SEQUENTIAL ANALYSIS, 20(3), 147–163 (2001)

BAYES STOPPING RULES IN A CHANGE-POINT MODEL WITH A RANDOM HAZARD RATE

Michael I. Baron

Department of Mathematical Sciences, University of Texas at Dallas, Richardson, TX 75083-0688

ABSTRACT

In the Bayes sequential change-point problem, an assumption of a fully known prior distribution of a change-point is usually impracticable. At every moment, one often knows only the discrete hazard function, that is, the probability of a change occurring before the next observation is collected, given that it has not occurred so far. In the randomized model, the observed or predicted values of the hazard function are assumed to form a Markov chain. Under these assumptions, the optimal change-point detection stopping rules are derived for two popular loss functions introduced in Shiryaev (1978) and Ritov (1990). Derivations are based on the theory of optimal stopping of Markov sequences.

Key Words: Change-point; Hazard function; Markov sequence; Optimal stopping; Payoff function; Sequential detection

1. INTRODUCTION

This work concerns efficient detection of a change in the distribution of sequentially obtained observations $\{X_1, X_2, \ldots\}$. The classical results of Lorden (1971) and Moustakides (1986) prove optimality of the CUSUM detection scheme in terms of minimizing the mean delay subject to the fixed mean time between false alarms. That is, let ν be the change point, and $\rho(x) = g(x)/f(x)$ be the likelihood ratio of the post-change and the pre-change distributions. The CUSUM stopping rule, derived by Page (1954) from the Wald's sequential probability ratio test, is defined as

$$\tau(h) = \inf\left\{n : \max_{1\le k<n} \sum_{i=k+1}^{n} \log \rho(X_i) > h\right\}. \tag{1}$$

This stopping rule minimizes the "worst" conditional mean delay

$$\operatorname{ess\,sup} E_\nu\left\{(T - \nu + 1)^+ | X_1, \ldots, X_\nu\right\}$$

under the constraint on the mean time between false alarms $E_{\nu=\infty}T$ (Moustakides (1986)). Further, Ritov (1990) shows that $\tau(h)$ is Bayes with respect to the risk function

$$R_1(T, \nu) = C_1 P\{T < \nu\} + C_2 E(T - \nu)^+ - C_3 E \min\{T, \nu\} \tag{2}$$

and the least favorable prior distribution of ν. This prior distribution, varying according to the sequence $X_1, X_2, \ldots$, is given by conditional probabilities

$$P\left\{\nu = n | \nu \ge n, X_1, \ldots, X_n\right\} = p\left(1 - \max_{1\le k\le n} \prod_{i=k+1}^{n} \rho(X_i)\right)^+. \tag{3}$$

We note that in the majority of applications change-points (breakdowns, disorders, etc.) occur at random times, according to some distribution. This distribution, however, is barely close to the least favorable prior (3). For most of *abrupt* changes, $\{\nu = n\}$ is independent of $X_1, \ldots, X_{n-1}$. In the case of *gradual* changes, the likelihood ratio $\rho(X_j)$ often increases when j is approaching ν. This phenomenon, caused by the latest observations, collected shortly before the change-point, already showing some patterns of the post-change distribution, justifies the use of *warning rules* in statistical process control (Montgomery (1997)). However, the change-point that behaves according to (3) occurs at n with a positive probability only if $\rho(X_{n-1}) < 1$, $\rho(X_{n-1})\rho(X_{n-2}) < 1$, etc. Certainly, this is the nature of the most favorable distribution. According to (3), the change point occurs when the most recent observations are unlikely to come from the post-change distribution. However, this rarely happens in practice.

In view of this, (1) represents a conservative stopping rule, which one may opt to use if the (prior) distribution of a change-point is completely unknown. If some information about the future occurrence of a change is available, one would prefer to use it, while (1) becomes of secondary interest.

However, as one of the results, we will show that the CUSUM stopping time (1) is optimal for a broader than (3) class of prior distributions.

We also note that an assumption of a fully known prior is usually impracticable, especially in the sequential setting. At every moment, one often knows only the current value of the discrete hazard function,

$$\phi_n = P\{\nu = n \mid \nu \geq n\}.$$

Note that if the values of ϕ_k for $k \leq n$ are the only ones available at the time n, then one cannot use the backward induction methods of finding the optimal stopping times because it is impossible to compute the conditional risk induced by collecting X_{n+1}, given $X_1, \ldots, X_n$ (see Ghosh, Mukhopadhyay and Sen (1997), Section 5.2, or Govindarajulu (1981), Section 3.12). Also, our search will go beyond the truncated stopping rules.

In general, ϕ_n may be any probabilities. From the identity

$$\prod_{k=m}^{n}(1 - \phi_k) = P\{\nu \geq n + 1 \mid \nu \geq m\},$$

one obtains that

$$\prod_{k=0}^{\infty}(1 - \phi_k) = P\{\nu = \infty\}. \tag{4}$$

Therefore, if $\prod_{k=0}^{\infty}(1 - \phi_k) > 0$, it is possible to observe no change points in the entire sequence.

Based on the situation, one may assume a suitable model for ϕ_n. Examples include the geometric model with constant ϕ_n for the memoryless distribution of ν, and the logistic model

$$\phi_n = \frac{e^{a+bn}}{1 + e^{a+bn}}, \tag{5}$$

which for $b > 0$ describes the case of an anticipated change-point following a long stable period of observation. Also, ϕ_n may be determined from the results of an experiment, or from an observation of another random variable. For example, the probability of a climate change may be predicted based on the pollution level and the concentration of greenhouse gases. In this case, we will treat ϕ_n as random variables, presumably possessing the Markov property.

Under the geometric model $\phi_n \equiv q$, the problem was studied and solved in Shiryaev (1963) and Shiryaev (1978). Shiryaev's stopping rule

$$\inf\left\{n : \sum_{k=1}^{n} q^{k-n-1} \prod_{i=k}^{n} \rho(X_i) > h\right\} \tag{6}$$

is Bayes under the risk function

$$R_2(T, \nu) = \lambda E(T - \nu)^+ + P\{T < \nu\}. \tag{7}$$

Pollak (1985) shows that the limiting procedure, as $q \to 1$, is asymptotically Bayes risk efficient, and derives an almost minimax rule (see also Lai (1995) for an overview).

The geometric-type prior distribution of ν corresponds to the case when change points occur according to a Poisson process. One of the consequences is the memoryless property. That is, occurrence of a change point at any time does not depend on the past history of the process, in particular, the time that elapsed since the previous change point. Although this may be a reasonable assumption in quality control, it is not the case in many other applications, when it is meaningful to consider other prior distributions (Shiryaev (1978), p. 194).

In clinical trials, after a sufficiently long stable period of treatment, a patient is expected to show a significant improvement. In learning and problem solving, after a long period of progress, a new microdevelopmental sequence is anticipated to start soon. In actuarial science, the claims are modeled by a Markov process rather than a process with independent increments (Bühlmann (1970)). The fact that ϕ_n is often a non-constant, sometimes random function of n justifies the search of optimal stopping times other than (6).

For the described change point hazard rate model, we derive a stopping rule $T(\alpha)$ from the sequential Bayes test, similarly to the CUSUM procedure derived from the Wald's SPRT. Bayes optimality of $T(\alpha)$ under the risk functions (2) and (7) is shown in Section 3. The methodology being used is essentially an application of Shiryaev's theory of the optimal stopping of Markov sequences. Section 4 proposed application of our results to the problem of detecting the beginning of influenza epidemics.

2. SEQUENTIAL BAYES TEST AND A CLASS OF STOPPING RULES

We consider a general sudden change-point model

$$X_i = U_i I\{\nu \geq i\} + V_i I\{\nu < i\},$$

where $U_1,\ U_2,\ \ldots$ and $V_1,\ V_2,\ \ldots$ are two independent sequences of random variables, from distributions F and G respectively, with densities f and g and the likelihood ratio $\rho = g/f$. Independently of $\{U_i\}$ and $\{V_i\}$, the change-point ν has a prior distribution $\pi_n = P\{\nu = n\}$. However, it is not completely known at any moment, and only the up-to-date values of the discrete hazard function $\phi_k = P\{\nu = k \mid \nu \geq k\}$, $k \leq n$, are available at time n.

A *random hazard rate* model may also be considered, where conditional probabilities $P\{\nu = n | \nu \geq n\}$ are results of some random experiment $\mathcal{E}_n$, independent of $U_1, U_2, \ldots$ and $V_1, V_2, \ldots$. By the time n, one observes X_k and $\phi_k = P\{\nu = k \mid \nu \geq k,\ \mathcal{E}_k\}$ for $k \leq n$, whereas U_i, V_i, ν, π_k and $\phi_k, k > n$, are unobservable.

In order to extend our consideration beyond truncated stopping rules, we assume that $\phi_n < 1$ for all n a.s., that is, there is no deterministic bound for the change point. In practice, a sequence may often contain multiple change points. We assume here that at least one change point occurs with the probability one since one starts collecting data, and $\{\phi_n\}$ describes the distribution of the first such change point. The goal is to find an optimal sequential algorithm for detecting the change-point under these assumptions.

We begin with the sequential Bayes test of a no-change null hypothesis $H_0 : \nu > n$ against the alternative $H_1 : \nu \leq n$ for $n = 1, 2, \ldots$. Under the significance level α, the Bayes test rejects the null hypothesis if $\Pi_n > 1 - \alpha$, where

$$\begin{aligned}
\Pi_n &= P\{\nu \leq n | X_1, \ldots, X_n, \pi_0, \ldots, \pi_n\} \\
&= \frac{\sum_{k \leq n} \pi_k \mathcal{L}(X_1, \ldots, X_n \mid \nu = k)}{\sum_{k=0}^{\infty} \pi_k \mathcal{L}(X_1, \ldots, X_n \mid \nu = k)} \\
&= \frac{\sum_{k \leq n} \pi_k \rho(X_{k+1}) \cdots \rho(X_n)}{\sum_{k<n} \pi_k \rho(X_{k+1}) \cdots \rho(X_n) + 1 - \sum_{k<n} \pi_k}.
\end{aligned} \tag{8}$$

Let the stopping rule $T(\alpha)$ be the first time when the hypothesis H_0 is rejected. Then

$$\begin{aligned}
T(\alpha) &= \inf\{n : \Pi_n > 1 - \alpha\} \\
&= \inf\left\{n : \sum_{k<n} \pi_k \prod_{k+1}^{n} \rho(X_i) > \frac{1-\alpha}{\alpha} \sum_{k \geq n} \pi_k \right\}.
\end{aligned} \tag{9}$$

In terms of the discrete hazard function,

$$T(\alpha) = \inf\left\{n : \sum_{k<n} \frac{\phi_k}{1-\phi_k} \prod_{k+1}^{n} \frac{\rho(X_i)}{1-\phi_i} > \frac{1-\alpha}{\alpha}\right\}, \tag{10}$$

because $\phi_n = \pi_n / \sum_{k \geq n} \pi_k$.

For the logistic model (5), one obtains

$$T(\alpha) = \inf\left\{n : \sum_{k<n} e^{bk} \prod_{i=k+1}^{n} (1+e^{a+bi})\rho(X_i) > \frac{1-\alpha}{\alpha e^a}\right\}. \tag{11}$$

The case of a geometric-type distribution $\pi_n = Cq^n$ is a special case of (5) with $a = \mathrm{logit}(q) = \log(q/1-q)$ and $b = 0$. In this case, the rule (11) is equivalent to the Shiryaev's stopping rule (6) with the threshold $h = (1-\alpha)/(\alpha(1-q))$. For the non-informative generalized prior $\pi_n = const$, (9) becomes the Shiryaev–Roberts procedure (Roberts (1966)). In the next section, a modification of (9) (with a possibly non-constant α) is proved to be the Bayes stopping rule with respect to the risk functions (2) and (7). The CUSUM procedure, minimax for the risk (2), also belongs to the class of stopping rules (9).

3. BAYES STOPPING RULES

In this section, we find the Bayes stopping rules for the change-point detection under the risk functions (2) and (7). Derivation of Bayes stopping rules follows the ideas of Theorem 4.7 of Shiryaev (1978), generalizing it from the geometric case to any type of a prior distribution which unfolds according to the random discrete hazard rate model. The only assumption is made that $\{\phi_n\}$ form a homogeneous Markov sequence, which essentially means the homogeneity of experiments $\mathcal{E}_n$.

In the following lemmas, we choose a suitable Markov sequence $\{Y_n\}$ determining the risks (2) and (7) and apply Theorem 2.23 of Shiryaev (1978) about the optimal stopping of Markov sequences. In the sequel, for any non-random $n \geq 1$, $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$ is the sigma-field generated by $X_1, \ldots, X_n$, $\mathcal{F}_n^{\phi}$ is the sigma-field generated by $\phi_0, \ldots, \phi_n$, and $\mathcal{F}_n^* = \sigma(\mathcal{F}_n \times \mathcal{F}_{n+1}^{\phi})$. For any Markov stopping time T, $\mathcal{F}_T$ is the sigma-field consisting of all events A with the property $A \cap \{T \leq n\} \in \mathcal{F}_n$. Similarly, define $\mathcal{F}_T^{\phi}$ and $\mathcal{F}_T^*$. For the generality of notations, let $\mathcal{F}_0, \mathcal{F}_0^*$, etc., be the degenerate sigma-fields.

Lemma 1. (*Markov property*) *Suppose that* $\{\phi_n\}$ *is a homogeneous Markov sequence. Then* $Y_n = \{\Pi_n, \phi_{n+1}\}$ *is a homogeneous Markov sequence.*

Proof: Consider arbitrary $\mathcal{F}^*_{n+1}$-measurable real sets A_1 and A_2, and let $B = \{u/(1-u) \mid u \in A_1\}$ and

$$C = C(\Pi_n, \phi_{n+1}) = \left\{ \left(u - \frac{\phi_{n+1}}{1-\phi_{n+1}} \right) \frac{(1-\phi_{n+1})(1-\Pi_n)}{\Pi_n} \middle| u \in B \right\}.$$

From (8),

$$\frac{\Pi_{n+1}}{1-\Pi_{n+1}} = \frac{\Pi_n}{1-\Pi_n} \frac{\rho(X_{n+1})}{1-\phi_{n+1}} + \frac{\phi_{n+1}}{1-\phi_{n+1}}.$$

Hence,

$$\begin{aligned} P\{\Pi_{n+1} \in A_1 \mid \mathcal{F}^*_n\} &= P\left\{ \frac{\Pi_{n+1}}{1-\Pi_{n+1}} \in B \middle| \mathcal{F}^*_n \right\} \\ &= P\left\{ \frac{\Pi_n}{1-\Pi_n} \frac{\rho(X_{n+1})}{1-\phi_{n+1}} + \frac{\phi_{n+1}}{1-\phi_{n+1}} \in B \middle| \mathcal{F}^*_n \right\} \\ &= P\{\rho(X_{n+1}) \in C \mid \mathcal{F}^*_n\} \\ &= \int_{\{\rho(u)\in C\}} f(u) P\{\nu > n | \mathcal{F}^*_n\} du \\ &\quad + \int_{\{\rho(u)\in C\}} g(u) P\{\nu \le n | \mathcal{F}^*_n\} du \\ &= \int_{\{\rho(u)\in C\}} \{f(u)(1-\Pi_n) + g(u)\Pi_n\} du \end{aligned} \tag{12}$$

The latter is a function of Y_n. Therefore, by the Markovian property of ϕ_n and conditional independence of Π_{n+1} and ϕ_{n+2}, given $\mathcal{F}^*_n$,

$$\begin{aligned} &P\{\Pi_{n+1} \in A_1, \phi_{n+2} \in A_2 \mid \mathcal{F}^*_n\} \\ &\quad = P\{\Pi_{n+1} \in A_1 \mid Y_n\} P\{\phi_{n+2} \in A_2 \mid \phi_{n+1}\} \\ &\quad = P\{Y_{n+1} \in A_1 \times A_2 \mid Y_n\}. \end{aligned} \tag{13}$$

Since equality (13) holds for all rectangular sets $A_1 \times A_2$, it can be extended to all the Borel sets, which proves the Markovian property of Y_n. □

The next lemma shows that for a stopping rule T and suitably chosen functions $\eta(\cdot)$ and $\zeta(\cdot)$, the risks (2) and (7) have the form

$$R(T) = E\left\{\sum_{n<T} \eta(Y_n) - \zeta(Y_T)\right\}. \tag{14}$$

A stopping time T is called *proper* if $P\{T = \infty\} = 0$.

Lemma 2. (*Bayes risks in terms of Markov sequences*) *Let* $R_i(T) = EL_i(T, \nu)$, $i = 1, 2$, *be the Bayes risks in* (2) *and* (7), and R(T) *be defined by (14). Assume that* $\prod_0^\infty (1 - \phi_n)$ *diverges to 0 a.s. and* $E(\nu) = EE\{\nu \mid \phi_1, \phi_s, \ldots\} < \infty$. *Then for any proper stopping time* T,

(i) if $\eta(\pi, \cdot) = (C_2 + C_3)\pi - C_3$, $\zeta(\pi, \cdot) = C_1(\pi - 1)$, then $R(T) = R_1(T)$;
(ii) if $\eta(\pi, \cdot) = \lambda\pi$, $\zeta(\pi, \cdot) = \pi - 1$, then $R(T) = R_2(T)$.

Proof: It suffices to prove three equalities,

$$E\,\Pi_T = P\{\nu \le T\}, \tag{15}$$

$$E\sum_{n=0}^{T-1} \Pi_n = E(T - \nu)^+, \tag{16}$$

$$E\sum_{n=0}^{T-1} (1 - \Pi_n) = E\min\{T, \nu\}. \tag{17}$$

The first equality is trivial, because

$$E\,\Pi_T = E\,P\big\{\nu \le T \mid T, \mathcal{F}_T, \mathcal{F}_T^\phi\big\} = P\{\nu \le T\}.$$

The others follow from the Optional Stopping Theorem (see Theorem 2.3.2 of Ghosh, Mukhopadhyay and Sen (1997), Theorem 12 of Shiryaev (1978), Theorem 2.3 of Chow, Robbins and Siegmund (1991)). Consider a sequence of σ-fields $\bar{F}_N = \sigma(\mathcal{F}_N \times \mathcal{F}_N^\phi)$, with respect to which

$$\begin{aligned}\chi_N &= \sum_{n=0}^{N-1} \big(\Pi_n - P\{\nu \le n \mid \bar{\mathcal{F}}_N\}\big)\\ &= \sum_{n=0}^{N-1} \big(P\big\{\nu > n \mid \bar{\mathcal{F}}_N\big\} - P\big\{\nu > n \mid \bar{\mathcal{F}}_n\big\}\big)\end{aligned}$$

is a martingale, because

$$E|\chi_N| \le 2\sum_{n=0}^{\infty} P\{\nu > n\} = 2E(\nu) < \infty$$

and

$$E\big\{\chi_N \mid \bar{F}_{N-1}\big\} = \sum_{n=0}^{N-1}\big(P\big\{\nu > n \mid \bar{\mathcal{F}}_{N-1}\big\} - P\big\{\nu > n \mid \bar{\mathcal{F}}_n\big\}\big) = \chi_{N-1}.$$

Similarly, $E|\chi_T| \le 2E(\nu) < \infty$ for a stopping time T, and

$$\left|\int_{\{T>t\}} \chi_t dP\right| \le E|\chi_t| P\{T > t\} \le 2E(\nu)P\{T > t\} \to 0,$$

as $t \to \infty$. Therefore, by the Optional Stopping Theorem, $E\chi_T = E\chi_1 = 0$, and

$$\begin{aligned}
E\sum_{n=0}^{T-1}\Pi_n &= E\sum_{n=0}^{T-1} P\big\{\nu \le n \mid \mathcal{F}_T\big\} \\
&= E\sum_{n=0}^{T-1}\sum_{k=0}^{n} P\big\{\nu = k \mid \mathcal{F}_T\big\} \\
&= E\sum_{k=0}^{T-1}(T-k)P\big\{\nu = k \mid \mathcal{F}_T\big\} \\
&= EE\big\{(T-\nu)I_{\nu<T} \mid \mathcal{F}_T\big\} = E(T-\nu)^+,
\end{aligned}$$

which proves (16).

The sequence χ_N can also be written as

$$\chi_N = \sum_{n=0}^{N-1}\big((1-\Pi_n) - P\big\{\nu > n \mid \mathcal{F}_N\big\}\big).$$

Thus, applying the Optional Stopping Theorem once again,

$$\begin{aligned}
E\sum_{n=0}^{T-1}(1-\Pi_n) &= E\sum_{n=0}^{T-1} P\big\{\nu > n \mid \mathcal{F}_T\big\} \\
&= E\sum_{n=0}^{T-1}\sum_{k=n+1}^{\infty} P\big\{\nu = k \mid \mathcal{F}_T\big\} \\
&= E\sum_{k=1}^{\infty}\min\{k, T\}P\big\{\nu = k \mid \mathcal{F}_T\big\} = E\min\{\nu, T\},
\end{aligned}$$

which proves (17) and concludes the proof of the Lemma. □

According to Lemmas 1 and 2, the problem of finding the Bayes stopping rules under the risk functions (2) and (7) reduces to the problem of optimal stopping of Markov sequences. For each of the considered risks, let $s(x) = -\inf_T R(T)$ be the corresponding *payoff function*. Theorem 2.23 of Shiryaev (1978) states that under the conditions guaranteed by Lemmas 1 and 2,

$$T^* = \inf\{n \geq 0 \mid s(Y_n) = \zeta(Y_n)\} \tag{18}$$

is the Bayes stopping rule if the risk function $R(T)$ has the form (14). It also characterizes the payoff function as the solution of the equation

$$s(y) = \max\{\zeta(y),\ E\{s(Y_{n+1}) \mid Y_n = y\} - \eta(y)\} \tag{19}$$

and as the limit

$$s(y) = \lim_{N\to\infty} Q^N \zeta(y), \tag{20}$$

where Q is an operator defined as

$$Qw(y) = \max\{w(y),\ E\{w(Y_{n+1}) \mid Y_n = y\} - \eta(y)\}. \tag{21}$$

We apply this theorem to the risks $R_i(T)$, the corresponding payoff functions $s_i(x)$, $i = 1, 2$, and the Markov sequence specified in Lemma 1.

Lemma 3. *(Convexity) If $h(\pi, \phi)$ is a non-strictly convex function in π on* $[0, 1]$ *for any fixed values of the other arguments, then so is the function*

$$\mathcal{T}h(\pi, \phi) = E\{h(Y_{n+1}) \mid \pi_n = \pi, \phi_{n+1} = \phi\}.$$

Proof: It follows from (8) that

$$\Pi_{n+1} = \frac{\Pi_n \rho(X_{n+1}) + \phi_{n+1}(1 - \Pi_n)}{\Pi_n \rho(X_{n+1}) + 1 - \Pi_n}. \tag{22}$$

Then, conditioning on $\Pi_n = \pi$ and $\phi_{n+1} = \phi$, one obtains,

$$\begin{aligned}\mathcal{T}h(\pi, \phi) &= P\{\nu \leq n \mid \Pi_n = \pi\} E\{h(Y_{n+1}) \mid \Pi_n = \pi, \phi_{n+1} = \phi, \nu \leq n\} \\ &\quad + P\{\nu > n \mid \Pi_n = \pi\} E\{h(Y_{n+1}) \mid \Pi_n = \pi, \phi_{n+1} = \phi, \nu > n\} \\ &= \pi E_G\{h(Y_{n+1}) \mid \pi, \phi\} + (1 - \pi) E_F\{h(Y_{n+1}) \mid \pi, \phi\}\end{aligned}$$

$$= E_F\{\pi\rho(X_{n+1})h(\Pi_{n+1}, \phi_{n+2}) + (1-\pi)h(\Pi_{n+1}, \phi_{n+2}) \mid \pi, \phi\}$$
$$= E_F\left\{(\pi\rho(X_{n+1}) + 1 - \pi)h\left(\frac{\pi\rho(X_{n+1}) + (1-\pi)\phi}{\pi\rho(X_{n+1}) + 1 - \pi}, \phi_{n+2}\right) \Bigm| \phi_{n+1} = \phi\right\}. \tag{23}$$

In the case when $h(\cdot)$ is twice differentiable, it is straightforward to obtain the second derivative of the right-hand side of (23),

$$\frac{\partial^2 \mathcal{T}h(\pi, \phi)}{\partial\pi^2} = E_F\left\{\left(\frac{\rho(X_{n+1})(1-\phi)}{\pi\rho(X_{n+1}) + 1 - \pi}\right)^2 \frac{\partial^2 h(\pi, \phi_2)}{\partial\pi^2} \mid \phi_1 = \phi\right\} \geq 0,$$

because $\partial^2 h(\pi, \phi)/\partial\pi^2 \geq 0$ for any ϕ. Thus, $\mathcal{T}h(\pi, \phi)$ is convex in its first argument. Every convex function h on [0, 1] can be uniformly approximated by twice differentiable convex functions h_j, so that

$$\lim_{j\to\infty} \max_{[0;1]} |h(\cdot, \phi) - h_j(\cdot, \phi)| = 0$$

for any $\phi \in [0, 1]$. Then

$$\lim_{j\to\infty} \max_{\pi\in[0;1]} |\mathcal{T}h(\pi, \phi) - \mathcal{T}h_j(\pi, \phi)|$$
$$\leq \lim_{j\to\infty} \max_{\pi\in[0;1]} E\{|h(\Pi_{n+1}, \phi) - h_j(\Pi_{n+1}, \phi)| \mid \pi, \phi\} = 0,$$

that is, $\mathcal{T}h$ is represented as the uniform limit of convex functions. Hence, $\mathcal{T}h$ is convex. □

The functions ζ defined in Lemma 2 are convex, so that Lemma 3 applies. It shows that $(\mathcal{T}\zeta - \eta)$ is also convex in its first argument, because η is linear. This yields the convexity of $Q\zeta$, which is defined in (21) as the maximum of two convex functions. For the same reason, if $Q^N\zeta$ is convex for $N \geq 1$, then $Q^{N+1}\zeta$ is convex. By induction, $Q^N\zeta$ is convex for all $N \geq 1$. Therefore, $s = \lim_N Q^N\zeta$ is convex in its first argument for both considered risk functions. The following lemma shows that roots of the equation $s(x) = \zeta(x)$ defining the optimal stopping time in (18) form a neighborhood of $\pi = 1$.

Lemma 4. *(Payoff function) For any $a > b \geq 0$, $c > 0$, let $\eta(y) = a\pi - b$ and $\zeta(y) = c(\pi - 1)$, where π is the first components of a vector y, and let Y_n be the Markov sequence defined in Lemma 1. If $\nu < \infty$ a.s., then there exists $\pi^* = \pi^*(\phi) \in [0, 1)$ such that $s(y) = \zeta(y) \Leftrightarrow \pi \geq \pi^*$. If $\phi = \phi_n$ for some n, then π^* is strictly positive a.s.*

Proof: Consider the payoff at $\pi = 1$ and $\pi = 0$. From (22), if $\Pi_n = 1$, then $\Pi_{n+1} = 1$. Therefore, for any function $w(\pi, \phi)$ with the property $w(1, \phi) \equiv 0$,

$$\begin{aligned} Qw(1,\phi) &= \max\bigl(w(1,\phi),\ E\bigl\{w(\Pi_{n+1}, \phi_{n+2}) \mid \Pi_n = 1, \phi_{n+1} = \phi\bigr\} - \eta(1,\phi)\bigr) \\ &= (-\eta(1,\phi))^+ \equiv 0. \end{aligned}$$

By induction, $Q^N w(1,\phi) \equiv 0$ for all $N \geq 1$, $s(1,\phi) = \lim Q^N \zeta(1,\phi) \equiv 0$ and $\mathcal{T}s(1,\phi) \equiv 0$.

From (22), if $\Pi_n = 0$, then $\Pi_{n+1} = \phi_{n+1}$. Also, from (19), $s(\pi,\phi) \geq \zeta(\pi,\phi)$ for all (π,ϕ). Therefore,

$$\begin{aligned} s(0,\phi) &= \mathcal{T}s(0,\phi) - \eta(0,\phi) \\ &= E\bigl\{s(\phi, \phi_{n+2}) \mid \phi_{n+1} = \phi\bigr\} + b \geq c\phi - c + b \geq -c. \end{aligned} \tag{24}$$

Equality in (24) holds if and only if $\phi = b = 0$, in which case

$$s(0,0) = E\bigl\{s(0,\phi_{n+2}) \mid \phi_{n+1} = 0\bigr\} = -c,$$

so that $s(0,\phi_{n+2}) \equiv -c$, i.e. $\phi_{n+2} = 0$. However, since ϕ_n form a homogeneous Markov sequence, this implies that $P\bigl\{\phi_{n+1} = 0 \mid \phi_n = 0\bigr\} = 1$ for all n. Hence, if $\phi = 0 = \phi_n$ for some n, then $\phi_k = 0$ a.s. for all $k \geq n$. This implies convergence of $\prod_0^\infty (1 - \phi_n)$ and according to (4), it contradicts the almost sure finiteness of ν. Thus, $s(0,\phi_n) > -c$ a.s. for any observed ϕ_n.

As a result, we have that $\mathcal{T}s(0,\phi) - \eta(0,\phi) \geq \zeta(0,\phi)$ and $\mathcal{T}s(1,\phi) - \eta(1,\phi) = b - a < \zeta(1,\phi)$. By continuity, there exists $\pi^* \in [0,1)$, such that $\mathcal{T}s(\pi^*,\phi) - \eta(\pi^*,\phi) = \zeta(\pi^*,\phi)$. The uniqueness of π^* is guaranteed by the convexity of $\mathcal{T}s - \eta$. And if $\phi = \phi_n$ for some n, then $\pi^* > 0$. □

The form of optimal stopping rules for the risks (2) and (7) follows from Lemma 4. According to (18), the Bayes stopping rule is

$$T^* = \inf\bigl\{n \geq 0 \mid \Pi_n \geq \pi^*\bigr\}. \tag{25}$$

where π^* is the point guaranteed by Lemma 4. Note that T^* coincides with the stopping rule (9), obtained from the sequential Bayes test, for $\alpha = 1 - \pi^*$.

We formulate the obtained results below.

Theorem 1. *Assume a random hazard rate change-point model with the homogeneous Markov hazard rate sequence $\{\phi_n\}$ and finite $E(\nu)$.*

(i) For any λ, there exists $0 < \alpha_1 < 1$, such that the stopping time $T(\alpha_1)$ is Bayes under the loss function (7).

(ii) For any positive C_1, C_2 and C_3, there exists $0 < \alpha_2 < 1$, such that the stopping time $T(\alpha_2)$ is Bayes under the loss function (2).

Theorem 1 also generalizes the known result about the optimality of the CUSUM procedure.

Corollary 1. *The CUSUM stopping time (1) is Bayes with respect to the risk function (2) or (7) if and only if Π_n is a non-decreasing function of $R_n = \max_{k<n} \rho(X_{k+1})1 \cdots \rho(X_n)$.*

Proof: From the form of stopping rules (1) and (9), it is clear that $\tau(h) = T(\alpha)$ a.s. for suitable h and α if and only if Π_n and R_n are non-decreasing functions of each other. If this condition holds, $\tau(h)$ is Bayes, because by Theorem 1, $T(\alpha)$ is Bayes. □

Notice that the least favorable prior distribution (3) guarantees that Π_n is a non-decreasing function of R_n. Indeed, let $p_n = P\{\nu < n \mid \bar{\mathcal{F}}_n\}$. According to Lemma 1 of Ritov (1990), for the prior distribution (3),

$$p_n = \frac{pR_n}{1 - p + pR_n}, \tag{26}$$

which is a non-decreasing function of R_n (notice that in Ritov (1990), a change point is defined as $\nu - 1$). Then,

$$\begin{aligned}\Pi_n &= P\{\nu < n \mid \bar{\mathcal{F}}_n\} + P\{\nu \geq n \mid \bar{\mathcal{F}}_n\} P\{\nu = n \mid \nu \geq n, \bar{\mathcal{F}}_n\} \\ &= p_n + (1 - p_n)(1 - R_n)^+,\end{aligned}$$

from where $\Pi_n \equiv p$ for $R_n \leq 1$, and $\Pi_n = p_n$ for $R_n \geq 1$. Hence, Π_n is a non-decreasing function of R_n.

4. EXAMPLE: DETECTING THE BEGINNING OF AN INFLUENZA EPIDEMIC PERIOD

Currently influenza causes more morbidity and mortality in the United States than AIDS. Each year it accounts for 10 000 to 40 000 fatalities, nearly 200 000 hospitalizations, and about 70 million work-loss days, with the associated annual cost of \$12 billion (Neuzil et al, 1999). The rate of influenza diagnoses and influenza related hospitalizations significantly increases during epidemic periods.

Recent studies (Medina et al, 1997, Peters et al, 2000) confirmed strong association between influenza epidemics and such factors as weather conditions, air pollution, pollen, ozone level, and others. For example, the likelihood of an influenza epidemic increases during a passage of a cold front followed by a high pressure system. At the same time, the influenza virus dies more rapidly with high humidity and vigorous air movement (White and Hertz-Picciotto, 1985). Based on the combination of factors, one can determine "favorable" and "unfavorable" days for the beginning of an influenza epidemic period.

Statistically, daily count for influenza diagnoses X can be modeled by a Poisson distribution with a parameter λ_1 during influenza epidemics and $\lambda_0 < \lambda_1$ during inter-epidemic periods. Thus, the time ν that marks the beginning of an influenza epidemic is a change point from Poisson (λ_0) distribution to Poisson (λ_1). Suppose that the prior probability for an epidemic period to start on an unfavorable day is p_0, whereas this probability is $p_1 > p_0$ on a favorable day. Further, unfavorable days are followed by favorable days with the probability q_0, and favorable days are followed by unfavorable days with the probability q_1. Thus, the discrete hazard rate ϕ_n of the change point forms a homogeneous Markov chain with the states p_0 and p_1 and the transition probability matrix

$$\begin{pmatrix} 1-q_0 & q_0 \\ q_1 & 1-q_1 \end{pmatrix}.$$

Based on this model, observed influencing factors, and the daily number of influenza diagnoses, we can now apply the methods of Section 3 and find the Bayes stopping rule signalling the beginning of an influenza epidemic.

First, parameters p_0, p_1, q_0, q_1, λ_0, and λ_1 are estimated from the historical data. Obviously, they vary from one region to another. For example, q_0 (q_1) can be estimated by the reciprocals of the average number of consecutive favorable (unfavorable) days in the area. Then, one computes $Q^N(\pi, \phi)$ recursively for $\pi \in [0, 1]$, $\phi \in \{p_0, p_1\}$, and $N = 0, 1, 2, \ldots$. Equation (21) is used here, and $\mathcal{T}w(\pi, \phi) = E\{w(Y_{n+1}) \mid \pi, \phi\}$ is evaluated as

$$\begin{aligned}\mathcal{T}w(\pi, p_0) = E_{\lambda_0}\bigg\{&(1-q_0)w\left(\frac{\pi\rho + (1-\pi)p_0}{\pi\rho + 1 - \pi}, p_0\right) \\ &+ q_0 w\left(\frac{\pi\rho + (1-\pi)p_0}{\pi\rho + 1 - \pi}, p_1\right)\bigg\}(\pi\rho + 1 - \pi)\end{aligned}$$

and

$$\begin{aligned}\mathcal{T}w(\pi, p_1) = E_{\lambda_0}\bigg\{&(1-q_1)w\left(\frac{\pi\rho + (1-\pi)p_1}{\pi\rho + 1 - \pi}, p_1\right) \\ &+ q_1 w\left(\frac{\pi\rho + (1-\pi)p_1}{\pi\rho + 1 - \pi}, p_0\right)\bigg\}(\pi\rho + 1 - \pi),\end{aligned}$$

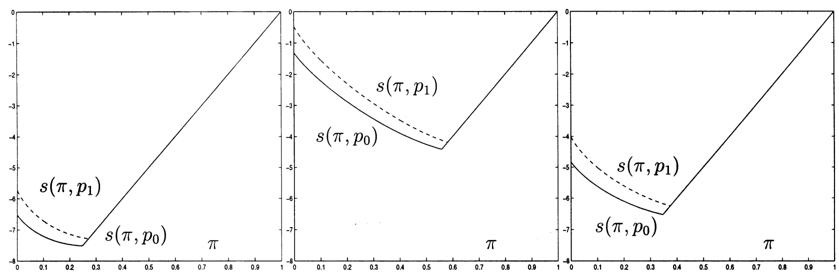

Figure 1. Payoff functions for the study of influenza epidemics.

$\rho = (\lambda_1/\lambda_0)^X \exp\{\lambda_0 - \lambda_1\}$, and X is a Poisson (λ_0) random variable. Sufficiently large number of iterations yields the form of the payoff function $s(\pi, \phi)$. Finally, one finds

$$\pi^*(p_i) = \min\{\pi \mid s(\pi, p_i) = \zeta(\pi, p_i)\} = \arg\min_{\pi} s(\pi, p_i)$$

and constructs the Bayes stopping rule T^* according to (25). As shown in Section 3, T^* is also the result of a sequential Bayes test of the no-change null hypothesis with the significance level $\alpha = 1 - \pi^*$.

Graphs of the payoff function $s(\pi, p_i)$ for $i = 0, 1$, $p_0 = 0.002$, $p_1 = 0.02, q_0 = 0.08$, $q_1 = 0.15$, $\lambda_0 = 30$, $\lambda_1 = 36$ and the risk function

$$R(T, \nu) = 10P\{T < \nu\} + 3E(T - \nu)^+ - 0.1E\min\{T, \nu\} \tag{27}$$

are depicted on Figure 1, left. We have $\pi^*(0.002) = 0.272$ and $\pi^*(0.02) = 0.249$. One can now compute the stopping rule

$$T^* = \inf\{n \geq 0 \mid \Pi_n \geq \pi^*\},$$

which is Bayes with respect to the chosen risk function. At the same time, T^* is the result of Bayes tests with the significance level $1 - \pi^*(\phi_{T+1})$.

Replacing the term $3E(T-\nu)^+$ in (27) by $2E(T-\nu)^+$ reduces the loss caused by the delay. Therefore, the corresponding Bayes stopping rule has a larger mean delay in favor of a smaller probability of a false alarm. This is achieved by a stronger condition of stopping, that is, by larger values of π^*, $\pi^*(0.002) = 0.582$ and $\pi^*(0.02) = 0.559$. The payoff function for this case is depicted on Figure 1, middle.

Replacing $\lambda_1 = 36$ by $\lambda_1 = 37$ implies a more significant change of the Poisson parameter and reduction of the mean delay. However, the balance between the mean delay and the frequency of false alarms is determined by (27). Larger values of π^*, $\pi^*(0.002) = 0.375$ and $\pi^*(0.02) = 0.348$, restore this balance. The corresponding payoff function is depicted on Figure 1, right.

ACKNOWLEDGMENT

This research was supported by the NSF grant, award #9818959.

REFERENCES

H. Bühlmann. *Mathematical Methods in Risk Theory*. Springer, New York, 1970.

Y. S. Chow, H. Robbins and D. Siegmund. *The Theory of Optimal Stopping*. Dover Publications, New York, 1991.

M. Ghosh, N. Mukhopadhyay and P. K. Sen. *Sequential Estimation*. Wiley, New York, 1997.

Z. Govindarajulu. *The Sequential Statistical Analysis of Hypothesis Testing, Point and Interval Estimation, and Decision Theory*. American Sciences Press, Columbus, Ohio, 1981.

T. L. Lai. Sequential change-point detection in quality control and dynamical systems. *J. R. Statist. Soc. B*, 57:613–658, 1995.

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

S. Medina, A. Le Tertre, P. Quénel, Y. Le Moullec, P. Lameloise, J. C. Guzzo, B. Festy, R. Ferry, W. Dab. Air pollution and doctors' house calls: Results from the ERPURS system for monitoring the effects of air pollution on public health in Greater Paris, France, 1991–1995. *Environmental Research*, 75:73–84, 1997.

D. C. Montgomery. *Introduction to Statistical Quality Control, Third Ed.* Wiley, New York, 1997.

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

K. M. Neuzil, G. W. Reed, E. F. Mitchel, M. R. Griffin. Influenza-associated morbidity and mortality in young and middle-aged women. *J. Amer. Medical Assoc.*, 281:901–907, 1999.

E. S. Page. Continuous inspection schemes. *Biomterika*, 41:100–115, 1954.

A. Peters, J. Skorkovsky, F. Kotesovec, J. Brynda, C. Spix, H. E. Wichmann, and J. Heinrich. Associations between mortality and air pollution in Central Europe. *Environmental Health Perspectives*, 108:283–287, 2000.

M. Pollak. Optimal detection of a change in distribution. *Ann. Statist.*, 13:206–227, 1985.

Y. Ritov. Decision theoretic optimality of CUSUM procedure. *Ann. Statist.*, 18:1464–1469, 1990.

S. W. Roberts. A comparison of some control chart procedures. *Technometrics*, 8:411–430, 1966.

A. N. Shiryaev. On optimal method in earliest detection problems. *Theory Probab. Appl.*, 8:26–51, 1963.

A. N. Shiryaev. *Optimal Stopping Rules*. Springer-Verlag, New York, 1978.

M. R. White and I. Hertz-Picciotto. Human health: Analysis of climate related to health. In *Characterization of information requirements for studies of* CO_2 *effects: Water resources, agriculture, fisheries, forests and human health*, ed. M.R. White, pp. 172–206, Washington, D.C.: Department of Energy, 1985.

Received July 2000
Revised March 2001
Recommended by S. Zacks