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

BY

P. J. BICKEL and J. A. YAHAV

TECHNICAL REPORT NO. 25
NOVEMBER 10, 1967

PREPARED UNDER THE AUSPICES OF
NATIONAL SCIENCE FOUNDATION GRANT GP-5705

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
ON AN A.P.O. RULE IN SEQUENTIAL ESTIMATION
WITH QUADRATIC LOSS(*)

by

Peter J. Bickel
University of California, Berkeley

and

Joseph A. Yahav**
Tel-Aviv University, Israel

1. Introduction.

In [1] we considered the following problem. We are given a sequence \( \{Y_n, n \geq 1\} \) of random variables on a probability space \((\Omega, \mathcal{F}, P)\) and that \( Y_n \) is measurable \( \mathcal{F}_n \) where \( \{\mathcal{F}_n\} \) is an increasing sequence of sigma fields. We suppose

\[
(1.1) \quad P[Y_n > 0] = 1
\]

\[
(1.2) \quad nY_n \xrightarrow{a.s.} V
\]

where \( P(V > 0) = 1 \).

Let

\[
(1.3) \quad X_n(c) = Y_n + nc.
\]

We are interested in finding stopping times \( t \) (with respect to \( \mathcal{F}_n \)) which in some way minimize \( X_t(c) \). In [1] we proposed the

(*) Supported in part by Grant GP-5705 of the National Science Foundation, while the authors were visiting the Department of Statistics, Stanford University.

(**) Supported in part by a grant from O.A.R. (European Office)
Following rule which we called \( \tilde{t}(c) \): Stop for the first \( n \geq 1 \), such that \( Y_n(n+1)^{-1} \leq c \). We showed in [1] that this rule is A.P.O. and in [2] we showed that \( \tilde{t}(c) \) is asymptotically optimal. In this paper we examine how close \( X_{\tilde{t}(c)}(c) \) is to \( \inf_n X_n(c) \) as \( c \to 0 \).

Typically, \( Y_n \) represents the posterior risk (variance) in the problem of estimating an unknown parameter with quadratic loss on the basis of \( n \) observations \( z_1, \ldots, z_n \) when a suitable prior is imposed on the parameter. For details, and regularity conditions see [2] and [3]. For convenience in what follows, we will consider the abstract problem as presented in (1.1)-(1.3), rather than the estimation problem which motivates our titles.

2. The Pointwise Difference Between the Performance of the Bayes Rule and the A.P.O. Rule.

Throughout this section we assume that

\[
Y_n = Vn^{-1} + R_n, \quad V > 0
\]

where

\[
\limsup_n n(n \log \log n)^{1/2}|R_n| < \infty.
\]

This condition is satisfied under regularity conditions as has been shown in theorem 2.5 of [3].

**Theorem 2.1.** If (2.1), (2.2) hold then for the stopping rule \( \tilde{t}(c) \) we have,

\[
X_{\tilde{t}(c)}(c) - \inf_n X_n(c) = o(c^{3/4} - \epsilon)
\]
for every $\epsilon > 0$.

Proof. Let us denote

\begin{equation}
    X(c) = \inf_n X_n(x)
\end{equation}

\begin{equation}
    X(c) = \min\{X^{(1)}(c), X^{(2)}(c)\}
\end{equation}

where

\begin{equation}
    X^{(1)}(c) = \min_{1 \leq n \leq n_c} X_n(c), \quad X^{(2)}(c) = \min_{n_c < n} X_n(c),
\end{equation}

and

\begin{equation}
    n_c = [vc^{-1}]^{1/2} \rho(c)
\end{equation}

where

\begin{equation}
    \rho(c) \to 0
\end{equation}

at a rate which is $o(c^{1/2})$. Then

\begin{equation}
    X(c) \geq \min(2[vc]^{1/2}, X^{(2)}(c)I_{X^{(1)}(c) > 2[vc]^{1/2}} + cI_{X^{(1)}(c) \leq 2[vc]^{1/2}})
\end{equation}

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

\begin{equation}
    [X(c) - 2[vc]^{1/2}]^- \leq [X^{(2)}(c) - 2[vc]^{1/2}]^- I_{X^{(1)}(c) > 2[vc]^{1/2}}^+
\end{equation}

\begin{equation}
    + |2[vc]^{1/2} - c| I_{X^{(1)}(c) \leq 2[vc]^{1/2}}
\end{equation}

\begin{equation}
    \leq [X^{(2)}(c) - 2[vc]^{1/2}]^- + (c + 2[vc]^{1/2})I_{X^{(1)}(c) \leq 2[vc]^{1/2}}
\end{equation}
By Lemma 2.1 of [1] and (2.14) of [2] \( \pi_{(1)}(c) \leq 2[Vc]^{1/2} \) for \( c \) sufficiently small so it is enough to consider \( [x^{(2)}(c)-2[Vc]^{1/2}]^- \).

Let,

\[
(2.11) \quad u^c_\lambda = -[\inf_{n \geq n_c} n^\lambda R_n]^{-1}, \quad 1 < \lambda < 3/2.
\]

Choose \( c_o \) so that \( Vn^{-1} + nc + n^{-\lambda} u^c_\lambda \) is positive for all \( c \leq c_o \).

We can do this since by (2.2), (2.11) \( u^c_\lambda \uparrow 0 \). Now we define

\[
(2.12) \quad z_n(c) = Vn^{-1} + nc + n^{-\lambda} u^c_\lambda.
\]

Then

\[
(2.13) \quad x^{(2)}(c) \geq \inf_n z_n(c), \quad c \leq c_o.
\]

Define \( n_o(c) \) to be the first \( m \) such that

\[
(2.14) \quad z_m(c) = \inf_n z_n(c).
\]

Define

\[
(2.15) \quad \hat{y}_n = Vn^{-1} + n^{-\lambda} u^c_\lambda(c_o).
\]

Then \( z_n(c) \geq \hat{y}_n \) and \( \hat{y}_n \) satisfies the conditions of theorem 2.1 of [1]. Therefore

\[
(2.16) \quad n_o(c) [\frac{c}{V}]^{1/2} \rightarrow 1 \text{ a.s.}
\]

By (2.13) for \( c \) sufficiently small and any \( \epsilon > 0 \)

\[
(2.17) \quad x^{(2)}(c) \geq 2[Vc]^{1/2} + [(1-\epsilon)V^{1/2}]^{-\lambda} c^{\lambda/2} u^c_\lambda(c_o)
\]
From (2.17), (2.2) and (2.10) we have

\[(2.18) \quad X(c) = 2[Vc]^{1/2} + o(c^{3/4 - \epsilon}) \, .\]

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

\[(2.19) \quad \frac{X_{\tilde{t}(c)}}{\tilde{t}(c)} (c) \leq 2c \tilde{t}(c) + c \]

and

\[(2.20) \quad \frac{Y_{\tilde{t}(c)-1}}{\tilde{t}(c)-1} > c \tilde{t}(c) \, .\]

Therefore by (2.1) we have

\[(2.21) \quad (\tilde{t}(c)-1) \frac{V}{(\tilde{t}(c)-1)^R} + (\tilde{t}(c)-1)^R \frac{c \tilde{t}^2(c) - c \tilde{t}(c)}{\tilde{t}(c)-1} > c \tilde{t}^2(c) - c \tilde{t}(c) \, .\]

since \( \tilde{t}(c) \not\to ([1]) \) by (2.2) for \( \epsilon > 0 \), there exists \( M_\epsilon \) possibly depending on the sample sequence such that,

\[(2.22) \quad V + M_\epsilon (\tilde{t}(c)-1)^\epsilon - 1/2 \geq c \tilde{t}^2(c) - c \tilde{t}(c) \, .\]

By [1] \( c\tilde{t}^2(c) \to V \) a.s. Hence for suitable \( M_{\epsilon} \) we have

\[(2.23) \quad c\tilde{t}^2(c) \leq V + M_{\epsilon}^{l/4} - \epsilon/2 \, .\]

Finally,

\[(2.24) \quad c\tilde{t}(c) \leq [Vc]^{1/2}(1+M_{\epsilon}^{l/4} - \epsilon/2)^{1/2} \leq [Vc]^{1/2}(1+M_\epsilon^{l/4} - \epsilon/2) \, .\]

Then (2.24) and (2.19) establish,

\[(2.25) \quad \frac{X_{\tilde{t}(c)}}{\tilde{t}(c)} (c) \leq 2[Vc]^{1/2} + o(c^{3/4 - \epsilon}) \, .\]

combining (2.27) and (2.18) 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

\( C_1: \ Y_n \) is an expectation decreasing martingale, with respect to the
\( \sigma \) field.

\( C_2(\lambda): \) If

\[
(3.1) \quad U_\lambda = -[\inf_n n^\lambda R_n]^{-},
\]

then

\[
(3.2) \quad \mathbb{E}[|U_\lambda|^\lambda] < \infty
\]

for some \( \lambda > 1 \)

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

\[
(3.3) \quad \sup_n n^{-b} \mathbb{E}(Y_n^{-b}) < \infty.
\]

\( C_4: \) 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,

\[
(3.4) \quad \mathbb{E}(X(c)-[Vc]^{1/2})^{-} = O\left(\frac{1}{2} + \frac{1}{2} + \frac{\lambda-1)b(b+(\lambda-1)]^{-1}}{2} \right)
\]
Proof. We use the breakup of $X(c)$ given by (2.5) and (2.6). We begin with

Lemma 3.2. If conditions (2.1) and (2.2) and $C_\lambda(\lambda)$ hold, then

$$E(X(2)(c)-[Vc]^{1/2}) \leq E[|U_\lambda|V^{-\lambda/2}] \lambda \left[\frac{\lambda-1}{2}\right] \rho(c)(1-\lambda) + \left[\frac{\lambda-1}{\lambda+1}\right]^2$$

Proof of lemma 3.1. Recall that,

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

Let

$$Q_\lambda^c(x,\omega) = Vx^{-1} + cx + U_\lambda(\omega)x^{-\lambda}$$

and suppose $x^\lambda_c(\omega)$ is the smallest $x \geq n_c$ for which $Q_\lambda^c(x,\omega)$ achieves its minimum in the range $x \geq n_c$. Define the variable $\Delta$ by $x = [Vc^{-1}]^{1/2}(1+\Delta)$ and let $\Delta_\lambda^c(\omega)$ correspond to $x^\lambda_c(\omega)$. Note that $\Delta_\lambda^c < 0$, since $Vx^{-1}+cx$ achieves its minimum for $\Delta = 0$, and $U_\lambda \leq 0$. Consider,

$$\frac{\partial Q_\lambda^c}{\partial x} = -c(1+\Delta)^{-2} + c - \lambda U_\lambda c^{-\lambda} \left[\frac{\lambda+1}{2}\right] V^{-\lambda}$$

and

$$H(\Delta) = (2\Delta + \Delta^2)(1+\Delta)^{-1}, \Delta > -1$$

Then,

$$\text{sgn} \frac{\partial Q_\lambda^c}{\partial x} = \text{sgn}(H(\Delta) - \lambda U_\lambda c^{-\lambda} \left[\frac{\lambda-1}{2}\right] V^{-\lambda}$$

Moreover,
\[(3.11) \quad H' (\Delta) = (1+\Delta)^{\lambda-2} \{2 + 2(\lambda+1)\Delta + (\lambda+1)\Delta^2 \} \]

and hence, for \(-1 < \Delta \leq 0\), \(H' (\Delta) \leq 0\) according as \(\Delta \leq -1 + [\frac{(\lambda-1)}{(\lambda+1)}]^{1/2} \).

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

(i) If \(\frac{\lambda-1}{2} \leq \frac{\lambda+1}{2} \leq H(-1+\frac{\lambda-1}{\lambda+1})^{1/2}\) then \(\frac{\partial Q^\lambda_c}{\partial x} \geq 0\) for all \(x > 0\).

(ii) If \(\frac{\lambda-1}{2} > \frac{\lambda+1}{2} > H(-1+\frac{\lambda-1}{\lambda+1})^{1/2}\) then there exist

\(0 < x_1 < x_2\), such that \(\frac{\partial Q^\lambda_c(x,\omega)}{\partial x} = 0\) for \(x = x_1, x_2\), \(x_1\) is a local maximum of \(Q^\lambda_c\), \(x_2\) is a local minimum of \(Q^\lambda_c\) and \(x_1 < [Vc^{-1}]^{1/2} [\frac{\lambda-1}{\lambda+1}]^{1/2} < x_2\). Of course, \(x_1\) and \(x_2\) are the only local extrema of \(Q^\lambda_c\) for \(x > 0\).

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

Clearly, the second of these eventualities must hold if there exists an \(x > n_c\) such that \(Q^\lambda_c(x,\omega) \leq Q^\lambda_c(n_c,\omega)\), and hence in particular if,

\[(3.12) \quad Q^\lambda_c(n_c,\omega) \geq [Vc]^{1/2} \geq Q^\lambda_c([Vc^{-1}]^{1/2},\omega) \].

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

\[(3.13) \quad U_\lambda \geq \frac{\lambda+1}{2} \rho(c)^\lambda \frac{\lambda-1}{2} \left[ \frac{1}{\rho(c)} - \left(\frac{\rho(c)-1}{\rho(c)}\right)^2 \right] \]

If \(\rho \leq \frac{1}{2}, \quad \frac{(\rho-1)^2}{\rho} \geq \frac{1}{4\rho} \). Let \(A_c = \{\omega: U_\lambda(\omega) \geq \frac{\lambda+1}{4} \left(\frac{1}{c^2 \rho(c)^{1-\lambda}}\right)\}. \]
From (3.6), (i), (ii), (3.12) and (3.13) we see that on \( A_c \),

\[
X^{(2)}(c) \geq \inf_n (V_n^{-1} + nc) + \frac{U_\lambda}{x_2} \frac{\lambda}{\chi^2} = 2[Vc]^{1/2} + \frac{U_\lambda}{x_2} \frac{\lambda}{\chi^2}.
\]

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

\[
(X^{(2)}(c) - [Vc]^{1/2}) \leq |U_\lambda| [Vc^{-1} - \frac{\lambda - 1}{\chi^2}]^{-\frac{\lambda}{2}} + [Vc]^{1/2} I_{A_c'}
\]

where \( I_{A_c} \) is the indicator of the event \( A \) and ' denotes complementation. Now,

\[
E[V^{1/2} I_{A_c'}] = \int_{A_c'} V^{1/2} dP
\]

\[
\leq \left[ \int_{A_c'} |U_\lambda| V^{-\frac{\lambda}{2}} dP \right] \left[ 4 c^2 \rho(c)^{-1} \right]^{\lambda - 1}
\]

\[
\leq E[|U_\lambda| V^{\lambda - 1}] \left[ 4 c^2 \rho(c)^{1 - \lambda} \right].
\]

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

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

\[
E[V^{1/2} I_{X^{(1)}(c) \leq 2[Vc]^{1/2}}] \leq s^{1/2} P[X^{(1)}(c) \leq 2[Vc]^{1/2}].
\]

But

\[
P[X^{(1)}(c) \leq K] = P[Y_n \leq K - nc \text{ for some } 1 \leq n \leq n_c]
\]

\[
\leq P[Y_n \leq K - nc \text{ for some } 1 \leq n \leq s^{1/2} c^{-1/2} \rho(c)].
\]
\[ P[Y_n^{-b} \geq [K-nc]^{-b} \text{ for some } 1 \leq n \leq s^{1/2} c^{-1/2} \rho(c)] . \]

Now, \( C_1 \) implies that \( Y_n^{-b} \) is an expectation increasing non-negative martingale. We recall Chow's [5] 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, then \( E(Z_n) \leq d_n \) which are monotone increasing

\[ (3.19) \quad P[Z_n \geq c_n \text{ for some } 1 \leq n \leq m] \leq \frac{d_1}{c_1} + \sum_{k=2}^{n} \frac{d_k - d_{k-1}}{c_k} c_k^{-1} . \]

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

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

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

\[ (3.21) \quad E(Y^{1/2} I_{X^{(1)}(c) \leq 2[Vc]^{1/2}}) \leq s^{b+1} (2s^{1/2} c^{1/2})^{-b} \rho^b(c) \]

\[ \sim 2^{-b} s^{1/2} \rho^b(c) . \]

Using (2.9) and combining lemma 3.2 and (3.21) we get, under the conditions of the theorem
\[ (3.22) \quad E[X(c)-2[Vc]^{1/2}] \leq c^2 E[|U_{\lambda}|^{1/2}] \cdot \frac{\lambda}{4^2} \frac{(1-\lambda)^{\lambda}}{\rho \lambda^{1+\lambda}} \left( \frac{\lambda}{\lambda+1} \right)^2 + c^{1/2} \frac{1}{\rho} \frac{\lambda}{2} \frac{(\lambda-1)^{b+1}}{2} \frac{1}{[\rho(c)]^b}(1+o(1)). \]

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

We now replace the unpleasant condition \( C_4 \) by \( C_4' \).

\( \textbf{C}_4' \). 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

\[ (3.23) \quad E[X(c)-2[Vc]^{1/2}] = o(c^{1/2} \cdot \rho(b+\lambda-1)^{-1}-\varepsilon) \]

for every \( \varepsilon > 0 \).

\textbf{Proof}. It clearly suffices to show,

\[ (3.24) \quad E(V^{1/2} I_{[X(1)(c) \leq 2[Vc]^{1/2}]}) = o(\rho^{b-\varepsilon}(c)) \]

for every \( \varepsilon > 0 \).

Now

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

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

\[ (3.26) \quad P[X(1)(c) \leq 2[Vc]^{1/2}] = o(\rho^{b-\varepsilon}(c)) \]

for every \( \varepsilon > 0 \). On the other hand,
(3.27) \[ P[X^{(1)}(c) \leq 2[Vc]^{1/2}] \leq \sum_{k=1}^{\infty} P[X^{(1)}(c) \leq 2[kc]^{1/2}, k-1 \leq V \leq k]\]

\[ \leq \sum_{k=1}^{\infty} P[X_1(k,c) \leq 2[kc]^{1/2}, (k-1) \leq V \leq k]\]

where \( X_1(k,c) = \inf_{n \leq k}^{1/2} \sigma_{\rho(c)}(Y_{n+nc}). \)

Again by Hölder's inequality,

(3.28) \[ P[X_1(k,c) \leq 2[kc]^{1/2}, (k-1) \leq V \leq k]\]

\[ \leq \frac{1}{2^{r-1}} \leq P^r[(k-1) \leq V \leq k] P^r [X_1(k,c) \leq 2[kc]^{1/2}]\]

using (3.19) we see that,

(3.29) \[ P[X_1(k,c) \leq 2[kc]^{1/2}] \leq 2^{-b} \sigma^2 (1+o(1)).\]

Hence,

(3.30) \[ P[X^{(1)}(c) \leq 2[Vc]^{1/2}] \leq \left[\frac{\sigma(c)}{2}\right]^r \cdot \sum_{k=1}^{\infty} 2^{r} P^{r}[(k-1) \leq V \leq k]\]

The last sum is finite for every \( r \) by \( C^+_4 \) and the theorem follows.

4. **An Upper Bound for the Bayes Risk of \( t(c) \).**

We again use the representation (2.1). We will require the following condition;

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

(4.1) \[ \tilde{W}_n = \sup_n \tilde{R}_n^+ \]

then
(4.2) \[ E\left(W_{\lambda}V^{-\frac{\lambda}{2}}\right) < \infty. \]

Note that \( C_{2}(\lambda) \) and \( D_{\lambda} \) are equivalent to requiring,

(4.3) \[ E\left(V^{-\frac{\lambda}{2}}\sup_{n}n^{\lambda}|R_{n}|\right) < \infty. \]

We have,

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

(4.4) \[ E(\tilde{X}_{\tilde{t}(c)} - 2[Vc]^{1/2}) = o(c^{\lambda}). \]

**Proof.** Since \( Y_{\tilde{t}(c)} \leq c(\tilde{t}(c)+1) \) it suffices to show that,

(4.5) \[ E(c\tilde{t}(c) - [Vc]^{1/2}) = o(\max(c^{\lambda},c)). \]

Now, defining \( Y_{o} = 0 \), and \( R_{o} = 0 \)

(4.6) \[ \tilde{Y}_{\tilde{t}(c)-1} \geq c(\tilde{t}(c)-1) \]

and hence,

(4.7) \[ c\tilde{t}(c) - [Vc]^{1/2} + c\tilde{t}(c) - [Vc]^{1/2} + R_{\tilde{t}(c)-1} \geq \tilde{t}(c). \]

Note that

(4.8) \[ R^{+}_{\tilde{t}(c)-1} \leq W_{\lambda}(\tilde{t}(c)-1)^{-\lambda} \]

Define,

\[ B_{c} := \{\tilde{t}(c) \leq [Vc]^{1/2} + 1\} \]
Then,

\[(4.9) \quad E(c_t(c)) \leq \int_{B_c} [Vc]^{1/2} dP + c + \int_{B_c} \tilde{c}(c) dP.\]

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

\[(4.10) \quad \int_{B_c} \tilde{c}(c) dP \leq \int_{B_c} ([Vc]^{1/2} + 3c + V^{-\lambda/2} \omega_{\lambda} c^{\lambda/2}) dP\]

The theorem follows.

In the Bayesian estimation situation with quadratic loss \([3]\) \(y_n\) can be typically represented as

\[(4.11) \quad y_n = \frac{s_n(\theta)}{n^2} + Q_n\]

where,

\[(4.12) \quad s_n(\theta) = \sum_{i=1}^{n} u(z_i, \theta)\]

\[(4.13) \quad E_g(u(z_i, \theta)) = V(\theta)\]

and \(Q_n = o(n^{-3/2})\).

If in fact, for every \(\epsilon, \epsilon' > 0\)

\[(4.14) \quad E\left\{ \frac{\sup_{n} n^{2-\epsilon} Q_n}{V^{1-\epsilon'}} \right\} < \infty\]

then one can show,

\[(4.15) \quad E(X_{t(c)}(c) - [Vc]^{1/2})^+ = o(c^{1-\epsilon})\]

for every \(\epsilon > 0\).
5. **Examples.**

I. **Estimation of normal mean.**

We wish to estimate $\mu$ with quadratic loss on the basis of $z_1, \ldots, z_n, \ldots$ where the $z_i$ are independent $\mathcal{N}(\mu, 1)$ and $\mu$ has a prior $\mathcal{N}(\mu_0, \sigma_0^2)$ distribution. In this case it is easy to compute,

\[(5.1) \quad Y_n = (n + \sigma^2)^{-1}\]

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 $(c^{-1/2} + 1 - e^{-c})$. Similarly $\tilde{\tau}(c)$ takes $\frac{1}{2}(-\ln\sigma^2 + ((1 - \sigma^2)^2 + 4c^{-1})^{1/2})$ observations and $|N(c) - \tilde{\tau}(c)| = o(c)$.

II. **Estimation in the binomial case.**

We wish to estimate $p$ with quadratic loss on the basis of $z_1, \ldots, z_n$, 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,

\[(5.2) \quad f_{a,b}(p) = \frac{\Gamma(a)\Gamma(c)}{\Gamma(a+c)} p^{a-1}(1-p)^{c-1} \quad a, c > 0, \ 0 < p < 1.\]

In this case we have,

\[(5.3) \quad Y_n(z_1, \ldots, z_n) = \frac{s_n + a}{n} \left(\frac{n + \frac{n}{2}}{n + (a+c)(n+z_1 + a + c + 1)}\right)\]

where

\[(5.4) \quad s_n = \sum_{i=1}^{n} z_i.\]
Then,

\[
Y_n = \frac{pq}{n} - \frac{n(n + (a+c))^2}{[n(n + (a+c) + 1)]^{-1}} + \frac{[3(a+c) + 1]n^2 + (a+c)(3(a+c) + 2)n + (a+c)^2(a+c + 1)pq}{[n^2 + (a+c) + 1]}^{-1}\left\{n - 2p + (c-a)(S_n - np) - (S_n - np)^2\right\}.
\]

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

The argument uses the following lemma due to Chung recently proved in a simple fashion by Brillinger in [4].

**Lemma.** Let \( \{Z_i\} \) be independent and identically distributed with mean 0. Let \( T_n = \sum_{i=1}^{n} Z_i \). Suppose \( \mathbb{E}|Z_1|^\beta < \infty \) for \( \beta > 2 \). Then, there exists a numerical constant \( K(\beta) \) such that,

\[
\mathbb{E}|T_n|^\beta \leq \frac{\beta}{n^2} K(\beta) \mathbb{E}|Z_1|^\beta.
\]

From this we can deduce a lemma communicated to us by G. Simons,

**Lemma 5.1.** If \( \alpha > \frac{\beta}{2} \), \( \mathbb{E}|Z_1|^{\beta + \epsilon} < \infty \), \( \beta > 2 \), then,

\[
\mathbb{E}(\sup\limits_n |T_n|^{\beta - \alpha}) \leq K_2(\beta) \mathbb{E}|Z_1|^{\beta + \epsilon}
\]

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

**Proof.** Note that,

\[
\mathbb{E}(\sup\limits_n |T_n|^{\beta - \alpha}) \leq \sum_{k=0}^{\infty} \mathbb{P}[\sup\limits_n |T_n|^{\beta - \alpha} \geq k].
\]

Now,

\[
\mathbb{P}[\sup\limits_n |T_n|^{\beta - \alpha} \geq k] = \mathbb{P}[|T_n|^{\beta + \epsilon} \geq (kn^{-\alpha})^{\frac{\beta + \epsilon}{\beta}} \text{ for some } n].
\]
By (3.19) since $|T_n|^{\beta+\epsilon}$ is an expectation increasing martingale, and applying Chung's lemma we get

(5.10) $\mathbb{P}[|T_n|^{\beta+\epsilon} \geq (kn^\alpha)^{\beta+\epsilon} \text{ for some } n]$

$\leq K(\beta) k^{\beta} \mathbb{E} |Z_1|^{\beta+\epsilon} \sum_{n=2}^{\infty} \frac{n^{2-(n-1)\beta}}{\alpha^{\beta+\epsilon}}$

$\leq K_2(\beta) k^{\beta} \mathbb{E} |Z_1|^{\beta+\epsilon}$

where $K_2(\beta)$ is suitably adjusted to include the sum of the series on the right (convergent for $\alpha > \frac{\beta}{2}$). Q.E.D.

Applying this lemma 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 (5.5) we see that

(5.11) $\mathbb{E}[n Y_n]^{-b} \sim n^{2b} \mathbb{E}[(S_n + a)(n-S_n + c)]^{-b}$.

Simplifying we get

(5.12) $\mathbb{E}[n^{2b} (S_n + a)^{-(n-S_n + c)^{-b}}] = \int_0^1 \left\{ \sum_{k=0}^{n} \binom{n}{k} \left( \frac{n}{k+a} \right)^b \left( \frac{n}{n-k+c} \right)^b p^b (1-p)^{n-k} \right\} \frac{\Gamma(a+c)}{\Gamma(a)\Gamma(c)} p^{a-1} (1-p)^{c-1} dp$

(5.13) $\mathbb{E}[n^{2b} (S_n + a)^{-(n-S_n + c)^{-b}}] \approx K_3 \sum_{k=0}^{n} \left( \frac{n}{k+a} \right)^b \left( \frac{n}{n-k+b} \right)^b \frac{k^{a-1}(n-k)^{c-1}}{a+c}$
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, $\sup_n \mathbb{E}[n^{2b}(S_n+a)^{-b}(n-S_n+c)^{-b}] < \infty$ 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


