m]} > \beta^{-1}[c^{-(\beta+1)^{-1}}m]]$$

where $[x]$ is the integer part of x.
Applying Markov's inequality,

$$(3.12) \qquad P\{Y_{[c^{-(\beta+1)^{-1}}m]} > \beta^{-1}c[c^{-(\beta+1)^{-1}}m]\} \leqq \alpha\beta c^{-1}[c^{-(\beta+1)^{-1}}m]^{-(1+\beta)}$$

where, $\alpha = \sup_n E(n^\beta Y_n)$.

Now,

$$(3.13) \qquad m^{1+\beta}c^{-1}[c^{-(\beta+1)^{-1}}m]^{-(1+\beta)} \leqq (1 - c^{-(\beta+1)^{-1}}m^{-1})^{-(1+\beta)} \leqq (1 - m^{-1})^{-(1+\beta)}.$$

We conclude,

$$(3.14) \qquad a_m \leqq \min(1, \alpha\beta(\tfrac{1}{2}m)^{-(\beta+1)})$$

and the theorem follows for $t'(c)$.

To prove the result for $\bar{t}(c)$, we note that since,

$$(3.15) \qquad \bar{t}(c) \leqq t'(c)$$

and

$$(3.16) \qquad c^{-(\beta+1)^{-1}}\bar{t}(c) \to (V\beta)^{(\beta+1)^{-1}} \quad \text{a.s.},$$

(3.9) implies that,

$$(3.17) \qquad c^{-(\beta+1)^{-1}}E(\bar{t}(c)) \to \beta^{(\beta+1)^{-1}}E(V).$$

Now,

$$(3.18) \qquad Y_{\bar{t}(c)} \leqq [1 - (1 - (\bar{t}(c) + 1)^{-1})^\beta]^{-1}c \leqq [1 - (1 - (t'(c) + 1)^{-1})^\beta]^{-1}c,$$

and

$$(3.19) \qquad c^{(\beta+1)^{-1}}[1 - (1 - (t'(c) + 1)^{-1})^\beta]^{-1} \to \beta^{-1}(V\beta)^{(\beta+1)^{-1}} \quad \text{a.s.}$$

But

$$(3.20) \qquad P[c^{(\beta+1)^{-1}}[1 - (1 - (t'(c) + 1)^{-1})^\beta]^{-1} > m] \leqq \min(1, \alpha\beta\gamma(\tfrac{1}{2}m)^{-(1+\beta)}) \qquad \text{for} \qquad 0 < c \leqq \tfrac{1}{2},$$

where

$$\gamma = \sup \{[(1 - x)^{-1/\beta} - 1]^{\beta+1} x^{-(\beta+1)}, 0 < x \leqq \tfrac{1}{2}\}.$$

The theorem is proved for $\tilde{t}$ by another application of Lemma 3.2. □

REMARK. This method may also be applied to obtain conditions for the asymptotic optimality of t' and $\tilde{t}$ even when $K(n) \nsim n$. However, whereas in the Bayesian inference situation the condition that $E(n^\beta Y_n)$ be bounded is easily interpretable and verifiable, this is no longer the case in general. In fact it seems more practical to attack the structure of the stopping rule using the special properties of the statistical situation as was done for instance by Kiefer and Sacks in [9] for testing.

We conclude with a counterexample which indicates that if $K(n) = n$ some condition such as (3.2) cannot be dispensed with.

Let $Y_n = (n^{-1} - n^{-2})V + an^{-2}$ where a is a fixed positive constant and V is a positive random variable such that,

$$E(V^{\frac{1}{2}}) = \infty. \tag{3.21}$$

Then $nY_n \to V$ and Y_n satisfies the conditions of Theorem 2.1. In particular if $K(n) = n$, the rule t' is asymptotically pointwise optimal. However, t' is not optimal. In fact,

$$E[t'(c)] = \infty \quad \text{for} \quad c < 2a \tag{3.22}$$

since, $t'(c) > 1$ if $c < 2a$ and in that case,

$$c[t'(c)]^2 \geqq \tfrac{1}{2}V. \tag{3.23}$$

On the other hand the rule which stops on the first observation has finite risk.

4. Bayes sequential estimation. As usual we assume that we observe in succession $z_1, z_2, \cdots$ a sequence of independent identically distributed random variables whose distribution has a density function $f(z, \boldsymbol{\theta})$ with respect to some σ finite measure μ on R for each $\boldsymbol{\theta} \in \Theta$. Let $P_{\boldsymbol{\theta}}$ denote the measure induced by the $\{z_i\}$ on R^∞ endowed with its product Borel field. We take Θ—the parameter space to be an open subset of R^k and endow Θ with the Borel field. We are given a prior density ψ with respect to Lebesgue measure on Θ viz. a nonnegative Borel measurable function on Θ such that,

$$\int_\Theta \psi(\boldsymbol{\theta})\, d\boldsymbol{\theta} = 1. \tag{4.1}$$

Suppose that we are interested in estimating $g(\boldsymbol{\theta})$ where g is real valued and take D—the decision space to be R. In order to specify that this is a point estimation situation we consider a loss function $l(\theta, d)$ such that,

$$l(\boldsymbol{\theta}, d) = \tilde{l}(|g(\boldsymbol{\theta}) - d|) \tag{4.2}$$

where $\tilde{l}$ is monotonely strictly increasing in s and $\tilde{l}(s) = 0$. A sequential procedure δ in this situation then consists of a stopping variable t and an estimate $g\hat{\theta}(z_1, \cdots, z_t)$. If the cost of the nth observation is $[K(n) - K(n - 1)]c$ the overall risk of δ is given by,

$$
\begin{aligned}
R_\delta(c) &= \textstyle\int_\Theta [E_{\boldsymbol\theta}(l(\boldsymbol\theta, g\hat\theta(z_1, \cdots, z_t))) + cE_{\boldsymbol\theta}(K(t))]\psi(\boldsymbol\theta)\, d\boldsymbol\theta \\
&= \textstyle\sum_{n=1}^\infty \int\int_{[t=n]} l(\boldsymbol\theta, g\hat\theta(z_1, \cdots, z_n)) \prod_{i=1}^n f(z_i, \theta)\mu(dz_i)\psi(\boldsymbol\theta)\, d\boldsymbol\theta \\
&\quad + c \textstyle\int_\Theta E_{\boldsymbol\theta}(K(t))\psi(\boldsymbol\theta)\, d\boldsymbol\theta
\end{aligned} \tag{4.3}
$$

where the subscript $\boldsymbol\theta$ for E indicates that computation is carried out when $\boldsymbol\theta$ is true. For simplicity we assume we must take at least one observation. It readily follows (cf [3]) that if there exists for every n, a measurable estimate $g\theta'(z_1, \cdots, z_n)$ such that,

$$
\begin{aligned}
&\textstyle\int l(\boldsymbol\theta, (g\theta')(z_1, \cdots, z_n)) \prod_{i=1}^n f(z_i, \boldsymbol\theta)\psi(\boldsymbol\theta)\, d\theta \\
&\qquad = \min \textstyle\int l(\boldsymbol\theta, d) \prod_{i=1}^n f(z_i, \boldsymbol\theta)\psi(\boldsymbol\theta)\, d\boldsymbol\theta
\end{aligned} \tag{4.4}
$$

then we need only consider δ of the form $(t, g\theta'(z_1, \cdots, z_t))$.

Define,

$$Y_n = \textstyle\int l(\boldsymbol\theta, (g\theta')(z_1, \cdots, z_n))\psi(\boldsymbol\theta \mid z_1, \cdots, z_n)\, d\boldsymbol\theta, \tag{4.5}$$

where

$$\psi(\boldsymbol\theta \mid z_1, \cdots, z_n) = \textstyle\prod_{i=1}^n f(z_i, \boldsymbol\theta)[\int \prod_{i=1}^n f(z_i, \boldsymbol\tau)\psi(\boldsymbol\tau)\, d\boldsymbol\tau]^{-1} \tag{4.6}$$

if the denominator of the right hand side is positive and 0 otherwise.

Then, the problem of finding the optimal (Bayes) procedure $\delta(c)$ reduces to finding that $t(c)$ (if any exists) such that,

$$E(Y_{t(c)}) + cE(K(t(c))) = \min\{E(Y_s) + cE(K(s)) : s \in S\}. \tag{4.7}$$

S here denotes the set of all stopping rules and for any X measurable with respect to the z_i sequence,

$$E(X) = \textstyle\int E_{\boldsymbol\theta}(X)\psi(\boldsymbol\theta)\, d\boldsymbol\theta. \tag{4.8}$$

In accordance with [5] we now make the following further assumptions which we break up into:

(1) *Assumptions about* $\tilde l$.

(a) $\tilde l$ is continuously differentiable, $\tilde l'(t) > 0$ for $t > 0$.

(b) There exist $r < \infty$, such that,

$$\limsup_{t\to\infty} t^{-r}\tilde l'(t) < \infty.$$

(c) There exists $s \geqq 0, \gamma > 0$ such that,

$$\lim_{t\to 0} t^{-s}\tilde l'(t) = \gamma.$$

(2) *Assumptions about* ψ:

(a) ψ is positive, continuous and bounded on Θ.

(b) $\int |\theta_i|^r\psi(\boldsymbol\theta)\, d\boldsymbol\theta < \infty$, where $\boldsymbol\theta = (\theta_1, \cdots, \theta_k)$, and where r is given in 1b.

(3) *Assumptions about f*: Let $\log f(z, \boldsymbol{\theta}) = \Phi(z, \boldsymbol{\theta})$. We assume,

(a) $\partial^2\Phi(z, \boldsymbol{\theta})/\partial\theta_i\partial\theta_j$ is finite and continuous in $\boldsymbol{\theta}$ for almost all z(a.e. μ) and all i, j if $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_k)$.

(b) $E_{\boldsymbol{\theta}}\, (\sup \{[\Phi(z_1, \mathbf{s}) - \Phi(z_1, \boldsymbol{\theta})] : \|\mathbf{s} - \boldsymbol{\theta}\| \geqq \epsilon, \mathbf{s} \in \Theta\}) < 0$ for all $\boldsymbol{\theta} \in \Theta$ and $\epsilon > 0$. $\| \ \|$ here denotes the usual Euclidean norm.

(c) $E_{\boldsymbol{\theta}}\, \{\sup |\partial^2\Phi(z_1, \mathbf{s})/\partial\theta_i\partial\theta_j| : \|\mathbf{s} - \boldsymbol{\theta}\| < \epsilon, \mathbf{s} \in \Theta\} < \infty$, for all $\boldsymbol{\theta}$, some $\epsilon(\boldsymbol{\theta}) > 0$.

Assumptions 3(a) and (c) imply,

$$E_{\boldsymbol{\theta}}(\partial\Phi(z_1, \boldsymbol{\theta})/\partial\theta_i) = 0 \qquad \text{for} \quad i = 1, \cdots, k. \tag{4.9}$$

Denote the covariance matrix of

$$(\partial\Phi(z_1, \boldsymbol{\theta})/\partial\theta_1, \cdots, \partial\Phi(z_1, \boldsymbol{\theta})/\partial\theta_k) \qquad \text{by} \quad A(\boldsymbol{\theta}),$$

and the matrix whose ijth entry is $E_{\boldsymbol{\theta}}(\partial\Phi(z_1, \boldsymbol{\theta})/\partial\theta_i\partial\theta_j)$ by $A^*(\boldsymbol{\theta})$. (4.9) and Assumptions 3(a) and (c) imply,

$$A(\boldsymbol{\theta}) = -A^*(\boldsymbol{\theta}). \tag{4.10}$$

We assume,

(d) $A(\boldsymbol{\theta})$ is positive definite for all $\boldsymbol{\theta}$.

(4) *Assumptions about g*:

(a) For simplicity suppose $(\partial g(\boldsymbol{\theta})/\partial\theta_i)$ is continuous in $\boldsymbol{\theta}$ for $i = 1, \cdots, k$. Then

$$\text{grad}\, g(\boldsymbol{\theta}) = (\partial g(\boldsymbol{\theta})/\partial\theta_1, \cdots, \partial g(\boldsymbol{\theta})/\partial\theta_k) \tag{4.11}$$

is a total differential and we suppose,

(4.12) (b) $\text{grad}\, g(\boldsymbol{\theta}) \neq \mathbf{0}$ for all $\boldsymbol{\theta}$,

$$\limsup_{\boldsymbol{\theta}\to\infty} |g(\boldsymbol{\theta})| \left(\textstyle\sum_{i=1}^k |\theta_i|\right)^{-1} < \infty.$$

(5) *"Joint" assumptions*: We suppose the equation,

$$\int_{[d>g(\boldsymbol{\theta})]} \tilde{l}'(d - g(\tilde{\theta}))\psi(\boldsymbol{\theta} \mid z_1, \cdots, z_n)\, d\boldsymbol{\theta} = \int_{[d<g(\boldsymbol{\theta})]} \tilde{l}'(g(\boldsymbol{\theta}) - d)\psi(\boldsymbol{\theta} \mid z_1, \cdots, z_n)\, d\boldsymbol{\theta} \tag{4.13}$$

has a measurable solution $d = (g\theta_n')(z_1, \cdots, z_n)$. This is true for instance if $\tilde{l}$ is convex.

In [5] it is shown that, under these assumptions,

$$Y_n = \int l(\boldsymbol{\theta}, (g\theta_n')(z_1, \cdots, z_n))\psi(\boldsymbol{\theta} \mid z_1, \cdots, z_n)\, d\boldsymbol{\theta} \tag{4.14}$$

and

$$n^{(s+1)/2}Y_n \to [A_g(\boldsymbol{\theta})]^{(s+1)/2}\gamma(s + 1)^{-1}\mu_{s+1} \qquad \text{a.s. } P_{\boldsymbol{\theta}} \tag{4.15}$$

where

$$A_g(\boldsymbol{\theta}) = [\text{grad}\, g(\boldsymbol{\theta})]A^{-1}(\boldsymbol{\theta})\, [\text{grad}\, g(\boldsymbol{\theta})]'$$

and μ_k is the kth absolute moment of the standard normal distribution. For convenience we can now state our immediate conclusions as a theorem. Let $\beta = (s + 1)/2$.

THEOREM 4.1. (i) *If assumptions* (1)–(5) *hold,* $n^{-(\beta+1)} = o(\Delta K(n))$ *and* $\Delta^2 K \geqq 0$, *then the rules* $\bar{t}$, t' *are asymptotically pointwise optimal in the Bayes estimation situation.*

(ii) *If, furthermore,* $\sup_n E(n^\beta Y_n) < \infty$, $K(n) = n$, *then the rules* $\bar{t}$, t' *are asymptotically optimal.* (*Here, as throughout the paper, we assume the rules* $\bar{t}$, t' *are defined in the context of the problem with the appropriate* Y_n, K *etc.*)

COROLLARY 4.2. *Suppose there exists a sequence of estimates* $g\hat{\theta}_n$ *such that,*

$$\sup n^\beta \int_\Theta E_\theta(l(\theta, g\hat{\theta}_n))\psi(\theta)\, d\theta < \infty \tag{4.16}$$

and assumptions (a)–(d) *hold. Then if* $K(n) = n$ *the rules* $\bar{t}$, t' *are asymptotically optimal.*

PROOF. Immediate from Theorem 4.1 on applying the definition of the Bayes risk $E(Y_n)$, (viz. (4.3)).

We now give some sufficient requirements for the conditions of the corollary to hold. Our approach is based on the "method of moments."

I. Let $v = (v_1, \cdots, v_k)$ be a measurable map from R^k to R^k. Define,

$$E_\theta(v(z_1, \cdots, z_k)) = (v_1^*(\theta), \cdots, v_k^*(\theta)) = \mathbf{v}^*(\theta). \tag{4.17}$$

(a) Suppose that $\mathbf{v}^*$ exists and is finite for all θ.

(b) Suppose that $\mathbf{v}^*$ is 1-1.

(c) Let the range of $\mathbf{v}^*$, $R(\mathbf{v}^*)$ be convex.

Denote the inverse by $\theta(\mathbf{v}^*)$ on the range of $\mathbf{v}^*$.

(d) Assume that the matrix,

$$J(\theta) = \|\partial v_i^*(\theta)/\partial\theta_j\| \tag{4.18}$$

exists, possesses an inverse $J^{-1}(\theta)$ for all θ such that, if $J^{-1}(\theta) = \|j_{ir}(\theta)\|$,

$$\text{(e)} \quad \sup\{|j_{ir}(\theta)| : 1 \leqq i \leqq k,\ 1 \leqq r \leqq k,\ \theta \in \Theta\} = M_1 < \infty, \tag{4.19}$$

(f) $P_\theta[v(z_1, \cdots, z_k) \in R(\mathbf{v}^*)] = 1$ for all θ.

II. We suppose,

$$\sup\{|\partial g(\theta)/\partial\theta_i| : 1 \leqq i \leqq k,\ \theta \in \Theta\} = M_2 < \infty. \tag{4.20}$$

We now prove,

THEOREM 4.3. *Suppose* $\beta \geqq (r + 1)/2$. *Let* $\beta' \geqq \max(1, \beta)$. *If assumptions* I *and* II *hold and moreover,*

$$\sum_{i=1}^k \int_\Theta E_\theta |v_i(z_1, \cdots, z_k) - v_i^*(\theta)|^{2\beta'}\, d\theta < \infty \tag{4.21}$$

then (4.16) *holds. Hence, if* $K(n) = n$ *the rules* $\bar{t}$, t' *are asymptotically optimal.*

PROOF. Note that I implies the range of $\mathbf{v}^*$ is open. Define,

$$(g\hat{\theta}_n)(z_1, \cdots, z_n) = g(\theta^{-1}[N^{-1} \textstyle\sum_{j=0}^{N(n)-1} v(z_{jk+1}, \cdots, z_{(j+1)k})], \tag{4.22}$$

where $n = Nk + r$ with $0 \leqq r < k$.

We note that the definition is valid (a.s. $P_{\boldsymbol{\theta}}$) by (c) and (f).

Upon applying the general Taylor formula ([8], p. 186), we get

$$\begin{aligned}(g\hat{\theta}_n)(z_1, \cdots, z_n) - g(\boldsymbol{\theta}) \\ = \textstyle\int_0^1 (1-\xi)^{p-1}/(p-1)!J^{-1}(\boldsymbol{\theta} + (\boldsymbol{\theta}(\hat{v}_N) - \boldsymbol{\theta})\xi) \\ \times [\operatorname{grad} g(\boldsymbol{\theta} + (\boldsymbol{\theta}(\hat{v}_N) - \boldsymbol{\theta})\xi)]' \, d\xi \times (\hat{v}_N - \mathbf{v}^*(\boldsymbol{\theta}))\end{aligned} \tag{4.23}$$

where $\hat{v}_N = N^{-1} \sum_{j=0}^{N-1} v(z_{jk+1}, \cdots, z_{(j+1)k})$ and $'$ as usual denotes matrix transposition.

We immediately obtain,

$$\begin{aligned}|g\hat{\theta}_n(z_1, \cdots, z_n) - g(\boldsymbol{\theta})| \\ \leqq M_1M_2 \textstyle\sum_{i=1}^k |N^{-1} \sum_{j=0}^{N-1} v_i(z_{jk+1}, \cdots, z_{(j+1)k}) - v_i^*(\boldsymbol{\theta})|.\end{aligned} \tag{4.24}$$

Hence,

$$\begin{aligned}n^\beta E_{\boldsymbol{\theta}}[l(\boldsymbol{\theta}, g\hat{\theta}_n)] \leqq M_3E_{\boldsymbol{\theta}}(|g\boldsymbol{\theta} - g\hat{\theta}_n|^{2\beta})n^\beta \\ \leqq M_1M_2M_3M_4(n/N)^\beta[\textstyle\sum_{i=1}^k E_{\boldsymbol{\theta}}(|N^{\frac{1}{2}}[\sum_{j=0}^{N-1} v_i(z_{jk+1}, \cdots, z_{(j+1)k}) \\ - v_i^*(\boldsymbol{\theta})|^{2\beta})]\end{aligned} \tag{4.25}$$

by (4.24).

We now appeal to a special case of an inequality due to Chung ((3.3), p. 348, [7]) recently rediscovered by Brillinger [6].

THEOREM. (Chung). *Let U_i be independently distributed with* $E(U_i) = 0$. *Then, if* $S_n = \sum_{i=1}^n U_i$, $\beta' \geqq 1$,

$$E|S_n|^{2\beta'} \leqq n^{\beta'-1}M_5(\beta') \textstyle\sum_{i=1}^n E|U_i|^{2\beta'}. \tag{4.26}$$

If the U_i are identically distributed we immediately get

$$E|S_n|^{2\beta'} \leqq n^{\beta'}M_5(\beta')E|U_1|^{2\beta'}. \tag{4.26(a)}$$

(4.25), (4.26)(a) and (4.21) establish the theorem. Another easy and useful theorem along the same lines is,

THEOREM 4.4. *If an unbiased estimate $\hat{g}(z_1, \cdots, z_l)$ of $g(\boldsymbol{\theta})$ exists, assumptions* I, II, (1)–(5) *hold and*

$$\textstyle\int_\Theta E_{\boldsymbol{\theta}}|\hat{g}(z_1, \cdots, z_l) - g(\boldsymbol{\theta})|^{2\beta'}\psi(\boldsymbol{\theta}) \, d\boldsymbol{\theta} < \infty \tag{4.27}$$

then (4.16) *is satisfied.*

PROOF. Immediate upon applying Chung's inequality.

Applications of Theorems 4.3 and 4.4 are numerous. Thus, we can conclude from Theorem 4.4 that in estimating location in a location and scale parameter family for any prior the rules $\bar{t}$, t' are asymptotically optimal if the population has moments of order at least $2\beta'$. Similarly in estimating scale, moments of order $4\beta'$ suffice as may be seen on applying Theorem 4.3 suitably.

Estimation of p and functions of p in binomial and negative binomial situations, of λ and functions of λ in the Poisson model readily fits into our model. Other common exponential family situations can also be dealt with.

5. Independence of the prior and minimaxity. In this section we investigate to what extent the procedure $(g\theta_{t'}, t')_\psi$ (where the subscript indicates dependence on ψ) remains asymptotically optimal under different priors. We define an equivalence relation on the set of all densities satisfying the conditions of Section 4 by, $\psi^* \equiv \psi$ if and only if there exists $M < \infty$ such that $\psi(\boldsymbol{\theta}) \leqq M\psi^*(\boldsymbol{\theta})$ and $\psi^*(\boldsymbol{\theta}) \leqq M\psi(\boldsymbol{\theta})$ for all $\boldsymbol{\theta}$. We have,

THEOREM 5.1. *If the problem and ψ satisfy conditions* (1)–(5) *of Section 4 and the condition of Theorem* 4.1 (ii), *then* $(g\theta_{t'}, t')_\psi$ *is asymptotically optimal for all* $\psi^* \equiv \psi$.

PROOF. Let

$$Y_n^* = \int_\Theta l(\boldsymbol{\theta}, (g\theta_n')(z_1, \cdots, z_n))\psi^*(\boldsymbol{\theta} \mid z_1, \cdots, z_n)\, d\boldsymbol{\theta} \tag{5.1}$$

where $g\boldsymbol{\theta}'$ is the Bayes estimate of $g(\boldsymbol{\theta})$ *for the prior* ψ. It follows from [5] that,

$$n^\beta Y_n^* \to [A_g(\boldsymbol{\theta})]^\beta\gamma(2\beta)^{-1}\mu_{2\beta} = V(\boldsymbol{\theta}) \qquad \text{a.s. } P_{\boldsymbol{\theta}}. \tag{5.2}$$

Since as we have already seen,

$$c^{(\beta+1)^{-1}}t'(c) \to (V(\boldsymbol{\theta})\beta)^{(\beta+1)^{-1}} \qquad \text{a.s. } P_{\boldsymbol{\theta}}, \tag{5.3}$$

we have

$$c^{-(\beta/(\beta+1))}Y^*_{t'(c)} \to (V(\boldsymbol{\theta})\beta)^{(\beta+1)^{-1}}\beta^{-1}. \tag{5.4}$$

Now the risk of the procedure $(g\theta'_{t'}, t')_\psi$ if ψ^* is the true prior, is precisely,

$$\int_\Theta [E_{\boldsymbol{\theta}}(Y^*_{t'(c)}) + cE_{\boldsymbol{\theta}}(t'(c))]\psi^*(\boldsymbol{\theta})\, d\boldsymbol{\theta}. \tag{5.5}$$

Therefore, to prove the theorem in view of (3.6) we need only show,

$$\sup_{c>0} c^{-(\beta(1+\epsilon)/(\beta+1))}\, [\int_\Theta \{E_{\boldsymbol{\theta}}[Y^*_{t'(c)}]^{1+\epsilon} + cE_{\boldsymbol{\theta}}(t'(c))^{1+\epsilon}\}\psi^*(\boldsymbol{\theta})\, d\boldsymbol{\theta} < \infty. \tag{5.6}$$

Now by assumption there exists $M < \infty$ such that,

$$\psi^*(\boldsymbol{\theta} \mid z_1, \cdots, z_n) \leqq M^2\psi(\boldsymbol{\theta} \mid z_1, \cdots, z_n). \tag{5.7}$$

Hence,

$$Y^*_{t(c)} \leqq M^2Y_{t'(c)} \leqq M^2c\beta^{-1}(t'(c) + 1), \tag{5.8}$$

and to establish (5.6) we need only show,

$$\sup_{c>0} \int_\Theta E_{\boldsymbol{\theta}}\{c^{(1+\epsilon)/(\beta+1)}[t'(c)]^{1+\epsilon}\}\psi(\boldsymbol{\theta})\, d\boldsymbol{\theta} < \infty. \tag{5.9}$$

But (5.9) follows readily by the argument of Theorem 3.1 if we choose $\epsilon < \beta$. □

REMARK. It is plausible to conjecture that having ψ^*/ψ bounded should suffice for the theorem to hold. It is easy to see that in this case the expected number of observations behaves properly but we can unfortunately say little about the behavior of the stopping risk under ψ^*. Unlike the testing problem when the stopping risk contribution is negligible (as shown by Kiefer and Sacks) the stopping risk in estimation is a component that cannot be neglected. It would be interesting to know if under our conditions a rule can be constructed which would be optimal asymptotically for all ψ^* with ψ^*/ψ bounded.

In the rest of this section we consider rules which are asymptotically minimax in the sense of Wald [12]. These rules as we shall see are asymptotically Bayes whatever be the true prior. Their existence is proved, however, only under very strong regularity conditions which are difficult to check.

Let $\mathfrak{D}$ be the class of all nonrandomized sequential estimation procedures in the model of Section 4. Thus if $\delta \in \mathfrak{D}$, $\delta = ((g\hat{\theta})(z_1, \cdots, z_t), t)$ where $g\hat{\theta}$ is some measurable estimate of $g(\boldsymbol{\theta})$ based on t observations and t is a stopping rule. Following Wald we say that $\delta_c = ((g\hat{\theta})(z_1, \cdots, z_{s(c)}), s(c))$ is asymptotically minimax if

$$(5.10) \qquad \lim_{c\to 0} \sup R_c(\boldsymbol{\theta}, \delta_c)[\inf_{\delta \in \mathfrak{D}} \sup_{\boldsymbol{\theta}} R_c(\boldsymbol{\theta}, \delta)]^{-1} = 1$$

where

$$(5.11) \qquad R_c(\boldsymbol{\theta}, \delta) = E_{\boldsymbol{\theta}}(l(\boldsymbol{\theta}, (g\hat{\theta})(z_1, \cdots, z_t))) + cE_{\boldsymbol{\theta}}(t).$$

(We assume $K(n) = n$.)

We begin by proving an assertion of Wald's ((3.17) of [12]) for our more general case.

LEMMA 5.2. *Suppose Assumptions* (1), (3) *and* (4) of *Section* 3 *hold. Moreover, assume that* $\tilde{l}$ *is convex, and that* $A(\boldsymbol{\theta})$ *is continuous in* $\boldsymbol{\theta}$ (*in the sense of convergence of entries*). *Then,*

$$(5.12) \quad \liminf_c [\sup_{\boldsymbol{\theta}} R_c(\boldsymbol{\theta}, \delta)]c^{-(\beta/\beta+1)} \geqq \sup_{\boldsymbol{\theta}} V^{(\beta+1)^{-1}}(\boldsymbol{\theta})\beta^{(\beta+1)^{-1}}(1 + \beta^{-1}).$$

PROOF.

$$(5.13) \qquad \sup_{\boldsymbol{\theta}} R_c(\boldsymbol{\theta}, \delta) \geqq \int_{\Theta} R_c(\boldsymbol{\theta}, \delta)\psi(\boldsymbol{\theta})\, d\boldsymbol{\theta} \qquad \text{for all } \psi.$$

If ψ satisfies Assumption (2) of Section 3, then $\tilde{l}$ and ψ automatically satisfy Assumption 5. Let $g\hat{\theta}'$ denote the Bayes estimate, Y_n denote the stopping risk etc. Then, from (5.13), if $\delta = (g\hat{\theta}(z_1, \cdots, z_t), t)$, we get,

$$(5.14) \qquad \sup_{\boldsymbol{\theta}} R_c(\boldsymbol{\theta}, \delta) \geqq E(Y_{t(c)}) + cE(t(c))$$

where E corresponds to averaging over all $\boldsymbol{\theta}$ with respect to ψ. Applying (3.7) we conclude,

$$(5.15) \quad \liminf_c c^{-(\beta/\beta+1)} \sup_{\boldsymbol{\theta}} R_c(\boldsymbol{\theta}, \delta) \geqq [\int_{\Theta} V^{(\beta+1)^{-1}}(\boldsymbol{\theta})\psi(\boldsymbol{\theta})\, d\boldsymbol{\theta}]\beta^{(\beta+1)^{-1}}(1 + \beta^{-1})$$

if ψ satisfies Assumption (2).

Maximizing the right hand side over all such ψ we obtain the right hand side of (5.13) since $V(\boldsymbol{\theta})$ is continuous by our assumptions. □

Following Wald we now make the following assumptions:

A 1. Assumptions 3(a), (d) and (4) of Section 4 hold.

A 2. $\inf \{\mathbf{x}A(\boldsymbol{\theta})\mathbf{x}': \|\mathbf{x}\| = 1, \boldsymbol{\theta} \in \Theta\} > 0$, where $\mathbf{x}$ denotes a vector in R^k and $\| \ \|$ is its length.

A 3. $A(\boldsymbol{\theta})$ is uniformly continuous in $\boldsymbol{\theta}$ in the sense of convergence of entries.

A 4. $\lim_{\epsilon\to 0} E_{\boldsymbol{\theta}}[\sup\{|\partial^2\Phi(z_1, \mathbf{s})/\partial\theta_i\partial\theta_j - \partial^2\Phi(z_1, \boldsymbol{\theta})/\partial\theta_i\partial\theta_j|: \|\mathbf{s} - \boldsymbol{\theta}\| < \epsilon, \mathbf{s} \in \Theta\}] = 0$ uniformly in $\boldsymbol{\theta}$ for all i, j.

A 5. $\sup \{E_{\boldsymbol{\theta}}(|\partial\Phi(z_1, \boldsymbol{\theta})/\partial\theta_i|^{2+\delta}): 1 \leqq i \leqq k, \boldsymbol{\theta} \in \Theta\} < \infty$ for some $\delta > 0$.

A 6. Assumption II of Section 4 holds, grad $g(\boldsymbol{\theta})$ is uniformly continuous in $\boldsymbol{\theta}$, and $\inf \{\sum_{i=1}^{k} |\partial g(\boldsymbol{\theta})/\partial\theta_i|: \boldsymbol{\theta} \in \Theta\} > 0$.

A 7. Uniformly strongly consistent maximum likelihood estimates $\hat{\boldsymbol{\theta}}_n$ of $\boldsymbol{\theta}$ exist.

A 8. Let $t(c)$ be any sequence of stopping rules such that for a real valued function $N(c, \boldsymbol{\theta})$ with $\lim_{c\to 0} N(c, \boldsymbol{\theta}) = \infty$ uniformly in $\boldsymbol{\theta}$, and a positive function $\epsilon(c) \to 0$ as $c \to 0$,

$$P_{\boldsymbol{\theta}}[|t(c)\,[N(c, \boldsymbol{\theta})]^{-1} - 1| \geqq \epsilon(c)] = 1 \qquad \text{uniformly in } \boldsymbol{\theta}.$$

Then we have,

THEOREM 5.3. *Under assumptions* A 1 — A 8,

$$\lim_{c\to 0} P_{\boldsymbol{\theta}}[[t(c)]^{\frac{1}{2}}(g(\hat{\theta}_{t(c)}) - g(\boldsymbol{\theta})) < \lambda[Ag(\boldsymbol{\theta})]^{-\frac{1}{2}}] = \Phi(\lambda)$$

uniformly in λ *and* $\boldsymbol{\theta}$ *where* Φ *is the standard normal cumulative.*

PROOF. We sketch the argument, referring the reader to [12] Theorem 4.1 for details.

$$\begin{aligned} (5.16) \qquad & [t(c)]^{\frac{1}{2}}(g(\hat{\boldsymbol{\theta}}_{t(c)}) - g(\boldsymbol{\theta})) \\ &= \textstyle\int_0^1 [(1 - \xi)^{p-1}/(p-1)!][\text{grad } g(\boldsymbol{\theta} + (\hat{\boldsymbol{\theta}}_{t(c)} - \boldsymbol{\theta})\xi) \\ &\quad - \text{grad } g(\boldsymbol{\theta})]\cdot[t(c)]^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}_{t(c)} - \boldsymbol{\theta})'\, d\xi \\ &\quad + [\text{grad } g(\boldsymbol{\theta})]\cdot[t(c)]^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}_{t(c)} - \boldsymbol{\theta})' \end{aligned}$$

where $\cdot$ denotes matrix multiplication and $'$ transposing. A 7 and A 8 imply that

$$(5.17) \qquad P_{\boldsymbol{\theta}}[|\hat{\boldsymbol{\theta}}_{t(c)} - \boldsymbol{\theta}| \geqq \epsilon] \to 0$$

as $c \to 0$ uniformly in $\boldsymbol{\theta}$ for all $\epsilon > 0$, viz., $\hat{\boldsymbol{\theta}}_{t(c)}$ tends to $\boldsymbol{\theta}$ in $P_{\boldsymbol{\theta}}$ probability uniformly in $\boldsymbol{\theta}$. Now, by exactly imitating the argument of Wald in Theorem 4.1 of [12] and using again the multivariate Taylor theorem we get,

$$\begin{aligned} (5.18) \qquad & [N(c, \boldsymbol{\theta})]^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}_{t(c)} - \boldsymbol{\theta})[A^*(\boldsymbol{\theta}) + E_1(c)] \\ &= N^{-\frac{1}{2}} \textstyle\sum_{i=1}^{N+\epsilon(c)N} (\partial\Phi(z_i, \boldsymbol{\theta})/\partial\theta_1, \cdots, \partial\Phi(z_i, \boldsymbol{\theta})/\partial\theta_k) \end{aligned}$$

where $E_1(c)$ (a matrix) tends to 0 in $P_{\boldsymbol{\theta}}$ probability uniformly in $\boldsymbol{\theta}$.

We claim that,

$$(5.19) \qquad E_1(c)[A^*(\boldsymbol{\theta})]^{-1} \to 0 \qquad \text{in } P_{\boldsymbol{\theta}} \text{ probability uniformly in } \boldsymbol{\theta}.$$

(5.19) clearly holds if

$$(5.20) \qquad \sup \{\langle [A^*(\boldsymbol{\theta})]^{-1}\rangle : \boldsymbol{\theta} \in \Theta\} < \infty$$

where

$$\langle A \rangle = \sup_{i,j} |a_{ij}| \quad \text{if} \quad A = \|a_{ij}\|.$$

But for positive definite matrices A,

$$\langle A^{-1} \rangle \leqq [\lambda(A)]^{-1} = [\inf \{\mathbf{x}A\mathbf{x}' : \|\mathbf{x}\| = 1\}]^{-1} \tag{5.21}$$

where $\lambda(A)$ is the minimal eigenvalue of A, and (5.19) follows from A 2. Another criterion equivalent to A 2 is

$$\sup \{[|\det A_{ii}(\boldsymbol{\theta})|][\det A(\boldsymbol{\theta})]^{-1} : 1 \leqq i \leqq k, \boldsymbol{\theta} \,\varepsilon\, \Theta\} < \infty \tag{5.22}$$

where $A_{ii}(\boldsymbol{\theta})$ is the principal minor corresponding to the ith entry on the diagonal of $A(\boldsymbol{\theta})$ and "det" denotes determinant. This is an easy consequence of the Laplace form of the inverse of a matrix and the invariance of the trace.

(5.18), (5.19), Assumption A 5 and the central limit theorem imply that if $\mathbf{v} = (v_1, \cdots, v_k)$

$$P_{\boldsymbol{\theta}}[\mathbf{v}[t(c)]^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}_{t(c)} - \boldsymbol{\theta})' < \lambda[\mathbf{v}A^{-1}(\boldsymbol{\theta})\mathbf{v}']^{-\frac{1}{2}}] \to \Phi(\lambda) \tag{5.23}$$

uniformly in λ, $\boldsymbol{\theta}$, and $\mathbf{v}$ such that,

$$\alpha^{-1} \leqq \|\mathbf{v}\| \leqq \alpha \tag{5.24}$$

for $0 < \alpha < \infty$ fixed.

Using assumption A 6, (5.16) and (5.23) we conclude that $\{t(c)\}^{\frac{1}{2}}(g(\hat{\boldsymbol{\theta}}_n) - g(\boldsymbol{\theta}))$ has the same behaviour asymptotically as $\{t(c)\}^{\frac{1}{2}}$ grad $g(\boldsymbol{\theta}) . (\hat{\boldsymbol{\theta}}_{t(c)} - \boldsymbol{\theta})'$ uniformly in $\boldsymbol{\theta}$. The theorem now follows from (5.23) and Assumptions A 1 and A 6.

Define a decision procedure $\delta_c^* = (g(\hat{\boldsymbol{\theta}}_{t^*}), t^*)$ as follows:

Stop at the first n for which,

$$V(\hat{\boldsymbol{\theta}}_n)\beta[n(n + 1)]^{-1} \leqq c \tag{5.25}$$

and then estimate $g(\boldsymbol{\theta})$ by $g(\hat{\boldsymbol{\theta}}_n)$. $V(\boldsymbol{\theta})$ is defined in (5.2).

We postulate,

A 9

$$\sup \{E_{\boldsymbol{\theta}}|g(\hat{\boldsymbol{\theta}}^*_{t(c)}) - g(\boldsymbol{\theta})|^{2\beta(1+\epsilon)}c^{-\beta(1+\epsilon)/(\beta+1)}\, V^{-(1+\epsilon)/(\beta+1)}(\boldsymbol{\theta}) : \boldsymbol{\theta} \,\varepsilon\, \Theta\} < \infty$$

for some $\epsilon > 0$. Moreover $\beta \geqq \frac{1}{2}(r + 1)$.

The following theorem holds.

THEOREM 5.4. *Under Assumptions* A 1–A 9 *the procedure* δ_c^* *is asymptotically Bayes for all priors* Ψ *satisfying Assumption* 2 *of Section* 3 *and is asymptotically minimax.*

PROOF. Both assertions follow from,

$$c^{-\beta/(\beta+1)}R_c(\boldsymbol{\theta}, \delta_c^*) \to V^{(\beta+1)^{-1}}(\boldsymbol{\theta})\beta^{(\beta+1)^{-1}}(1 + \beta^{-1}) \tag{5.26}$$

uniformly in $\boldsymbol{\theta}$, in view of Lemma 5.2 and (3.5). (5.26) is an easy consequence of Theorem 5.3 and A' 9. □

Clearly other asymptotically minimax and asymptotically Bayes rules may be constructed. However all seem to involve a version of A 9 as a regularity condition, and this is, of course, highly undesirable since this requirement involves "stopped" random variables.

The "Bayes" rules discussed in the previous sections are for instance asymptotically minimax and Bayes for all priors under a version of A 9 and a "uniform in θ" version of Assumptions (1)–(5) of Section 4.

6. Concluding remarks. Our techniques clearly generalize to the case of estimation of a vector parameter. Another type of asymptotic approach has been considered by Robbins [10] and more recently by Starr [11] in the problem of sequentially estimating the mean of a normal distribution with unknown mean and variance. The fundamental difference between their approach and the one we have considered is that they prove optimality properties for a single procedure as the parameter approaches an asymptote, while our interest centers on a sequence of procedures which "become" globally optimal. The distinction is the same as that between "locally most powerful" and "asymptotically most powerful" tests in classical statistical theory.

The theory of Bayesian sequential confidence intervals would seem to offer an interesting field for the further extension and application of our methods. The classical case has, of course, already been thoroughly treated by Anscombe [1] and [2].

REFERENCES

[1] ANSCOMBE, F. (1952). Large sample theory of sequential estimation. *Proc. Camb. Phil. Soc.* **48** 600–607.

[2] ANSCOMBE, F. (1953). Sequential estimation. *J. Roy. Stat. Soc. B* **15** 1–29.

[3] ARROW, K., BLACKWELL, D. and GIRSHICK, M. A. (1949). Bayes and minimax solutions of sequential decision problems. *Econometrica* **17** 213–244.

[4] BICKEL, P. J. and YAHAV, J. A. (1965). Asymptotically pointwise optimal procedures in sequential analysis. To appear in *Proc. Fifth Berk. Symp. Math.Statist.Prob.*

[5] BICKEL, P. J. and YAHAV, J. A. (1966). Asymptotic theory of Bayes solutions. Technical Report, Department of Statistics, Stanford Univ.

[6] BRILLINGER, D. (1962). A note on the rate of convergence of a mean. *Biometrika* **49** 574–576.

[7] CHUNG, K. L. (1951). The strong law of large numbers. *Proc. Second Berk. Symp. Prob. Statist.* Univ. California Press.

[8] DIEUDONNE, J. (1961). *Foundations of Modern Analysis.* Interscience, New York.

[9] KIEFER, J. and SACKS, J. (1963). Asymptotically optimal sequential inference and design. *Ann. Math. Statist.* **34** 705–750.

[10] ROBBINS, H. (1959). Sequential estimation of the mean of a normal population. *Probability and Statistics* (H. Cramer Volume). Almquist and Wicksell, Uppsala.

[11] STARR, N. (1966). Efficiency of a sequential procedure for estimating the mean. *Ann. Math. Statist.* **37** 1173–1185.

[12] WALD, A. (1951). Asymptotic minimax solutions of sequential point estimation problems. *Proc. Second Berkeley Symp. Math. Statist. Prob.* Univ. of California Press.