The Annals of Statistics
1990, Vol. 18, No. 3, 1464–1469

DECISION THEORETIC OPTIMALITY OF THE CUSUM PROCEDURE

BY Y. RITOV

The Hebrew University of Jerusalem

Suppose $X_1, X_2, \ldots$ are independent random variables such that for some unknown ν, each of $X_1, \ldots, X_{\nu-1}$ is distributed according to F_0, while $X_\nu, X_{\nu+1}, \ldots$ are all distributed according to F_1. We prove a result of Moustakides that claims that the CUSUM procedures are optimal in the sense of Lorden. We do that by proving that the procedures are Bayes for some stochastic mechanism of generating ν.

1. Introduction. Let us assume that $X_1, X_2, \ldots$ are independent random variables, $X_1, \ldots, X_{\nu-1}$ have a common distribution F_0, while $X_\nu, X_{\nu+1}, \ldots$ are all distributed according to F_1. We assume that F_0 and F_1 are known, ν is unknown and $X_1, X_2, \ldots$ are observed sequentially. We wish to find a stopping time N that detects the change point ν as soon as possible. Let $\mathbf{F}_n$ be the σ field generated by $\{X_1, \ldots, X_n\}$.

We define the optimality of a stopping time in the sense of Lorden (1971). That is, we consider the conditional expectation of the loss function given the least favorable event before the change point. One possible formal definition of the problem is: look for a procedure that minimizes $\sup_\nu$ ess-sup $E_\nu\{(N - \nu + 1)^+ | \mathbf{F}_{n-1}\}$ subject to $E_\infty(N) \le \gamma$, where E_ν is the expectation operator when the change point is at time ν. It was proved in Moustakides (1986) that the CUSUM, or Page, procedures are optimal in that sense. We consider a slightly different version of this problem, which is more standard from the decision theory point of view. As a consequence, we obtain an alternative proof of Moustakides' (1986) result. We believe that our proof is instructive, since we prove that, in some sense, the CUSUM procedure is a Bayes procedure, a fact that is not mentioned in Moustakides (1986). To be exact, we consider the situation as a sequential stochastic game: The statistician chooses a stopping time N, while nature chooses the change time ν, and both $1_{\{N > n\}}$ and $1_{\{\nu > n+1\}}$ should be a (random) measurable function of $X_1, X_2, \ldots, X_n$.

2. Main results. Let $U_1, U_2, \ldots, V_1, V_2, \ldots$ and $W_1, W_2, \ldots$ be independent random variables, all defined on the same probability space, $U_i \sim U(0,1)$, $V_i \sim F_0$ and $W_i \sim F_1$, $i = 1, 2, \ldots$.

Let $\mathbf{N_c}$ be the class of all random variables ν such that $1_{\{\nu=1\}}$ is a measurable function of U_1 and $1_{\{\nu=n\}}$ is a measurable function of $1_{\{\nu<n\}}$, U_n and

Received June 1988; revised June 1989.
AMS 1980 *subject classifications*. 62L10, 62L15.
Key words and phrases. Page procedures, sequential games, sequential analysis, CUSUM test, SPRT.

$V_1, \dots, V_{n-1}$. Given $\nu \in \mathbf{N_c}$, define $X_n = 1_{\{\nu > n\}} V_n + 1_{\{\nu \le n\}} W_n$. The class of all stopping times for the sequence $X_1, X_2, \dots$ is $\mathbf{N_s}$. For any $N \in \mathbf{N_s}$, $E_0 N$ is the expectation of N applied to $W_1, W_2, \dots$ and $E_\infty N$ is the expectation of N applied to $V_1, V_2, \dots$.

Let $L_n = dF_1/dF_0(X_n)$ be the likelihood ratio. For simplicity, we assume that L_n is a continuous random variable both under $\nu \le n$ and $\nu > n$. Let $S_0 = 0$ and $S_n = L_n \max\{1, S_{n-1}\}$ for $n > 0$. A CUSUM stopping time with threshold A, $N_A \in \mathbf{N_s}$, is defined by $N_A = \inf\{n: S_n \ge A\}$.

We consider the loss function

$$l(\nu, N) = C_1 1_{\{N < \nu\}} - C_2 \min\{N, \nu - 1\} + C_3(N - \nu + 1)^+. \tag{1}$$

This loss function is quite reasonable. It pays to use the machine as long as possible before the change. On the other hand, one pays for false alarms and for using the machine after the change.

A strategy for nature is a specification of $P(\nu = n | \nu \ge n,\ X_1, \dots, X_{n-1})$, $n = 1, 2, \dots$. Let $\nu_p (0 \le p \le 1)$ be defined by

$$P(\nu_p = n | \nu_p \ge n, \mathbf{F}_{n-1}) = p(1 - S_{n-1})^+.$$

Note that since the change may happen at any n such that $S_{n-1} < 1$, $P(\nu_p < \infty) = 1$.

Let $p_0 = 0$ and for $n \ge 1$, define p_n by

$$\frac{p_n}{1 - p_n} = \frac{p}{1 - p} S_n.$$

LEMMA 1.

$$P(\nu_p \le n | \mathbf{F}_n) = p_n.$$

PROOF. First,

$$\frac{P(\nu_p \le 1 | \mathbf{F}_1)}{1 - P(\nu_p \le 1 | \mathbf{F}_1)} = \frac{P(\nu_p = 1)}{1 - P(\nu_p = 1)} L_1 = \frac{p}{1 - p} S_1.$$

Thus the claim is true for $n = 1$. We continue by induction. Assume that $P(\nu_p \le n - 1 | \mathbf{F}_{n-1}) = p_{n-1}$. Clearly,

$$\begin{aligned} P(\nu_p \le n | \mathbf{F}_{n-1}) &= P(\nu_p \le n - 1 | \mathbf{F}_{n-1}) \\ &\quad + \{1 - P(\nu_p \le n - 1 | \mathbf{F}_{n-1})\} P(\nu_p = n | \mathbf{F}_{n-1}, \nu_p \ge n) \\ &= p_{n-1} + (1 - p_{n-1}) P(\nu_p = n | \mathbf{F}_{n-1}, \nu_p \ge n). \end{aligned}$$

Hence, the definitions of p_{n-1}, ν_p and some algebra now yield

$$\begin{aligned}\frac{P(\nu_p \le n|\mathbf{F}_{n-1})}{1-P(\nu_p \le n|\mathbf{F}_{n-1})} &= \frac{p_{n-1}+(1-p_{n-1})p(1-S_{n-1})^+}{(1-p_{n-1})\{1-p(1-S_{n-1})^+\}} \\ &= \frac{p_{n-1}}{1-p_{n-1}}\frac{1}{1-p(1-S_{n-1})^+}+\frac{p(1-S_{n-1})^+}{1-p(1-S_{n-1})^+} \\ &= \frac{p}{1-p}\left[\frac{S_{n-1}+(1-p)(1-S_{n-1})^+}{1-p(1-S_{n-1})^+}\right] \\ &= \frac{p}{1-p}\max\{1, S_{n-1}\}.\end{aligned}$$

Hence

$$\begin{aligned}\frac{P(\nu_p \le n|\mathbf{F}_n)}{1-P(\nu_p \le n|\mathbf{F}_n)} &= \frac{P(\nu_p \le n|\mathbf{F}_{n-1})}{1-P(\nu_p \le n|\mathbf{F}_{n-1})}L_n \\ &= \frac{p}{1-p}\max\{1, S_{n-1}\}L_n \\ &= \frac{p}{1-p}S_n \\ &= \frac{p_n}{1-p_n}.\end{aligned}$$

Lemma 1 follows. □

Our main result is as follows.

PROPOSITION 1. *Suppose $C_1 - C_2 \ge C_3$. Let A be the unique solution of*

$$C_1 - C_2 E_\infty N_A = C_3 E_0 N_A. \tag{2}$$

Then for some $p \in [0, 1]$,

$$El(\nu_p, N_A) = \inf_{N \in \mathbf{N}_c} \sup_{\nu \in \mathbf{N}_s} El(\nu, N) = \sup_{\nu \in \mathbf{N}_c} \inf_{N \in \mathbf{N}_s} El(\nu, N).$$

PROOF. Suppose that nature uses a ν_p strategy. The optimal strategy of the statistician should be based only on the *a* posteriori probability that the change has already occurred, i.e., given $N \ge n$, the decision $N = n$ should be a function only of $P(\nu_p \le n|\mathbf{F}_n)$. But $P(\nu_p \le n|\mathbf{F}_n) = p_n$, by Lemma 1. We prove now that the optimal procedure is a threshold one, i.e., stop the first time p_n, or equivalently S_n, is above a given level. We call the interval $k, k+1, \dots, l$ a cycle if $p_k = p$ and $l = \inf\{n > k,\ p_n = p\}$. Since all cycles are alike, any stopping rule δ that is a function of p_n only is characterized by $t_0(\delta), q_0(\delta), (t_1(\delta), q_1(\delta))$, i.e., the expected length of the cycle and the probability that the change will be declared during the cycle, respectively, under the assumption that the change point will be after the end of the cycle (was before its beginning). Note that $t_i(\delta)$ is the expected time until either a change is declared or $p_n = p$. Now, the risk of the procedure is the expected value of the loss during the cycle multiplied by the expected number of cycles, or

$$\frac{(1-p)(C_1 q_0(\delta) - C_2 t_0(\delta)) + pC_3 t_1(\delta)}{(1-p)q_0(\delta) + pq_1(\delta)}.$$

To find the class of optimal procedures we find the minimizer of the numerator subject to a fixed value of $(1-p)q_0(\delta) + pq_1(\delta)$. Using a standard Lagrange multiplier argument we can consider the loss function

$$(1-p)(C_1 q_0(\delta) - C_2 t_0(\delta)) + pC_3 t_1(\delta) + \lambda[(1-p)q_0(\delta) + pq_1(\delta)]$$

for some Lagrange multiplier λ.

Now this loss function is linear in p and therefore the risk function is concave. Since we certainly should stop immediately if $p = 1$, the optimal procedure should be a threshold one. This last step of the argument is equivalent to the proof of the optimality of the SPRT procedure [cf. Lehmann (1959) pages 104–106]. We obtained that the CUSUM is an optimal answer to ν_p.

But how does nature choose p? Clearly, both $E_0(N_A)$ and $E_\infty(N_A)$ are increasing functions of A, converging to 1 as $A \to 0$ and to ∞ as $A \to \infty$. Hence (2) can be solved.

Moreover, let $A(p)$ be the optimal value of A when nature chooses the change point according to ν_p. Then $E_\infty\{N_{A(p)}\}$ ranges continuously from 1 to ∞ as p ranges from 1 to 0. Therefore nature can pick a value of p for which $A(p) = A$, where A satisfies (2), i.e., $C_1 - C_2 E_\infty\{N_{A(p)}\} = C_3 E_0\{N_{A(p)}\}$.

Suppose the statistician uses a CUSUM procedure with this A. We claim that the suggested strategy for nature [ν_p with $A(p) = A$] will maximize the expected loss: nature can choose either $\nu = \infty$ or $\nu < \infty$. If $\nu = \infty$ is picked, then the expected loss will be $C_1 - C_2 E_\infty\{N_A\}$. Otherwise, nature may choose any $n \le N_A$. But, if $S_{n-1} \le 1$, then the expected future loss is $C_3 E_0\{N_A\}$. It is smaller than that when $S_{n-1} > 1$, since the expected time from n to N_A is monotone in S_{n-1} (note that S_m is stochastically monotone in S_n for any $m \ge n$). Nature should therefore not pick any n if $S_{n-1} > 1$ (it will be better to pick ∞) and can randomize between n and infinity if $\nu \ge n$ and $S_{n-1} \le 1$. Hence $C_3 E_0\{N_A\}$ is the maximum that nature can ensure and the suggested procedure guarantees it.

We conclude that the above pair of strategies is a saddle point of the game. □

We prove now the main result of Moustakides (1986).

PROPOSITION 2. *A CUSUM procedure with* $A > 1$ *minimizes* $\sup_\nu$ ess-sup $E_\nu\{(N - \nu + 1)^+ | \mathbf{F}_{n-1}\}$ *among all* $N \in \mathbf{N_s}$ *such that* $E_\infty N \ge E_\infty N_A$.

PROOF. Suppose that for some $N \in \mathbf{N_s}$, $E_\infty N \geq E_\infty N_A$ and $\sup_\nu$ ess-sup $E_\nu\{(N - \nu + 1)^+ | \mathbf{F}_{n-1}\} < E_0 N_A$. Then there exists $N' \in \mathbf{N_s}$ that satisfies both inequalities strictly [e.g., $P(N' = N + 1) = 1 - P(N' = N) = \varepsilon$ for some $\varepsilon > 0$ and the randomization is independent of everything else].

Let (ν_p, N_A) be the saddle point of the game with loss function

$$l^{(p)}(\nu, N) = 1_{\{N < \nu\}} - C_2^{(p)} \min\{N, \nu - 1\} + C_3^{(p)}(N - \nu + 1)^+.$$

Note that by (2),

$$C_2^{(p)} E_\infty N_A = 1 - C_3^{(p)} E_0 N_A. \tag{3}$$

Now,

(4)

$$E \min\{N_A, \nu_p - 1\} \to E_\infty N_A \quad \text{and} \quad E \min\{N', \nu_p - 1\} \to E_\infty N' \quad \text{as } p \to 0,$$

$$\begin{aligned} E(N_A - \nu_p + 1)^+ &= P(N_A \geq \nu_p) E(N_A - \nu_p + 1 | N_A \geq \nu_p) \\ &= P(N_A \geq \nu_p) E_0 N_A \end{aligned} \tag{5}$$

and by the definition of N',

$$\begin{aligned} E(N' - \nu_p + 1)^+ &= P(N' \geq \nu_p) E\big(N' - \nu_p + 1 | N' \geq \nu_p\big) \\ &\leq P(N' \geq \nu_p) \sup_\nu \text{ess-sup } E_\nu\big\{(N' - \nu + 1)^+ | \mathbf{F}_{n-1}\big\} \\ &< P(N' \geq \nu_p) \sup_\nu \text{ess-sup } E_\nu\big\{(N_A - \nu + 1)^+ | \mathbf{F}_{n-1}\big\} \\ &= P(N' \geq \nu_p) E_0 N_A. \end{aligned} \tag{6}$$

Now,

$$\begin{aligned} 1 - El^{(p)}(\nu_p, N') &= P(N' \geq \nu_p) + C_2^{(p)} E \min\{N', \nu_p - 1\} - C_3^{(p)} E(N' - \nu_p + 1)^+ \\ &> P(N' \geq \nu_p) + C_2^{(p)} E \min\{N', \nu_p - 1\} \\ &\qquad - C_3^{(p)} P(N' \geq \nu_p) E_0 N_A \quad [\text{by (6)}] \\ &= P(N' \geq \nu_p)\big(1 - C_3^{(p)} E_0 N_A\big) + C_2^{(p)} E \min\{N', \nu_p - 1\} \\ &= C_2^{(p)}\big[P(N' \geq \nu_p) E_\infty N_A + E \min\{N', \nu_p - 1\}\big] \quad [\text{by (3)}]. \end{aligned} \tag{7}$$

Similar expressions but with equalities throughout, hold for $l(\nu_p, N_A)$:

$$1 - El^{(p)}(\nu_p, N_A) = C_2^{(p)}\big[p(N_A \geq \nu_p) E_\infty N_A + E \min\{N_A, \nu_p - 1\}\big]. \tag{8}$$

Since $P(N' \geq \nu_p) \to 0$ and $P(N_A \geq \nu_p) \to 0$ as $p \to 0$, we obtain from (4) and (8) that

$$\lim_{p \to 0} \big\{1 - El^{(p)}(\nu_p, N_A)\big\} / C_2^{(p)} = E_\infty N_A, \tag{9}$$

while (7) implies that

$$\liminf_{p \to 0} \{1 - El^{(p)}(\nu_p, N')\}/C_2^{(p)} \ge E_\infty N' > E_\infty N_A. \tag{10}$$

We conclude from (9) and (10) that for p small enough, $El^{(p)}(\nu_p, N') < El^{(p)}(\nu_p, N_A)$. This is a contradiction since N_A is Bayes. □

Acknowledgment. I wish to express my thanks to Moshe Pollak for long discussions of this problem and for a referee who helped to correct the proof.

REFERENCES

LEHMANN, E. L. (1959). *Testing Statistical Hypotheses*. Wiley, New York.

LORDEN, G. (1971). Procedures for reacting to a change in distribution. *Ann. Math. Statist.* **42** 1897–1908.

MOUSTAKIDES, G. V. (1986). Optimal stopping times for detecting a change in distribution. *Ann. Statist.* **14** 1379–1388.

SIEGMUND, D. (1985). *Sequential Analysis, Tests and Confidence Intervals*. Springer, New York.

DEPARTMENT OF STATISTICS
UNIVERSITY OF PENNSYLVANIA
PHILADELPHIA, PENNSYLVANIA 19104–6302