Sankhyā : The Indian Journal of Statistics
1985, Volume 47, Series A, Pt. 1, pp. pp. 36-46.

MAXIMIN OPTIMAL STOPPING FOR NORMALLY DISTRIBUTED RANDOM VARIABLES

By JOSEPH D. PETRUCCELLI
Worcester Polytechnic Institute, Massachusetts

SUMMARY. The following optimal stopping problem is considered : $X_1, .., X_n$ i.i.d. $N(\mu, \sigma^2)$ random variables, (μ, σ^2) unknown, are observed sequentially. If observation is terminated immediately after X_j is observed a payoff of $(X_j - \mu)/\sigma$ is obtained. A best invariant stopping rule $\delta(n)$ is found and is shown to be maximin. In addition it is shown that $\lim_{n \to \infty} EX_{\delta(n)}/EX_{[n]} = 1$ where $X_{[n]} = \max\{X_1, .., X_n\}$.

1. Introduction

1.1. *Preliminaries.* Let $X_1, X_2, \ldots,$ be i.i.d. random variables having a continuous cumulative distribution function (c.d.f.) F. The X's are observed sequentially and we may terminate observation at any step j with a resulting payoff of X_j.

Moser (1956) and Guttman (1960) found the optimal stopping rule—the rule which maximizes EX_τ over all stopping rules τ—when F is known and a maximum of n observations are possible (see also Gilbert and Mosteller, 1966). This rule has the form

$$\gamma(n) = \min\{n, \min_{k \geqslant 1} \{k : X_k > R(n-k)\}\} \quad \ldots \quad (1.1)$$

with the R's given by

$$R(1) = EX_1, R(k+1) = E(X_1 I(X_1 > R(k))) + R(k)F(R(k)), \; k \geqslant 1. \quad \ldots \quad (1.2)$$

The above is known as the full information case. Other researchers have considered various partial information cases, in which it is known only that F is a member of a certain family of distributions, $\mathcal{F}$.

Sahaguchi (1961) and DeGroot (1968), for example, considered the problem in which $\mathcal{F}$ is the location parameter family of $N(\mu, 1)$ distributions. In their formulation there was a known cost function for taking observations and the payoff for taking the k-th observation X_k was X_k itself. The approach of both investigators was Bayesian, with a normal prior assumed for μ.

More recently Stewart (1978) studied the problem in which $\mathcal{F}$ is the family of all uniform distributions, $U[\alpha, \beta]$, $\alpha < \beta$. His approach was also

AMS (1980) *subject classification :* 60G40; 62L15; 62C20.
Key words and phrases : Optimal stopping; Invariance.

Bayesian with a conjugate prior assumed for (α, β). Samuels (1979, 1981) found minimax stopping rules, which turn out to be best invariant stopping rules, for this same problem with loss $(X_k-\alpha)/(\beta-\alpha)$.

In this paper we consider the optimal stopping problem in which $\mathcal{F}$ is the family of all normal distributions $\{F_{\mu,\sigma}\}_{\mu\in\mathcal{R}, \sigma>0}$ ($F_{\mu,\sigma}$ being the $N(\mu, \sigma^2)$ c.d.f.) and $(X_k-\mu)/\sigma$ is the payoff for choosing the k-th observation when $F_{\mu,\sigma}$ is the c.d.f. of the X's. In Section 2 a best invariant stopping rule $\delta(n)$ is derived when a maximum of n observations is possible. This rule is then shown to be maximin.

In Section 3 it is shown that one can do as well asymptotically with the maximin rule as one can do with complete availability of observations. That is, for any underlying normal distribution $F_{\mu,\sigma}$

$$\lim_{n\to\infty} (EX_{\delta(n)}/EX_{[n]}) = 1$$

where $X_{[n]} = \max\{X_1, \ldots, X_n\}$.

1.2. *Notation.* $\Phi(\cdot)$ is the standard normal c.d.f. $\psi(n, \cdot)$ is the c.d.f. of Student's t distribution with n degrees of freedom. $\bar{X}_k$ is the sample mean of $X_1, \ldots, X_k$.

$$S_k^2 = k^{-1} \sum_{j=1}^{k} (X_j-\bar{X}_k)^2$$

$X_1, X_2, \ldots \sim$ i.i.d. $F \in \mathcal{F}$ are the observations. $\boldsymbol{X}_n = (X_1, \ldots, X_n)$; $\boldsymbol{X}_n'$ is the transpose of $\boldsymbol{X}_n$.

For a parametric family of distributions $\mathcal{F} = \{F_\theta\}_{\theta\in\Theta}$, E_θ, P_θ will denote expectation and probability taken with respect to F_θ.

2. The stopping rule

2.1. *A best invariant rule.* We assume :

(i) $\mathcal{F} = \{F_{\mu,\sigma}\}_{\mu\in\mathcal{R}, \sigma>0}$ where $F_{\mu,\sigma}$ is the $N(\mu, \sigma^2)$ c.d.f.

(ii) There are n available observations, $X_1, \ldots, X_n \sim$ i.i.d. $F_{\mu,\sigma}$.

(iii) The gain to the observer for choosing X_k is

$$G((\mu, \sigma), \boldsymbol{X}_n, k) = (X_k-\mu)/\sigma.$$

Note that the gain is invariant under location and scale changes.

Let

$$Z_2 = (X_2 - \bar{X}_2)/S_2 = 1 \quad \text{if } X_2 > X_1$$
$$= -1 \quad \text{if } X_2 < X_1$$

and for $k \geqslant 3$,

$$Z_k = (X_k - \bar{X}_{k-1})/S_{k-1}.$$

Then $\boldsymbol{Z}_j = (Z_2, \ldots, Z_j)$, $j \geqslant 2$, is a maximal invariant, after $X_1, \ldots, X_j$ have been observed, for the decision problem described. Further the Z's have the desirable property of being mutually independent.

Let

$$T(n, 2, \boldsymbol{Z}_2) = E_{\mu,\sigma}(X_2 \mid Z_2) = \sigma\pi^{-1/2} + \mu, \; Z_2 = +1$$
$$= -\sigma\pi^{-1/2} + \mu, \; Z_2 = -1$$

and for $k = 3, \ldots, n$ let

$$T(n, k, \boldsymbol{Z}_k) = E_{\mu,\sigma}(X_k \mid \boldsymbol{Z}_k) = \mu + \sigma c(k) Z_k/(1 + Z_k^2/k)^{1/2},$$

where $c(k) = (2(k-1))^{1/2}\Gamma(k/2)/k\Gamma((k-1)/2)$.

Define

$$V(n, n, \boldsymbol{Z}_n) = T(n, n, \boldsymbol{Z}_n)$$
$$V(n, k, \boldsymbol{Z}_k) = \max\{T(n, k, \boldsymbol{Z}_k),\; E_{\mu,\sigma}(V(n, k+1, \boldsymbol{Z}_{k+1}) \mid \boldsymbol{Z}_k)\}$$
$$k = 2, \ldots, n-1.$$

A best invariant stopping rule will choose the first observation X_k for which

$$T(n, k, \boldsymbol{Z}_k) > E_{\mu,\sigma}(V(n, k+1, \boldsymbol{Z}_{k+1}) \mid \boldsymbol{Z}_k). \quad \ldots \quad (2.1)$$

(2.1) can be shown to be the rule: "Choose X_2 if $X_2 > X_1$ and $\pi^{-1/2} > M(n, 2)$; otherwise choose X_k if k is the smallest integer exceeding 2 for which

$$Z_k > M(n, k)/(c^2(k) - M^2(n, k)/k)^{1/2}\text{''}.$$

The $M(n, k)$ are defined recursively by:
$M(n, n-1) = 0$

$$M(n, l) = M(n, l+1)\psi(l-1, (l-1)^{1/2}M(n, l+1)/((l+1)c^2(l+1)$$
$$-M^2(n, l+1))^{1/2}) + (l/2\pi(l+1))^{1/2}(((l+1)c^2(l+1)$$
$$-M^2(n, l+1))/(l+1)c^2(l+1))^{(l-1)/2} \quad \text{if } M^2(n, l+1) < (l+1)c^2(l+1)$$
$$= M(n, l+1) \qquad \text{otherwise.}$$

In addition, if $\tau(n)$ is the best invariant rule described above, then

$$E_{\mu,\sigma}(X_{\tau(n)}-\mu)/\sigma = M(n, 2) \qquad \text{if } \pi^{-1/2} \leqslant M(n, 2)$$
$$= (M(n, 2)+\pi^{-1/2})/2 \quad \text{if } \pi^{-1/2} > M(n, 2).$$

2.2. *When is a best invariant rule maximum ?* In view of the Hunt-Stein theorem it is natural to ask if the best invariant rules derived above are maximin. That is, if for $\tau(n)$ given in Section 2.1 and Δ_n the collection of all stopping rules based on $\{X_i\}_{i=1}^k$, $1 \leqslant k \leqslant n$,

$$E_{\mu,\sigma}(X_{\tau(n)}-\mu)/\sigma = \sup_{\delta \epsilon \Delta_n} \inf_{(\mu,\sigma)\epsilon \mathcal{R}\times\mathcal{R}^+} E_{\mu,\sigma}(X_\delta-\mu)/\sigma.$$

We assume a location-scale parameter family of distributions $\mathcal{F} = \{F_{\mu,\sigma}\}_{\substack{\mu\epsilon\mathcal{R} \\ \sigma>0}}$, where $F_{\mu,\sigma}(x) = F_{0,1}((x-\mu)/\sigma)$. We further assume that $F_{\mu,\sigma}$ has a density $f_{\mu,\sigma}$. The observations $X_1, \ldots, X_n$ are sequentially and independently observed from an unknown distribution in $\mathcal{F}$.

Let $G((\mu, \sigma), \boldsymbol{X}_n, i)$ be a nonnegative function which denotes the gain for choosing X_i when $\boldsymbol{X}_n = (X_1, \ldots, X_n)$ is the vector of observations and $F_{\mu,\sigma}$ is the underlying distribution. Assume G is invariant under location-scale changes. It follows from a more general version of the Hunt-Stein theorem (Kiefer, 1957) that if G is bounded then a best invariant stopping rule will be maximin. Samuels (1979) gives a simplified derivation of Kiefer's result under the condition that $\mathcal{F}$ be the family of uniform distributions. By modifying his proof (see Appendix A) we obtain the following generalization :

Lemma 2.1 : *If there is a* $g : \mathcal{R}^n \rightarrow \mathcal{R}^+$ *for which* $G((0, 1), \boldsymbol{x}_n, i)g(\boldsymbol{x}_n)$ *is bounded for each* $i = 2, \ldots, n$ *and* $\prod_{i=1}^{n} f_{0,1}(x_i)/g(\boldsymbol{x}_n)$ *is integrable on* $\mathcal{R}^n$, *then for each stopping rule* $\tau \geqslant 2$ *based on* $X_1, \ldots, X_n$ *and any* $\varepsilon > 0$, *there is an invariant stopping rule* τ' *for which* $E_{0,1}(G((0, 1), \boldsymbol{X}_n, \tau'))$

$$\geqslant \inf_{(\mu,\sigma)\epsilon\mathcal{R}\times\mathcal{R}^+} E_{\mu,\sigma}(G((\mu, \sigma), \boldsymbol{X}_n, \tau)-\varepsilon.$$

It follows immediately

Theorem 2.1 : *If the conditions of Lemma* 2.1 *are satisfied and if there is a best invariant stopping rule* τ *among those taking at least two observations, then* τ *is maximin.*

Consider now the special choice of $\mathcal{F}$ as the family of all normal distributions, with gain function $G((\mu, \sigma), \boldsymbol{X}_n, i) = (X_i-\mu)/\sigma$. We may choose

$g(\boldsymbol{x}_n) = \exp\{-(\boldsymbol{x}_n\boldsymbol{x}_n')/4\}$ to satisfy the conditions of Lemma 2.1, and thus of Theorem 2.1. It is intuitively clear—and it can be shown rigorously (see Appendix B)—that for $n \geqslant 3$ no reasonable stopping rule will ever choose X_1. If $n = 2$ then either the rule $\tau \equiv 1$ or $\tau \equiv 2$ or any randomized combination of the two is maximin. Therefore

Corollary 2.1 : *If $\mathcal{F}$ is the family of all normal distributions and if $G((\mu, \sigma), \boldsymbol{X}_n, i) = (X_i - \mu)/\sigma$ then for $n \geqslant 2$ the best invariant rule of Section 2.1 is maximin.*

3. Asymptotics

We assume without loss of generality that $X_1, \ldots, X_n$ are standard normal random variables. We will prove :

Theorem 3.1 : *Suppose $\mathcal{F}$ is the family of all normal distributions and let $\tau(n)$ be the best invariant (maximin) stopping rule of Section 2.1. Let $X_{[n]} = max\{X_1, \ldots, X_n\}$. Then*

$$\lim_{n \to \infty} (EX_{\tau(n)}/EX_{[n]}) = 1.$$

Proof : Assume $X_1, X_2, \ldots$ i.i.d. $N(0, 1)$ and let Z_k, $k \geqslant 2$ be as defined in Section 2.1. Let $\varepsilon > 0$.

Define the invariant stopping rule

$$\begin{aligned} \sigma^*(n) &= \min\{n/2 < k \leqslant n : Z_k > (1-\varepsilon)\sqrt{2\log n}\} && \text{if such } k \text{ exists} \\ &= n+1 && \text{otherwise.} \end{aligned}$$

Let $\sigma(n) = \min\{\sigma^*(n), n\}$. It is well known that

$$X_{[n]}/\sqrt{2\log n} \xrightarrow{p} 1.$$

This, plus results of Resnick and Tompkins (1973) implies

$$X_{[n]}/\sqrt{2\log n} \xrightarrow{\text{a.s.}} 1.$$

Since $\bar{X}_k \xrightarrow{\text{a.s.}} 0$ and $S_k \xrightarrow{\text{a.s.}} 1$ one can show that

$$\max_{n/2 \leqslant k \leqslant n} Z_k/\sqrt{2\log n} \xrightarrow{\text{a.s.}} 1.$$

Now $$E(X_{\sigma(n)})/\sqrt{2\log n} \geqslant (1-\varepsilon) \int_{\{\sigma^*(n) \leqslant n\}} S_{\sigma^*(n)-1} dP$$

$$+(\sqrt{2\log n})^{-1} \int_{\{\sigma^*(n) \leqslant n\}} \bar{X}_{\sigma^*(n)-1} dP + (\sqrt{2\log n})^{-1} \int_{\{\sigma^*(n) > n\}} \bar{X}_n dP \quad \ldots (3.1)$$

By Fatou's lemma

$$\varliminf_{n\to\infty} \int_{\{\sigma^*(n)\leqslant n\}} S_{\sigma^*(n)-1}dP$$

$$> \int_{\Omega} \varliminf_{n\to\infty} I_{\{\sigma^*(n)<n\}} S_{\sigma^*(n)-1}dP = 1.$$

$$\left| \int_{\{\sigma^*(n)<n\}} \bar{X}_{\sigma^*(n)-1}dP \right|$$

$$\leqslant \int_{\Omega} |\bar{X}_{\sigma^*(n)-1}|dP \leqslant E(\sup_{n/2\leqslant k\leqslant n} |\bar{X}_k|) \leqslant (2/n)E(\sup_{1\leqslant k\leqslant n} |V_k|)$$

where $V_k = \sum_{i=1}^{k} X_i$.

Using Kolmogorov's inequality

$$P(\sup_{1\leqslant k\leqslant n} |V_k| > \nu) \leqslant E(V_k^2)/\nu^2,$$

we have $$E(\sup_{1\leqslant k\leqslant n} |V_k| \leqslant 1+\int_1^{\infty} (E(V_k^2)/\nu^2)d\nu. \quad = n+1.$$

Thus the middle term of (3.1) converges to 0.

$$\left| \int_{\{\sigma^*(n)>n\}} \bar{X}_n dP \right| \leqslant E(|\bar{X}_n|) \to 0 \text{ as } n\to\infty.$$

Thus $$\varliminf_{n\to\infty} E(X_{\sigma(n)})/\sqrt{2\log n} \geqslant 1-\varepsilon.$$

and the theorem is proved. □

By computation we obtain the following values for the quantities, $EX_{\gamma(n)}$, $\Phi(EX_{\gamma(n)})$, $EX_{\tau(n)}$ and $\Phi(EX_{\tau(n)})$:

TABLE 1. EXPECTED GAINS AND QUANTILES OF EXPECTED GAINS FOR FULL INFORMATION AND BEST INVARIANT STOPPING RULES

n	$EX_{\gamma(n)}$	$\Phi(EX_{\gamma(n)})$	$EX_{\tau(n)}$	$\Phi(EX_{\tau(n)})$
2	.39890	.65500	0	.50000
3	.62970	.73560	.28209	.61106
4	.79037	.78535	.44496	.67183
5	.91263	.81928	.54071	.70565
10	1.27620	.89906	.94844	.82855
50	2.01387	.97799	1.82391	.95692
100	2.28879	.98895	2.14987	.98422
500	2.85043	.99782	2.79111	.99737
1000	3.06626	.99892	3.02684	.99876
1500	3.18666	.99928	3.15593	.99920
2000	3.24607	.99941	3.24407	.99941

If $U_n = n(1-\Phi(EX_{\tau(n)}))$ then the relation $U_n = e^{2.66n^{-.3667}}$ gives a good fit to these values for $1000 \leqslant n \leqslant 2000$. If anything the fit indicates that for larger n the true value of U_n will be less than $e^{2.66n^{-3667}} \simeq 1+2{\cdot}66/n.3667$. Thus it seems that $\lim_{n\to\infty} U_n = 1$ (note that it is clear that $\underline{\lim}_{n\to\infty} U_n \geqslant 1$), which is the same asymptotic value as for $n(1-\Phi(EX_{[n]}))$.

This compares with an analogous result when $\mathscr{F}$ is taken to be the family of all uniform distributions. If for observations $X_1, \ldots, X_n$ from some distribution $U[\alpha, \beta]$ in $\mathscr{F}$, $\pi(n)$ is a maximin stopping rule, Samuels (1981) has shown that

$$\lim_{n\to\infty} n(1-E_{\alpha,\beta}(X_{\pi(n)}-\alpha)/(\beta-\alpha)) = 3{\cdot}478.$$

Appendix A

Proof of Lemma 2.1 : This proof is an extension of Samuel's (1979) proof to the case of an unbounded gain function and of an arbitrary location-scale parameter family $\mathscr{F}$ satisfying the conditions described.

For any stopping rule $\tau \geqslant 2$ based on the observations $X_1, \ldots, X_n$ and any $(\mu, \sigma) \in \Theta = \mathcal{R}\times\mathcal{R}^+$ (the parameter space) define an invariant stopping rule

$$\tau_{\mu,\sigma}(\boldsymbol{x}_n) = \tau(\mu+\sigma(\boldsymbol{x}_n-x_1)/|x_2-x_1|).$$

Here the sum of a vector and a scalar has the meaning

$$\boldsymbol{x}_n \pm x = (x_1\pm x, x_2\pm x, \ldots, x_n\pm x).$$

Let ν be any probability measure on the Borel sets of Θ. Let

$$\xi_\tau(\mu, \sigma) = E_{\mu,\sigma}(G((\mu, \sigma), \boldsymbol{x}_n, \tau)).$$

If τ is an invariant stopping rule then $\xi_\tau(\mu, \sigma)$ is constant in τ and σ and will be denoted ξ_τ.

We have

$$\inf_\Theta \xi_\tau(\mu, \sigma) \leqslant \int_\Theta \xi_\tau(\mu, \sigma)\nu(d\mu, d\sigma),$$

$$\int_\Theta \xi_{\tau_{\mu,\sigma}}\nu(d\mu, d\sigma) \leqslant \sup_\Theta \xi_{\tau_{\mu,\sigma}}.$$

For any two elements of Θ, (μ, σ), (μ', σ'), define

$$(\mu, \sigma) \circ (\mu', \sigma') = (\mu+\mu'\sigma, \sigma\sigma')$$

and for any subset $H \subset \Theta$ let

$$H \circ (\mu', \sigma') = \{(\mu, \sigma) \circ (\mu', \sigma') : (\mu, \sigma) \in H\}.$$

Let $\pi_0(d\mu, d\sigma) = \sigma^{-1}d\mu d\sigma$. π_0 is an infinite right invariant measure on Θ; that is for any $H \subseteq \Theta$, $\pi_0(H \circ (\mu, \sigma)) = \pi_0(H)$. Lemma 2.1 will be proved if we can show

$$\lim_{m\to\infty} \int_{\Theta} (\xi_\tau(\mu, \sigma) - \xi_{\tau_{\mu,\sigma}})\pi_m(d\mu, d\sigma) = 0 \qquad \ldots \quad (A.1)$$

where $\pi_m(H) = \pi_0(HH_m)/\pi_0(H_m)$ and the H_m are increasing subsets of Θ with $\pi_0(H_m) < \infty$. Define for $z_n \in \mathcal{R}^n$,

$$H_i(z_n) = \{(\mu, \sigma) \in \Theta : \tau(\mu + \sigma z_n) = i\}.$$

Then $(a, b) \in \Theta$ implies

$$H_i(z_n) \circ (a, b) = \{(\mu, \sigma) \in \Theta : \tau(\mu + \sigma(z_n - a)/b) = i\}.$$

Now $$\xi_\tau(\mu, \sigma) = \sum_{i=2}^{n} E_{0,1}(G((0, 1), \boldsymbol{X}_n, i)I(\tau(\mu + \sigma \boldsymbol{X}_n) = i))$$

so that

$$\int_{\Theta} \xi_\tau(\mu, \sigma)\nu(d\mu, d\sigma)$$

$$= \int_{\Theta} \nu(d\mu, d\sigma) \int_{\mathcal{R}^n} \sum_{i=2}^{n} G((0, 1), \boldsymbol{x}_n, i)I(\{\boldsymbol{x}_n : \tau(\mu + \sigma \boldsymbol{x}_n) = i\})$$

$$\times \prod_{j=1}^{n} f_{(0,1)}(x_j)dx_j$$

$$= \sum_{i=2}^{n} \int_{\mathcal{R}^n} G((0, 1), \boldsymbol{x}_n, i)\nu(H_i(\boldsymbol{x}_n)) \prod_{j=1}^{n} f_{(0,1)}(x_j)dx_j. \qquad \ldots \quad (A.2)$$

$$\int_{\Theta} \xi_{\tau_{\mu,\sigma}}\nu(d\mu, d\sigma)$$

$$= \int_{\Theta} \nu(d\mu, d\sigma) \int_{\mathcal{R}^n} \sum_{i=2}^{n} G((0, 1), \boldsymbol{x}_n, i)$$

$$\times I(\{\boldsymbol{x}_n : \tau(\mu + \sigma(\boldsymbol{x}_n - x_1)/|x_2 - x_1|) = i\}) \prod_{j=1}^{n} f_{(0,1)}(x_j)dx_j$$

$$= \sum_{i=2}^{n} \int_{\mathcal{R}^n} G((0, 1), \boldsymbol{x}_n, i)\nu(H_i(\boldsymbol{x}_n) \ (x_1, |x_2 - x_1|))$$

$$\times \prod_{j=1}^{n} f_{(0,1)}(x_j)dx_j. \qquad \ldots \quad (A.3)$$

The interchange in integration in both (A.2) and (A.3) is possible since by assumption

$$G((0, 1), \boldsymbol{x}_n, i) \prod_{j=1}^{n} f(x_j)$$

$$= [G((0, 1), \boldsymbol{x}_n, i)g(\boldsymbol{x}_n)] \left[\prod_{j=1}^{n} f(x_j)/g(\boldsymbol{x}_n)\right],$$

the first factor being bounded, the second integrable on $\mathcal{R}^n$.

If we could find a right invariant ν we would be done. However, since all right invariant measures are infinite we must resort to showing (A.1).

Let $\quad H_m = \{(\mu, \sigma) \in \Theta : |\mu| \leqslant A_m,\ 0 < b_m < \sigma \leqslant B_m < \infty\}$

with $\quad A_m \uparrow \infty,\ b_m \downarrow 0,\ B_m \uparrow \infty$ and $B_m/\log B_m = o(A_m)$.

Let π_m be as described above. Then from (A.2) and (A.3)

$$\int_{\Theta} |\xi_\tau(\mu, \sigma) - \xi_{\tau_{\mu,\sigma}}| \pi_m(d\mu, d\sigma)$$

$$= \sum_{i=2}^{n} \int_{\mathcal{R}^n} G((0, 1), \boldsymbol{x}_n, i) |\pi_m(H_i(\boldsymbol{x}_n)) - \pi_m(H_i(\boldsymbol{x}_n) \circ (x_1, |x_2 - x_1|))|$$

$$\times \prod_{j=1}^{n} f_{(0,1)}(x_j) dx_j. \qquad \ldots \quad (A.4)$$

If $h(\boldsymbol{x}_n) = (x_1, |x_2 - x_1|)$, then

$$\begin{aligned}\pi_0(\{H_i(\boldsymbol{x}_n) \circ h(\boldsymbol{x}_n)\} \cap H_m) = {} & \pi_0(H_i(\boldsymbol{x}_n) H_m \circ h(\boldsymbol{x}_n)) \\ & + \pi_0(\{(H_i(\boldsymbol{x}_n) - H_m) \circ h(\boldsymbol{x}_n)\} \cap H_m) \\ & - \pi_0(H_i(\boldsymbol{x}_n) H_m \circ h(\boldsymbol{x}_n) - H_m).\end{aligned}$$

From this we can show that (A.4) is bounded by

$$\sum_{i=2}^{n} \sup_{\boldsymbol{x}_n \in \mathcal{R}^n} \{G((0, 1), \boldsymbol{x}_n, i) g(\boldsymbol{x}_n)\}$$

$$\times \int_{\mathcal{R}^n} (\pi_0(H_m \Delta (H_m \circ h(\boldsymbol{x}_n)) / \pi_0(H_m)) \left(\prod_{j=1}^{n} f_{(0,1)}(x_j) / g(\boldsymbol{x}_n) \right)$$

$$\times \prod_{j=1}^{n} dx_j.$$

Further we can show that

$$\pi_0(H_m \Delta H_m \circ h(\boldsymbol{x}_n)) / \pi_0(H_m) \to 0 \text{ as } m \to \infty$$

for all $\boldsymbol{x}_n$ with $|x_2 - x_1| > 0$.

This completes the proof. □

Appendix B

Theorem B.1 : *Let $n \geqslant 2$. For the optimal stopping problem defined in Section 2.1, any stopping rule that allows a choice of X_1 can be beaten in the maximin sense by the stopping rule $\tau_1 \equiv 1$.*

Proof : The proof is an extension to the family of all normal distributions and to the unbounded gain function $G((\mu, \sigma), \boldsymbol{x}_n, i) = (x_i - \mu)/\sigma$ of

Samuels' (1979) proof of the same result when the family is that of all uniform distributions and the gain functions are bounded. It is clear that Theorem B.1 could be generalized to other distributional families and unbounded gain functions.

The key step in the proof is the following measure theoretic result :

If m is Lebesgue measure on $\mathcal{R}$, $A \subset \mathcal{R}$ a measurable set with $0 < m(A) < \infty$, $0 < \varepsilon < 1$, then there is a finite interal I such that $m(A \cap I) > (1-\varepsilon)m(I)$ (see Halmos, 1950, p. 68).

Let τ be a stopping rule which for some (μ, σ) has positive probability of taking observation X_1. This means if $A = \{x_1 : \tau(\boldsymbol{x}_n) = 1\}$, then $m(A) > 0$. Let $\{\varepsilon_l\}_{l=1}^{\infty}$ be a sequence, $\varepsilon_l \downarrow 0$. Corresponding to each ε_l find an $I_l = (a_l, b_l)$ such that

$$m(A_\cap I_l) > (1-\varepsilon_l)m(I_l).$$

We may also find $\{(\mu_l, \sigma_l) \in \mathcal{R} \times \mathcal{R}^+, l \geqslant 1\}$ such that

$$P_{(\mu_l, \sigma_l)}(X_1 \in I_l) \to 1 \text{ as } l \to \infty.$$

In particular we choose $\mu_l = (a_l+b_l)/2$, $\sigma_l = \varepsilon_l^{1/2}(b_l-a_l)$. Then for this choice of (μ_l, σ_l) and $Z \sim N(0, 1)$

$$P_{(\mu_l, \sigma_l)}(\tau = 1) > 1-P(|Z| > (4\epsilon_l)^{-1/2})-\epsilon_l^{1/2}/(2\pi)^{1/2} \to 1 \text{ as } l \to \infty.$$

Now

$$\begin{aligned} & E_{\mu_l, \sigma_l}((x_\tau-\mu_l)/\sigma_l) \\ &= \sum_{k=1}^{n} E_{\mu_l, \sigma_l}(((X_k-\mu_l)/\sigma_l)I(\tau(\boldsymbol{x}_n) = k)) \\ &= (2\pi)^{-n/2} \sum_{k=1}^{n} \int_{\mathcal{R}^n} z_k I_{A_{k,l}}(\boldsymbol{z}_n) e^{-\boldsymbol{z}_u \boldsymbol{z}_n'/2} \prod_{j=1}^{n} dz_j \end{aligned}$$

where $A_{k,l} = \{\boldsymbol{x}_n : \tau(\sigma_l \boldsymbol{x}_n + \mu_l) = k\}$. Let $Z_1, \ldots, Z_n$ be i.i.d. $N(0, 1)$ random variables and define for $l \geqslant 1$

$$Y_l = \sum_{k=1}^{n} Z_k I_{A_{k,l}}(\boldsymbol{Z}_l).$$

It is easy to see that

$$Y_l \xrightarrow{p} Z_1 \quad \text{as} \quad l \to \infty.$$

Hence there is a subsequence $\{Y_{l_i}\}$ which converges almost surely to Z_1. Further $|Y_l|$ is bounded by

$$\sum_{k=1}^{n} [|Z_k|^2 I(\{\boldsymbol{Z}_n : |Z_k| \geqslant 1\}) + I(\{\boldsymbol{Z}_n : |Z_k| < 1\})]$$

which is integrable with respect to the density

$$f(z_n) = (2\pi)^{-n/2}e^{-z_n z_n'/2}.$$

Thus by dominated convergence

$$EY_{l_i} \to EZ_1.$$

This shows that

$$\inf_{\mu, \sigma} E_{\mu, \sigma}((X_\tau - \mu)/\sigma \leqslant E_{\mu, \sigma}((X_{\tau_1} - \mu)/\sigma) = 0$$

and so proves the theorem. □

Acknowledgement. The author thanks an anonymous referee for greatly simplifying the proof of Theorem 3.1.

References

DeGroot, Morris H. (1968) : Some problems of optimal stopping. *J. Roy. Stat. Soc.*, Series B, **30**, 108-122.

Gilbert, John P. and Mosteller, Frederick (1966) : Recognizing the maximum of a sequence. *Jour. Amer. Stat. Assoc.* **61**, 35-73.

Guttman, Irwin (1960) : On a problem of L. Moser. *Canadian Mathematical Bulletin*, **3**, 35-39.

Halmos, Paul R. (1950) : *Measure Theory*, Princeton, N. J., Van Nostrand.

Kiefer, J. (1957) : Invariance, minimax sequential estimation, and continuous time processes. *Ann. Math. Statist.*, **28**, 573-601.

Moser, L. (1956) : On a problem of Cayley. *Scripta Mathematica*, **22**, 289-292.

Resnick, S. I., and Tomkins, R. J. (1973) : Almost sure stability of maxima, *Jour. Applied Probability*, **10**, 387-401.

Sakaguchi, M. (1961) : Dynamic programming of some sequential sampling design, *Journal of Mathematical Analysis and Applications*, **2**, 446-466.

Samuels, Stephen M. (1979) : Minimax stopping rules when the underlying distribution is uniform, Purdue University Department of Statistics ***Mimeograph Series*** 79-21.

——— (1981) : Minimax stopping rules when the underlying distribution is uniform, *Jour. Amer. Stat. Assoc.* **76**, 188-197.

Stewart, Theodor J. (1978) : Optimal selection from a random sequence with learing of the underlying distribution, *Jour. Amer. Statist. Assoc.* **73**, 775-780.

Paper received : October, 1983.