The Annals of Statistics
1986, Vol. 14, No. 4, 1379–1387

OPTIMAL STOPPING TIMES FOR DETECTING CHANGES IN DISTRIBUTIONS

BY GEORGE V. MOUSTAKIDES

Institut de Recherche en Informatique et Systemes Aleatoires

It is shown that Page's stopping time is optimal for the detection of changes in distributions, in a well defined sense. This work is a generalization of an existing result where it was shown that Page's stopping time is optimal asymptotically.

1. Introduction. Let us assume that $X_1, X_2, \ldots$ are independent and identically distributed random variables that are observed sequentially. Let also $X_1, \ldots, X_{m-1}$ have distribution function F_0 and $X_m, X_{m+1}, \ldots$ distribution function $F_1 \neq F_0$. The two distributions are known, but the time of change m is assumed unknown. We are interested in finding stopping times that will detect the change with a delay as small as possible. Let P_m denote the true distribution of $X_1, X_2, \ldots$ when the change occurs at m and E_m the expectation under this distribution. We allow m to take the value infinity, denoting by this the case where no change occurs. Let $\mathscr{F}_n$, $n \geq 1$, be the σ-algebra generated by $\{X_1, X_2, \ldots, X_n\}$. We also consider an auxiliary σ-algebra $\mathscr{Y}$ and we will assume that we can extend the measures P_m on $\mathscr{Y}$ in such a way that for every real $p \in [0,1]$ we can generate an event in $\mathscr{Y}$ that has probability p for every P_m. Define $\mathscr{Y}_0 = \mathscr{Y}$ and $\mathscr{Y}_n = \mathscr{F}_n \cup \mathscr{Y}$. Allowable stopping times (s.t.) will be all those s.t. that have the form $N = 0$ with probability $(1-p)$ and $N = N'$ with probability p where N' is any s.t. measurable with respect to the filtration $\{\mathscr{F}_n\}_{n \geq 1}$ and with the randomization p being done before any observation is taken.

We define optimality of a s.t. in the sense of Lorden [3]. That is, if N is a s.t. define

$$D_m(N) = \operatorname{ess\,sup} E_m\{[N - m + 1]^+ / \mathscr{Y}_{m-1}\}, \qquad m \geq 1, \tag{1}$$

$$D(N) = \sup_{m \geq 1} D_m(N). \tag{2}$$

Thus we consider the conditional expectation of the delay over those events before the change occurs that least favor the detection of the change. We would like to minimize $D(N)$ over those s.t. from the allowable class that satisfy the following constraint on the rate of false detections:

$$E_\infty\{N\} \geq \gamma > 0. \tag{3}$$

Our goal in the next section will be to prove that Page's s.t. is optimal in the above sense. Let us first define this s.t. For simplicity we will assume that F_0 and F_1 are mutually absolutely continuous. Let $l(x)$ denote the Radon–Nikodym

Received January 1985; revised October 1985.
AMS 1980 *subject classification*. Primary 62E.
Key words and phrases. Stopping times, detection of changes, sequential analysis.

derivative of F_1 with respect to F_0. We will assume that $l(X_1)$ has no atoms with respect to P_∞. Let us define the following sequence of random variables:

$$S_0 = 0, \qquad S_n = \max\{S_{n-1}, 1\} l(X_n), \qquad n \geq 1. \tag{4}$$

We define Page's s.t. N_P as follows:

$$N_P = \inf\{n \geq 1: S_n \geq \mu\}, \tag{5}$$

where μ is a nonnegative constant and the infimum of the empty set is infinity. Page's s.t. is defined a little differently here than it is usually defined in the literature. N_P is usually defined as the first n for which $T_n = \max\{S_n, 1\}$ exceeds μ. The two definitions are equivalent when $\mu > 1$, but there is a difference when $\mu < 1$. When $\mu < 1$, with the definition that uses T_n we stop at $n = 1$, but with the definition in (5) this is not the case. As we will see in the next section there exists a nontrivial range of values of γ for which $\mu \in (0, 1]$. With the following lemma we give some properties of the sequences S_n and T_n that will be used later.

LEMMA 1. *Let* $T_n = \max\{S_n, 1\}$. *For any* $n > m \geq 1$ *and for fixed* $\{X_{m+1}, \dots, X_n\}$, *the quantity* S_n *is a nondecreasing function of* T_m. *Also* T_n *can be written as*

$$T_n = \sum_{j=1}^{n+1} [1 - S_{j-1}]^+ \prod_{k=j}^{n} l(X_k), \tag{6}$$

where we define $\prod_{k+1}^{k} = 1$.

PROOF. The property that S_n is a nondecreasing function of T_m (for fixed $\{X_{m+1}, \dots, X_n\}$) can be proved by induction and using the definition in (4). To prove (6), we can see from (4) that

$$T_n = \max\{S_n, 1\} = S_n + [1 - S_n]^+. \tag{7}$$

If we use (4), (7) and induction we can easily show (6). □

A very important consequence of the monotonicity of S_n with respect to T_m is that on the event $\{N_P \geq m\}$ the s.t. N_P is nonincreasing with T_{m-1}; thus, the essential supremum in (1) is achieved for $T_{m-1} = 1$ (or $S_{m-1} \leq 1$). This means that restarting Page's s.t. at m gives the worst conditional average delay and thus from stationarity all $D_m(N_P)$ are equal. The s.t. N_P is thus an equalizer rule, a very important property for proving its optimality.

2. Optimal stopping time. Notice first that for $\gamma > 0$ we have $D(N) \geq 1$. This is true because with $E_\infty\{N\} \geq \gamma > 0$ it is not possible to stop a.s. at $n = 0$ and thus we will take at least one sample. With this remark we have that for $1 \geq \gamma > 0$, the optimal s.t. (say N_0) is {stop at $n = 0$ with probability $1 - \gamma$ otherwise stop at $n = 1$}. This yields $D(N_0) = 1$ and $E_\infty\{N_0\} = \gamma$. We now consider the case $1 < \gamma < \infty$. With the next lemma we will show that in order to find the optimal s.t. it is enough to limit ourselves to a smaller class of s.t.

LEMMA 2. *In order to minimize* $D(N)$ *over the s.t. that satisfy* (3) *it is enough to consider those s.t. that satisfy* (3) *with equality.*

PROOF. From the way we defined our class of s.t., we have that $N = 0$ with probability $(1 - p)$ and $N = N'$ with probability p, where N' is measurable with respect to $\{\mathscr{F}_n\}$. If $p > 0$ then we have that $D_m(N) = D_m(N')$ for every $m > 1$ and thus $D(N) = D(N')$. Notice also that $E_\infty\{N\} = pE_\infty\{N'\}$, where the product $pE_\infty\{N'\}$ is defined as being zero when $p = 0$. If $E_\infty\{N\} = \infty$ then $p > 0$ and $E_\infty\{n'\} = \infty$. We can always find a large enough integer K such that if we define $M' = \min\{N', K\}$ to have $pE_\infty\{M'\} > \gamma$. If now M is the randomization of M' with probability p then $D(M) = D(M') \le D(N') = D(N)$. It is thus enough to consider s.t. that have finite $E_\infty\{N\}$. Let now $\infty > E_\infty\{N\} = pE_\infty\{N'\} > \gamma$; this means $p > 0$ and $E_\infty\{N'\} < \infty$. By defining a new s.t. M that is equal to N' with randomization probability $p' = \gamma/E_\infty\{N'\} < p$ we have $E_\infty\{M\} = \gamma$ and $D(M) = D(N') = D(N)$. This concludes the proof. □

In the following lemma we introduce a lower bound for $D(N)$ that we will use as our performance measure instead of $D(N)$.

LEMMA 3. *For any s.t. N satisfying* $0 < E_\infty\{N\} < \infty$ *we have that*

$$D(N) \ge \frac{E_\infty\{\sum_{k=0}^{N-1}\max\{S_k, 1\}\}}{E_\infty\{\sum_{k=0}^{N-1}[1 - S_k]^+\}} = \bar{D}(N), \tag{8}$$

where we define $\sum_{k=j}^{j-1} = 0$. *We have equality in* (8) *when* $N = N_p$.

PROOF. Let $I(A)$ denote the indicator function of the event A. Define

$$B_m(N) = E_m\{[N - m + 1]^+/\mathscr{Y}_{m-1}\}. \tag{9}$$

Because the event $\{N \ge k\}$ is $\mathscr{Y}_{k-1}$ measurable we have

$$\begin{aligned} B_m(N) &= \sum_{k=m}^{\infty} E_m\{I(N \ge k)/\mathscr{Y}_{m-1}\} \\ &= \sum_{k=m}^{\infty} E_\infty\left\{\left[\prod_{j=m}^{k-1} l(X_j)\right] I(N \ge k)/\mathscr{Y}_{m-1}\right\} \\ &= E_\infty\left\{\left[\sum_{k=m}^{N}\prod_{j=m}^{k-1} l(X_j)\right]\bigg/\mathscr{Y}_{m-1}\right\}. \end{aligned} \tag{10}$$

Notice that in (1) $D_m(N)$ was defined as the essential supremum of $B_m(N)$. Since for every $m > 1$ we have $D(N) \ge D_m(N)$ using (10), this yields

$$\begin{aligned} &E_\infty\{[1 - S_{m-1}]^+ I(N \ge m)\}D(N) \\ &\quad \ge E_\infty\{[1 - S_{m-1}]^+ I(N \ge m)B_m(N)\} \\ &\quad = E_\infty\left\{I(N \ge m)\sum_{k=m}^{N}[1 - S_{m-1}]^+\prod_{j=m}^{k-1} l(X_j)\right\}. \end{aligned} \tag{11}$$

When $N = N_P$ we have equality in (11). This is true because $D(N_P) = D_m(N_P)$ for every m and, because as we said in the introduction, the essential supremum of $B_m(N_P)$ is achieved on the event $\{N_P \geq m\} \cap \{S_{m-1} \leq 1\}$ and that $[1 - S_{m-1}]^+ I(N_P \geq m)$ may be nonzero only on this event. Summing now (11) for all $m \geq 1$, interchanging summations and expectations and using (6), the last term of the inequality in (11) gives

$$\begin{aligned}
&\sum_{m=1}^{\infty} E_\infty\left\{I(N \geq m) \sum_{k=m}^{N} [1 - S_{m-1}]^+ \prod_{j=m}^{k-1} l(X_j)\right\} \\
&\quad = E_\infty\left\{\sum_{m=1}^{N} \sum_{k=m}^{N} [1 - S_{m-1}]^+ \prod_{j=m}^{k-1} l(X_j)\right\} \\
(12)\qquad &\quad = E_\infty\left\{\sum_{k=1}^{N} \left[\sum_{m=1}^{k} [1 - S_{m-1}]^+ \prod_{j=m}^{k-1} l(X_j)\right]\right\} \\
&\quad = E_\infty\left\{\sum_{k=1}^{N} T_{k-1}\right\} = E_\infty\left\{\sum_{k=0}^{N-1} T_k\right\} = E_\infty\left\{\sum_{k=0}^{N-1} \max\{S_k, 1\}\right\}.
\end{aligned}$$

For the first term in (11) we have that

$$(13)\qquad \sum_{m=1}^{\infty} E_\infty\{I(N \geq m)[1 - S_{m-1}]^+\} = E_\infty\left\{\sum_{m=0}^{N-1} [1 - S_m]^+\right\}.$$

The quantity in (13) is less than $E_\infty\{N\}$, thus finite. For $N \geq 1$ we also have that

$$(14)\qquad \sum_{m=0}^{N-1} [1 - S_m]^+ \geq 1;$$

thus, the quantity in (14) is no less than the probability $P_\infty\{N \geq 1\}$, which is nonzero since by assumption we have $E_\infty\{N\} > 0$. Thus, we have shown (8). □

In order now to show that Page's s.t. is optimal it is enough to show that among all s.t. that satisfy $E_\infty\{N\} = \gamma$, Page's s.t. is the one that minimizes $\overline{D}(N)$. Specifically N_P minimizes $\overline{D}(N)$ by simultaneously minimizing its numerator and maximizing its denominator. One can now see why it was necessary to limit ourselves to the class $E_\infty\{N\} = \gamma$. If we had condition (3) instead, this gives $N = \infty$ as optimal for the denominator. Since from now on all events will be considered with respect to the measure P_∞, for simplicity we drop the subscript ∞. With the following theorem we show that Page's s.t. is optimal for a whole class of optimization problems.

THEOREM 1. *Let $\infty > \gamma > 1$. If $\varphi(z)$ is a continuous nonincreasing function, well defined for all $z \geq 0$ with $\varphi(0)$ bounded, then Page's s.t. satisfies*

$$(15)\qquad \sup_N E\left\{\sum_{k=0}^{N-1} \varphi(S_k)\right\} = E\left\{\sum_{k=0}^{N_P-1} \varphi(S_k)\right\}$$

for all s.t. N that satisfy $E\{N\} = \gamma$.

Using Theorem 1 we can easily show that N_P minimizes $\overline{D}(N)$. With $\varphi(z) = -\max\{z, 1\}$, Theorem 1 shows that N_P minimizes the numerator of $\overline{D}(N)$ and, with $\varphi(z) = [1 - z]^+$, that it maximizes the denominator. By assumption we have that $\varphi(z)$ is bounded from above by $\varphi(0) < \infty$. Without loss of generality we can assume that $\varphi(z)$ is also bounded from below. This is true because if $\varphi(z)$ is not bounded from below we can always define a new function $\varphi'(z) = \max\{\varphi(z), A\}$, where $A \le \varphi(\mu)$ (μ is the threshold for N_P). We will thus have

$$E\left\{\sum_{k=0}^{N-1} \varphi(S_k)\right\} \le E\left\{\sum_{k=0}^{N-1} \varphi'(S_k)\right\}. \tag{16}$$

Notice that we have equality in (16) when $N = N_P$. This is true because the two sums in (16) are taken up to $N - 1$; thus, for $N = N_P$, S_k is always in the region where $\varphi(S_k) = \varphi'(S_k)$. Clearly if N_P maximizes the right-hand side of (16), it also maximizes the left-hand side. We will thus assume that $\varphi(z)$ is bounded and let $D < \infty$ be a bound.

Before going to the proof of Theorem 1 we give some definitions and present certain results that will be useful for this proof. Let us denote by ν_r the time of the rth entry of S_n in the set $[0, 1]$, i.e., $\nu_0 = 0$ and, for $r \ge 1$,

$$\nu_r = \inf\{n > \nu_{r-1}\colon S_n \le 1\}. \tag{17}$$

In the next lemma we present a property of the time ν_1.

LEMMA 4. *All finite moments of the time ν_1 exist.*

PROOF. Since for $a \ge 0$ the function a^s is a convex function of s we have for $0 \le s \le 1$ that $a^s \le (1 - s) + sa$, with equality if and only if $a = 1$. If now $a = l(X_1)$ and we define $\alpha(s) = E\{[l(X_1)]^s\}$, we conclude that $\alpha(s) \le 1$. Since $l(X_1)$ is not a constant equal to unity, there exists s_0 that satisfies $\alpha(s_0) < 1$. Notice now that

$$P\{\nu_1 = k\} \le P\{S_{k-1} > 1\} = P\left\{\prod_{j=1}^{k-1} l(X_j) > 1\right\}$$
$$= P\left\{\prod_{j=1}^{k-1} \left[l(X_j)\right]^{s_0} > 1\right\} \le \left[\alpha(s_0)\right]^{k-1}.$$

To show that all the moments exist, we have

$$E\{\nu_1^j\} = \sum_{k=1}^{\infty} k^j P(\nu_1 = k) \le \sum_{k=1}^{\infty} k^j \left[\alpha(s_0)\right]^{k-1} < \infty,$$

and this concludes the proof. □

In order to solve the constrained optimization problem defined in (15) using the Lagrange multiplier technique, we will reduce it to an unconstrained optimization problem. Let $S_0 = x \ge 0$ and, for any real λ define

$$V(x, \lambda) = \sup_N E\left\{\sum_{k=0}^{N-1} \left[\varphi(S_k) - \lambda\right]\right\}, \tag{18}$$

where now the supremum is taken over all s.t. N. The function $V(x,\lambda)$ is nonnegative and can take the value infinity. We are interested in finding for which values of x and λ $V(x,\lambda)$ is finite. With the next lemma we can see the behavior of $V(x,\lambda)$ with respect to x when λ is fixed.

LEMMA 5. *Let $V(x,\lambda)$ be defined by* (18); *then $V(x,\lambda)$ is finite if and only if $V(0,\lambda)$ is finite.*

PROOF. Notice first that using Lemma 1 and the monotonicity of $\varphi(z)$ we conclude that $V(x,\lambda)$ is nonincreasing in x for fixed λ. With this property the if part is easy to show since $V(x,\lambda) \le V(0\,\lambda) < \infty$. For the only if part now, assume that $V(x,\lambda) < \infty$ for some $x = x_0$. For $x \ge x_0$ we then have $V(x,\lambda) \le V(x_0,\lambda) < \infty$. We will now show that we also have $V(0,\lambda) < \infty$. Let N_0 denote Page's s.t. with threshold x_0. For any s.t. N define $N' = \min\{N, N_0\}$; then

$$\begin{aligned}
&E\left\{\sum_{k=0}^{N-1}[\varphi(S_k)-\lambda]\right\}\\
&\quad = E\left\{\sum_{k=0}^{N'-1}[\varphi(S_k)-\lambda]\right\} + E\left\{\sum_{k=N'}^{N-1}[\varphi(S_k)-\lambda]\right\}\\
&\quad \le (D+|\lambda|)E\{N_0\} + \sum_{j=1}^{\infty}\sup_{N\ge j}E\left\{\sum_{k=j}^{N-1}[\varphi(S_k)-\lambda]I(N_0=j)\right\}\\
&\quad \le (D+|\lambda|)E\{N_0\} + V(x_0,\lambda) < \infty
\end{aligned}$$

and this concludes the proof. □

From this lemma we have that for fixed λ, $V(x,\lambda)$ is either finite for all values of x or it is infinite. With the following theorem we identify the range of values of λ for which $V(0,\lambda)$ (and thus $V(x,\lambda)$) is finite.

THEOREM 2. *Let $\xi_1 = \sum_{k=1}^{\nu_1}\varphi(S_k)$ and $\lambda_0 = E\{\xi_1\}/E\{\nu_1\}$.*

(a) *If $\infty > \lambda > \lambda_0$ and $S_0 = x \ge 0$, then we have*

$$E\left\{\sup_n\left[\sum_{k=0}^{n}[\varphi(S_k)-\lambda]\right]^+\right\} < \infty.$$

(b) *If $\xi_1 - \lambda_0\nu_1$ is not a constant, then $V(0,\lambda) \to \infty$ as $\lambda \to \lambda_0+$.*

PROOF. Condition (a) is sufficient to ensure existence of $V(x,\lambda)$ ([6], page 69). First notice that $|E\{\xi_1\}| \ge DE\{\nu_1\}$ (where D is a bound for $\varphi(z)$), thus λ_0 is finite. To show (a) for every $x \ge 0$, it is enough to show it (using Lemma 1) for $x = 0$. Consider the sequence $\nu_1, \nu_2, \dots$ defined in (17); it goes to infinity a.s. Define $\xi_r = \sum_{\nu_{r-1}+1}^{\nu_r}\varphi(S_k)$ and $\eta_r = \nu_r - \nu_{r-1}$. Using the strong Markov property of the sequence $\{S_n\}$, we have that the two sequences $\{\xi_r\}$ and $\{\eta_r\}$ are i.i.d. sequences of random variables. Let now $\nu_{r-1} < n \le \nu_r$; then,

$$\sum_{k=0}^{n}[\varphi(S_k)-\lambda] \le \sum_{k=1}^{r}[\xi_k - \lambda\eta_k] + \eta_r D', \tag{19}$$

where $D' = D + |\lambda|$. The sequence $\{\omega_k\}$ with $\omega_k = \xi_k - \lambda\eta_k$ is also an i.i.d. sequence. For $\lambda > \lambda_0$ we have that $E\{\omega_1\} < 0$. Let $\delta > 0$ be such that $E\{\omega_1\} + \delta < 0$; then, from (19) we have

$$\begin{aligned}\left[\sum_{k=0}^{n}[\varphi(S_k)-\lambda]\right]^+ &\le \left[\sum_{k=1}^{r}[\omega_k+\delta]\right]^+ + [\eta_r D' - \delta r]^+ \\ &\le \left[\sum_{k=1}^{r}[\omega_k+\delta]\right]^+ + D'\eta_r I(\eta_n \ge \delta' r),\end{aligned} \tag{20}$$

where $\delta' = \delta/D'$. From (20) we have

$$\begin{aligned}&E\left\{\sup_n\left[\sum_{k=0}^{n}[\varphi(S_k)-\lambda]\right]^+\right\} \\ &\quad\le E\left\{\sup_r\left[\sum_{k=1}^{r}[\omega_k+\delta]\right]^+\right\} + D'\sum_{r=1}^{\infty}E\{\eta_r I(\eta_r \ge \delta' r)\}.\end{aligned} \tag{21}$$

For the first term in (21) to be finite, sufficient conditions are $E\{\omega_1\} + \delta < 0$ and $E\{([\omega_1 - E\{\omega_1\}]^+)^2\} < \infty$ ([2], page 92). The first condition is satisfied; we can easily show that the second is also satisfied using the properties that $\varphi(z)$ is bounded, that the second moment of ν_1 is finite (Lemma 4) and the fact that for any real a we have $(a^+)^2 \le a^2$. To show that the second term in (21) is finite notice that since η_r is distributed as ν_1 we have

$$\sum_{r=1}^{\infty}E\{\eta_r I(\eta_r \ge \delta' r)\} \le \frac{E\{\nu_1^3\}}{(\delta')^2}\sum_{r=1}^{\infty}\frac{1}{r^2} < \infty.$$

To prove (b), notice first that $V(0,\lambda)$ is nonincreasing in λ; thus, the limit of $V(0,\lambda)$ exists (being possibly infinity). Consider now s.t. that can stop only at the instances ν_r and the decision whether to stop or continue at ν_r is made by using the random variables $\xi_1,\dots,\xi_r$ and $\eta_1,\dots,\eta_r$. For this case we have

$$\sum_{k=0}^{\nu_R-1}[\varphi(S_k)-\lambda] \ge \sum_{k=1}^{R}[\xi_k - \lambda\eta_k] - 2D'.$$

This yields

$$V(0,\lambda) \ge E\left\{\sum_{k=1}^{R}[\xi_k - \lambda\eta_k]\right\} - 2D'. \tag{22}$$

Consider now a s.t. R with $E\{R\} < \infty$, from (22) taking limits with respect to λ gives

$$\lim_{\lambda\to\lambda_0+} V(0,\lambda) \ge E\left\{\sum_{k=1}^{R}[\xi_k - \lambda_0\eta_k]\right\} - 2D'. \tag{23}$$

The random variable $\xi_1 - \lambda_0\eta_1$ has zero mean; thus, from [2, page 27], the right-hand side in (23) can be made arbitrarily large. This concludes the proof of Theorem 2. □

We are now ready to prove Theorem 1.

PROOF OF THEOREM 1. Notice that the first term in the sum in (15) is $\varphi(S_0) = \varphi(0)$. In general, the value $\varphi(0)$ is not attainable by $\varphi(S_k)$ for $k > 0$ because $\varphi(S_k) \le \varphi(l(X_k)) \le \varphi(z_0)$, where $z_0 = \operatorname{ess\,inf}\{l(X_1)\}$. This will be, for example, the case when $\varphi(z) = [1 - z]^+$ and $z_0 > 0$. The quantity $\varphi(z_0)$ will play a role in our proof. We distinguish two cases

(1) $\varphi(z_0) = \varphi(\mu)$. For this case we have

$$E\left\{\sum_{k=0}^{N-1} \varphi(S_k)\right\} \le [\varphi(0) - \varphi(z_0)]P\{N > 0\} + \varphi(z_0)E\{N\}$$
$$\le [\varphi(0) - \varphi(z_0)] + \varphi(z_0)\gamma,$$

with equality when $N = N_P$ because for $k > 0$ and $z_0 \le S_k \le \mu$ we have $\varphi(S_k) = \varphi(z_0)$.

(2) $\varphi(z_0) > \varphi(\mu)$. Let $\lambda > \lambda_0$. For every s.t. N, $E\{\sum_{k=0}^{N-1}[\varphi(S_k) - \lambda]\}$ is linear in λ; thus, $V(x, \lambda)$ being the supremum over N of this expression, is a convex function of λ. If we denote $[x] = \max\{x, 1\}$, the function $E\{V([x]l(X_1), \lambda)\}$ will also be convex and thus continuous in λ.

Consider now the equation

$$\beta(\lambda) = \varphi(\mu) - \lambda + E\{V([\mu]l(X_1), \lambda)\} = 0.$$

The function $\beta(\lambda)$ is continuous in λ. Since $V(x, \varphi(0)) = 0$, we have $\beta(\varphi(0)) = \varphi(\mu) - \varphi(0) < 0$ and $\beta(\lambda) \to \infty$ as $\lambda \to \lambda_0 +$ (Theorem 2(b)). Thus, there exists a λ^* with $\varphi(0) > \lambda^* > \lambda_0$ that satisfies $\beta(\lambda^*) = 0$. Let now $\lambda = \lambda^*$. From Theorem 2(a) and [6, page 69] we have that $V(x, \lambda^*)$ exists and satisfies

$$V(x, \lambda^*) = [\varphi(x) - \lambda^* + E\{V([x]l(X_1), \lambda)\}]^+.$$

From [6, page 74] we have that the optimum is to stop when $V(S_n, \lambda^*) = 0$. Since $\varphi(x)$ and $E\{V([x]l(X_1), \lambda)\}$ are nonincreasing functions of x, this will be the case when $S_n \ge \mu$, i.e., Page's s.t. N_P. □

COMMENTS. It is very difficult in general to relate explicitly γ to μ, though there is a range of values of γ where this is possible. Let us consider the case $\mu \le 1$. For this case N_P is equivalent to $N_P = \inf\{n: l(X_n) \ge \mu\}$. In other words, given that there is no stop before n, we have that $S_k \le 1$ for $k < n$. Indeed if for some k we had $S_k > 1$, then we would also have $S_k > \mu$, thus having a stop at k, a contradiction. For this case the expectation of N_P under P_1 and P_∞ is $E_i\{N_P\} = [P_i\{l(X_1) \ge \mu\}]^{-1}$, $i = 0, \infty$. Thus, for

$$1 < \gamma \le [P_\infty\{l(X_1) \ge 1\}]^{-1},$$

the relation between γ and μ is given by $P_\infty\{l(X_1) \ge \mu\} = \gamma^{-1}$. For other values of γ the integral equation defined in Page's paper [4] can be used, but clearly this is a more complicated situation. For approximations see [7]. In the introduction we assumed that $l(X_1)$ has no atoms. In the general case we have to modify the s.t. N_P by including a randomization whether to continue or stop every time we have $S_n = \mu$.

The approach we have followed here is non-Bayesian. The criterion $D(N)$ that we used takes into account only the worst possible situation before a change occurs; thus, it may be considered as conservative. In [5], following again a non-Bayesian approach, the criterion $D(N) = \sup_m E_m\{(N - m)/N \geq m\}$ is used. The optimal s.t. obtained for this criterion is again of the form "stop when a statistic R_n exceeds a threshold." The statistic R_n satisfies the recursion $R_n = (1 + R_{n-1})l(X_n)$. Notice the similarity with the statistic S_n used for N_P; both can be written in the form $Z_n = \omega(Z_{n-1})l(X_n)$, where $\omega(z)$ is a univariate function satisfying $\omega(z) \geq 1$. Clearly S_n is more conservative with respect to past information (included in S_{n-1}) because it does not take it into account if it is not important enough. This does not happen with R_n. In any case, Page's s.t. is very popular and widely used in practice because it has the important property of combining detection and estimation. Specifically, the largest $n \leq N_P$ for which $S_n \leq 1$ is the maximum likelihood estimate of the change time m using all observations up to time N_P. Finally for Bayesian approaches see [1] and [6, page 193].

Acknowledgments. I would like to thank Dr. A. Benveniste and Professor Le Page for helpful discussions.

REFERENCES

[1] BOJDECKI, T. and HOSZA, J. (1984). On a generalized disorder problem. *Stochastic Process. Appl.* **18** 349–359.
[2] CHOW, Y. S., ROBBINS, H. and SIEGMUND, D. (1971). *Great Expectations: The Theory of Optimal Stopping*. Houghton Mifflin, Boston.
[3] LORDEN, G. (1971). Procedures for reacting to a change in distribution. *Ann. Math. Statist.* **42** 1897–1908.
[4] PAGE, E. S. (1954). Continuous inspection schemes. *Biometrika* **41** 100–115.
[5] POLLAK, M. (1985). Optimal detection of a change in distribution. *Ann. Statist.* **13** 206–227.
[6] SHIRYAYEV, A. N. (1978). *Optimal Stopping Rules*. Springer, New York.
[7] SIEGMUND, D. (1975). Error probabilities and average sample number of the sequential probability ratio test. *J. Roy. Statist. Soc. Ser. B* **37** 394–401.

INSTITUT DE RECHERCHE EN INFORMATIQUE
ET SYSTEMES ALEATOIRES
AVENUE DU GENERAL LECLERC
35042 RENNES CEDEX
FRANCE