The Annals of Mathematical Statistics
1969, Vol. 40, No. 2, 417–426

ON AN A.P.O. RULE IN SEQUENTIAL ESTIMATION WITH QUADRATIC LOSS[1]

BY PETER J. BICKEL AND JOSEPH A. YAHAV[2]

University of California, Berkeley and Tel-Aviv University, Israel

1. Introduction and summary. Consider the problem of Bayesian sequential estimation of a real parameter θ with quadratic loss and fixed cost c per observation. It is well known (cf. [1], [2]) that, under simple regularity conditions, this problem reduces to the following one.

If $Z_1, Z_2, \cdots, Z_n, \cdots$ are the observations (independent and identically distributed given θ) let

$$Y_n = \operatorname{Var}(\theta \mid Z_1, \cdots, Z_n), \tag{1.1}$$

the posterior variance of θ, and,

$$X_n(c) = Y_n + nc. \tag{1.2}$$

The problem is then to find a stopping time $s(c)$ such that $E(X_{s(c)}(c)) = \inf \{E(X_t(c)) : t \in T\}$ where T is the set of all stopping times. In general, although $s(c)$ can usually be shown to exist finding it in explicit form is difficult.

In [2] we proposed the following stopping time $\bar{t}(c)$ for this problem: "Stop as soon as $Y_n \leqq c(n + 1)$". We showed in [2] (generalized in [3]) that under some regularity conditions this rule is asymptotically pointwise optimal (A.P.O.) i.e.,

$$\lim_{c \to 0} X_{\bar{t}(c)}(c)[X(c)]^{-1} = 1 \tag{1.3}$$

a.s. where,

$$X(c) = \inf_n X_n(c). \tag{1.4}$$

In fact, we proved that,

$$X_{\bar{t}(c)}(c) = 2c^{\frac{1}{2}}V^{\frac{1}{2}}(\theta) + o(c^{\frac{1}{2}}) \quad \text{a.s.} \tag{1.5}$$

and,

$$X_{\bar{t}(c)}(c) - X(c) = o(c^{\frac{1}{2}}) \quad \text{a.s.} \tag{1.6}$$

where $V(\theta)$ is the reciprocal of the Fisher information number. Later, (in [3]) we showed, under some additional conditions that $\bar{t}(c)$ is asymptotically optimal i.e., that,

$$\lim_{c \to 0} [E(X_{s(c)}(c))][E(X_{\bar{t}(c)}(c))]^{-1} = 1, \tag{1.7}$$

Received 27 November 1967.

[1] Supported in part by Grant GP-5705 of the National Science Foundation, while the authors were visiting the Department of Statistics, Stanford University, and by the Office of Naval Research, Contract Nonr N00014-67-A-0114-0004.

[2] Supported in part by a grant from O.A.R. (European Office).

and in fact, that

$$E(X(c)) = 2c^{\frac{1}{2}}E(V(\theta)) + o(c^{\frac{1}{2}}) \tag{1.8}$$

and

$$E(X_{\tilde{t}(c)}(c)) - E(X(c)) = o(c^{\frac{1}{2}}). \tag{1.9}$$

In this paper we seek to refine the term $o(c^{\frac{1}{2}})$ in (1.5)–(1.6) and (1.8)–(1.9). Our analysis, as in our previous work, is based on looking at the asymptotic properties of Y_n. We showed in [2] and [4] that,

$$Y_n = V(\theta)\, n^{-1} + R_n \tag{1.10}$$

where $R_n = o(n^{-1})$ a.s. In [4] we further showed that, under suitable conditions,

$$Y_n = V(\theta)n^{-1} + S_n(\theta)n^{-2} + R_n' \tag{1.11}$$

a.s. where $R_n' = o(n^{-3/2})$ and

$$S_n(\theta) = \sum_{i=1}^{n} W_i(\theta) \tag{1.12}$$

where the W_i are independent and identically distributed with mean 0 given θ.

If $W_1(\theta)$ has a second moment and is non degenerate the law of the iterated logarithm enables us to conclude that,

$$R_n = O(n^{-3/2}[\log \log n]^{\frac{1}{2}}) \quad \text{a.s.} \tag{1.13}$$

This suggests Theorem 2.1 which asserts that if (1.13) holds then,

$$X_{\tilde{t}(c)}(c) - X(c) = o(c^{3/4-\epsilon}) \tag{1.14}$$

a.s. for all $\epsilon > 0$. The analogues of (1.8) and (1.9) pose greater difficulty. In Section 3 we show that (Theorems 3.1, 3.2),

$$E(X(c) - 2[V(\theta)c]^{\frac{1}{2}}) = \max\,(o(c^{\frac{1}{2}+\delta(\lambda,b)-\epsilon}), O(c)), \tag{1.15}$$

for every $\epsilon > 0$ where,

$$\delta(\lambda, b) = \tfrac{1}{2}(\lambda - 1)b(b + (\lambda - 1))^{-1} \tag{1.16}$$

and b and λ depend on the problem. (Typically $\lambda = \frac{3}{2}$.) On the other hand, in Section 4 we establish, (Theorem 4.1),

$$E(X_{\tilde{t}(c)}(c) - 2[V(\theta)c]^{\frac{1}{2}})^{+} = \max\,(O(c^{\lambda/2}), O(c)), \tag{1.17}$$

for every $\epsilon > 0$ where again typically $\lambda = \frac{3}{2}$. Finally in Section 5 we apply our general results to two special situations.

(i) Estimating the mean of a normal distribution with a normal prior.

(ii) Estimating p on the basis of binomial trials with a beta prior.

In case (i) our conditions yields $O(c)$ in both (1.15) and (1.17) and this is best possible. In (ii) when for instance we have a uniform prior the best $\lambda = \frac{3}{2}$ and the best $b = 1$ and we therefore get $o(c^{-3/4\epsilon})$ for every $\epsilon > 0$ in (1.15) and $o(c^{2/3-\epsilon})$ for every $\epsilon > 0$ in (1.17). We do not believe these are best possible. A further analysis of (1.11) would seem to be required for anything better.

2. The pointwise difference between the performance of the Bayes rule and the A.P.O. rule. We will throughout use the representation (1.10) suppressing the θ in $V(\theta)$. In fact, in accordance with [3] we will not require that the Y_n originate in the estimation problem but merely that they be a sequence of random variables such that Y_n is measurable $\mathfrak{F}_n$ where $\{\mathfrak{F}_n\}$ is an increasing sequence of sigma fields, and that $P[Y_n > 0] = 1$. The V in (1.10) is then also supposed to be positive with probability 1.

THEOREM 2.1. *If* (1.13) *holds then for the stopping rule* $\bar{t}(c)$ *we have,*

$$X_{\bar{t}(c)}(c) - X(c) = o(c^{3/4-\epsilon}) \quad \text{a.s.} \tag{2.1}$$

for every $\epsilon > 0$.

PROOF. Let us write.

$$X(c) = \min\{X^{(1)}(c), X^{(2)}(c)\} \tag{2.2}$$

where

$$X^{(1)}(c) = \min_{1 \leqq n \leqq n_c} X_n(c), \qquad X^{(2)}(c) = \min_{n_c < n} X_n(c), \tag{2.3}$$

and

$$n_c = [Vc^{-1}]^{\frac{1}{2}}\rho(c) \tag{2.4}$$

where

$$\rho(c) \to 0 \tag{2.5}$$

at a rate which is not $o(c^{\frac{1}{2}})$. Then

$$X(c) \geqq \min(2[Vc]^{\frac{1}{2}}, X^{(2)}(c))I_{\{X^{(1)}(c) > 2[Vc]^{\frac{1}{2}}\}} + cI_{\{X^{(1)}(c) \leqq 2[Vc]^{\frac{1}{2}}\}} \tag{2.6}$$

where I_A is the usual indicator function of the event A. It follows that

$$\begin{aligned} &[X(c) - 2[Vc]^{\frac{1}{2}}]^- \\ &\leqq [X^{(2)}(c) - 2[Vc]^{\frac{1}{2}}]^- I_{\{X^{(1)}(c) > 2[Xc]^{\frac{1}{2}}\}} + |2[Vc]^{\frac{1}{2}} - c| I_{\{X^{(1)}(c) \leqq [Vc]^{\frac{1}{2}}\}} \\ &\leqq [X^{(2)}(c) - 2[Vc]^{\frac{1}{2}}]^- + (c + 2[Vc]^{\frac{1}{2}}) I_{\{X^{(1)}(c) \leqq 2[Vc]^{\frac{1}{2}}\}}. \end{aligned} \tag{2.7}$$

By Lemma 2.1 of [1] and (2.14) of [2] $I_{\{X^{(1)}(c) \leqq 2[Vc]^{\frac{1}{2}}\}} = 0$ for c sufficiently small so it is enough to consider $[X^{(2)}(c) - 2[Vc]^{\frac{1}{2}}]^-$. Let,

$$U_\lambda^{(c)} = -[\inf_{n \geqq n_c} n^\lambda R_n]^{-1}, \qquad 1 < \lambda < \tfrac{3}{2}. \tag{2.8}$$

Choose c_0 so that $Vn^{-1} + nc + n^{-\lambda}U_\lambda^{(c_0)}$ is positive for all $c \leqq c_0$. We can do this since by (1.13), (2.8), $U_\lambda^{(c)} \uparrow 0$. Now we define

$$Z_n(c) = Vn^{-1} + nc + n^{-\lambda}U_\lambda^{(c_0)}. \tag{2.9}$$

Then

$$X^{(2)}(c) \geqq \inf_n Z_n(c), \qquad c \leqq c_0. \tag{2.10}$$

Define $n_0(c)$ to be the first m such that

$$Z_m(c) = \inf_n Z_n(c). \tag{2.11}$$

Define

$$\tilde{Y}_n = Vn^{-1} + n^{-\lambda} U_\lambda^{(c_0)}. \tag{2.12}$$

Then $Z_n(c) = \tilde{Y}_n + nc$ and $\tilde{Y}_n$ satisfies the conditions of Theorem 2.1 of [1]. Therefore

$$n_0(c)[c/V]^{\frac{1}{2}} \to 1 \quad \text{a.s.} \tag{2.13}$$

By (2.10) for c sufficiently small and any $\epsilon > 0$

$$X^{(2)}(c) \geqq 2[Vc]^{\frac{1}{2}} + [(1-\epsilon)V^{\frac{1}{2}}]^{-\lambda} c^{\lambda/2} U_\lambda^{(c_0)}. \tag{2.14}$$

From (2.14), (1.13) and (2.7) we have

$$X(c) = 2[Vc]^{\frac{1}{2}} + o(c^{3/4-\epsilon}). \tag{2.15}$$

We now consider $\tilde{t}(c)$. By definition

$$X_{\tilde{t}(c)}(c) \leqq 2c\tilde{t}(c) + c \tag{2.16}$$

and

$$Y_{\tilde{t}(c)-1} > c\tilde{t}(c). \tag{2.17}$$

Therefore by (1.10) we have

$$(\tilde{t}(c)-1)V(\tilde{t}(c)-1)^{-1} + (\tilde{t}(c)-1)R_{(\tilde{t}(c)-1)} > c\tilde{t}^2(c) - c\tilde{t}(c). \tag{2.18}$$

Since $\tilde{t}(c) \uparrow \infty$ ([1]) by (1.13) for $\epsilon > 0$, there exists M_ϵ possibly depending on the sample sequence such that,

$$V + M_\epsilon(\tilde{t}(c)-1)^{\epsilon-\frac{1}{2}} \geqq c\tilde{t}^2(c) - c\tilde{t}(c). \tag{2.19}$$

By [1] $c\tilde{t}^2(c) \to V$ a.s. Hence for suitable M_ϵ' we have

$$c\tilde{t}^2(c) \leqq V + M_\epsilon' c^{\frac{1}{4}-\epsilon/2}. \tag{2.20}$$

Finally,

$$c\tilde{t}(c) \leqq [Vc]^{\frac{1}{2}}(1 + M_\epsilon' V^{-1} c^{\frac{1}{4}-\epsilon/2})^{\frac{1}{2}} \leqq [Vc]^{\frac{1}{2}}(1 + M_\epsilon'' c^{\frac{1}{4}-\epsilon/2}). \tag{2.21}$$

Then (2.21) and (2.16) establish,

$$X_{\tilde{t}(c)}(c) \leqq 2[Vc]^{\frac{1}{2}} + o(c^{3/4-\epsilon}). \tag{2.22}$$

Combining (2.22) and (2.15) the theorem is established.

3. A lower bound for the Bayes risk in estimation. We continue to use the general notation of Section 2. The following conditions will be required by our main theorem, in addition to our general conditions on the Y_n.

C_1: Y_n is an expectation decreasing martingale, with respect to the σ fields $\mathfrak{F}_n$.

$C_2(\lambda)$: If

$$U_\lambda = -[\inf_n n^\lambda R_n]^-, \tag{3.1}$$

then

$$E[|U_\lambda|V^{-\frac{1}{2}\lambda}] < \infty \tag{3.2}$$

for some $\lambda > 1$.

$C_3(b)$: For some $b > 0$,

$$\sup_n n^{-b}E(Y_n^{-b}) < \infty. \tag{3.3}$$

C_4 : Ess. sup. $V < \infty$.

As is well known C_1 is always satisfied if Y_n is the Bayes posterior risk, and in particular is satisfied for estimation with quadratic loss. We have,

THEOREM 3.1. *If C_1, $C_2(\lambda)$, $C_3(b)$ and C_4 are satisfied, then,*

$$E(X(c) - [Vc]^{\frac{1}{2}})^- = O(c^{\frac{1}{2}+\min(\delta(\lambda,b),\frac{1}{2})}). \tag{3.4}$$

PROOF. We use the breakup of $X(c)$ given by (2.2) and (2.3). We begin with

LEMMA 3.2. *If our general conditions and $C_2(\lambda)$ hold, then*

$$\begin{aligned} E(X^{(2)}(c) - [Vc]^{\frac{1}{2}})^- \\ \leqq E\{|U_\lambda|V^{-\frac{1}{2}\lambda}\}c^{\frac{1}{2}\lambda}[4^{\frac{1}{2}(\lambda-1)}\rho(c)^{(1-\lambda)} + ((\lambda-1)/(\lambda+1))^{\frac{1}{2}\lambda}]. \end{aligned} \tag{3.5}$$

PROOF OF LEMMA 3.2. Recall that,

$$X^{(2)}(c) \geqq \inf_{n\geqq n_c}[Vn^{-1} + nc + n^{-\lambda}U_\lambda]. \tag{3.6}$$

Let

$$Q_c^\lambda(x, \omega) = Vx^{-1} + cx + U_\lambda(\omega)x^{-\lambda} \tag{3.7}$$

and suppose $x_c^\lambda(\omega)$ is the smallest $x \geqq n_c$ for which $Q_c^\lambda(x, \omega)$ achieves its minimum in the range $x \geqq n_c$. Define the variable Δ by $x = [Vc^{-1}]^{\frac{1}{2}}(1 + \Delta)$ and let $\Delta_c^\lambda(\omega)$ correspond to $x_c^\lambda(\omega)$. Note that $\Delta_c^\lambda < 0$, since $Vx^{-1} + cx$ achieves its minimum for $\Delta = 0$, and $U_\lambda \leqq 0$. Consider,

$$\partial Q_c^\lambda/dx = -c(1+\Delta)^{-2} + c - \lambda U_\lambda c^{\frac{1}{2}(\lambda+1)}V^{-\frac{1}{2}(\lambda+1)}c^{\frac{1}{2}(\lambda+1)} \tag{3.8}$$

and

$$H(\Delta) = (2\Delta + \Delta^2)(1+\Delta)^{\lambda-1}, \qquad \Delta > -1. \tag{3.9}$$

Then,

$$\operatorname{sgn} \partial Q_c^\lambda/\partial x = \operatorname{sgn}(H(\Delta) - \lambda U_\lambda c^{\frac{1}{2}(\lambda-1)}V^{-\frac{1}{2}(\lambda+1)}). \tag{3.10}$$

Moreover,

$$H'(\Delta) = (1+\Delta)^{\lambda-2}\{2 + 2(\lambda+1)\Delta + (\lambda+1)\Delta^2\} \tag{3.11}$$

and hence, for $-1 < \Delta \leqq 0$, $H'(\Delta) \lesseqgtr 0$ according as

$$\Delta \lesseqgtr -1 + [(\lambda-1)/(\lambda+1)]^{\frac{1}{2}}.$$

Using, (3.12) and (3.10) we see that

(i) If $\lambda U_\lambda c^{\frac{1}{2}(\lambda-1)} V^{-\frac{1}{2}(\lambda+1)} \leqq H(-1 + [(\lambda - 1)/(\lambda + 1)]^{\frac{1}{2}})$ then $\partial Q_c^\lambda / \partial x \geqq 0$ for all $x > 0$.

(ii) If $\lambda U_\lambda c^{\frac{1}{2}(\lambda-1)} V^{-\frac{1}{2}(\lambda+1)} > H(-1 + [(\lambda - 1)/(\lambda + 1)]^{\frac{1}{2}})$ then there exist $0 < x_1 < x_2$, such that $\partial Q_c^\lambda(x, \omega)/dx = 0$ for $x = x_1, x_2$, x_1 is a local maximum of Q_c^λ, x_2 is a local minimum of Q_c^λ and $x_1 < [Vc^{-1}]^{\frac{1}{2}}((\lambda - 1)/(\lambda + 1))^{\frac{1}{2}} < x_2$. Of course, x_1 and x_2 are the only local extrema of Q_c^λ for $x > 0$.

From (i) and (ii) it follows that either $x_c^\lambda = n_c$ or $x_c^\lambda = x_2$ (where x_2, of course, depends on c, λ and ω).

Clearly, the second of these eventualities must hold if there exists an $x > n_c$ such that $Q_c^\lambda(x, \omega) \leqq Q_c^\lambda(n_c, \omega)$, and hence in particular if,

$$Q_c^\lambda(n_c, \omega) \geqq [Vc]^{\frac{1}{2}} \geqq Q_c^\lambda([Vc^{-1}]^{\frac{1}{2}}, \omega). \tag{3.12}$$

The first inequality, of (3.12) holds if and only if,

$$U_\lambda \geqq V^{\frac{1}{2}(\lambda+1)} \rho(c)^\lambda c^{\frac{1}{2}(\lambda-1)}[-(\rho(c) - 1)^2/\rho(c)] \tag{3.13}$$

If $\rho \leqq \frac{1}{2}$, $(\rho - 1)^2/\rho \geqq (4\rho)^{-1}$. Let $A_c = \{\omega: U_\lambda(\omega) \geqq -\frac{1}{4}V^{\frac{1}{2}(\lambda+1)}(c^{\frac{1}{2}}\rho(c)^{-1})^{1-\lambda}\}$. From (3.6), (i), (ii), (3.12) and (3.13) we see that on A_c,

$$X^{(2)}(c) \geqq \inf_n (Vn^{-1} + nc) + U_\lambda x_2^{-\lambda} = 2[Vc]^{\frac{1}{2}} + U_\lambda x_2^{-\lambda}. \tag{3.14}$$

Decomposing $X^{(2)}(c)$ according to A_c and using (3.14) and (ii) we see that,

$$(X^{(2)}(c) - [Vc]^{\frac{1}{2}}) \leqq |U_\lambda|[Vc^{-1}((\lambda - 1)/(\lambda + 1))]^{-\frac{1}{2}\lambda} + [Vc]^{\frac{1}{2}} I_{A_c'} \tag{3.15}$$

where I_A is the indicator of the event A and $'$ denotes complementation. Now,

$$\begin{aligned} E(V^{\frac{1}{2}} I_{A_c'}) &= \textstyle\int_{A_c'} V^{\frac{1}{2}}\, dP \\ &\leqq [\textstyle\int_{A_c'} |U_\lambda| V^{-\frac{1}{2}\lambda}\, dP][4\, c^{\frac{1}{2}}\rho(c)^{-1}]^{\lambda-1} \\ &\leqq E[|U_\lambda| V^{-\frac{1}{2}\lambda}] 4^{\frac{1}{2}(\lambda-1)} c^{\frac{1}{2}(\lambda-1)} \rho(c)^{1-\lambda}. \end{aligned} \tag{3.16}$$

The lemma follows from (3.15) and (3.16).

We now analyze, $E(V^{\frac{1}{2}} I_{[X^{(1)}(c) \leqq 2[Vc]^{\frac{1}{2}}]})$. Using C_4, let ess sup $V = s$

$$E(V^{\frac{1}{2}} I_{[X^{(1)}(c) \leqq [Vc]^{\frac{1}{2}}]}) \leqq s^{\frac{1}{2}} P[X^{(1)}(c) \leqq 2s^{\frac{1}{2}}c^{\frac{1}{2}}]. \tag{3.17}$$

But

$$\begin{aligned} &P[X^{(1)}(c) \leqq K] \\ &\quad = P[Y_n \leqq K - nc \quad \text{for some} \quad 1 \leqq n \leqq n_c] \\ &\quad \leqq P[Y_n \leqq K - nc \quad \text{for some} \quad 1 \leqq n \leqq s^{\frac{1}{2}}c^{-\frac{1}{2}}\rho(c)] \\ &\quad = P[Y_n^{-b} \geqq [K - nc]^{-b} \quad \text{for some} \quad 1 \leqq n \leqq s^{\frac{1}{2}}c^{-\frac{1}{2}}\rho(c)]. \end{aligned} \tag{3.18}$$

Now, C_1 implies that Y_n^{-b} is an expectation *increasing* nonnegative martingale. We recall Chow's [6] generalization of the Hajek-Renyi inequality which states that if Z_n is a nonnegative expectation increasing martingale, c_n is a nondecreasing sequence of constants, and $E(Z_n) \leqq d_n$ which are monotone increasing, then

(3.19) $P[Z_n \geqq c_n \text{ for some } 1 \leqq n \leqq m] \leqq d_1/c_1 + \sum_{k=2}^{n} [d_k - d_{k-1}]c_k^{-1}$.

Substituting $Y_n^{-b} = Z_n$, $[K - nc]^{-b} = c_n$, and $m = s^{\frac{1}{2}}c^{-\frac{1}{2}}\rho(c)$, we get using (3.18)

(3.20) $P[X^{(1)}(c) \leqq K] \leqq [K - c]^{-b}E(Y_m^{-b}) \leqq m^b[K - c]^{-b} \sup_n E(nY_n)^{-b}$.

After some simplification we get from (3.17) and (3.20) with $K = 2s^{\frac{1}{2}}c^{\frac{1}{2}}$,

(3.21) $E(V^{\frac{1}{2}}I_{[X^{(1)}(c) \leqq 2[Vc]^{\frac{1}{2}}]}) \leqq s^{\frac{1}{2}(b+1)}(2s^{\frac{1}{2}} - c^{\frac{1}{2}})^{-b}\rho^b(c) \sim 2^{-b}s^{\frac{1}{2}}\rho^b(c)$.

Using (2.7) and combining Lemma 3.2 and (3.21) we get, under the conditions of the theorem

$$(3.22)\quad E[X(c) - 2[Vc]^{\frac{1}{2}})^{-} \leqq c^{\frac{1}{2}\lambda}E\{|U_\lambda|V^{-\frac{1}{2}\lambda}\}[4^{\frac{1}{2}(\lambda-1)}\rho^{\frac{1}{2}(1-\lambda)}(c) + ((\lambda - 1)/(\lambda + 1))^{\frac{1}{2}\lambda} + c^{\frac{1}{2}}s^{\frac{1}{2}}[\rho(c)/2]^b(1 + o(1)).$$

It is an easy exercise in the calculus to see that an optimal choice of $\rho(c)$ is $\rho(c) \sim c^{(\lambda-1)[b+(\lambda-1)]^{-1}}$ which yields the theorem.

We now replace the unpleasant condition C_4 by

C_4'. All moments of V are finite.

Using C_4' we can obtain the weaker,

THEOREM 3.3. *If C_1, $C_2(\lambda)$, $C_3(b)$, and C_4' are satisfied, then*

$$(3.23)\quad E(X(c) - 2[Vc]^{\frac{1}{2}})^{-} = \max(o(c^{\frac{1}{2}+\delta(\lambda,b)-\epsilon}), O(c))$$

for every $\epsilon > 0$.

Proof. It clearly suffices to show,

$$(3.24)\quad E(V^{\frac{1}{2}}I_{[X^{(1)}(c) \leqq 2[Vc]^{\frac{1}{2}}]}) = o(\rho^{b-\epsilon}(c))$$

for every $\epsilon > 0$.

Now

$$(3.25)\quad E(V^{\frac{1}{2}}I_{[X^{(1)}(c) \leqq 2[Vc]^{\frac{1}{2}}]}) \leqq E^{r^{-1}}(V^{\frac{1}{2}r})P^{(r-1)/r}[X^{(1)}(c) \leqq 2[Vc]^{\frac{1}{2}}]$$

by Hölder's inequality for every $r > 1$. Using C_4' we see that (3.24) follows if,

$$(3.26)\quad P[X^{(1)}(c) \leqq 2[Vc]^{\frac{1}{2}}] = o(\rho^{b-\epsilon}(c))$$

for every $\epsilon > 0$. On the other hand,

$$(3.27)\quad \begin{aligned} &P[X^{(1)}(c) \leqq 2[Vc]^{\frac{1}{2}}] \\ &\quad \leqq \sum_{k=1}^{\infty} P\{X^{(1)}(c) \leqq 2[kc]^{\frac{1}{2}}, k - 1 \leqq V \leqq k\} \\ &\quad \leqq \sum_{k=1}^{\infty} P[X_1(k, c) \leqq 2[kc]^{\frac{1}{2}}, (k - 1) \leqq V \leqq k] \end{aligned}$$

where $X_1(k, c) = \inf_{n \leqq k^{\frac{1}{2}}c^{-\frac{1}{2}}\rho(c)}(Y_n + nc)$.

Again by Hölder's inequality,

$$(3.28)\quad P\{X_1(k, c) \leqq 2[kc]^{\frac{1}{2}}, (k - 1) \leqq V \leqq k\} \leqq P^{r^{-1}}[(k - 1) \leqq V \leqq k]P^{(r-1)/r}[X_1(k, c) \leqq 2[kc]^{\frac{1}{2}}].$$

Using (3.19) we see that,

(3.29) $$P[X_1(k, c) \leqq 2[kc]^{\frac{1}{2}}] \leqq 2^{-b}k^{\frac{1}{2}}\rho^b(c)(1 + o(1)).$$

Hence,

(3.30) $$P[X^{(1)}(c) \leqq 2[Vc]^{\frac{1}{2}}] \leqq [\rho(c)/2]^{b(r-1)r^{-1}} \cdot \sum_{k=1}^{\infty} k^{(r-1)/2r} P^{r^{-1}}[(k-1) \leqq V \leqq k].$$

The last sum is finite for every r by C_4' and the theorem follows.

4. An upper bound for the Bayes risk of $\bar{t}(c)$. We again use the representation (1.10). We will require the following condition

$D(\lambda)$: If $(\lambda > 1)$,

(4.1) $$W_\lambda = \sup_n n^\lambda R_n^{+}$$

then

(4.2) $$E\{W_\lambda V^{-\frac{1}{2}\lambda}\} < \infty.$$

Note that $C_2(\lambda)$ and D_λ are equivalent to requiring,

(4.3) $$E\{V^{-\frac{1}{2}\lambda} \sup_n n^\lambda |R_n|\} < \infty.$$

We have,

THEOREM 4.1. *If $D(\lambda)$ holds, then,*

(4.4) $$E(X_{\bar{t}(c)}(c) - 2[Vc]^{\frac{1}{2}})^{+} = O(\max (c^{\lambda/2}, c)).$$

PROOF. Since $Y_{\bar{t}(c)} \leqq c(\bar{t}(c) + 1)$ it suffices to show that,

(4.5) $$E(c\bar{t}(c) - [Vc]^{\frac{1}{2}})^{+} = O(\max (c^{\lambda/2}, c)).$$

Now, defining $Y_0 = 0$, and $R_0 = 0$

(4.6) $$Y_{\bar{t}(c)-1} \geqq c(\bar{t}(c) - 1)$$

and hence,

(4.7) $$c\bar{t}^{-1}(c) + 2c + V\bar{t}^{-1}(c) + R_{\bar{t}(c)-1} \geqq c\bar{t}(c).$$

Note that

(4.8) $$R^{+}_{\bar{t}(c)-1} \leqq W_\lambda(\bar{t}(c) - 1)^{-\lambda}.$$

Define,

$$B_c = \{\bar{t}(c) \leqq [Vc]^{\frac{1}{2}} + 1\}.$$

Then,

(4.9) $$E(c\bar{t}(c)) \leqq \int_{B_c} [Vc]^{\frac{1}{2}}\, dP + c + \int_{B_c'} c\bar{t}(c)\, dP.$$

Applying (4.7) and (4.8) to the second part of (4.9) we get

$$\int_{B_c'} c\bar{t}(c)\, dP \leqq \int_{B_c'} \{[Vc]^{\frac{1}{2}} + 3c + V^{-\lambda/2}W_\lambda c^{\lambda/2}\}\, dP. \tag{4.10}$$

The theorem follows.

In the Bayesian estimation situation if in the representation (1.11) we have for every $\epsilon, \epsilon' > 0$,

$$E\{\sup_n [n^{2-\epsilon}R_n']V^{-(1-\epsilon')}\} < \infty \tag{4.11}$$

then one can show,

$$E(X_{\bar{t}(c)}(c) - 2[Vc]^{\frac{1}{2}})^+ = o(c^{1-\epsilon}) \tag{4.12}$$

for every $\epsilon > 0$.

5. Examples.

I. *Estimation of normal mean.* We wish to estimate μ with quadratic loss on the basis of $z_1, \cdots, z_n, \cdots$ where the z_i are independent $\mathfrak{N}(\mu, 1)$ and μ has a prior $\mathfrak{N}(\mu_0, \sigma^2)$ distribution. In this case it is easy to compute,

$$Y_n \equiv (n + \sigma^{-2})^{-1} \tag{5.1}$$

and a direct computation yields that the Bayes rule is a fixed sample size rule taking $N(c)$ observations where $N(c)$ is one of the natural numbers closest to $(c^{\frac{1}{2}}\sigma^{-1} - \sigma^{-2})$. Similarly $\bar{t}(c)$ takes $\frac{1}{2}\{-(1 + \sigma^{-2}) + ((1 - \sigma^{-2})^2 + 4c^{-1})^{\frac{1}{2}}\}$ observations and $|N(c) - \bar{t}(c)| = O(c)$.

II. *Estimation in the binomial case.* We wish to estimate p with quadratic loss on the basis of $z_1, \cdots, z_n$, where the z_i are independent and take on the value 1 with probability p and 0 with probability $1 - p$, $0 < p < 1$. We put a beta (a, c) prior distribution on p, that is we suppose p has density,

$$f_{a,b}(p) = (\Gamma(a)\Gamma(c)/\Gamma(a + c))^{-1}p^{a-1}(1 - p)^{c-1} \quad a,c > 0, \quad 0 < p < 1. \tag{5.2}$$

In this case we have,

$$Y_n(z_1, \cdots, z_n) = (S_n + a)(n - S_n + c)/[n + (a + c)]^2(n + a + c + 1) \tag{5.3}$$

where

$$S_n = \sum_{i=1}^n z_i. \tag{5.4}$$

Then,

$$\begin{aligned} Y_n = pqn^{-1} &- [n(n + (a + c))^2(n + (a + c) + 1)]^{-1}\{[3(a + c) + 1]n^2 \\ &+ (a + c)(3(a + c) + 2)n + (a + c)^2(a + c + 1)\}pq \\ &+ [(n + (a + c))^2(n + (a + c) + 1)]^{-1} \\ &\quad \cdot\{n - 2p + (c - a)\}(S_n - np) - (S_n - np)^2\}. \end{aligned} \tag{5.5}$$

We now check that $C_2(\lambda)$ and $D(\lambda)$ are satisfied for every $\lambda < \frac{3}{2}$.

The following result has been established in [5]. (Similar results have appeared in [7] and elsewhere).

THEOREM. *Let Z_i be independent and identically distributed with mean* 0. *Let* $T_n = \sum_{i=1}^n Z_i$. *Then, if* $\alpha > \beta/2, \beta \geqq 2$,

$$E(\sup_n n^{-\alpha} |T_n|^\beta) \leqq K_2(\beta) E |Z_1|^\beta, \tag{5.6}$$

where $K_2(\beta)$ *is a numerical constant.*

Applying this theorem to the R_n defined by (5.5) our initial statements about $C_2(\lambda)$ and $D(\lambda)$ are verified. We now show that $C_3(b)$ holds for $b < \min(a, c)$.

From (5.3) we see that

$$E[nY_n]^{-b} \sim n^{2b} E[(S_n + a)(n - S_n + c)]^{-b}. \tag{5.7}$$

Simplifying we get

$$\begin{aligned} &E[n^{2b}(S_n + a)^{-b}(n - S_n + c)^{-b}] \\ &= \int_0^1 \{\textstyle\sum_{k=0}^n \binom{n}{k} (n/(k + a))^b (n/(n - k + c))^b P^b (1 - P)^{n-k}\} \\ &\qquad \cdot (\Gamma(a + c)/\Gamma(a)\Gamma(c)) P^{a-1}(1 - P)^{c-1}\, dP, \end{aligned} \tag{5.8}$$

$$\begin{aligned} &E[n^{2b}(S_n + a)^{-b}(n - S_n + c)^{-b}] \\ &= \textstyle\sum_{k=0}^n \binom{n}{k} (n/(k + a))^b (n/(n - k + c))^b (\Gamma(a + c)/\Gamma(a)\Gamma(c)) \\ &\qquad \cdot (\Gamma(a + k)\Gamma(n - k + c)/\Gamma(n + a + c)) \\ &\sim K_3 \textstyle\sum_{k=0}^n (n/(k + a))^b (n/(n - k + b))^b k^{a-1}(n - k)^{c-1} n^{-(a+c)} \end{aligned} \tag{5.9}$$

where K_3 is a constant. The right hand side of (5.13) converges to $K_3 \int x^{a-b-1}(1 - x)^{c-b-1}\, dx$. Hence, $\sup_n E[n^{2b}(S_n + a)^{-b}(n - S_n + c)^{-b}] < \infty$ if and only if $b < \min(a, c)$, which establishes our assertion about $C_3(b)$.

Similar arguments may be used to deal with estimation of the Poisson parameter with gamma prior, and the gamma scale parameter with gamma prior and other cases of a similar nature.

REFERENCES

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

[2] BICKEL, P. J. and YAHAV, J. A. (1965). Asymptotically pointwise optimal procedures in sequential analysis. *Proc. Fifth Berkeley Symp. Prob. Statist.* **1** Univ. of California Press.

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

[4] BICKEL, P. J. and YAHAV, J. A. (1967). Some contributions to the asymptotic theory of Bayes solutions. To appear in *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete.*

[5] BICKEL, P. J. (1968). A Hajek-Renyi extension of Levy's inequality and some applications. *Acta. Math. Acad. Sci. Hungar.* (to appear).

[6] CHOW, Y. S. (1960). A martingale inequality and the law of large numbers. *Proc. Amer. Math. Soc.* **11** 107–111.

[7] TEICHER, H. (1967). A dominated ergodic type theorem. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete.* **8** 113–116.