The Annals of Statistics
1985, Vol. 13, No. 2, 676–688

ASYMPTOTIC LOCAL MINIMAXITY IN SEQUENTIAL POINT ESTIMATION

BY MICHAEL WOODROOFE[1]

The University of Michigan

Let $X_1, X_2, \cdots$ be i.i.d. random variables with mean θ and finite, positive variance σ^2, depending on unknown parameters $\omega \in \Omega$. The problem addressed is that of finding a stopping time t for which the risk $R_A(t, \omega) = E_\omega\{A\gamma_0^2(\omega)(\bar{X}_t - \theta)^2 + t\}$ is as small as possible (in a suitable sense), where $A > 0$, γ_0 is a positive function on Ω, and $\bar{X}_t = (X_1 + \cdots + X_t)/t$. For fixed (nonrandom) sample sizes, $2\sqrt{A}(\gamma_0\sigma)$ is a lower bound for $R_A(n, \omega)$, $n \geq 1$; and the regret of a stopping time t is defined to be $r_A(t, \omega) = R_A(t, \omega) - 2\sqrt{A}(\gamma_0\sigma)$. The main results determine an asymptotic lower bound, as $A \to \infty$, for the minimax regret $M_A(\Omega_0) = \inf_t \sup_{\omega\in\Omega_0} r_A(t, \omega)$ for neighborhoods Ω_0 of arbitrary parameter points $\omega_0 \in \Omega$. The bound is obtained for multiparameter exponential families and the nonparametric case. The bound is attained asymptotically by an intuitive procedure in several special cases.

1. Introduction. Let F_ω, $\omega \in \Omega$, denote a family of probability distributions in $R = (-\infty, \infty)$, each of which has a finite mean and variance

$$\theta = \int_{-\infty}^{\infty} x dF_\omega(x) \quad \text{and} \quad \sigma^2 = \int_{-\infty}^{\infty} x^2 \, dF_\omega(x) - \theta^2.$$

The dependence of θ and σ^2 on ω is suppressed in the notation. Next, let $X_1, X_2, \cdots$ be independent random variables with common distribution F_ω for some unknown $\omega \in \Omega$. Suppose that $X_1, X_2, \cdots$ may be observed sequentially, that one must cease to observe the process at some (possibly random) time n, and that if observation is terminated at time n, then one incurs the loss

$$L_A(n, \omega) = A\gamma_0^2(\omega)(\bar{X}_n - \theta)^2 + n, \tag{1}$$

where $A > 0$, γ_0 is a positive function on Ω, and $\bar{X}_n = (X_1 + \cdots + X_n)/n$. Thus, the formulation assumes that the population mean θ is to be estimated by the sample mean $\bar{X}_n$ in which case the only remaining problem is to select an appropriate sample size n. This might be appropriate if F_ω, $\omega \in \Omega$, were the family of all nondegenerate normal distributions, the family of all distributions with a finite mean and variance, or a one-parameter exponential family with its natural parameters, since $\bar{X}_n$ is the maximum likelihood estimator of θ in each of these cases. The loss (1) has been so normalized that each observation costs one unit. The factor $A\gamma_0^2(\omega)$ determines the importance of estimation error relative to the cost of a single observation. The main results, Theorems 1 and 2, require $A \to \infty$.

Received May 1984; revised October 1984.

[1] Research supported by the National Science Foundation under MCS 8101897.

AMS 1980 *subject classification.* Primary 62L12.

Key words and phrases. Weighted squared error loss, exponential families, the nonparametric case, regret, Bayes risk, the Minimax Theorem, the Martingale Convergence Theorem.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Statistics*.
www.jstor.org

Recall that a stopping time t is a positive integer or $+\infty$ valued random variable for which the event $\{t = n\}$ is determined (measurably) by $X_1, \cdots, X_n$ for each $n = 1, 2, \cdots$. If t is any stopping time, then the risk incurred by using t is defined to be

$$R_A(t, \omega) = E_\omega\{L_A(t, \omega)\}, \quad \omega \in \Omega, \quad A > 0, \tag{2}$$

where E_ω denotes the expectation when $X_1, X_2, \cdots$ are i.i.d. with common distribution F_ω. If t is a fixed (nonrandom) sample size, say $t = n$, then the risk is

$$R_A(n, \omega) = (A/n)\gamma_0^2(\omega)\sigma^2 + n \geq 2\sqrt{A}\gamma(\omega) \tag{3}$$

where $\gamma(\omega) = \gamma_0(\omega)\sigma$. Moreover, there is equality in (3) iff $n = \sqrt{A}\gamma(\omega)$, which is unknown in many examples. The regret of a stopping time t is defined to be the additional risk incurred by using t instead of the best fixed sample size, namely

$$r_A(t, \omega) = R_A(t, \omega) - 2\sqrt{A}\gamma(\omega), \quad \omega \in \Omega, \quad A > 0. \tag{4}$$

For the case of unknown γ, Robbins (1959) suggested sequential procedures of the form

$$\tau = \tau_A = \inf\{n \geq m: n > \sqrt{A}\hat{\gamma}_n\} \tag{5}$$

where $\hat{\gamma}_n = \hat{\gamma}n(X_1, \cdots, X_n)$, $n \geq 1$, is a consistent sequence of positive estimators of $\gamma = \gamma(\omega)$ and $m \geq 1$ is an initial sample size. This suggestion has prompted substantial interest in comparing the risk of τ with $2\sqrt{A}\gamma(\omega)$, the risk of the best fixed sample size procedure. For example, when F_ω, $\omega \in \Omega$, is the family of all nondegenerate normal distributions, $\gamma_0 = 1$, and $\hat{\sigma}_n^2$, $n \geq 2$, is the sequence of maximum likelihood estimators, Starr (1966) shows that $R_A(\tau, \omega)/2\sqrt{A}\sigma \to 1$ as $A \to \infty$ for all $\omega \in \Omega$ iff $m \geq 3$, and Woodroofe (1977) shows that $r_A(\tau, \omega) \to \frac{1}{2}$ as $A \to \infty$ for all $\omega \in \Omega$, if $m \geq 6$. Recently, Martinsek (1983a) shows that $r_A(\tau, \omega)$ has a limit $c(\omega)$ in the nonparametric case (under some additional, technical conditions) and notes that $c(\omega)$ may have arbitrarily large negative values for some values of ω. The latter two papers contain many additional references.

While there is substantial research which compares τ to the best fixed sample size procedure, there is less which compares τ to other sequential procedures; and much of this work compares Bayesian analogues of τ to optimal Bayesian procedures. Woodroofe (1981) and Rehalia (1983) provide examples of such comparisons and may be consulted for further references. Bickel and Yahav (1968), Cabilio (1977), and Vardi (1979a) combine Bayesian and frequentist ideas in their comparisons. Wald and Wolfowitz (1951) derive minimax procedures, including τ, under strong regularity conditions.

Here an optimality property called asymptotic local minimax regret is investigated. Suppose that Ω is a metric space and let

$$M_A(t, \Omega_0) = \sup_{\omega\in\Omega_0} r_A(t, \omega) \tag{6}$$

and

$$M_A(\Omega_0) = \inf_t M_A(t, \Omega_0) \tag{7}$$

for stopping times t and subsets $\Omega_0 \subset \Omega$. Then a family $t = t_A$, $A > 0$, of stopping times is said to have *asymptotic local minimax regret* iff

$$\lim_{\Omega_0 \downarrow \omega_0} \limsup_{A \to \infty} [M_A(t, \Omega_0) - M_A(\Omega_0)] = 0, \quad \forall \omega_0 \in \Omega_0, \tag{8}$$

in which Ω_0 denote neighborhoods of ω_0. This definition is similar in spirit to that of Hajek (1972); but the implementation differs by focusing on regret instead of risk. The two notions are contrasted below.

When γ is continuous and $M_A(\Omega_0)$ is bounded below by $o(\sqrt{A})$ for all sufficiently small spheres, as in Theorem 1 below, it is easily seen that a family $t = t_A$, $A > 0$, with asymptotic local minimax regret must also have asymptotic local minimax risk; that is,

$$\lim_{\Omega_0 \downarrow \omega_0} \limsup_{A \to \infty} \sup_{\omega \in \Omega_0} R_A(t, \omega) / \inf_s \sup_{\omega \in \Omega_0} R_A(s, \omega) = 1$$

for all $\omega_0 \in \Omega$. The converse fails, however. So, the notion of asymptotic local minimax regret provides a finer comparison of families $t = t_A$, $A > 0$, than does asymptotic local minimax risk. Roughly, regret measures the second-order properties of stopping times, while risk measures their first-order properties. Here it is relevant that units in (8) are the original ones, the cost of a single observation.

Using (a slightly different notion of) regret in conjunction with the minimax principle in order to reduce the focus on worst cases has been advocated by several authors. See, for example, Berger (1980, Section 5.4.5).

The main results of this paper determine an asymptotic lower bound for $M_A(\Omega_0)$ for appropriate subsets $\Omega_0 \subset \Omega$. This bound is derived in Section 2 for multiparameter exponential families and in Section 4 for the nonparametric case. It is hoped that these results provide some insight into the phenomena of negative regrets, discovered by Martinsek (1983a); in any case, they provide bounds on the extent of the phenomena. The bounds are shown to be sharp and τ is shown to have asymptotic local minimax regret in some special cases in Section 3. No general result on the asymptotic optimality of τ or other procedures is included, but Section 6 outlines some approaches.

2. The parametric case.

To determine an asymptotic lower bound for $M_A(\Omega_0)$, it is convenient to consider some related Bayesian optimal stopping problems. If π is a prior distribution for which $\int_\Omega \gamma \, d\pi < \infty$, let

$$\bar{r}_A(t, \pi) = \int_\Omega r_A(t, \omega) \, d\pi(\omega) \tag{9}$$

and

$$\bar{r}_A(\pi) = \inf_t \bar{r}_A(t, \pi)$$

for stopping times t and $A > 0$. Then a simple application of the Minimax Theorem yields

$$M_A(\Omega_0) = \sup\{\bar{r}_A(\pi) \colon \pi(\Omega_0) = 1\}$$

for all $A > 0$ and all compact $\Omega_0 \subset \Omega$. The program is to determine the limit of $\bar{r}_A(\pi)$ as $A \to \infty$ for a large class of prior distributions π. Maximizing with respect to π then provides an asymptotic lower bound for $M_A(\Omega_0)$ for compact $\Omega_0 \subset \Omega$.

Computing $\bar{r}_A(\pi)$ for fixed $A > 0$ and π is an optimal stopping problem. In fact,

$$\bar{r}_A(t, \pi) = E^\pi\{A\gamma_0^2(\omega)(\theta - \bar{X}_n)^2 + n - 2\sqrt{A}\gamma(\omega)\}$$

where E^π denotes expectation with respect to a probability measure P^π under which ω has distribution π and $X_1, X_2, \cdots$ are conditionally i.i.d. with common distribution F_ω, given ω. The infimum in (9) is attained by Theorem 4.5′ of Chow, Robbins, and Siegmund (1971, page 82).

The asymptotic lower bound is first derived when F_ω, $\omega \in \Omega$, is a multiparameter exponential family and γ_0 is a smooth function. Specifically, the following is assumed: for some $k \geq 1$ and some sigma-finite measure H in R^k,

$$dF_\omega(x) = \exp\{\omega \cdot Y(x) - \psi(\omega)\}\, dH(x)$$

for $-\infty < x < \infty$ and $\omega = (\omega_1, \cdots, \omega_k) \in \Omega$, where $Y = (Y_1, \cdots, Y_k)$ is a Borel measurable transformation from $(-\infty, \infty)$ into R^k and $\omega \cdot Y = \omega_1 Y_1 + \cdots + \omega_k Y_k$; Ω is the interior of the natural parameter space;

$$Y_1(x) = x, \quad -\infty < x < \infty; \tag{10}$$
$$\mathscr{I}(\omega) = [(\partial^2\psi(\omega)/\partial\omega_i\partial\omega_j)\colon i, j = 1, \cdots, k]$$

is nonsingular for all $\omega \in \Omega$; and γ_0 is a twice continuously differentiable, positive function on Ω. *These assumptions are made for the remainder of this section* and are not repeated in the statements of the lemmas and theorem. By (10), the mean and variance of $X = Y_1$ are

$$\theta = (\partial/\partial\omega_1)\psi(\omega) \quad \text{and} \quad \sigma^2 = (\partial^2/\partial\omega_1{}^2)\psi(\omega), \quad \omega \in \Omega;$$

and $\bar{X}_n$ is the maximum likelihood estimator of θ, whenever one exists.

The following lemma is crucial to the analysis of $\bar{r}_A(\pi)$.

LEMMA 1. *If π has a twice continuously differentiable density ξ_0 with compact support in Ω, and if $\xi_0^\#$ is defined by*

$$\xi_0^\#(\omega) = [1/\xi_0(\omega)][\partial^2(\gamma_0^2\xi_0)/\partial\omega_1^2]\mathscr{I}\{\xi_0(\omega) > 0\}, \tag{11}$$

then

$$E^\pi\{\gamma_0^2(\omega)(\theta - \bar{X}_n)^2 \mid X_1, \cdots, X_n\} \tag{12}$$
$$= (1/n)E^\pi\{\gamma^2(\omega) \mid X_1, \cdots, X_n\} + (1/n)^2 E^\pi\{\xi_0^\#(\omega) \mid X_1, \cdots, X_n\}$$

*w.p.*1. (P^π) *for all* $n \geq 1$.

PROOF. Letting $C_n = C_n(X_1, \cdots, X_n)$ denote a normalizing constant, one has w.p.1.

$$
\begin{aligned}
\text{LHS}(12) &= \frac{1}{C_n}\int_\Omega \gamma_0^2(\theta - \bar{X}_n)^2 \exp[n\omega\cdot \bar{Y}_n - n\psi(\omega)]\xi_0(\omega)\,d\omega \\
&= \frac{-1}{nC_n}\int_\Omega \frac{\partial}{\partial\omega_1}\exp[n\omega\cdot\bar{Y}_n - n\psi(\omega)][(\theta-\bar{X}_n)\gamma_0^2\xi_0]\,d\omega \qquad (13)\\
&= \frac{1}{nC_n}\int_\Omega \exp[n\omega\cdot\bar{Y}_n - n\psi(\omega)]\left\{\frac{\partial}{\partial\omega_1}[(\theta-\bar{X}_n)\gamma_0^2\xi_0]\right\}d\omega
\end{aligned}
$$

by an integration by parts in which one integrates first with respect to ω_1. Then

$$
\begin{aligned}
\text{RHS}(13) &= \frac{1}{nC_n}\int_\Omega \gamma_0^2\sigma^2\exp[n\omega\cdot\bar{Y}_n - n\psi(\omega)]\xi_0\,d\omega \\
&\quad + \frac{1}{nC_n}\int_\Omega (\theta-\bar{X}_n)\exp[n\omega\cdot\bar{Y}_n - n\psi(\omega)]\frac{\partial}{\partial\omega_1}(\gamma_0^2\xi_0)\,d\omega \\
&= \text{RHS}(12)
\end{aligned}
$$

by a second integration by parts on the second integral.

With π as in Lemma 1, define martingales U_n, $n \ge 1$, V_n, $n \ge 1$, and M_n, $n \ge 1$, by

$$U_n = E^\pi\{\gamma^2(\omega) \mid X_1, \cdots, X_n\}$$

$$V_n = E^\pi\{\gamma(\omega) \mid X_1, \cdots, X_n\}$$

and

$$M_n = E^\pi\{\xi_0^\#(\omega) \mid X_1, \cdots, X_n\}$$

for $n \ge 1$. Then, by conditioning on $X_1, \cdots, X_t$, the Bayes regret $\bar{r}_A(t, \pi)$ of a stopping time t with respect to π may be written

$$
\begin{aligned}
\bar{r}_A(t, \pi) &= E^\pi\{(A/t)U_t + (A/t^2)M_t + t - 2\sqrt{A}V_t\} \\
&= E^\pi\{(A/t)(U_t - V_t^2) + (A/t^2)M_t + (1/t)(\sqrt{A}V_t - t)^2\}. \qquad (14)
\end{aligned}
$$

Let Π denote the class of prior distributions π for which π has a twice continuously differentiable density ξ_0 with compact support in Ω and $\inf_\omega \xi_0^\#(\omega) > -\infty$. Observe that Π contains distributions which peak about any given $\omega_0 \in \Omega$. In fact, if $\Omega_0 = [a_1, b_1] \times \cdots \times [a_k, b_k]$ is any nondegenerate rectangle in Ω, then Π contains all distributions with densities

$$\xi_0(\omega) = \frac{1}{C}\prod_{i=1}^k h_m\left(\frac{\omega_i - a_i}{b_i - a_i}\right) \Big/ \gamma_0^2(\omega), \quad \omega \in \Omega,$$

where

$$h_m(x) = x^m(1-x)^m I_{(0,1)}(x), \quad -\infty < x < \infty,$$

C is a normalizing constant, and $m > 2$.

LEMMA 2. *If $\pi \in \Pi$, then there are stopping times $s = s(A, \pi)$ which minimize $\bar{r}_A(t, \pi)$ with respect to t for $A > 0$ and satisfy the following: $s/\sqrt{A} \to \gamma$ in P^π-probability as $A \to \infty$; and there is a $\delta = \delta(\pi) > 0$ for which $s \geq \delta\sqrt{A}$ w.p.1 (P^π) for all $A > 0$.*

Lemma 2 is proved in Section 5. The convergence (which actually holds w.p.1) is intuitive, in view of the behavior of the best fixed sample size.

THEOREM 1. *If $\pi \in \Pi$, then*

$$\liminf_{A\to\infty} \bar{r}_A(\pi) \geq \int_\Omega \Gamma(\omega)\, d\pi(\omega), \tag{15}$$

where

$$\Gamma = \gamma^{-2}(\Delta\gamma)\mathscr{I}^{-1}\cdot(\Delta\gamma) + \gamma_0^2((\partial^2/\partial\omega_1{}^2)\gamma^{-2}) \tag{16}$$

and Δ denotes gradient.

PROOF. Let $s = s(A, \pi)$ be as in Lemma 2. Then there is a sequence $A_1, A_2, \cdots$ along which $A \to \infty$, the lim inf in (15), is attained, and $s/\sqrt{A} \to \gamma$ w.p.1 (P^π). Attention is restricted to such a sequence.

The three terms on the right side of (14) may be considered separately, with $t = s$. The last is nonnegative. For the second, $As^{-2}M_s \to \gamma^{-2}\xi_0^\#$ by Lemma 2 and the Martingale Convergence Theorem; and $As^{-2}M_s$ is uniformly integrable, since As^{-2} is bounded and M_n, $n \geq 1$, is a uniformly integrable martingale. For the first term, observe that $U_n - V_n^2$ is the conditional variance of $\gamma(\omega)$ given $X_1, \cdots, X_n$. Let $\hat{\omega}_n$ denote the maximum likelihood estimator of ω (which exists for all sufficiently large n w.p.1). Then the conditional distribution of $\sqrt{n}(\omega - \hat{\omega}_n)$ converges weakly to a normal distribution with mean vector 0 and covariance matrix $\mathscr{I}^{-1}(\omega_0)$ as $n \to \infty$ w.p.1 (P_{ω_0}) for all $\omega_0 \in \Omega$. See, for example, Bickel and Yahav (1969). So, $\liminf_{n\to\infty} n(U_n - V_n^2) \geq (\Delta\gamma)\mathscr{I}^{-1} \cdot (\Delta\gamma)$ w.p.1 (P^π) by Fatou's Lemma. Combining these observations with another application of Fatou's lemma yields

$$\liminf_{A\to\infty} \bar{r}_A(\pi) \geq \int_\Omega [\gamma^{-2}\Delta\gamma\mathscr{I}^{-1} \cdot \Delta\gamma + \gamma^{-2}\xi_0^\#]\xi_0\, d\omega \tag{17}$$

and the right sides of (15) and (17) are equal by an integration by parts on the second term in (17).

COROLLARY 1. *If Ω_0 is an open set with compact closure $\bar{\Omega}_0 \subset \Omega$, then*

$$\liminf_{A\to\infty} M_A(\Omega_0) \geq \sup_{\omega\in\Omega_0} \Gamma(\omega). \tag{18}$$

PROOF. The limit inferior of $M_A(\Omega_0)$ is at least as big as the supremum of the right side of (15) over $\pi \in \Pi$ for which $\pi(\Omega_0) = 1$; and the latter is easily seen to be the right side of (18).

When $\gamma_0 = 1/\sigma$, so that the best fixed sample size $\sqrt{A}$ is known, there is asymptotically no gain from sequential sampling.

COROLLARY 2. *If* $\gamma_0 = 1/\sigma$, *then* $\liminf_{A\to\infty} M_A(\Omega_0) \geq 0$ *for all open* Ω_0 *with compact closure in* Ω.

PROOF. When $\gamma_0 = 1/\sigma$, γ is constant and $\Gamma(\omega) = 0$ for all $\omega \in \Omega$.

3. Examples. In this section Γ is computed and compared with the asymptotic regret of τ for several special cases which have been considered in the literature.

EXAMPLE 1. *The Poisson case.* If F_ω is the Poisson distribution with mean $\theta = e^\omega$ and if $\gamma_0 = 1/\theta^\alpha$, where $-\infty < \omega < \infty$ is unknown and $-\infty < \alpha < \infty$, then Corollary 1 is applicable with

$$\Gamma(\omega) = 5(\alpha - 1/2)^2/\theta, \quad -\infty < \omega < \infty.$$

Thus, negative asymptotic local minimax regrets are not possible in this example, and large regrets must occur when θ is small and $\alpha \neq 1/2$. Vardi (1979b) shows that $r_A(\tau, \omega) = O(1)$ and $A \to \infty$ uniformly in $\theta \geq \varepsilon$ for any positive ε (for an appropriate version of τ).

EXAMPLE 2. *The exponential case.* If F_ω is the exponential distribution with mean $\theta = 1/|\omega|$ and if $\gamma_0 = 1/\theta^\alpha$, where $-\infty < \omega < 0$ is unknown and $-\infty < \alpha < \infty$, then Corollary 1 is applicable with $\gamma = \theta^{1-\alpha}$ and

$$\Gamma(\omega) = \Gamma_\alpha = (1 - \alpha)(3 - 5\alpha), \quad -\infty < \omega < 0.$$

Thus, Γ_α is negative for $3/5 < \alpha < 1$.

If $\hat{\gamma}_n = \bar{X}_n^{1-\alpha}$, $n \geq 1$, the regret of τ depends on A and θ only through $\lambda = A\theta^{2-2\alpha}$, say $r_A(\tau, \omega) = r_\lambda(\tau, 1)$. When $\alpha = 0$ and $m > 2$, Theorem 3.2 of Woodroofe (1977) asserts that $r_\lambda(\tau, 1) \to 3 = \Gamma_0$ as $\lambda \to \infty$; and a straightforward extension of this result shows that $r_\lambda(\tau, 1) \to \Gamma_\alpha$ as $\lambda \to \infty$, if $m > 2 - 2\alpha$ for any $\alpha < 1$. So, in this example with $\alpha < 1$; the bound of Corollary 1 is sharp; τ has asymptotic local minimax regret; and negative regrets are possible uniformly over $\theta \geq \varepsilon$ for any positive ε, if $3/5 < \alpha < 1$.

EXAMPLE 3. *The* $N(\theta, 1)$ *case.* If F_ω is the normal distribution with mean $\theta = \omega$ and unit variance, and if $\gamma_0(\omega) = 1/\omega$ for $-\infty < \omega < \infty$, then Corollary 1 is not directly applicable, since γ_0 is not continuous at 0. However, if Ω_0 is a finite interval for which $0 \notin \bar{\Omega}_0$, then $M_A(\Omega_0)$ is unaffected by the behavior of $\gamma_0(\omega)$ for ω near 0, and it is easily seen that Corollary 1 is valid in this case. In fact, Corollary 1 is valid when $0 \in \Omega_0$ too. In this case $\Gamma(\omega) = 3/\omega^2$; and a result of Martinsek (1983b) shows that $r_A(\tau, \omega) \to \Gamma(\omega)$ as $A \to \infty$ for each fixed $\omega \neq 0$.

EXAMPLE 4. *The $N(\theta, \sigma^2)$ case.* If F_ω is the normal distribution with unknown mean θ and unknown variance σ^2, where $-\infty < \omega_1 = \theta/\sigma^2 < \infty$ and $\omega_2 = -\tfrac{1}{2}\sigma^2 < 0$, and if $\gamma_0(\omega) = 1/\sigma^\alpha$, where $-\infty < \alpha < \infty$, then Corollary 1 is applicable with $\Gamma(\omega) = \Gamma_\alpha = (1 - \alpha)^2/2$ for all $\omega \in \Omega$, after some algebra. As in Example 2, if $\hat{\gamma}_n$ is the maximum likelihood estimator of γ for all $n \geq 2$, then the risk of τ depends on A and ω only through

$$\lambda = A\sigma^{2-2\alpha}, \quad \text{say} \quad r_A(\tau, \omega) = r_\lambda[\tau; (0, 1)];$$

and an easy extension of Theorem 3.1 of Woodroofe (1977) shows that

$$\lim_{\lambda\to\infty} r_\lambda[\tau, (0, 1)] = \Gamma_\alpha, \quad \text{if} \quad m > 3 - 2\alpha \quad \text{and} \quad \alpha < 1.$$

Thus, τ is asymptotically minimax in this case.

4. The nonparametric case. In this section the common distribution function of $X_1, X_2, \cdots$ is assumed to belong to a class $\mathscr{F}$ which satisfies the following conditions: every $F \in \mathscr{F}$ has a finite positive variance and a finite fourth moment; no $F \in \mathscr{F}$ is supported by two or fewer points; if $F_0 \in \mathscr{F}$, then there is an $\delta = \delta(F_0) > 0$ for which $\mathscr{F}$ contains the exponential family

$$dF_\omega(x) = \exp\{\omega_1 x + \omega_2 x^2 - \psi(\omega_1, \omega_2)\}\, dF_0(x), \quad -\infty < x < \infty, \tag{19}$$

where $\omega = (\omega_1, \omega_2)$, $|\omega_1| < \delta$, and $-\delta < \omega_2 < 0$. For example, the class of all continuous distribution functions with a finite fourth moment satisfies these conditions.

The information matrix of the family (19) is nonsingular, if F_0 is not supported by one or two points.

THEOREM 2. *Suppose that $\gamma_0(F) = 1/\sigma^\alpha$, where $\sigma^2 = \sigma^2(F)$ denotes the variance of F for $F \in \mathscr{F}$ and $-\infty < \alpha < \infty$. Let $F_0 \in \mathscr{F}_0 \subset \mathscr{F}$ and suppose that $\mathscr{F}_0$ contains the exponential family* (19) *for sufficiently small $\delta > 0$. Then*

$$\liminf_{A\to\infty} M_A(\mathscr{F}_0) \geq (1 - \alpha)[2 + (2 - \alpha)\rho_0^2 - \tfrac{1}{4}(3 + \alpha)\kappa_0],$$

where

$$\begin{aligned} \rho_0 &= \sigma_0^{-3} \int_{-\infty}^{\infty} (x - \mu_0)^3\, dF_0(x), \\ \kappa_0 &= \sigma_0^{-4} \int_{-\infty}^{\infty} (x - \mu_0)^4\, dF_0(x) - 1, \end{aligned} \tag{20}$$

and μ_0 and σ_0^2 denote the mean and variance of F_0.

PROOF. Define Γ by (16) for the exponential family (19) and $\gamma_0(\omega) = 1/\sigma^\alpha(\omega)$, where $\sigma^2(\omega)$ denotes the variance of F_ω. Then

$$\liminf_{A\to\infty} M_A(\mathscr{F}_0) \geq \limsup_{\omega\to 0} \Gamma(\omega) = \Gamma^*,$$

say, by Theorem 1 applied to the family (19) with the interior of its natural parameter space. In this case $\gamma = \sigma^{1-\alpha}$ and

$$\Gamma = \tfrac{1}{4}(1-\alpha)^2(\Delta \log \sigma^2)\mathscr{I}^{-1} \cdot (\Delta \log \sigma^2) + \sigma^{-2\alpha}(\partial^2/\partial\omega_1^2)\sigma^{2\alpha-2},$$

where $\sigma^2 = \sigma^2(\omega)$ and $\mathscr{I}$ denotes the Fisher information matrix. Let

$$\psi_{ij} = \psi_{ij}(\omega_1, \omega_2) = \partial^{i+j}\psi(\omega_1, \omega_2)/\partial^i\omega_1\partial^j\omega_2$$

for $i, j = 0, 1, 2, \cdots$ and $\omega_2 < 0$. Then $\sigma^2 = \psi_{20}$ and $\mathscr{I} = [\psi_{ij}\colon i + j = 2]$, so that $\Gamma(\omega)$ may be expressed in terms of ψ_{ij} with $i \le 4$ and $j \le 2$. The expression for Γ may be simplified by using the relations

$$\psi_{11} = \sigma^3\rho + 2\theta\sigma^2, \qquad \psi_{02} = \sigma^4\kappa + 4\theta\sigma^3\rho + 4\theta^2\sigma^2,$$

$$\psi_{21} = \sigma^4\kappa + 2\theta\sigma^3\rho, \qquad \psi_{30} = \sigma^3\rho, \quad \text{and} \quad \psi_{40} = \sigma^4(\kappa - 2),$$

where $\theta = \theta(\omega)$ denotes the mean of F_ω and $\rho = \rho(\omega)$ and $\kappa = \kappa(\omega)$ are defined by (20) with F_ω in place of F_0; and one finds that

$$\Gamma(\omega) = (1-\alpha)[2 + (2-\alpha)\rho^2 - \tfrac{1}{4}(3+\alpha)\kappa],$$

after some algebra. The theorem then follows by letting $\omega \to 0$.

COROLLARY 3. *If* $\gamma_0 = 1$, *then* $\liminf_{A\to\infty} M_A(\mathscr{F}_0) \ge 2 + 2\rho_0^2 - 3\kappa_0/4$.

When $\gamma_0 = 1$, Theorem 1 of Martinsek (1983a) asserts that $r_A(\tau, F_0) \to 2 + 2\rho_0^2 - 3\kappa_0/4$ as $A \to \infty$ for all nonlattice F with more than eight moments, provided that $m = m_A \to \infty$ at a suitable rate and σ^2 is estimated appropriately.

5. Proof of Lemma 2. In this section F_ω, $\omega \in \Omega$, is assumed to be an exponential family which satisfies the conditions listed in Section 2.

Recall that Lemma 2 makes two assertions about stopping times $s = s(A, \pi)$ which minimize $\bar{r}_A(t, \pi)$ with respect to t, where $A > 0$ and $\int_\Omega \gamma\, d\pi < \infty$: if $\pi \in \Pi$, then $s/\sqrt{A} \to \gamma(\omega)$ in P^π-probability as $A \to \infty$, and $s/\sqrt{A} > \delta > 0$ w.p.1 (P^π) for all $A \ge 1$ for some $\delta = \delta(\pi) > 0$.

To establish the convergence, one first notes that $\bar{r}_A(\pi) = o(\sqrt{A})$ as $A \to \infty$, if π has compact support. This follows from Theorem 4.1 of Bickel and Yahav (1968). It may also be established directly by considering stopping times of the form $t = t_A = \max\{m, [\sqrt{A}V_m]\}$, where $m \sim A^{1/4}$, $V_m = E^\pi\{\gamma(\omega) \mid X_1, \cdots, X_m\}$, and $[\cdot]$ denotes the greatest integer function; in fact, $\bar{r}_A(\pi) \le \bar{r}_A(t, \pi) = o(\sqrt{A})$ as $A \to \infty$ when π has compact support, since then $\gamma + \gamma^{-1}$ is bounded on the support of π and $V_m \to \gamma$ w.p.1 (P^π) by the Martingale Convergence Theorem. It then follows from (14) and the Martingale Convergence Theorem that $E^\pi\{s^{-1}(\sqrt{A}V_s - s)^2\} = o(\sqrt{A})$ as $A \to \infty$, when $\pi \in \Pi$; and this requires $s/\sqrt{A} \to \gamma$ in P^π-probability, when $\pi \in \Pi$.

The proof of the second assertion in Lemma 2 uses the following.

LEMMA 3. *Suppose that* π *has a continuous density with compact support. Then, for each* $n \ge 1$, *there is a* $\delta_n = \delta_n(\pi) > 0$ *for which*

$$E^\pi\{\gamma_0^2(\omega)(\theta - \bar{X}_n)^2 \mid X_1, \cdots, X_n\} \ge \delta_n, \quad \textit{w.p.1}. \tag{21}$$

PROOF. The proof of the lemma is given for the case $k \geq 2$. The case $k = 1$ is similar, and simpler.

Let $\mathscr{X}$ denote the convex support of H; let Ω_0 denote the support of ξ_0, the density of π; and let Ω_1 and Ω_2 denote the projections of Ω_0 onto the spaces of ω_1 and $\omega_2, \cdots, \omega_k$. For fixed $n \geq 1$, $(x_1, \cdots, x_n) \in \mathscr{X}^n$, and $(\omega_2, \cdots, \omega_k) \in \Omega_2$, let

$$\xi_n'(\omega_1) = (1/C_n')\exp\{n\omega_1\bar{x}_n - n\psi(\omega_1, \omega_2, \cdots, \omega_k)\}\xi_0(\omega_1, \cdots, \omega_k)$$

for $\omega_1 \in \Omega_1$, where C_n' is a normalizing constant and $\bar{x}_n = (x_1 + \cdots + x_n)/n$. Then ξ_n' is a version of the conditional density of ω_1, given $X_1 = x_1, \cdots, X_n = x_n$ and $\omega_2, \cdots, \omega_k$; and ξ_n' depends on $x_1, \cdots, x_n$ only through $\bar{x}_n$. Now, the left side of (21) is the integral of

$$Q(\bar{x}_n; \omega_2, \cdots, \omega_k) = \int_{\Omega_1} \gamma_0^2(\omega)(\theta - \bar{x}_n)^2\xi_n'(\omega_1)\, d\omega_1 \tag{22}$$

with respect to the conditional distribution of $\omega_2, \cdots, \omega_k$, given $X_1 = x_1, \cdots, X_n = x_n$. So, it suffices to show that the right side of (22) has a positive infimum over $\bar{x}_n \in \mathscr{X}$ and $(\omega_2, \cdots, \omega_k) \in \Omega_2$.

Let Θ and Θ_0 denote the images of Ω and Ω_0 under the mapping $\theta = \theta(\omega) = \partial\psi(\omega)/\partial\omega_1$. Then Θ is open, Θ_0 is compact, and $\Theta_0 \subset \Theta \subset \mathscr{X}$. So, there is a compact $\Theta_1 \subset \Theta$ for which $\Theta_0 \subset \text{int}(\Theta_1) \subset \Theta_1 \subset \Theta$. Now, the right side of (22) is positive for each $\bar{x}_n$ and $\omega_2, \cdots, \omega_k$ and is jointly continuous in its arguments; so, it has a positive infimum on the set where $\bar{x}_n \in \Theta_1$ and $(\omega_2, \cdots, \omega_k) \in \Omega_2$. Moreover, $|\theta - \bar{x}_n|$ has a positive infimum on the set where $\bar{x}_n \notin \Theta_1$; so, Q has a positive infimum on the set, $\bar{x}_n \notin \Theta_1$ and $(\omega_2, \cdots, \omega_k) \in \Omega_2$ too. The lemma follows.

To complete the proof of Lemma 2, first observe that $s = s(A, \pi)$ also minimizes the Bayes risk $\bar{R}_A(t, \pi) = E^\pi\{A\gamma_0^2(\omega)(\theta - \bar{X}_t)^2 + t\}$ with respect to t for $A > 0$ and π for which $\int_\Omega \gamma d\pi < \infty$, since $\bar{R}_A(t, \pi)$ and $\bar{r}_A(t, \pi)$ differ by a term which does not depend on t; and let $\bar{R}_A(\pi) = \inf_t \bar{R}_A(t, \pi)$ denote the minimum Bayes risk. Next, let $W_n(A, \pi) = E^\pi\{A\gamma_0^2(\omega)(\theta - \bar{X}_n)^2\}$ and let π_n^* denote the posterior distribution of given $X_1, \cdots, X_n$ for prior distributions π and $n \geq 1$. If $E^\pi\{\theta^2\gamma_0^2(\omega)\} < \infty$, then

$$s = \inf\{n \geq 1\colon W_n(A, \pi_n^*) \leq \bar{R}_A(\pi_n^*)\} \tag{23}$$

minimizes $\bar{R}_A(t, \pi)$ with respect to t for all $A > 0$. See Theorems 4.2 and 5.2 of Chow, Robbins, and Siegmund (1971).

Now suppose that $\pi \in \Pi$; let Ω_0 denote the support of π; and let B denote an upper bound for $\gamma(\omega)$ for $\omega \in \Omega_0$. Then, by considering fixed sample sizes, one finds easily that $\bar{R}_A(\rho) \leq 2\sqrt{AB} + 1$ for all prior distributions ρ with support Ω_0. In particular, $\bar{R}_A(\pi_n^*) \leq 2\sqrt{AB} + 1$ w.p.1 (P^π) for all $n \geq 1$. Since $W_n(A, \pi_n^*) \geq A\delta_n$ for some positive constants $\delta_n = \delta_n(\pi)$ for each $n \geq 1$, it follows that there are (nonrandom) integers N_A for which $N_A \to \infty$ and $s \geq N_A$ w.p.1 (P^π) for all $A \geq 1$. Next, from (23), one has $s > n$ whenever

$$E^\pi\{W_{n+1}(A, \pi_{n+1}^*) \mid X_1, \cdots, X_n\} - W_n(A, \pi_n^*) < -1; \tag{24}$$

and by Lemma 1 of Section 2, the inequality (24) may be rewritten

$$\frac{A}{n(n+1)}\left[U_n + \frac{2n+1}{n(n+1)} M_n\right] > 1 \tag{25}$$

where U_n and M_n are as in (14). Now, there are $b > 0$ and $0 < B < \infty$ for which $U_n \geq b$ and $M_n \geq -B$ w.p.1 (P^π) for all $n \geq 1$. So, the left side of (25) is at least $Ab/2n^2$ for all sufficiently large n. Thus, $s \geq \sqrt{(Ab/2)}$ for all sufficiently large A.

6. Concluding remarks. Examples 2, 3, and 4 and Corollary 3 suggest that τ may have asymptotic local minimax regret for a large class of models. Two approaches to this conjecture are described below.

One approach is to use nonlinear renewal theory to compute $r_A(\tau, \omega)$ for fixed ω and large A, as in Woodroofe (1977) and Martinsek (1983a). To yield asymptotic local minimaxity, this approach would have to establish limits uniformly whenever ω is near a given point ω_0; and this would require uniformity in the nonlinear renewal theorems of Lai and Siegmund (1977, 1979) and Lalley (1983). Uniform versions of the nonlinear renewal theorem have been obtained recently by Pollak (1984) and Zhang (1984). These may be sufficient to show that τ has asymptotic local minimax regret in nonlattice cases. Lattice cases present two additional difficulties, nonuniformity of convergence and possible asymptotic dependence between the undershoot and τ, properly normalized. There is less reason to believe that τ has asymptotic local minimax regret in such cases.

Another approach is to attempt to modify the arguments of Woodroofe (1981) and Rehalia (1983) to show that $\bar{r}_A(\tau, \pi) - \bar{r}_A(\pi) \to 0$ as $A \to \infty$, uniformly with respect to π in an appropriate class. Even for fixed π, such an extension presents technical difficulties, since the stopping times employed by the latter authors make explicit use of the prior information. If true for a large enough class of π, the result might be of interest, even without a statement of uniformity.

Still another approach to finding asymptotically optimal procedures is to abandon τ in favor of multistage procedures, following Hall's (1981) treatment of fixed width confidence intervals. For example, in the normal case with unknown θ and σ^2 and $\gamma_0 = 1$, Hall's (1981) techniques show that the following procedure has the same asymptotic properties as τ and, therefore, asymptotic local minimax regret: let $m = m_A \to \infty$ and $\delta = \delta_A \to 0$ as $A \to \infty$ in such a manner that $m \sim A^{1/3}$ and $\delta \sim m^{-1/4}$; let $\hat{\sigma}_n^2$ denote the maximum likelihood estimator of σ^2 for $n \geq 2$; and let

$$s = \max\{m, [(1-\delta)\hat{\sigma}_m\sqrt{A}] + 1\}$$

and

$$t = \max\{s, [\hat{\sigma}_s\sqrt{A}] + 1\},$$

where $[\cdot]$ denotes the greatest integer function. Then $r_A(t, \omega) \to ½$ and $A \to \infty$ for all ω.

There are a variety of possible refinements and generalizations of Theorems 1 and 2. An interesting one is to consider estimators of the form $\hat{\theta}_n = \bar{X}_n -$

$b_n(\hat{\omega}_n)/n$, $n \geq 1$, in the context of Theorem 1, where $\hat{\omega}_n$ denotes the maximum likelihood estimator of θ and b_n converge to a smooth limit b as $n \to \infty$. Such estimators have attracted substantial recent attention in the nonsequential case. See Ghosh, Sinha, and Wieand (1980) and the references cited there. In the sequential case, where $\bar{X}_t$ may be a biased estimator of ω, inclusion of the term $b_t(\hat{\omega}_t)/t$ may allow a reduction of the order of magnitude of the bias. So, reducing the bias reduces the mean squared error in some cases, but may increase it in others.

It seems likely that the approach developed here is extendable to parameters other than the mean, with the maximum likelihood estimator replacing $\bar{X}_n$, to smooth loss functions, and to more complicated data structures, such as regression models, models involving censored data, and models for which the densities do not form exponential families.

Acknowledgements. Thanks to Arthur Cohen and Adam Martinsek for helpful comments on an earlier draft of this paper. Thanks also to a referee for helpful comments and the reference to Zhang (1984).

REFERENCES

BERGER, J. (1980). *Statistical Decision Theory*. Springer, Berlin.

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

BICKEL, P. and YAHAV, J. (1969). Some contributions to the asymptotic theory of Bayes solutions. *Z. Wahrsch. verw. Gebiete* **11** 257–276.

CABILIO, P. (1977). Sequential estimation in Bernoulli trials. *Ann. Statist.* **5** 342–356.

CHOW, Y. S., ROBBINS, H. and SIEGMUND, D. (1971). *Great Expectations*. Houghton-Mifflin, New York.

GHOSH, J. K., SINHA, B. K. and WIEAND, H. S. (1980). Second order efficiency of the MLE with respect to any bounded bowl-shaped loss function. *Ann. Statist.* **8** 506–521.

HAJEK, J. (1972). Local asymptotic minimaxity and admissibility in estimation. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* **1** 175–194.

HALL, P. (1981). Asymptotic theory of triple sampling for sequential estimation of a mean. *Ann. Statist.* **11** 1229–1238.

LAI, T. L. and SIEGMUND, D. (1977). A non-linear renewal theory with applications to sequential analysis, I. *Ann. Statist.* **5** 946–954.

LAI, T. L. and SIEGMUND, D. (1979). A non-linear renewal theory with applications to sequential analysis, II. *Ann. Statist.* **7** 60–76.

LALLEY, S. (1982). Non-linear renewal theory for lattice random walks. *Sequential Analysis* **1** 193–206.

MARTINSEK, A. (1983a). Second order approximations to the risk of a sequential procedure. *Ann. Statist.* **11** 827–836.

MARTINSEK, A. (1983b). Sequential estimation with relative squared error loss. *Bull. Inst. Math Acad. Sinica* **11** 607–623.

POLLAK, M. (1984). On the asymptotic formula for the probability of a type I error of a mixture type power one test. Unpublished manuscript, Hebrew University.

REHALIA, M. (1984). *L^p asymptotic expansions for posterior moments with applications to sequential analysis*. Ph.D. thesis, Statistics, The University of Michigan.

ROBBINS, H. (1959). Sequential estimation of the mean of a normal population. In *Probability and Statistics* (the Harold Cramer volume), 235–245. Almquist and Wiksell, Uppsala.

STARR, N. (1966). On the asymptotic efficiency of a sequential procedure for estimating the mean.

Ann. Math. Statist. **37** 1173–1185.

VARDI, Y. (1979a). Asymptotic optimality of certain sequential estimators. *Ann. Statist.* **7** 1034–1039.

VARDI, Y. (1979b). Asymptotic optimal sequential estimation: the Poisson case. *Ann. Statist.* **7** 1040–1051.

WALD, A. and WOLFOWITZ, J. (1951). Asymptotic minimax solutions of sequential estimation problems. *Proc. Second Berkeley Symp. Math. Statist. Probab.* 1–11.

WOODROOFE, M. (1977). Second order approximations for sequential point and interval estimation. *Ann. Statist.* **5** 984–995.

WOODROOFE, M. (1981). A.P.O. rules are asymptotically non-deficient for estimation with squared error loss. *Z. Wahrsch. verw. Gebiete* **58** 331–341.

ZHANG, C. (1984). *Random Walk and Renewal Theory.* Ph.D. Thesis, Columbia University.

DEPARTMENT OF STATISTICS
UNIVERSITY OF MICHIGAN
ANN ARBOR, MICHIGAN 48109