ASYMPTOTICALLY POINTWISE OPTIMAL PROCEDURES IN SEQUENTIAL ANALYSIS

PETER J. BICKEL

and

JOSEPH A. YAHAV

UNIVERSITY OF CALIFORNIA, BERKELEY

1. Introduction

After sequential analysis was developed by Wald in the forties [5], Arrow, Blackwell, and Girshick [1] considered the Bayes problem and proved the existence of Bayes solutions. The difficulties involved in computing explicitly the Bayes solutions led Wald [6] to introduce asymptotic sequential analysis in estimation. Asymptotic in his sense, as for all subsequent authors, refers to the limiting behavior of the optimal solution as the cost of observation tends to zero. Chernoff [2] investigated the asymptotic properties of sequential testing. The testing theory was developed further by Schwarz [4] and generalized by Kiefer and Sacks [3]. This paper approaches the asymptotic theory from a slightly different point of view. We introduce the concept of "asymptotic pointwise optimality," and we construct procedures that are "asymptotically pointwise optimal" (A.P.O.) for certain rates of convergence [as $n \to \infty$] of the a posteriori risk. The rates of convergence that we consider apply under some regularity conditions to statistical testing and estimation with quadratic loss.

2. Pointwise optimality

Let $\{Y_n, n \geq 1\}$ be a sequence of random variables defined on a probability space (Ω, F, P) where Y_n is F_n measurable and $F_n \subset F_{n+1} \cdots \subset F$ for $n \geq 1$. We assume the following two conditions:

(2.1) $$P(Y_n > 0) = 1,$$

(2.2) $$Y_n \to 0, \qquad \text{a.s.}$$

Define

(2.3) $$X_n(c) = Y_n + nc \qquad \text{for} \quad c > 0.$$

Let T be the class of all stopping times defined on the σ-fields F_n. We say that $s \in T$ is "pointwise optimal" if

$$(2.4) \qquad P\left[\frac{X_s(c)}{X_t(c)} \leq 1\right] = 1 \qquad \text{for all} \quad t \in T.$$

Unfortunately, such s's usually do not exist except in essentially deterministic cases. Let us consider two examples of such situations:

$$(2.5) \qquad Y_n = \frac{V}{n}, \qquad\qquad \infty > V > 0,$$

$$(2.6) \qquad \frac{\log Y_n}{n} = U, \qquad\qquad -\infty < U < 0.$$

In these examples one easily sees that the pointwise optimal rule is given by the following:

Example 1: stop as soon as $\dfrac{V}{n(n-1)} \leq c$;

Example 2: stop as soon as $e^{nU} \leq \dfrac{c}{(1 - e^U)}$.

These examples will play a role in theorem 2.1. In nondeterministic cases one might hope that, under some conditions, we can get **A.P.O.** procedures. Let us define these more formally.

Abusing our notation, in a fashion long used in large sample theory, use the words "stopping rule" to also denote a function from $(0, \infty)$ to T, say $t(\cdot)$, $c \in (0, \infty)$, $t(c) \in T$. Now in analogy to our previous definition we say $s(\cdot)$ is **A.P.O.** if for any other $t(\cdot)$,

$$(2.7) \qquad \limsup_{c \to 0} \frac{X_{s(c)}(c)}{X_{t(c)}(c)} \leq 1, \qquad\qquad \text{a.s.}$$

Consideration of the deterministic case naturally leads us to hope for asymptotically pointwise optimal solutions in situations where the rate of convergence of Y_n stabilizes. This hope is fulfilled in the following theorem.

THEOREM 2.1.   (i) *If condition* (2.1) *holds and* $nY_n \to V$, *a.s. where* V *is a random variable such that* $P(V > 0) = 1$, *the stopping rule, which is determined by "stop the first time that* $(Y_n/n) \leq c$," *is* **A.P.O.**

(ii) *If condition* (2.1) *holds and* $(\log Y_n/n) \to U$, *a.s. where* U *is a random variable and* $P(U < 0) = 1$, *then the rules* (ii)a, "*stop the first time* $Y_n \leq c$" *and* (ii)b, "*stop the first time* $Y_n(1 - Y_n^{1/n}) \leq c$" *are* **A.P.O.**

PROOF.   Let $s_1(\cdot)$ be the stopping time defined by rule (i). Let $t(\cdot)$ be any other rule. Then

$$(2.8) \qquad \frac{X_{s_1}}{X_t} = \frac{Y_{s_1} + cs_1}{Y_t + ct} = \frac{\dfrac{Y_{s_1}}{s_1 c} + 1}{\dfrac{y_t}{s_1 c} + \dfrac{t}{s_1}} \leq \frac{2}{\dfrac{Y_t}{s_1 c} + \dfrac{t}{s_1}}.$$

It suffices to show that $\liminf_{c\to 0} (Y_t/s_1 c) + t/s_1 \geq 2$, a.s., but this follows upon remarking that $(x + 1/x) \geq 2$ for $x \geq 0$ and applying the following lemma.

LEMMA 2.1.   If $c_n \to 0$, $t(c_n)/s_1(c_n) \to x \geq 0$, $t(c_n)$ converges (possibly to $+\infty$) with probability 1, then $\liminf_{c_n \to 0} (Y_{t(c_n)}/c_n s_1(c_n)) \to 1/x$, a.s.

PROOF OF LEMMA.   It follows from the assumptions of the theorem and the definition of $s_1(c)$ that

$$(2.9) \qquad\qquad P[\lim_{c\to 0} s_1(c) = \infty] = 1.$$

Suppose first that $P[\lim_{c_n} t(c_n) < \infty] > 0$. On this set $\liminf_{c_n} Y_{t(c_n)} > 0$, and our lemma will follow in this case if we show that $c s_1(c) \to 0$, a.s. as $c \to 0$. We in fact will show the stronger

$$(2.10) \qquad\qquad c s_1^2(c) \to V, \qquad\qquad\qquad \text{a.s.}$$

This follows immediately from the inequalities

$$(2.11) \qquad\qquad \frac{Y_{s_1}}{s_1} \leq c < \frac{Y_{(s_1-1)}}{(s_1 - 1)},$$

$$(2.12) \qquad\qquad s_1 Y_{s_1} \leq s_1^2 c \leq \frac{s_1^2}{(s_1 - 1)^2}(s_1 - 1) Y_{(s_1-1)},$$

and (2.9).

The general case of the lemma, on the set $[t(c) \to \infty]$ is a consequence of the identity

$$(2.13) \qquad\qquad \frac{Y_t}{s_1 c} = \frac{s_1}{t}\frac{t Y_t}{c s_1^2}$$

and our assumptions.

We prove case (ii)a; case (ii)b follows similarly. Let $s_2(\cdot)$ be the rule defined by (ii)a, $t(\cdot)$ be any other stopping rule. Again we have,

$$(2.14) \qquad\qquad \frac{X_{s_2}}{X_t} = \frac{\dfrac{Y_{s_2}}{s_2 c} + 1}{\dfrac{Y_t}{s_2 c} + \dfrac{t}{s_2}}, \quad \text{and} \quad s_2 \to \infty, \qquad\qquad \text{a.s.}$$

But then, $Y_{s_2}/s_2 c \leq 1/s_2 \to 0$, a.s. In an analogous fashion to lemma 2.1, we use lemma 2.2.

LEMMA 2.2.   If $c_n \to 0$, $t(c_n)$ converges a.s. (possibly to $+\infty$), $t(c_n)/s_2(c_n) \to x < 1$, then, $Y_{t(c_n)}/c s_2(c_n) \to \infty$.

PROOF.   We prove first that

$$(2.15) \qquad\qquad \frac{s_2(c)}{|\log c|} \to \frac{1}{|U|}, \qquad\qquad\qquad \text{a.s.}$$

This is a consequence of inequalities,

$$(2.16) \qquad\qquad Y_{s_2} \leq c < Y_{(s_2-1)},$$

$$(2.17) \qquad\qquad \frac{\log Y_{s_2}}{s_2} \leq \frac{\log c}{s_2} < \frac{\log Y_{(s_2-1)}}{(s_2 - 1)}\frac{(s_2 - 1)}{s}.$$

Now suppose $t(c_n) \to \infty$, a.s. Then,

$$(2.18) \qquad \frac{\log y_t}{cs_2} = t\left(\frac{\log Y_t}{t} - \frac{s_2}{t}\left(\frac{\log s_2}{s_2} + \frac{\log c}{s_2}\right)\right).$$

Now,

$$(2.19) \qquad \frac{\log Y_t}{t} \to U, \qquad\qquad\qquad \text{a.s.,}$$

$$(2.20) \qquad \frac{\log c}{s_2} \to U, \qquad\qquad\qquad \text{a.s.,}$$

and

$$(2.21) \qquad \frac{s(c_n)}{t(c_n)} \to \frac{1}{x} > 1$$

by hypothesis and (2.15). Since $U < 0$, the result follows.

The theorem is proved.

COROLLARY 2.1.   *Let $N(c)$ be defined as any solution of $X_{N(c)}(c) = \inf_n X_n(c)$. Then, in both cases of theorem* 2.1,

$$(2.22) \qquad \lim_{c \to 0} \frac{X_{s(c)}(c)}{X_{N(c)}(c)} = 1.$$

PROOF.   Note that in the proof of the theorem, no use was made of the fact that the $t(c)$ is a stopping time.

REMARK.   In both cases it may readily be seen that $s_i(c)$ is strictly better than $t(c)$ if $t(c)/s(c) \not\to 1$, a.s. However, although in case (i) the converse holds, that is, $t(c)$ is also asymptotically pointwise optimal if $s(c)/t(c) \to 1$, a.s., this is not true necessarily in case (ii). However, as the existence of rule (ii)b indicates, here too there are many A.P.O. rules. We shall see more in the conclusion.

3. Sequential estimation with quadratic loss

The main theorem of this section states that for the one parameter exponential family (Koopman-Darmois, K-D), Bayes estimation with quadratic loss satisfies condition (i) of theorem 2.1, and therefore the rule given in theorem 2.1 (i) is A.P.O. This result can in fact be generalized to an arbitrary family of distributions under some regularity conditions. A theorem of this type will be stated at the end of this section. We give the proof only for the K-D family both for ease of exposition and because we hope to weaken the regularity conditions of our general theorem. Let $\{Z_i, i \geq 1\}$ be a sequence of independent identically distributed random variables having density function $f_\theta(z) = e^{q(\theta)T(z) - b(\theta)}$ with respect to some σ-finite nondegenerate measure μ on the real line endowed with the Borel σ-field where $q(\theta)$ and $T(z)$ are real-valued.

We let Θ, the parameter space, be the natural range of θ; that is,

$$\Theta = \left\{\theta \colon \int_{-\infty}^{\infty} e^{q(\theta)T(z)}\mu(dz) < \infty\right\}$$

and endow it also with the Borel σ-field and the usual topology.

It follows that Θ is an interval, finite or infinite, and we assume that (i) $q(\theta)$ possesses at least two continuous derivatives in the interior of Θ and (ii), $q'(\theta) \neq 0$.

It is known that under these conditions the following propositions hold:

(A) $E_\theta[T(Z_1)] = \dfrac{b'(\theta)}{q'(\theta)}$.

(B) If $q(\theta) = \theta$, then $b''(\theta) = \text{var}_\theta [T(Z_1)]$.

(C) Let $\Phi(\theta, z) = \log f_\theta(z)$ and $A(\theta) = E_\theta\left[\left(\dfrac{\delta\Phi(\theta, Z_1)}{\delta\theta}\right)\right]$; then $0 < A(\theta) = [q'(\theta)]^2 \text{var}_\theta [T(Z_1)] < \infty$;

(D) For θ_0 in the interior of Θ, the equation $\displaystyle\sum_{i=1}^{n} \dfrac{\delta\Phi(\theta, Z_i)}{\delta\theta} = 0$

has eventually a unique solution $\hat\theta_n$, the maximum likelihood estimate, and $\hat\theta_n \to \theta_0$, a.s. P_{θ_0} where P_{θ_0} is the measure induced on the space of all real sequences $\{z_1, z_2, \cdots\}$ by the density $f_{\theta_0}(z)$.

Let ν be a probability measure on Θ which has a continuous bounded density Ψ with respect to Lebesgue measure such that $\int \theta^2 \Psi(\theta)\, d\theta < \infty$. Consider the problem of estimating θ sequentially, where the loss on taking n observations and deciding $\theta = d$ is given by $nc + (d - \theta_0)^2$ when θ_0 is the true value of the parameter. The overall risk, $R(\bar\theta, t)$, for a sequential procedure consisting of a stopping rule t and estimator $\bar\theta(Z_1, \cdots, Z_t)$ is then given by,

$$(3.1) \qquad R(\bar\theta, t) = cE(t) + E[(\bar\theta(Z_1, \cdots, Z_t) - \theta)^2].$$

It follows from the results of Arrow, Blackwell, and Girshick that whatever be the choice of t, the optimal estimate given t is the conditional expectation of θ given the past, $E[\theta|Z_1, \cdots, Z_t] = \bar\theta_B$. Hence, finding optimal procedures for the sequential problem is equivalent to constructing optimal stopping rules for the sequence $\{X_n\}$ where $X_n = Y_n + nc$ and

$$(3.2) \qquad Y_n = E[(\theta - \bar\theta_B)^2|Z_1, \cdots, Z_n] = \text{var}\,(\theta|Z_1, \cdots, Z_n).$$

In order to find an **A.P.O.** rule by the method of theorem 2.1 (i), we have to show that $P(Y_n > 0) = 1$ and $nY_n \to V$, a.s. where $P(V > 0) = 1$.

THEOREM 3.1. *For the K-D family obeying assumptions* (i) *and* (ii), *we have* $P(Y_n > 0) = 1$ *and*

$$(3.3) \qquad\qquad\qquad nY_n \to 1/A(\theta).$$

PROOF. Since the a posteriori density exists with probability one, the conditional variance of θ is positive with probability one.

To show (3.3) we will establish

$$(3.4) \qquad P_{\theta_0}\{nE[(\theta - \bar\theta_n)^2|Z_1, \cdots, Z_n] \to 1/A(\theta_0)\} = 1$$

and

$$(3.5) \qquad n^{1/2}(E[\theta|Z_1, \cdots, Z_n] - \hat\theta_n) \to 0$$

with probability one, where $\hat{\theta}_n$ is the maximum likelihood esitmate of θ. The theorem readily follows from (3.4), (3.5), and the identity,

$$(3.6) \qquad \text{var } (\theta|Z_1, \cdots, Z_n) = E[(\theta - \hat{\theta}_n)^2|Z_1, \cdots, Z_n]$$
$$- [E(\theta|Z_1, \cdots, Z_n) - \hat{\theta}_n]^2.$$

Let us define, $\Psi^*(t|Z_1, \cdots, Z_n)$ to be the a posteriori density of $n^{1/2}(\theta - \hat{\theta}_n)$. Thus,

$$(3.7) \qquad \Psi^*(t|Z_1, \cdots, Z_n) = \exp\left\{\sum_{i=1}^{n} \Phi\left(\hat{\theta}_n + \frac{t}{\sqrt{n}}, Z_i\right)\right\}$$
$$\cdot \Psi\left(\frac{t}{\sqrt{n}} + \hat{\theta}_n\right)\left[n^{1/2} \int_{-\infty}^{\infty} \exp\left\{\sum_{i=1}^{n} \Phi(s, Z_i)\right\} \Psi(s)\, ds\right]^{-1}.$$

Equations (3.4) and (3.5) follow easily from

$$(3.8) \qquad P_{\theta_0}\left[\int_{-\infty}^{\infty} |t|^i|\Psi^*(t|Z_1, \cdots, Z_n) - \sqrt{A(\theta_0)}\phi(t\sqrt{A(\theta_0)})|\, dt \to 0\right] = 1$$

for $i = 1, 2$, where $\phi(x)$ is the standard normal density.

Define the random quantity

$$(3.9) \qquad \nu_n(t) = \exp\left\{\sum_{i=1}^{n} \Phi\left[\left(\hat{\theta}_n + \frac{t}{\sqrt{n}} Z_i\right) - \phi(\hat{\theta}_n, Z_i)\right]\right\}.$$

To prove (3.8), it suffices to show

$$(3.10) \qquad \int_{-\infty}^{\infty} |t|^i|\nu_n(t) - \sqrt{2\pi}\, \phi(\sqrt{A(\theta_0)}\, t)|\Psi\left(\hat{\theta}_n + \frac{t}{\sqrt{n}}\right) dt \to 0, \qquad \text{a.s. } P_\theta$$

for $i = 0, 1, 2$. To see this, note that by the case $i = 0$ we would have

$$(3.11) \qquad \int_{-\infty}^{\infty} \nu_n(t)\Psi\left(\hat{\theta}_n + \frac{t}{\sqrt{n}}\right) dt$$
$$- \int_{-\infty}^{\infty} \sqrt{2\pi}\, \phi(\sqrt{A(\theta_0)}\, t)\Psi\left(\hat{\theta}_n + \frac{t}{\sqrt{n}}\right) dt \to 0.$$

Now by the dominated convergence theorem, the boundedness and continuity of Ψ, and the consistency of $\hat{\theta}_n$, we have

$$(3.12) \qquad \int_{-\infty}^{\infty} \sqrt{2\pi}\, \phi(\sqrt{A(\theta_0)}\, t)\Psi\left(\hat{\theta}_n + \frac{t}{\sqrt{n}}\right) dt \to \frac{\Psi(\theta_0)\sqrt{2\pi}}{\sqrt{A(\theta_0)}}.$$

Let

$$(3.13) \qquad c_n = \int_{-\infty}^{\infty} \exp\left\{\sum_{i=1}^{n}\left(\Phi\left(\frac{s}{\sqrt{n}} + \hat{\theta}_n, Z_i\right) - \Phi(\hat{\theta}_n, Z_i)\right)\right\} \Psi\left(\frac{s}{\sqrt{n}} + \hat{\theta}_n\right) ds.$$

Then,

$$(3.14) \qquad \Psi^*(t|Z_1, \cdots, Z_n) = \frac{\nu_n(t)\Psi(\hat{\theta}_n + t/\sqrt{n})}{c_n},$$

and since Ψ^* is a probability density, we have

(3.15) $$c_n = \int_{-\infty}^{\infty} \nu_n(t) \Psi\left(\hat{\theta}_n + \frac{t}{\sqrt{n}}\right) dt.$$

Therefore,

(3.16) $$c_n \to \frac{\Psi(\theta_0)\sqrt{2\pi}}{\sqrt{A(\theta_0)}},$$

and the sufficiency of (3.10) for our result is clear.

Write (3.10) as,

(3.17) $$\int_{|t| < \delta^* \sqrt{n}} \Psi\left(\hat{\theta}_n + \frac{t}{\sqrt{n}}\right) |t|^i |\nu_n(t) - \sqrt{2\pi} \, \phi(\sqrt{A(\theta_0)} \, t)| \, dt$$

$$+ \int_{|t| \geq \delta^* \sqrt{n}} \Psi\left(\hat{\theta}_n + \frac{t}{\sqrt{n}}\right) |t|^i |\nu_n(t) - \sqrt{2\pi} \, \phi(\sqrt{A(\theta_0)} \, t)| \, dt.$$

We first establish the following lemma.

LEMMA 3.1.   *Under the above conditions,*

(3.18) $$A_n = \int_{|t| \geq \delta^* \sqrt{n}} \Psi\left(\hat{\theta}_n + \frac{t}{\sqrt{n}}\right) |t|^i \nu_n(t) \, dt \to 0, \qquad \text{a.s. } P_{\theta_0},$$

for $i = 0, 1, 2$, and all $\delta^* > 0$.

PROOF.   We change variables to $y = \frac{t}{\sqrt{n}}$. Then,

(3.19) $$A_n = n^{\frac{i+1}{2}} \int_{|y| \geq \delta^*} |y|^i \exp \sum_{i=1}^{n} \{\Phi(y + \hat{\theta}_n, Z_i) - \Phi(\hat{\theta}_n, Z_i)\} \Psi(\hat{\theta}_n + y) \, dy.$$

Define,

(3.20) $$H_n(y) = [q(\hat{\theta}_n + y) - q(\hat{\theta}_n)] \frac{1}{n} \sum_{i=1}^{n} T(Z_i) - [b(\hat{\theta}_n + y) - b(\hat{\theta}_n)].$$

Then, in our case, (3.14) reduces to

(3.21) $$A_n = n^{\frac{i+1}{2}} \int_{|y| \geq \delta^*} |y|^i \exp\{nH_n(y)\} \Psi(y + \hat{\theta}_n) \, dy.$$

By (D), for n sufficiently large the equation $H_n'(y) = 0$ has a unique solution given by $y = 0$, and moreover, 0 is then the unique local maximum of H_n.

Therefore we may conclude that

(3.22) $$\sup_{|y| \geq \delta^*} H_n(y) = \max\,(H_n(\delta^*), \qquad H_n(-\delta^*)) \leq -\mathrm{M} < 0,$$

eventually. Therefore,

(3.23) $$A_n \leq n^{\frac{i+1}{2}} \exp\,(-Mn) \int_{-\infty}^{\infty} |y|^i \Psi(y + \hat{\theta}_n) \, dy \to 0, \qquad \text{a.s. } P_{\theta},$$

since $\hat{\theta}_n$ are bounded a.s., which proves the lemma.

From lemma 3.1, the boundedness of Ψ, and the well-known properties of the normal distribution, it follows that

$$(3.24) \qquad \int_{|t| \geq \delta^* \sqrt{n}} \Psi \left(\hat{\theta}_n + \frac{t}{\sqrt{n}} \right) |t|^i |\nu_n(t) - \sqrt{2\pi} \, \phi(\sqrt{A(\theta_0)} \, t)| \, dt \to 0, \qquad \text{a.s. } P_{\theta_0}.$$

We finish the theorem with lemma 3.2.

LEMMA 3.2.   *Under the above conditions, there exists a $\delta^* > 0$ such that*

$$(3.25) \qquad B_n = \int_{|t| < \delta^* \sqrt{n}} |\nu_n(t) - \sqrt{2\pi} \, \phi(\sqrt{A(\theta_0)} \, t)| \, dt \to 0, \qquad \text{a.s. } P_{\theta_0}.$$

PROOF.   We expand $\log \nu_n(t)$ formally to get

$$(3.26) \qquad \log \nu_n(t) = \sum_{i=1}^{n} \left\{ \frac{\delta \Phi}{\delta \theta} (\hat{\theta}_n, Z_i) \frac{t}{\sqrt{n}} + \frac{\delta^2 \Phi}{\delta \theta^2} (\theta_n^*(t, Z_i), Z_i)) \frac{t^2}{2n} \right\}$$

where $\theta_n^*(t, Z_i)$ lies between $\hat{\theta}_n$ and $\hat{\theta}_n + t/\sqrt{n}$.

Of course, $\sum_{i=1}^{n} \delta \Phi / \delta \theta (\hat{\theta}_n, Z_i) = 0$ whenever this expression is valid. In our case, (3.26) is valid and simplifies to

$$(3.27) \qquad \log \nu_n(t) = \frac{t^2}{2} \left\{ q''(\theta_n^*(t)) \frac{1}{n} \sum_{i=1}^{n} T(Z_i) - b''(\theta_n^*(t)) \right\}.$$

Choose $\epsilon > 0$ so that $3\epsilon < A(\theta_0)$. Then, by the continuity of q'', there exists a $\delta^*(\epsilon)$ so that

$$(3.28) \qquad |q''(s) - q''(\theta_0)| < \frac{\epsilon |q'(\theta_0)|}{2|b'(\theta_0)|}$$

and $|b''(s) - b''(\theta_0)| < \epsilon$, for $|s - \theta_0| \leq \delta^*(\epsilon)$.

On the other hand, with probability one for n sufficiently large,

$$(3.29) \qquad \left| \frac{1}{n} \sum_{i=1}^{n} T(Z_i) - \frac{b'(\theta_0)}{q'(\theta_0)} \right| < \frac{\epsilon |q'(\theta_0)|}{2|b'(\theta_0)|},$$

and therefore, for such n, $|t/\sqrt{n}| \leq \delta^*(\epsilon)$, we have

$$(3.30) \qquad \left| \log \nu_n(t) - \frac{t^2}{2} \left\{ q''(\theta_0) \frac{b'(\theta_0)}{q'(\theta_0)} - b''(\theta_0) \right\} \right| < 3\epsilon.$$

But,

$$(3.31) \qquad q''(\theta_0) \frac{b'(\theta_0)}{q'(\theta_0)} - b'(\theta_0) = -A(\theta_0).$$

Equality (3.31) follows by double differentiation of the identity

$$(3.32) \qquad \int_{-\infty}^{\infty} e^{q(\theta)T(z) - b(\theta)} \mu(dz) = 1$$

and (C).

Therefore, $\nu_n(t) \leq \exp\{(3\epsilon - A(\theta_0))(t^2/2)\}$ for n sufficiently large, independent of t. But $\nu_n(t) - \sqrt{2\pi} \, \phi(\sqrt{A(\theta_0)} \, t) \to 0$ for each fixed t by (3.27) and (3.31). Applying the dominated convergence theorem, the lemma follows.

The theorem is now an immediate consequence since Ψ is bounded. For reference we now consider the general model and state theorem 3.2. Let Θ be an open subset of the line. Let $\{Z_i, i \geq 1\}$ be distributed according to $f_\theta(x)$, a density with respect to a σ-finite measure μ for $\theta \in \Theta$. Let Ψ be a probability

density on Θ with respect to Lebesgue measure satisfying the conditions of this section. We define $\Phi(\theta, x) = \log f_\theta(x)$ as before. We then have the following theorem.

THEOREM 3.2. *If*

(1) $\dfrac{\delta^2 \Phi}{\delta \theta^2} (\theta, x)$ *exists and is continuous for almost all x;*

(2) $E_\theta(\sup_{|s-\theta| \geq \epsilon} [\Phi(s, Z_1) - \Phi(\theta, Z_1)]) < 0$ *for almost all θ and all $\epsilon > 0$;*

(3) $E_\theta\left(\sup_{|s-\theta| < \epsilon} \left[\dfrac{\delta^2 \Phi}{\delta \theta^2} (s, Z_1) \right]^+ \right) < \infty$ *for some $\epsilon > 0$, for almost all θ;*

(4) $E_\theta\left(\dfrac{\delta^2 \Phi(\theta, Z_1)}{\delta \theta^2} \right) = -E_\theta \left[\dfrac{\delta \Phi(\theta, Z_1)}{\delta \theta} \right]^2;$

(5) *maximum likelihood estimates $\{\hat{\theta}_n\}$ of θ exist and are consistent;*

(6) Ψ *satisfies the condition of this section, is continuous, bounded, and $\int \theta^2 \Psi(\theta) \, d\theta < \infty$;*

then $nY_n \to \dfrac{1}{A(\theta)}$, a.s., where $A(\theta) = E_\theta\left(\left[\dfrac{\delta \Phi}{\delta \theta} (\theta, Z_1) \right]^2 \right)$—the Fisher information number, and $Y_n = \mathrm{var}\, (\theta | Z_1, \cdots, Z_n).$

4. Sequential testing

The main theorem of this section states that for the one parameter K-D family sequential Bayesian testing satisfies condition (ii) of theorem 2.1, and therefore the rules given in theorem 2.1 (ii)a, (ii)b are **A.P.O.**

Again we shall state a more general theorem at the end of the section whose proof will appear elsewhere.

Without loss of generality, we assume $\{Z_i, i \geq 1\}$ to be distributed according to the density $f_\theta(x) = e^{\theta T(x) - b(\theta)}$ with respect to some nondegenerate σ-finite measure μ. Let μ be as before and let ν be a probability measure on Θ such that ν assigns positive probability to any nonempty open subset of Θ.

As is customary in the testing problem, we have a decomposition of Θ into two disjoint Borel sets H and $\overline{H}$ (H complement), H being the hypothesis. We have a choice of two decisions (accepting or rejecting H); we pay no penalty for the right decision and incur a measurable loss $\ell(\theta) \geq 0$ when θ is the true parameter and we make the wrong decision. We assume that $\int \ell(\theta)\nu(d\theta) < \infty$. In addition, as usual, we pay $c > 0$ for each observation. The overall risk $R(\phi, t)$ for a sequential procedure consisting of a stopping rule t and randomized test $\phi(Z_1, \cdots, Z_t)$ is then given by

(4.1) $\qquad R(\phi, t) = cE(t) + E[\phi(Z_1, \cdots, Z_t)\ell(\theta)I_H(\theta)]$
$$+ E[(1 - \phi(Z_1, \cdots, Z_t))\ell(\theta)I_{\overline{H}}(\theta)]$$

where $I_A(\theta)$ is 1 if $\theta \in A$, and 0 otherwise. Again by [1], we can separate the

final decision problem from the stopping problem; that is, there is an obvious optimal choice of ϕ given t. We may now write $X_n = Y_n + nc$ where

$$(4.2) \qquad Y_n = \phi_B(Z_1, \cdots, Z_n)E[\ell(\theta)[I_H(\theta) - I_{\overline{H}}(\theta)]|Z_1, \cdots, Z_n]$$
$$+ E[\ell(\theta)I_{\overline{H}}(\theta)|Z_1, \cdots, Z_n],$$

and ϕ_B is the Bayes test given n observations. Now again the problem is to find optimal stopping rules for the process X_n. We will establish under some regularity conditions on ν that $\log Y_n/n \to U$ and $P(U < 0) = 1$. Let ν^* be the measure defined by $\nu^*(A) = \int_A \ell(\theta)\nu(d\theta)$. We have the following theorem.

THEOREM 4.1.   *Assume, in addition to the conditions given beforehand in this section:*

   (1) $0 < \nu^*(H) < \nu^*(\Theta)$,
   (2) $\nu(\hat{H}) = 0$ *where* $\hat{H}$ *is the boundary of* H, *and* $\nu(U) > 0$ *for all open* U, *and*
   (3) $\ell(\theta)$ *is strictly bounded away from zero outside of some compact* K. *Then,*

$$(4.3) \qquad \frac{\log Y_n}{n} \to (\nu^*) \operatorname*{ess\,sup}_{\theta \in \overline{H}} J(\theta, \theta_0)I_H(\theta_0) + (\nu^*) \operatorname*{ess\,sup}_{\theta \in H} J(\theta, \theta_0)I_{\overline{H}}(\theta_0) = B(\theta_0)$$

where

$$(4.4) \qquad J(\theta, \theta_0) = E_{\theta_0}(\Phi(\theta, Z_1) - \Phi(\theta_0, Z_1)),$$

and $I_A(\theta)$ *is the indicator function of* A.

   PROOF.   It is well known that $J(\theta, \theta_0) < 0$ if $P_\theta \neq P_{\theta_0}$. This observation and the following lemma will establish that $P[B(\theta_0) < 0] = 1$ in our case.

   LEMMA 4.1.   *In a K-D family as above,* $J(\theta, \theta_0)$ *is concave in* θ *with a unique maximum of* 0 *at* $\theta = \theta_0$.

   PROOF.   According to condition (A), $J(\theta, \theta_0) = (\theta - \theta_0)b'(\theta_0) + b(\theta_0) - b(\theta)$. Further, $J'(\theta_0, \theta_0) = 0$ and $J''(\theta, \theta_0) = -b''(\theta) < 0$ by (B). The lemma follows.

   To prove convergence of $\log Y_n/n$, it evidently suffices to consider $Y_n^{1/n}$ which is given by

$$(4.5) \qquad Y_n^{1/n} = 1 \bigg/ \left[\int_\Theta \exp nQ_n(\theta)\nu(d\theta)\right]^{1/n} \min \left\{ \left(\int_H \ell(\theta) \exp nQ_n(\theta)\nu(d\theta)\right)^{1/n}, \right.$$
$$\left. \left(\int_{\overline{H}} \ell(\theta) \exp nQ_n(\theta)\nu(d\theta)\right)^{1/n} \right\}$$

where $Q_n(\theta) = [1/n \sum_{i=1}^n T(Z_i)]\theta - b(\theta)$. Let $Q(\theta, \theta_0) = b'(\theta_0) - b(\theta)$.

   LEMMA 4.2.   *Let* $\{W_n\}$ *be a sequence of essentially bounded random variables such that* ess sup $|W_n - W| \to 0$. *Then* $E^{1/n}|W_n|^n \to$ ess sup W.

   PROOF.   By Minkowski's inequality,

$$(4.6) \qquad E^{1/n}|W|^n - E^{1/n}|W_n - W|^n \leq E^{1/n}|W_n|^n \leq E^{1/n}|W_n|^n + E^{1/n}|W_n - W|^n.$$

Since $E^{1/n}|W_n - W|^n \leq$ ess sup $|W_n - W|$, the lemma follows from the convergence of the L_n norm to the L_∞ norm. Q.E.D.

   To establish the theorem, it suffices to show that if $\nu^*(B) > 0$,

$$(4.7) \qquad \left\{\int_B [\exp Q_n(\theta)]^n \ell(\theta)\nu(d\theta)\right\}^{1/n} \to (\nu^*) \operatorname*{ess\,sup}_{\theta \in B} [\exp Q(\theta, \theta_0)],$$

and in particular,

$$(4.8) \qquad \left\{ \int_\Theta [\exp nQ_n(\theta)]\nu(d\theta) \right\}^{1/n} \to (\nu) \operatorname*{ess\,sup}_{\theta} [\exp Q(\theta, \theta_0)]$$

since then, a.s. P_{θ_0}, $Y_n^{1/n}$ converges to

$$(4.9) \qquad \frac{\min \{(\nu^*) \operatorname*{ess\,sup}_{\theta \in H} \exp Q(\theta, \theta_0), (\nu^*) \operatorname*{ess\,sup}_{\theta \in \bar{H}} \exp Q(\theta, \theta_0)\}}{(\nu^*) \operatorname*{ess\,sup}_{\theta \in \Theta} \exp Q(\theta, \theta_0)}.$$

Now $(\nu) \operatorname{ess\,sup} Q(\theta, \theta_0) = Q(\theta_0, \theta_0)$ by lemma 4.1 and condition (2), and $J(\theta, \theta_0) = Q(\theta, \theta_0) - Q(\theta_0, \theta_0)$. We prove (4.7). By lemma (4.1) and condition (3), there exists a compact K such that

$$(4.10) \qquad (\nu^*) \operatorname*{ess\,sup}_{\theta \in K \cap B} Q(\theta, \theta_0) = \operatorname*{ess\,sup}_{\theta \in B} Q(\theta, \theta_0)$$

and

$$(4.11) \qquad (\nu^*) \operatorname*{ess\,sup}_{\theta \in \bar{K} \cap B} Q(\theta, \theta_0) < \operatorname*{ess\,sup}_{\theta \in B} Q(\theta, \theta_0).$$

Now clearly $\nu^*(K \cap B) > 0$. Remark first that

$$(4.12) \qquad (\nu^*) \operatorname*{ess\,sup}_{\theta \in B} \exp Q_n(\theta) \geq \left[\int_B \exp nQ_n(\theta)\nu^*(d\theta) \right]^{1/n}$$

$$\geq \left[\int_{K \cap B} \exp nQ_n(\theta)\nu^*(d\theta) \right]^{1/n}.$$

But by lemma 4.2,

$$(4.13) \qquad \lim_n \left[\int_{K \cap B} \exp nQ_n(\theta)\nu^*(d\theta) \right]^{1/n}$$

$$= \lim_n \left[1/\nu^*(K \cap B) \cdot \int_{K \cap B} \exp nQ_n(\theta)\nu^*(d\theta) \right]^{1/n}$$

$$= (\nu^*) \operatorname*{ess\,sup}_{\theta \in K \cap B} \exp Q(\theta, \theta_0),$$

since $Q_n(\theta) \to Q(\theta, \theta_0)$, a.s. P_{θ_0} uniformly on K by the S.L.L.N. On the other hand, by lemma 4.1, (4.11), and the S.L.L.N.,

$$(4.14) \qquad (\nu^*) \operatorname*{ess\,sup}_{\theta \in B} \exp Q_n(\theta) \to (\nu^*) \operatorname*{ess\,sup}_{\theta \in B} \exp Q(\theta, \theta_0);$$

which completes the proof of the theorem.

As in section 3, we again state a general theorem without proof. Let $\{Z_i, i \geq 1\}$ be distributed according to a density $f_\theta(x)$ with respect to a σ-finite measure μ for $\theta \in \Theta \subset R^p$ for some p, Θ-Borel measurable. Let H be a measurable subset of Θ. Then let ν, $\ell(\theta)$, ν^*, Y_n, $J(\theta, \theta_0)$, $B(\theta)$, $\Phi(\theta, x)$, H be defined as before. We have the following theorem.

THEOREM 4.2. *Suppose that*

(1) $\nu(U) > 0$ *for any open set* U, $0 < \nu^*(H) < \nu^*(\Theta)$;
(2) $\nu^*(\hat{H}) = 0$;
(3) $\ell(\theta) \geq 0$ *and is strictly positive outside a compact;*

(4) $\Phi(\theta, x)$ *is continuous in θ for almost all x;*

(5) $E_{\theta_0}\{\sup_{\|s-t\| \leq \Delta(\theta_0)}|\Phi(s, Z_1) - \Phi(t, Z_1)|\} < \infty$ *for some* $\Delta(\theta_0) > 0$, *and for all* θ_0;

(6) $E_{\theta_0}[\Phi(s, Z_1)] > -\infty$ *for all s;*

(7) $E_{\theta_0}\{\sup_{\|\theta-\theta_0\| \geq K(\theta_0)} [\Phi(\theta, Z_1) - \Phi(\theta_0, Z_1)]\} \leq B(\theta_0)$ *for some* $K(\theta_0) < \infty$.

Then, $\log Y_n/n \to B(\theta_0)$, a.s. P_{θ_0}.

This theorem of course covers the multivariate as well as univariate K-D families and also many other examples. The reader will also note the by no means accidental resemblance of our conditions to those of Kiefer and Sacks in [3]. Of course, the conditions required to prove pointwise optimality are less stringent.

6. Conclusion

Some of the procedures suggested in this paper and similar **A.P.O.** procedures for estimation and testing have already appeared in the literature. Thus, Wald [6] proved that under some regularity conditions, similar to those of theorem 3.2, the following procedure is asymptotically minimax: "Stop the first time $(1/(n + 1)A(\hat{\theta}_n)) \leq c$" where $\hat{\theta}_n$ is the maximum likelihood estimate and $A(\theta)$ is as before. It is not difficult to verify that this procedure is asymptotically equivalent to the rule given in theorem 2.1 (i) since, under the conditions of theorem 3.2,

$$(6.1) \qquad \left|\frac{(n + 1)Y_n}{A(\hat{\theta}_n)} - 1\right| \to 0, \qquad \text{a.s.}$$

Schwarz [4] showed the procedure of theorem 2.1 (ii)a to have asymptotically the same shape as the optimal Bayes region for the exponential family under essentially the conditions of theorem 4.1. Kiefer and Sacks [3] extended his results to more general families and strengthened them. They proved, under some regularity conditions, that, in the presence of an indifference region between hypothesis and alternative, the procedure "Stop when Y_n is first $\leq c$" is asymptotically Bayes. It may be shown from their results that the Bayes optimal rule is **A.P.O.** as might be expected.

The procedure given in theorem 2.1 (ii)b seems to be "better" if an indifference region is not assumed. We are, at present, investigating the connection between **A.P.O.** rules and asymptotic Bayes solutions in this and other instances. The results of [3], [4], [6] give the reader some idea of what can be expected. It may be noted that the rules of Wald and one of the asymptotically Bayes rules proposed by Kiefer and Sacks which is **A.P.O.** are essentially independent of the choice of prior distribution. In general (because of dependence on large samples), the concept of asymptotic pointwise optimality seems to be "prior distribution free" a property which augurs well for its application to non-Bayesian and even nonparametric statistics. We hope to explore these questions also in subsequent papers.

Finally, it seems that the results of this paper are answers to interesting examples of a more general question. Suppose we are given a stochastic sequence of processes, $\{X_c(t)\}$ consisting of a deterministic component $\{D_c(t)\}$ and a noise component $\{N_c(t)\}$. Let $d(c)$ denote the time at which $\{D_c(t)\}$ reaches its minimum, and $o(c)$ denote the time at which $\{X_c(t)\}$ reaches its minimum and suppose that $o(c) \to \infty$ as $c \to 0$. Assume further that we can estimate $D_c(t)$ consistently from $X_c(t)$ by $\hat{D}_c(t)$, where consistency refers to the behavior of $\hat{D}$ as $c \to 0$. Let $\hat{d}(c)$ denote the approximation to $d(c)$ based on $\hat{D}_c(t)$. When is it true that $D_c(d(c)) \sim D_c(\hat{d}(c)) \sim X_c(o(c))$? Obviously, $N_c(t) \to 0$ as $c \to 0$, but further investigation is required. We intend to deal with this question in a forthcoming paper.

We would like to thank A. Dvoretzky for a remark which led to corollary 2.1.

REFERENCES

[1] K. J. ARROW, D. BLACKWELL, and M. A. GIRSHICK, "Bayes and minimax solutions of sequential decision problems," *Econometrica*, Vol. 17 (1949), pp. 213–244.
[2] H. CHERNOFF, "Sequential design of experiments," *Ann. Math. Statist.*, Vol. 30 (1959), pp. 755–770.
[3] J. KIEFER and J. SACKS, "Asymptotically optimum sequential inference and design," *Ann. Math. Statist.*, Vol. 34 (1963), pp. 705–750.
[4] G. SCHWARZ, "Asymptotic shapes of Bayes sequential testing regions," *Ann. Math. Statist.*, Vol. 33 (1962), pp. 224–236.
[5] A. WALD, *Sequential Analysis*, New York, Wiley, 1947.
[6] ———, "Asymptotic minimax solutions of sequential point estimation problems," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley and Los Angeles, University of California Press, 1950, pp. 1–11.