The Annals of Statistics
2017, Vol. 45, No. 4, 1375–1402
DOI: 10.1214/16-AOS1479

ON THE OPTIMALITY OF BAYESIAN CHANGE-POINT DETECTION

BY DONG HAN[1*], FUGEE TSUNG[2†] AND JINGUO XIAN[1*]

Shanghai Jiao Tong University and*
Hong Kong University of Science and Technology†

By introducing suitable loss random variables of detection, we obtain optimal tests in terms of the stopping time or alarm time for Bayesian change-point detection not only for a general prior distribution of change-points but also for observations being a Markov process. Moreover, the optimal (minimal) average detection delay is proved to be equal to 1 for any (possibly large) average run length to false alarm if the number of possible change-points is finite.

1. Introduction. Consider a series of observations $X_0, X_1, X_2, \dots,$ whose distribution may change at some point in time $\tau \geq 1$. The objective is to raise an alarm as soon as a change occurs, subject to a restriction on the rate of false alarms.

The need for abrupt changes to be detected quickly with a low false alarm rate in a stochastic system arises in a variety of applications, including quality control, segmentation of signals, biomedical signal processing, financial engineering and fault detection and diagnosis in complex structures.

In the classical setting of the problem, it is commonly assumed that the probability distribution of observations $X_0, X_1, X_2, \dots, X_n, \dots$ is known and is P_{v_0} for $n < \tau$, and that after the change-point τ, the distribution of $X_\tau, X_{\tau+1}, \dots$ becomes P_{v_1} which is also known. Without loss of generality, we can assume that the distributions P_{v_i} have density $p_{v_i}, i = 0, 1$. Here, $v_0, v_1 \in V$ are two parameters and V denotes the parameter space of v where $p_v \neq p_{v'}$ if and only if $v \neq v'$ for $v, v' \in V$. Note that the parameter v may not be the mean, deviation, etc., and is often used to distinguish the distributions. Let $\mathbf{P}_\tau$ and $\mathbf{E}_\tau$ be the probability distribution and the expectation of $\{X_\tau, X_{\tau+1}, \dots\}$, respectively, if a change occurs at time τ, where $\tau = 1, 2, \dots$. When $\tau = \infty$, that is, a change never occurs, the probability distribution and the expectation are denoted by $\mathbf{P}_\infty$ and $\mathbf{E}_\infty$, respectively for all observations $X_0, X_1, X_2, \dots, X_n, \dots$.

Generally speaking, any sequential test for change-point detection can be modeled as a stopping time or an alarm time $T \geq 1$ adapted to the filtration $\{\mathfrak{F}_t\}_{t\geq 0}$,

Received October 2015; revised August 2016.

[1]Supported by National Natural Science Foundation of China (11531001) and National Basic Research Program of China (973 Program, 2015CB856004).

[2]Supported by the RGC General Research Fund 619913.

MSC2010 subject classifications. Primary 62L10; secondary 62L15.

Key words and phrases. Optimal test, Bayesian change-point detection, Markov process.

where $\mathfrak{F}_t = \sigma\{X_k, 0 \le k \le t\}$ denotes the smallest σ-algebra with respect to which all of the random variables $X_0, \dots, X_t$ are measurable. We usually assume that $X_0 \equiv x_0$ (a constant) and $\mathfrak{F}_0 = \{\phi, \Omega\}$. The optimality of the stopping time usually means that the detection delay $(T - \tau + 1)^+$ measured in some sense is the smallest among all stopping times with a probability of false alarm $\mathbf{P}_\infty(T < \tau)$ no greater than a preset level $\alpha \in (0, 1)$, or among all stopping times with a false alarm rate no less than a given value $\gamma > 1$, that is, $\mathbf{E}_\infty(T) \ge \gamma$.

Three major optimal performance measures exist for detecting change-points. The first one is the Shiryaev [13, 15] procedure which can be written as

$$\inf_{T:\mathbf{P}_\infty(T<\tau)\le\alpha} \mathcal{J}_S(T)$$

or

$$\inf_{T:\mathbf{E}_\infty(T)\ge\gamma} \mathcal{J}_S(T),$$

where

$$\mathcal{J}_S(T) = \mathbf{E}\big((T - \tau + 1)|T \ge \tau\big) = \frac{\sum_{k=1}^\infty \rho_k \mathbf{E}_k (T - k + 1)^+}{\sum_{k=1}^\infty \rho_k \mathbf{P}_\infty(T \ge k)}, \tag{1}$$

and $\{\rho_k, k \ge 1\}$ is a known prior probability distribution of τ, that is, $\rho_k = \mathbf{P}(\tau = k), k \ge 1$. It is usually called Shiryaev's Bayesian change-point detection approach or the Bayesian formulation.

The second is Lorden's procedure [5]:

$$\inf_{T:\mathbf{E}_\infty(T)\ge\gamma} \mathcal{J}_L(T), \tag{2}$$

where $\mathcal{J}_L(T)$ is the worst average delay, that is,

$$\mathcal{J}_L(T) = \sup_{k\ge1} \operatorname{esssup}\big\{\mathbf{E}_k\big((T - k + 1)^+|\mathfrak{F}_{k-1}\big)\big\}.$$

The third one, proposed by Pollak [7], also considers the worst average delay $\mathcal{J}_P(T) = \sup_{t\ge1} \mathbf{E}_t(T - t + 1|T \ge t)$, and the optimal stopping time can be written as

$$\inf_{T:\mathbf{E}_\infty(T)\ge\gamma} \mathcal{J}_P(T). \tag{3}$$

Note that τ is deterministic and unknown in both Lorden's measure and Pollak's measure. Hence, the two measures are often called non-Bayesian detection formulations. Moustakides [6] introduced a general framework for capturing and better understanding the above three optimization criteria and showed that $\mathcal{J}_S(T) \le \mathcal{J}_P(T) \le \mathcal{J}_L(T)$.

Since this paper focus is on the optimality of the Bayesian detection problem, we will only recall the main results on the optimal or asymptotically optimal tests in the Bayesian detection procedure. Those who are interested in the results on the optimality of Lorden's measure and Pollak's measure may refer to [9, 11, 14] and [17].

Under the assumption that the observation process $\{X_k, k \geq 0\}$ is i.i.d. and the prior distribution of the change-point is geometric distribution, that is, $\rho_1 = \rho_0 + (1-\rho_0)\rho$, $\rho_k = (1-\rho_0)\rho(1-\rho)^{k-1}$ $(k \geq 2)$, where $0 \leq \rho_0 < 1$ and $0 < \rho < 1$, Shiryaev [13, 15] first proved that

$$S_\rho = \min\{n : \mathbf{P}(\tau \leq n | \mathfrak{F}_n) \geq a(\rho, \alpha)\} \tag{4}$$

is optimal, that is, it minimizes $\mathcal{J}_S(T)$ for an appropriately chosen threshold $a(\rho, \alpha) \in (0, 1)$ such that $\alpha = \mathbf{P}_\infty(S_\rho < \tau)$.

It follows from the Bayes rule that [10]

$$\mathbf{P}(\tau \leq n | \mathfrak{F}_n) = \frac{R_n(\rho)}{R_n(\rho) + 1/\rho},$$

where

$$R_n(\rho) = \frac{\rho_0}{(1-\rho_0)\rho} \prod_{k=1}^{n} \frac{\Lambda_k}{1-\rho} + \sum_{t=1}^{n} \prod_{k=t}^{n} \frac{\Lambda_k}{1-\rho} \tag{5}$$

for $n \geq 1$ and $\Lambda_k = p_{v_1}(X_k)/p_{v_0}(X_k)$. Hence, S_ρ is equivalent to the stopping time

$$T_S(\rho) = \min\{n : R_n(\rho) \geq c\}, \tag{6}$$

where c is a positive constant.

For a general prior distribution of the change-point τ, Chow, Robbins and Siegmund [1] have shown that the following test:

$$T_{\mathrm{CRS}} = \min\{n : \psi_n \geq \mathbf{E}(\zeta_{n+1} | \mathfrak{F}_n)\} \tag{7}$$

is optimal in the sense that $\mathbf{E}(\psi_{T_{\mathrm{CRS}}}) \geq \mathbf{E}(\psi_T)$ for all tests (stopping times) T such that $\mathbf{E}(\psi_T) < \infty$, where

$$\psi_n = c(1-\pi_n) + \sum_{k=0}^{n-1} (n-i)\pi_k^n,$$

$\pi_n = \mathbf{P}(\tau \leq n | \mathfrak{F}_n)$, $\pi_k^n = \mathbf{P}(\tau = k | \mathfrak{F}_n)$ and $\zeta_n = \text{ess sup}_{T \geq n}\{\mathbf{E}(\psi_T | \mathfrak{F}_n)\}$. Since the random variables $\zeta_n, n \geq 0$, in T_{CRS} are in general impossible to compute directly from the definition and, therefore, the test is difficult to use in practice, the above result has rarely been mentioned in the relevant literature.

When the distribution of the change-point τ is geometric, Yakir [20] generalized Shiryaev's result for the Markov chain case. Tartakovsky and Veeravalli [19] proved that the Shiryaev procedure is asymptotically optimal in the general non-i.i.d. case when the false alarm probability approaches zero. Tartakovsky [16] proposed an asymptotically optimal Bayesian change-point detection procedure for general non-i.i.d. models when the global false alarm probability $\mathbf{P}_\infty(T < \infty)$ approaches zero.

Note that for $\rho_0 = r\rho$ as $\rho \to 0$, where $r \geq 0$, Shiryaev's stopping time $T_S(\rho)$ converges to the following Shiryaev–Roberts [12, 13] test:

$$T_{\mathrm{SR}}(c) = \min\left\{n : r\prod_{k=1}^{n}\Lambda_k + \sum_{t=1}^{n}\prod_{k=t}^{n}\Lambda_k \geq c\right\} \tag{8}$$

for a positive threshold $c > 0$ and

$$\begin{aligned}\hat{\mathcal{J}}_S(T) &= \lim_{\rho\to 0}\frac{r\rho\mathbf{E}_1(T) + (1-r\rho)\sum_{k=1}^{\infty}\rho(1-\rho)^{k-1}\mathbf{E}_k(T-k+1)^+}{r\rho + (1-r\rho)\sum_{k=1}^{\infty}\rho(1-\rho)^{k-1}\mathbf{P}_\infty(T\geq k)}\\ &= \frac{r\mathbf{E}_1(T) + \sum_{k=1}^{\infty}\mathbf{E}_k(T-k+1)^+}{r+\mathbf{E}_\infty(T)}.\end{aligned} \tag{9}$$

$T_{\mathrm{SR}}(c)$ is often called the generalized Bayesian change-point detection. Pollak and Tartakovsky [8] proved that the Shiryaev−Roberts test is optimal in the measure (9), that is, $T_{\mathrm{SR}}(c_\gamma)$ minimizes $\hat{\mathcal{J}}_S(T)$ over all stopping times T satisfying $\mathbf{E}_\infty(T) \geq \gamma$, where $\mathbf{E}_\infty(T_{\mathrm{SR}}(c_\gamma)) = \gamma$. Furthermore, Tartakovsky, Pollak and Polunchenko [18] showed that the Shiryaev–Roberts procedure that starts either from a specially designed point r or from the random quasi-stationary point is third-order asymptotically optimal.

So far, in addition to the above two tests, S_ρ and $T_{\mathrm{SR}}(c)$ have been proved to be strictly (nonasymptotically) optimal. There is little research that deals with strictly optimal tests in Bayesian change-point detection for a general prior distribution of the change-point τ.

In this paper, by introducing suitable loss random variables of detection we obtain strictly optimal tests for Bayesian change-point detection not only for a general prior distribution of the change-point but also for observations $X_0, X_1, X_2, \ldots$ that form a Markov process. Our main contributions to the study of the Bayes optimality of stopping times are in the following four aspects:

(i) We propose and prove that the stopping time (see Theorem 1): $T^*(c^*) = \min\{n \geq 1 : Y_n \geq c_n(c^*)\rho_{n+1}\}$ is strictly optimal in the Bayesian formulation for a general prior distribution $\{\rho_k\}$ and an observation sequence $\{X_n, n \geq 0\}$ being a Markov process.

(ii) Considering simultaneously the probability of false alarm $P(T \leq \tau)$ and the average run length (ARL) $\mathbf{E}_\infty(T)$ to false alarm, where $P(T \leq \tau)$ does not exceed, or is not less than, a desired level $\alpha \in (0,1)$ and $\mathbf{E}_\infty(T)$ is "no worse" than $\gamma > 1$, we prove that the two stopping times (see Theorems 2 and 3): $T_\gamma(c,b) = \min\{n \geq 1 : Y_n \geq c_n(c)\rho_{n+1} + b\}$ and $T_\eta(c) = \min\{n \geq 1 : Y_n \geq c_n(c)\}$ are strictly optimal over all stopping times respectively in $D_\gamma = \{T \geq 1 : \mathbf{E}_\infty(T) \geq \gamma; \mathbf{P}(T < \tau) \leq \alpha^*\}$ and $D'_\gamma = \{T \geq 1 : \mathbf{E}_\infty(T) \geq \gamma; \mathbf{P}(T < \tau) \geq \alpha^*\}$.

(iii) By constructing a series of stopping times, we prove that

$$\inf_{T:\mathbf{E}_\infty(T)\geq\gamma}\mathcal{J}_S(T) = 1$$

for every finite ($N < \infty$) prior distribution $\{\rho_k, 1 \le k \le N\}$ and any (possibly large) $\gamma > 1$. This result is somewhat unexpected.

(iv) We choose another approach (different from Shiryaev's method) to prove the above results of optimality of stopping times since the posterior probabilities $\pi_n = \mathbf{P}(\tau \le n|\mathfrak{F}_n)$, $n \ge 1$, is not a Markov process when the prior distribution is not geometric and the observation sequence $\{X_n, n \ge 0\}$ is not mutually independent.

Section 2 presents three optimal tests subject to a given probability of false alarm, or a given ARL to false alarm, for a general prior distribution of change-point τ. In Section 3, for any finite possible change-points [$\rho_k = \mathbf{P}(\tau = k) = 0$ for all $k \ge N + 1$, $N < \infty$], we prove that the optimal (minimal) average detection delay is 1, that is, $\inf_{T:\mathbf{E}_\infty(T)\ge\gamma} \mathcal{J}_S(T) = 1$. Section 4 illustrates simulation results of comparing the detection performance of the three tests: $T_5(c)$, the CUSUM test $T_C(c)$ and EWMA test $T_E(c)$ for monitoring mean shifts in the special cases of τ: $\rho_k = 1/5$, $1 \le k \le 5$. We conclude the paper in Section 5.

2. Optimal tests for Bayesian change-point detection with infinite possible change-points. In this section, we consider the case where the change-point τ will take infinite possible nonnegative integer values, that is, $\mathbf{P}(\tau_k = k) = \rho_k > 0$ for all $k \ge 1$.

Let the observations $X_0, X_1, X_2, \dots$ be a time homogeneous Markov process taking values in some real number space $E \subseteq (-\infty, +\infty)$ with a known transition probability density $p_{v_0}(\cdot,\cdot)$ before the change-point $\tau \ge 1$, and with the possible post-change transition probability density $p_v(\cdot,\cdot)$ after the change-point, where $v \in V$ and V is a parameter space. We do not know the post-change transition probability density of the process until it is detected. But the possible change domain, and its boundary (including the size and form of the boundary), of the observation process may be determined from engineering knowledge, practical experience or statistical data. So we may assume that the unknown post-change transition probability density $p_v(\cdot,\cdot)$ lies in a closed domain $\Gamma(V)$, where $\Gamma(V) = \{p_v(\cdot,\cdot) : v \in V\}$, V is a closed set of parameters, and a probability distribution Q of V is known. If Q is unknown, we may assume that Q is subject to the equal probability distribution of V if V is finite, or, to the uniform distribution of V if V is a continuous region. For example, let $p_v(x, y) = (\sqrt{2\pi}\sigma)^{-1}e^{-(y-x-\mu)^2/2\sigma^2}$, where $v = (\mu, \sigma)$, μ and σ denote the mean and standard deviation, respectively. We may take the set V satisfying $V = \{(\mu, \sigma) : |\mu| \le \mu_1, 0 < \sigma \le \sigma_1\}$, where μ_1 and σ_1 are known; that is, the domain V and the set of possible post-change distributions, that is, $\Gamma(V)$ and the uniform distribution Q of V are known.

Let

$$p_{v_1}(\cdot,\cdot) = \int_V p_v(\cdot,\cdot)\, dQ(v).$$

The function $p_{v_1}(\cdot,\cdot)$ can be considered as a special, known post-change transition probability density of E.

For every $n \geq 1$, the joint probability $\mathbf{P}(X_1 \leq x_1, \dots, X_n \leq x_n) = F(x_1, \dots, x_n)$ can be written as

$$\begin{aligned} F(x_1, \dots, x_n) &= \sum_{k=1}^{n} \rho_k \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_{k-1}} \cdots \int_{-\infty}^{x_n} \prod_{j=1}^{k-1} p_{v_0}(y_{j-1}, y_j) \\ &\quad \times \prod_{j=k}^{n} p_{v_1}(y_{j-1}, y_j)\, dy_1 \cdots dy_{k-1} \cdots dy_n \\ &\quad + \left(1 - \sum_{k=1}^{n} \rho_k\right) F_\infty(x_1, \dots, x_n), \end{aligned}$$

where

$$F_\infty(x_1, \dots, x_n) = \int_{\infty}^{x_1} \cdots \int_{\infty}^{x_n} \prod_{j=1}^{n} p_{v_0}(y_{j-1}, y_j)\, dy_1 \cdots dy_n$$

and $y_0 = x_0$.

When the observations $X_1, \dots, X_n, \dots$ are independent with the probability density p_{v_0} before the change-point and the probability density p_{v_1} after the change-point, the joint probability $\mathbf{P}(X_1 \leq x_1, \dots, X_n \leq x_n) = F(x_1, \dots, x_n)$ can be written as

$$\begin{aligned} F(x_1, \dots, x_n) &= \sum_{k=1}^{n} \rho_k F_\infty(x_1, \dots, x_{k-1}) F_{v_1}(x_k, \dots, x_n) \\ &\quad + \left(1 - \sum_{k=1}^{n} \rho_k\right) F_\infty(x_1, \dots, x_n), \end{aligned}$$

where $F_\infty(x_1, \dots, x_k) = \prod_{j=1}^{k} F_\infty(x_k)$, $F_{v_1}(x_1, \dots, x_k) = \prod_{j=1}^{k} F_{v_1}(x_k)$ and the probability density functions of $F_\infty(x_k)$ and $F_{v_1}(x_k)$ are p_{v_0} and p_{v_1}, respectively.

For simplicity, we assume that $X_0 \equiv x_0 \in E$, $p_{v_1}(x, \cdot)/p_{v_0}(x, \cdot)$ has no atoms with respect to $\mathbf{P}_\infty$ for all $x \in E$, $p_{v_1}(x, y)/p_{v_0}(x, y) = 1$ when $p_{v_1}(x, y) = p_{v_0}(x, y) = 0$ and

$$p_{v_0}(x, y) > 0 \quad \text{if and only if} \quad p_{v_1}(x, y) > 0$$

for all $x, y \in E$.

In order to derive the optimal detection test, we need the following lemma which is a generalization of Theorem 1 given by Shiryaev and Zryumov in [14].

LEMMA 1. *For any finite stopping time T, that is, $\mathbf{P}_\infty(T < \infty) = 1$,*

$$\sum_{k=1}^{\infty} \rho_k \mathbf{E}_k (T-k+1)^+ = \mathbf{E}_\infty\left(\sum_{k=1}^{T}[Y_{k-1}+\rho_k]\right)$$

$$= \sum_{k=1}^{\infty} \mathbf{E}_\infty\big(Y_k I(T \ge k+1)\big) + \sum_{k=1}^{\infty} \rho_k \mathbf{P}_\infty(T \ge k), \tag{10}$$

where $I(\cdot)$ is the indicator function, the random process $\{Y_k, k \ge 0\}$ satisfies the recurrent equations:

$$Y_k = (Y_{k-1}+\rho_k)\Lambda_k, \qquad Y_0 = 0 \tag{11}$$

and $\Lambda_k = p_{v_1}(X_{k-1}, X_k)/p_{v_0}(X_{k-1}, X_k)$ for $k \ge 1$.

PROOF. Equation (10) can be obtained by way of proving Theorem 1 in [14]. Since

$$(T-k+1)^+ = \sum_{m=1}^{\infty} I(T-k \ge m) = \sum_{m=k}^{\infty} I(T \ge m),$$

where $T \ge m \in \mathfrak{F}_{m-1}$ and $m \ge k$, it follows that

$$\begin{aligned}
&\sum_{m=k}^{\infty} \mathbf{E}_k\big(I(T \ge m)\big)\\
&\quad = \sum_{m=k}^{\infty} \mathbf{E}_k\left(I(T \ge m)\prod_{j=1}^{k-1} p_{v_0}(X_{j-1}, X_j)\prod_{j=k}^{m-1} p_{v_1}(X_{j-1}, X_j)\right)\\
&\quad = \sum_{m=k}^{\infty} \mathbf{E}_\infty\left(I(T \ge m)\frac{\prod_{j=1}^{k-1} p_{v_0}(X_{j-1}, X_j)\prod_{j=k}^{m-1} p_{v_1}(X_{j-1}, X_j)}{\prod_{j=1}^{m-1} p_{v_0}(X_{j-1}, X_j)}\right)\\
&\quad = \sum_{m=k}^{\infty} \mathbf{E}_\infty\left(I(T \ge m)\frac{L_{m-1}}{L_{k-1}}\right),
\end{aligned}$$

where $\prod_{j=k}^{k-1} p_{v_1}(X_{j-1}, X_j) = 1$ and

$$L_n = \prod_{k=1}^{n} \frac{p_{v_1}(X_{k-1}, X_k)}{p_{v_0}(X_{k-1}, X_k)}, \qquad L_0 = 1.$$

Thus,

$$\begin{aligned}\sum_{k=1}^{\infty}\rho_k\mathbf{E}_k(T-k+1)^+ &= \mathbf{E}_\infty\left(\sum_{k=1}^{\infty}\rho_k\sum_{m=k}^{\infty}I(T\ge m)\frac{L_{m-1}}{L_{k-1}}\right)\\ &= \mathbf{E}_\infty\left(\sum_{k=1}^{\infty}\rho_k\sum_{m=k}^{T}\frac{L_{m-1}}{L_{k-1}}\right)\\ &= \mathbf{E}_\infty\left(\sum_{m=1}^{T}\sum_{k=1}^{m}\rho_k\frac{L_{m-1}}{L_{k-1}}\right) = \mathbf{E}_\infty\left(\sum_{m=1}^{T}[Y_{m-1}+\rho_m]\right).\end{aligned}$$

Furthermore, we have

$$\begin{aligned}\mathbf{E}_\infty\left(\sum_{m=1}^{T}[Y_{m-1}+\rho_m]\right) &= \mathbf{E}_\infty\left(\sum_{k=1}^{\infty}I(T=k)\left(\sum_{m=1}^{k}[Y_{m-1}+\rho_m]\right)\right)\\ &= \sum_{k=1}^{\infty}\mathbf{E}_\infty\big(Y_kI(T\ge k+1)\big)+\sum_{k=1}^{\infty}\rho_k\mathbf{P}_\infty(T\ge k).\end{aligned}$$

□

It follows from (1) and (10) that

$$\mathcal{J}_S(T)=1+\widetilde{\mathcal{J}}_S(T) \tag{12}$$

for any test (stopping time) T, where

$$\widetilde{\mathcal{J}}_S(T)=\frac{\sum_{k=1}^{\infty}\mathbf{E}_\infty(Y_kI(T\ge k+1))}{\sum_{k=1}^{\infty}\rho_k\mathbf{P}_\infty(T\ge k)}. \tag{13}$$

By (11), we know that $Y_n-Y_{n-1}-\rho_n=(\Lambda_n-1)(Y_{n-1}+\rho_n)$. Moreover,

$$\mathbf{E}_\infty(\Lambda_n|\mathfrak{F}_{n-1})=\mathbf{E}_\infty\left(\frac{p_{v_1}(X_{n-1},X_n)}{p_{v_0}(X_{n-1},X_n)}\Big|X_{n-1}\right)=1.$$

It follows that

$$\mathbf{E}_\infty\left(Y_n-\sum_{k=1}^{n}\rho_k\Big|\mathfrak{F}_{n-1}\right)=Y_{n-1}-\sum_{k=1}^{n-1}\rho_k,$$

that is, $\{Y_n-\sum_{k=1}^{n}\rho_k, n\ge 0\}$ is a martingales with $\mathbf{E}_\infty(Y_n)=\sum_{k=1}^{n}\rho_k$. By the optional sampling theorem of martingale (see [4], page 333) we can obtain the following lemma.

LEMMA 2. *For any finite stopping time T, if*

$$\mathbf{E}_\infty\left(\left|Y_T-\sum_{k=1}^{T}\rho_k\right|\right)<\infty,\qquad \lim_{n\to\infty}\mathbf{E}_\infty\left(\left|Y_n-\sum_{k=1}^{n}\rho_k\right|;T>n\right)=0, \tag{14}$$

then

$$\mathbf{E}_\infty(Y_T) = \mathbf{E}_\infty\left(\sum_{k=1}^{T} \rho_k\right) = \sum_{k=1}^{\infty} \rho_k \mathbf{P}_\infty(T \geq k) = \mathbf{P}(T \geq \tau). \tag{15}$$

We present a nonnegative loss random variable of detection, ξ_n, at time n in the following:

$$\xi_n = \sum_{k=1}^{n} Y_{k-1} + 1 + c\left(1 - \sum_{k=1}^{n} \rho_k\right)$$

for $n \geq 1$, where $c > 0$ is a constant. By Lemma 1, we have

$$\begin{aligned}
\mathbf{E}_\infty(\xi_T) &= \mathbf{E}_\infty\left(\sum_{k=1}^{T}[Y_{k-1} + \rho_k]\right) + (c+1)\left[1 - \mathbf{E}_\infty\left(\sum_{k=1}^{T} \rho_k\right)\right] \\
&= \sum_{k=1}^{\infty} \rho_k \mathbf{E}_k(T - k + 1)^+ + (c+1)\left[1 - \sum_{k=1}^{\infty} \rho_k \mathbf{P}_\infty(T \geq k)\right] \qquad (16) \\
&= \mathbf{E}(T - \tau + 1 | T \geq \tau)\mathbf{P}(T \geq \tau) + (c+1)\mathbf{P}(T < \tau).
\end{aligned}$$

The last equality follows from (1) and (15). If $c \leq 0$, the value of ξ_n will be smallest when $n = 1$, that is, $\mathbf{E}_\infty(\xi_T)$ will be smallest when $T \equiv 1$. So, we assume that $c > 0$.

Now we present a test $T^*(c)$:

$$T^*(c) = \min\{n \geq 1 : Y_n \geq c_n(c)\rho_{n+1}\},$$

where $\{c_n = c_n(c), n \geq 1\}$ is a series of positive random variable, measurable with respect to $\mathfrak{F}_n$, and depends on $c > 0$ for every $n \geq 1$.

In the following theorem, we not only give the expression of c_n but also prove that the test $T^*(c^*)$ is optimal in the Bayesian formulation.

THEOREM 1. *Let $0 < \alpha < 1 - \rho_1$ and $\rho_k > 0$ for all $k \geq 1$. Then there exists a positive number c^* and a series of positive random variables $c_n(c^*)$ such that $\mathbf{P}_\tau(T^*(c^*) < \tau) = \alpha$,*

$$c_n(c^*) = c^* + \mathbf{E}_\infty\left(\sum_{m=n+1}^{\infty} \frac{\rho_{m+1}}{\rho_{n+1}} I(B_{m,n+1})[c^* - \tilde{Y}_m] \Big| \mathfrak{F}_n\right) \tag{17}$$

for $n \geq 1$ and the test $T^*(c^*)$ is optimal in the Bayesian sense

$$\inf_{T \in D_\alpha} \mathcal{J}_S(T) = \mathcal{J}_S(T^*(c^*)), \tag{18}$$

where $D_\alpha = \{T \geq 1 : \mathbf{P}_\tau(T < \tau) \leq \alpha\}$, $\tilde{Y}_m = Y_m/\rho_{m+1}$ and $B_{m,n} = \{\tilde{Y}_k \leq c_k(c^*), n \leq k \leq m\}$. Moreover,

$$\mathcal{J}_S(T^*(c^*)) \leq 1 + c^*(1 - \rho_1). \tag{19}$$

In particular, if $c_n = c_1$ is a constant for all $n \geq 1$, then

$$\frac{1-\alpha}{\mathbf{E}_\infty(\rho_{T^*})} - 1 < c_1 \leq \frac{1-\alpha}{\mathbf{E}_\infty(\rho_{T^*+1})}, \tag{20}$$

where $T^* = T^*(c^*)$,

PROOF. Let $\tilde{\xi}_n = -\xi_n$ and define a subset $\hat{D}_\alpha$ of D_α as

$$\hat{D}_\alpha = \{T \geq 1 : \mathbf{P}_\tau(T < \tau) \leq \alpha, |\mathbf{E}_\infty(\tilde{\xi}_T)| < \infty\}.$$

We first prove that (18) holds for $\hat{D}_\alpha$. To this end, it needs to prove the following inequality:

$$\mathbf{E}_\infty(\xi_T) \geq \mathbf{E}_\infty(\xi_{T^*}) \quad \text{or} \quad \mathbf{E}_\infty(\tilde{\xi}_{T^*}) \geq \mathbf{E}_\infty(\tilde{\xi}_T)$$

for any finite test T with $\mathbf{E}_\infty(\xi_T) < \infty$.

By Lemma 3.2 in [1] (page 52), we only need to prove that for each $n \geq 1$

$$\mathbf{E}_\infty(\tilde{\xi}_{T^*}|\mathfrak{F}_n) \geq \tilde{\xi}_n \qquad \text{on } \{T^* > n\} \tag{21}$$

and

$$\mathbf{E}_\infty(\tilde{\xi}_T|\mathfrak{F}_n) \leq \tilde{\xi}_n \qquad \text{on } \{T^* = n, T > n\}. \tag{22}$$

In fact, by (21) and (22) we have

$$\begin{aligned}
\mathbf{E}_\infty(\tilde{\xi}_{T^*}) &= \int_{\{T^* \leq T\}} \tilde{\xi}_{T^*} + \int_{\{T^* > T\}} \tilde{\xi}_{T^*} \\
&= \sum_{n=1}^\infty \int_{\{T^*=n\leq T\}} \tilde{\xi}_{T^*} + \sum_{n=1}^\infty \int_{\{T=n<T^*\}} \mathbf{E}_\infty(\tilde{\xi}_{T^*}|\mathfrak{F}_n) \\
&\geq \sum_{n=1}^\infty \int_{\{T^*=n\leq T\}} \mathbf{E}_\infty(\tilde{\xi}_T|\mathfrak{F}_n) + \sum_{n=1}^\infty \int_{\{T=n<T^*\}} \tilde{\xi}_n = \mathbf{E}_\infty(\tilde{\xi}_T).
\end{aligned}$$

For every $N \geq 2$, we define N positive random variables, $c_{N,N}, c_{N-1,N}, \ldots, c_{1,N}$ in the following:

$$\begin{aligned}
c_{N,N} &= c, \\
c_{n,N} &= c + \frac{\rho_{n+2}}{\rho_{n+1}} \mathbf{E}_\infty\big(I\big(B_{n+1,n+1}(N)\big)[c_{n+1,N} - \tilde{Y}_{n+1}]|\mathfrak{F}_n\big), \\
c_{n,N} &= c + \mathbf{E}_\infty\left(\sum_{m=n+1}^N \frac{\rho_{m+1}}{\rho_{n+1}} I\big(B_{m,n+1}(N)\big)[c - \tilde{Y}_m]\Big|\mathfrak{F}_n\right)
\end{aligned} \tag{23}$$

for $1 \leq n \leq N$, where $\tilde{Y}_m = Y_n/\rho_{m+1}$ and $B_{m,n+1}(N) = \{\tilde{Y}_k \leq c_{k,N}, n+1 \leq k \leq m \leq N\}$.

Next, we show that $c_{n,N} \geq c$ for all $1 \leq n \leq N$. Obviously, $c_{N,N} = c \geq c$. Assume that $c_{n,N} \geq c$ for $n \leq N$. Since

$$\begin{aligned} c_{n-1,N} &= c + \frac{\rho_{n+1}}{\rho_n} \mathbf{E}_\infty \Biggl(I\bigl(B_{n,n}(N)\bigr) \mathbf{E}_\infty \Biggl(\Biggl[c - \tilde{Y}_n \\ &\qquad + \sum_{m=n+1}^{N} \frac{\rho_{m+1}}{\rho_{n+1}} I\bigl(B_{m,n+1}(N)\bigr)(c - \tilde{Y}_m) \Biggr] \Bigg| \mathfrak{F}_n \Biggr) \Bigg| \mathfrak{F}_{n-1} \Biggr) \\ &= c + \frac{\rho_{n+1}}{\rho_n} \mathbf{E}_\infty \bigl(I\bigl(B_{n,n}(N)\bigr)[c_{n,N} - \tilde{Y}_n] | \mathfrak{F}_{n-1} \bigr) \end{aligned}$$

and $I(B_{n,n}(N))[c_{n,N} - \tilde{Y}_n] \geq 0$, it follows that $c_{n-1,N} \geq c$. By the mathematical induction, we know that $c_{n,N} \geq c$ for all $1 \leq n \leq N$.

On the other hand, $c_{n,N} \leq c(1 + \rho_{n+1}^{-1} \sum_{m=n+1}^{\infty} \rho_{m+1})$ for all $N \geq 2$. This means that there exists a sub-sequence $\{N_k\}$ such that $c_{n,N_k} \to c_n$ as $k \to \infty$ for every $n \geq 1$ and, therefore, $B_{m,n}(N_k) \to B_{m,n} = \{\tilde{Y}_j \leq c_j, n \leq j \leq m\}$ as $k \to \infty$, and

$$\begin{aligned} c_n &= c + \frac{\rho_{n+2}}{\rho_{n+1}} \mathbf{E}_\infty \bigl(I\bigl(B_{n+1,n+1}[c_{n+1} - \tilde{Y}_{n+1}] | \mathfrak{F}_n \bigr)\bigr) \\ &= c + \mathbf{E}_\infty \Biggl(\sum_{m=n+1}^{\infty} \frac{\rho_{m+1}}{\rho_{n+1}} I(B_{m,n+1})[c - \tilde{Y}_m] \Bigg| \mathfrak{F}_n \Biggr). \end{aligned}$$

This is (17).

Hence,

$$\begin{aligned} &I(T^* > n) \mathbf{E}_\infty \bigl((\tilde{\xi}_{T^*} - \tilde{\xi}_n) | \mathfrak{F}_n \bigr) \\ &\quad = I(T^* > n) \sum_{m=n}^{\infty} \mathbf{E}_\infty \bigl(\bigl(I(T^* > m)[\tilde{\xi}_{m+1} - \tilde{\xi}_m] \bigr) | \mathfrak{F}_n \bigr) \\ &\quad = -\rho_{n+1} I(T^* > n) \\ &\qquad \times \Biggl[\tilde{Y}_n - c + \sum_{m=n+1}^{\infty} \mathbf{E}_\infty \frac{\rho_{m+1}}{\rho_{n+1}} \bigl(I(B_{m,n+1})(\tilde{Y}_m - c) | \mathfrak{F}_n \bigr) \Biggr] \\ &\quad = -\rho_{n+1} I(T^* > n)(\tilde{Y}_n - c_n) > 0. \end{aligned}$$

That is, (21) holds.

Let $T_N = \min\{T, N+1\}$ for each finite stopping time T. We first show that

$$\begin{aligned} &\mathbf{E}_\infty \Biggl(\sum_{m=n+1}^{N} \frac{\rho_{m+1}}{\rho_{n+1}} I(T_N > m)[c - \tilde{Y}_m] \Bigg| \mathfrak{F}_n \Biggr) \\ &\qquad \leq (c_{n,N} - c) I(T_N > n) \end{aligned} \tag{24}$$

for $1 \le n \le N$. Obviously, (24) holds for $n = N$. Let $n = N - 1$. We have

$$\begin{aligned}
&\frac{\rho_{N+1}}{\rho_N}\mathbf{E}_\infty\big(I(T_N > N)[c - \tilde{Y}_N]|\mathfrak{F}_{N-1}\big)\\
&\quad \le \frac{\rho_{N+1}}{\rho_N}\mathbf{E}_\infty\big(I(T_N > N)I\big(B_{N,N}(N)\big)[c - \tilde{Y}_N]|\mathfrak{F}_{N-1}\big)\\
&\quad \le I(T_N > N-1)\frac{\rho_{N+1}}{\rho_N}\mathbf{E}_\infty\big(I\big(B_{N,N}(N)\big)[c - \tilde{Y}_N]|\mathfrak{F}_{N-1}\big)\\
&\quad = (c_{N-1,N} - c)I(T_N > N-1).
\end{aligned}$$

Assume that (24) holds for $n \le N - 1$. It follows that

$$\begin{aligned}
&\mathbf{E}_\infty\left(\sum_{m=n}^{N}\frac{\rho_{m+1}}{\rho_n}I(T_N > m)[c - \tilde{Y}_m]\Big|\mathfrak{F}_{n-1}\right)\\
&\quad = \frac{\rho_{n+1}}{\rho_n}\mathbf{E}_\infty\big(I(T_N > n)[c - \tilde{Y}_n]|\mathfrak{F}_{n-1}\big)\\
&\qquad + \frac{\rho_{n+1}}{\rho_n}\mathbf{E}_\infty\left(\mathbf{E}_\infty\left[\sum_{m=n+1}^{N}\frac{\rho_{m+1}}{\rho_{n+1}}I(T_N > m)(c - \tilde{Y}_m)\Big|\mathfrak{F}_n\right]\Big|\mathfrak{F}_{n-1}\right)\\
&\quad \le \frac{\rho_{n+1}}{\rho_n}\mathbf{E}_\infty\big(I(T_N > n)[c - \tilde{Y}_n]|\mathfrak{F}_{n-1}\big)\\
&\qquad + \frac{\rho_{n+1}}{\rho_n}\mathbf{E}_\infty\big((c_{n,N} - c)I(T_N > n)|\mathfrak{F}_{n-1}\big)\\
&\quad = \frac{\rho_{n+1}}{\rho_n}\mathbf{E}_\infty\big([c_{n,N} - \tilde{Y}_n]I(T_N > n)|\mathfrak{F}_{n-1}\big)\\
&\quad \le \frac{\rho_{n+1}}{\rho_n}\mathbf{E}_\infty\big(I\big(B_{n,n}(N)\big)[c_{n,N} - \tilde{Y}_n]I(T_N > n)|\mathfrak{F}_{n-1}\big)\\
&\quad \le I(T_N > n-1)(c_{n-1,N} - c).
\end{aligned}$$

By the mathematical induction, we know that (24) holds for all $1 \le n \le N$. Taking $N = N_k$ and let $k \to \infty$ in (24), we have

$$\mathbf{E}_\infty\left(\sum_{m=n+1}^{\infty}\frac{\rho_{m+1}}{\rho_{n+1}}I(T > m)[c - \tilde{Y}_m]\Big|\mathfrak{F}_n\right) \le (c_n - c)I(T > n) \tag{25}$$

for all $n \ge 1$. Thus,

$$\begin{aligned}
&I\big(T^* = n\big)I(T > n)\mathbf{E}_\infty\big((\tilde{\xi}_T - \tilde{\xi}_n)|\mathfrak{F}_n\big)\\
&\quad = I\big(T^* = n\big)\sum_{m=n}^{\infty}\mathbf{E}_\infty\big(\big(I(T > m)[\tilde{\xi}_{m+1} - \tilde{\xi}_m]\big)|\mathfrak{F}_n\big)\\
&\quad = -\rho_{n+1}I\big(T^* = n\big)\Big[\, I(T > n)(\tilde{Y}_n - c)
\end{aligned}$$

$$+ \sum_{m=n+1}^{\infty} \mathbf{E}_\infty\left(\frac{\rho_{m+1}}{\rho_{n+1}} I(T > m)(\tilde{Y}_m - c)\Big|\mathfrak{F}_n\right)\Bigg]$$
$$= \rho_{n+1} I\big(T^* = n\big)\Bigg[-I(T > n)\tilde{Y}_n + I(T > n)c$$
$$+ \sum_{m=n+1}^{\infty} \mathbf{E}_\infty\left(\frac{\rho_{m+1}}{\rho_{n+1}} I(T > m)(c - \tilde{Y}_m)\Big|\mathfrak{F}_n\right)\Bigg]$$
$$\le \rho_{n+1} I\big(T^* = n\big)\big[-I(T > n)\tilde{Y}_n + I(T > n)c_n\big]$$
$$= \rho_{n+1} I\big(T^* = n\big) I(T > n)[c_n - \tilde{Y}_n] \le 0$$

since $c_n \le \tilde{Y}_n$ when $T^* = n$. That is, (22) holds for all $n \ge 1$.

From (21) and (22), it follows that $\mathbf{E}_\infty(\xi_T) \ge \mathbf{E}_\infty(\xi_{T^*})$ for any finite test T with $\mathbf{E}_\infty(\xi_T) < \infty$.

Since the probability

$$\mathbf{P}\big(T^*(c) \ge \tau\big) = \rho_1 + \sum_{k=2}^{\infty} \rho_k \mathbf{P}_\infty\big(T^*(c) \ge 2\big)$$

is continuous and strictly increasing in c, $\mathbf{P}(T^*(0) \ge \tau) = \rho_1 < 1-\alpha$ for $c = 0$ and $\mathbf{P}(T^*(c) \ge \tau) > 1 - \alpha$ for a large c, it follows that there exists a positive number c^* such that $\mathbf{P}(T^*(c^*) \ge \tau) = 1 - \alpha$ or $\mathbf{P}(T^*(c^*) < \tau) = \alpha$.

It follows from (16) and $\mathbf{E}_\infty(\xi_T) \ge \mathbf{E}_\infty(\xi_{T^*(c^*)})$ that

$$\begin{aligned}
&\mathbf{E}\big((T - \tau + 1)|T \ge \tau\big)\mathbf{P}(T \ge \tau) + \big(1 + c^*\big)\mathbf{P}(T < \tau) \\
&\quad \ge (1 - \alpha)\mathbf{E}\big(\big(T^*\big(c^*\big) - \tau + 1\big)|T^*\big(c^*\big) \ge \tau\big) + \big(1 + c^*\big)\alpha.
\end{aligned}$$

Furthermore, by (12), (13) and (21) we have

$$\begin{aligned}
\mathcal{J}_S\big(T^*\big(c^*\big)\big) &= \mathbf{E}\big(\big(T^*\big(c^*\big) - \tau + 1\big)|T^*\big(c^*\big) \ge \tau\big) \\
&= 1 + \frac{\sum_{k=1}^{\infty} \mathbf{E}_\infty(Y_k I(T^*(c^*) \ge k + 1))}{\sum_{k=1}^{\infty} \rho_k \mathbf{P}_\infty(T^*(c^*) \ge k)} \\
&= 1 + \frac{\sum_{k=1}^{\infty} \mathbf{E}_\infty((Y_k - c^*\rho_{k+1} + c^*\rho_{k+1}) I(T^*(c^*) \ge k + 1))}{\sum_{k=1}^{\infty} \rho_k \mathbf{P}_\infty(T^*(c^*) \ge k)} \\
&\le 1 + \frac{\sum_{k=1}^{\infty} c^*\rho_{k+1} \mathbf{P}_\infty(T^*(c^*) \ge k + 1)}{\sum_{k=1}^{\infty} \rho_k \mathbf{P}_\infty(T^*(c^*) \ge k)} \\
&= 1 + c^*\left(1 - \frac{\rho_1}{\sum_{k=1}^{\infty} \rho_k \mathbf{P}_\infty(T^*(c^*) \ge k)}\right) \le 1 + c^*(1 - \rho_1).
\end{aligned}$$

This is (19). Thus, if $\mathbf{E}((T - \tau + 1)|T \ge \tau) > 1 + c^*$, then

$$\mathbf{E}\big((T - \tau + 1)|T \ge \tau\big) > 1 + c^* \ge \mathbf{E}\big(\big(T^*\big(c^*\big) - \tau + 1\big)|T^*\big(c^*\big) \ge \tau\big).$$

If $\mathbf{E}((T-\tau+1)|T \geq \tau) < 1+c^*$, then $\mathbf{E}((T-\tau+1)|T \geq \tau)(1-x)+(1+c^*)x$ is increasing in $x = \mathbf{P}(T < \tau)$ and, therefore,

$$\begin{aligned}(1-\alpha)&\mathbf{E}\big((T-\tau+1)|T \geq \tau\big)+\big(1+c^*\big)\alpha \\ &\geq \mathbf{E}\big((T-\tau+1)|T \geq \tau\big)\big(1-\mathbf{P}(T<\tau)\big)+\big(1+c^*\big)\mathbf{P}(T<\tau) \\ &\geq (1-\alpha)\mathbf{E}\big(\big(T^*\big(c^*\big)-\tau+1\big)|T^*\big(c^*\big) \geq \tau\big)+\big(1+c^*\big)\alpha\end{aligned}$$

since $x = \mathbf{P}(T < \tau) \leq \alpha$. This means that

$$\mathbf{E}\big((T-\tau+1)|T \geq \tau\big) \geq \mathbf{E}\big(\big(T^*\big(c^*\big)-\tau+1\big)|T^*\big(c^*\big) \geq \tau\big).$$

Thus, (18) holds for $\hat{D}_\alpha$.

In order to prove that (18) also holds for D_α, we define a set of stopping times as

$$D_\alpha(M) = \big\{T_M = T \wedge M : \mathbf{P}_\tau(T < \tau) \leq \alpha\big\}$$

for every positive integer $M > 0$. Obviously, $D_\alpha(M) \subset \hat{D}_\alpha \subset D_\alpha$ and $\lim_{M\to\infty} D_\alpha(M) = D_\alpha$. This means that (18) holds for D_α.

Let $c_n = c_1$ is a positive constant for all $n \geq 1$. In order to estimate c_1 we will check that $T^* = T^*(c^*)$ satisfies the condition of Lemma 2.

Note that $|Y_n - Y_{n-1} - \rho_n| \leq (\Lambda_n + 1)(Y_{n-1} + \rho_n)$, $I(T^* \geq n) \in \mathfrak{F}_{n-1}$ and $\mathbf{E}_\infty(\Lambda_n|\mathfrak{F}_{n-1}) = 1$. We have

$$\begin{aligned}\mathbf{E}_\infty&\left(\left|Y_{T^*} - \sum_{k=1}^{T^*}\rho_k\right|\right) \\ &\leq \mathbf{E}_\infty\left(\sum_{n=1}^{T^*}|Y_n - Y_{n-1} - \rho_n|\right) \\ &\leq \mathbf{E}_\infty\left(\sum_{n=1}^{T^*}(\Lambda_n+1)(Y_{n-1}+\rho_n)\right) \\ &= \sum_{n=1}^{\infty}\mathbf{E}_\infty\big((\Lambda_n+1)(Y_{n-1}+\rho_n)I\big(T^* \geq n\big)\big) \\ &= \mathbf{E}_\infty\left(\sum_{n=1}^{\infty}I\big(T^* \geq n\big)\mathbf{E}_\infty\big((\Lambda_n+1)(Y_{n-1}+\rho_n)|\mathfrak{F}_{n-1}\big)\right) \\ &= 2\mathbf{E}_\infty\left(\sum_{n=1}^{\infty}(Y_{n-1}+\rho_n)I\big(T^* \geq n\big)\right) \\ &\leq 2(c_1+1)\sum_{n=1}^{\infty}\rho_n\mathbf{P}_\infty\big(T^* \geq n\big) < \infty,\end{aligned}$$

where the last inequality comes from $Y_{n-1}I(T^* \ge n) \le c_1\rho_n I(T^* \ge n)$. On the other hand,

$$\begin{aligned}\lim_{n\to\infty}\mathbf{E}_\infty\left(\left|Y_n-\sum_{k=1}^n\rho_k\right|, T^*>n\right)\\ \le \lim_{n\to\infty}\mathbf{E}_\infty\left(\sum_{k=1}^{T^*}(\Lambda_k+1)(Y_{k-1}+\rho_k), T^*>n\right)=0.\end{aligned}$$

This means that condition (14) of Lemma 2 holds. By the definition of T^*, we have $c_1\rho_T^* > Y_{T^*-1}$ and $Y_{T^*} \ge c_1\rho_{T^*+1}$. From Lemma 2, it follows that

$$\frac{1-\alpha}{\mathbf{E}_\infty(\rho_{T^*})}-1=\frac{\mathbf{E}_\infty(Y_{T^*-1})}{\mathbf{E}_\infty(\rho_{T^*})}<c_1\le\frac{\mathbf{E}_\infty(Y_{T^*})}{\mathbf{E}_\infty(\rho_{T^*+1})}=\frac{1-\alpha}{\mathbf{E}_\infty(\rho_{T^*+1})}. \qquad \square$$

REMARK 1. When the observation processes $\{X_n, n\ge 0\}$ are mutually independent, we know that $\{\tilde{Y}_m, m\ge 0\}$ is a Markov process. In this case, the positive random variables $\{c_n\}$ become a series of positive numbers which can be written by

$$\begin{aligned}c_n&=c+\frac{\rho_{n+2}}{\rho_{n+1}}\mathbf{E}_\infty\big(I(B_{n+1,n+1})[c_{n+1}-\tilde{Y}_{n+1}]|\tilde{Y}_n=c_n\big)\\&=c+\mathbf{E}_\infty\left(\sum_{m=n+1}^\infty\frac{\rho_{m+1}}{\rho_{n+1}}I(B_{m,n+1})[c-\tilde{Y}_m]\middle|\tilde{Y}_n=c_n\right).\end{aligned}$$

In fact, $\tilde{Y}_n \ge c+\frac{\rho_{n+2}}{\rho_{n+1}}\mathbf{E}_\infty(I(B_{n+1,n+1})[c_{n+1}-\tilde{Y}_{n+1}]|\tilde{Y}_n)$ if and only if $\tilde{Y}_n\ge c_n$ for $n\ge 1$.

Though it is very difficult to find the exact value of c^* and $c_n(c^*)$ in Theorem 1, as it depends on α, $p_{v_0}(\cdot,\cdot)$, $p_{v_1}(\cdot,\cdot)$ and $\{\rho_k\}$, one may estimate its upper and lower bounds by calculating the two numbers $\mathbf{E}_\infty(\rho_{T^*})$ and $\mathbf{E}_\infty(\rho_{T^*+1})$. In fact, we can obtain an approximate value $\tilde{c}^*$ of c^* by numerically calculating the expectation $\mathbf{E}_\infty(\sum_{k=1}^{T^*(\tilde{c}^*)}\rho_k)$ such that $1-\alpha=\mathbf{P}(T^*(\tilde{c}^*)\ge\tau)=\mathbf{E}_\infty(\sum_{k=1}^{T^*(\tilde{c}^*)}\rho_k)$ for a given $\alpha=\mathbf{P}(T^*(\tilde{c}^*)<\tau)$. Especially, when the observation processes $\{X_n, n\ge 0\}$ are mutually independent and $\rho_k=(1-\rho)\rho^{k-1}, k\ge 1$, we know that $\{\tilde{Y}_n, n\ge 1\}$ is a time homogeneous Markov process. Hence, $c_n=c_1$ is a constant and

$$\begin{aligned}c_1=c_n&=c+\rho\mathbf{E}_\infty\big(I(B_{n+1,n+1})[c_{n+1}-\tilde{Y}_{n+1}]|\tilde{Y}_n=c_n\big)\\&=c+\mathbf{E}_\infty\left(\sum_{m=1}^\infty\rho^m I(B_{m+1,2})[c-\tilde{Y}_{m+1}]\middle|\tilde{Y}_1=c_1\right)\end{aligned}$$

for all $n\ge 1$. Note that $1-\alpha=1-\mathbf{E}_\infty((1-\rho)^{T^*})$ and

$$\mathbf{E}_\infty(\rho_{T^*+1})=\sum_{k=1}^\infty\rho(1-\rho)^k\mathbf{P}_\infty(T^*=k)=\rho\alpha, \qquad \mathbf{E}_\infty(\rho_{T^*})=\frac{\rho\alpha}{1-\rho}.$$

It follows from Theorem 1 that

$$\frac{1-\alpha}{\mathbf{E}_\infty(\rho_{T^*})} - 1 = \frac{1-\alpha-\rho}{\rho\alpha} < c_1 \le \frac{1-\alpha}{\rho\alpha}.$$

For example, let $\alpha = 0.05$, $\rho = 0.7$, $\alpha = 0.1$, $\rho = 0.6$ and $\alpha = 0.2$, $\rho = 0.5$, we can obtain the estimations c_1', c_1'' and c_1''', respectively, in the following:

$$\frac{1-0.05-0.7}{0.05\times 0.7} \approx 7.14 < c_1' < 27.14 \approx \frac{1-0.05}{0.05\times 0.7},$$
$$\frac{1-0.1-0.6}{0.1\times 0.6} = 5 < c_1'' < 15 = \frac{1-0.1}{0.1\times 0.6},$$
$$\frac{1-0.2-0.5}{0.2\times 0.5} = 3 < c_1''' < 8 = \frac{1-0.2}{0.2\times 0.5}.$$

On the other hand, let $p_{v_0}(\cdot)$ and $p_{v_1}(x)$ be two densities of normal distributions $N(0,1)$ and $N(1,1)$, respectively. We can obtain the numerical simulation values $c_1' \approx 10.885$, $c_1'' \approx 6.301$ and $c_1''' \approx 3.352$ by running 10^6 repetitions.

Now we consider the optimal test with the restrictive condition $\mathbf{E}_\infty(T) \ge \gamma > 1$. Consider the loss random variable of detection, ϖ_n, at time n in the following:

$$\varpi_n = \sum_{k=1}^n Y_{k-1} + 1 + c\left(1 - \sum_{k=1}^n \rho_k\right) - bn, \tag{26}$$

where c, b are two nonnegative constants satisfying $c + b > 0$. Note that ϖ_n may be negative. By Lemma 1, we have

$$\begin{aligned}
&\mathbf{E}_\infty(\varpi_T)\\
&\quad= \mathbf{E}_\infty\left(\sum_{k=1}^T [Y_{k-1} + \rho_k]\right) + (c+1)\left[1 - \mathbf{E}_\infty\left(\sum_{k=1}^T \rho_k\right)\right] - b\mathbf{E}_\infty(T)\\
&\quad= \sum_{k=1}^\infty \rho_k \mathbf{E}_k(T-k+1)^+ + (c+1)\left[1 - \sum_{k=1}^\infty \rho_k \mathbf{P}_\infty(T \ge k)\right] - b\mathbf{E}_\infty(T)\\
&\quad= \mathbf{E}(T-\tau+1 | T \ge \tau)\mathbf{P}(T \ge \tau) + (c+1)\mathbf{P}(T < \tau) - b\mathbf{E}_\infty(T).
\end{aligned}\tag{27}$$

We present a test:

$$T_\gamma(c,b) = \min\{n \ge 1 : Y_n \ge c_n(c,b)\rho_{n+1} + b\}, \tag{28}$$

where $\{c_n(c,b), n \ge 1\}$ is a series of positive random variables satisfying $c_n(c,b) \in \mathfrak{F}_n$. Obviously, $T_\gamma(c,0) = T^*(c)$ when $b = 0$.

THEOREM 2. *Let $\rho_k > 0$ for all $k \ge 1$ and $\mathbf{E}_\infty(T_\gamma(c,b_0)) < \infty$ for $c > 0$ and some number $b_0 \ge 0$. Then there exist two numbers b^*, c^* and a series of positive* *random variables* $c_n(c^*, b^*)$ *such that* $\mathbf{E}_\infty(T_\gamma(c^*, b^*)) = \gamma$, $b^* \le \rho_1 \min\{c^*/(\gamma - 1), b_0, 1\}$,

$$c_n(c^*, b^*) = c^* + b^*/\rho_{n+1} + \mathbf{E}_\infty\left(\sum_{m=n+1}^{\infty} \frac{\rho_{m+1}}{\rho_{n+1}} I(B_{m,n+1})\big[c^* + b^*/\rho_{m+1} - \tilde{Y}_m\big] \Big| \mathfrak{F}_n\right)$$

for $n \ge 1$, *and* $T_\gamma(c^*, b^*)$ *is optimal in the following sense*:

$$\inf_{T \in D_\gamma} \mathcal{J}_S(T) = \mathcal{J}_S(T_\gamma(c^*, b^*)), \tag{29}$$

where

$$D_\gamma = \{T \ge 1 : \mathbf{E}_\infty(T) \ge \gamma; \mathbf{P}(T < \tau) \le \alpha^*\},$$

$B_{m,n+1} = \{\tilde{Y}_k \le c_k(c^*, b^*) + b^*/\rho_{n+1}, n+1 \le k \le m\}$ *and* $\alpha^* = \mathbf{P}(T_\gamma(c^*, b^*) < \tau)$. *Moreover*,

$$\mathcal{J}_S(T_\gamma(c^*, b^*)) \le 1 + c^*(1 - \rho_1) + b^*(\gamma - 1). \tag{30}$$

PROOF. It needs only to be proved that (29) holds for the following subset $\hat{D}_\gamma$ of D_γ:

$$\hat{D}_\gamma = \{T \ge 1 : \mathbf{E}_\infty(T) \ge \gamma; \mathbf{P}(T < \tau) \le \alpha^*; \mathbf{E}_\infty(|\varpi_T|) < \infty\}.$$

By the same way of proving Theorem 1, we can show that

$$\mathbf{E}_\infty(\varpi_T) \ge \mathbf{E}_\infty(\varpi_{T_\gamma(c,b)}) \tag{31}$$

for any finite test T and $\mathbf{E}_\infty(|\varpi_T|) < \infty$.

On the other hand, we know that $\mathbf{E}_\infty(T_\gamma(c, b))$ is continuous and strictly increasing in c and b with $\mathbf{E}_\infty(T_\gamma(0, 0)) = 1$ and

$$\mathbf{E}_\infty(T_\gamma(c, b)) \nearrow \infty$$

as $c \nearrow \infty$ for every fixed $b \ge 0$. This means that for every small $b^* \le b_0$, there exists a unique positive number $c^* > 0$ such that

$$\mathbf{E}_\infty(T_\gamma(c^*, b^*)) = \gamma.$$

Obviously, the number $c^* = c^*(b^*)$ is monotonically decreasing in b^*. Hence, we can take a small b^* such that $b^* \le \rho_1 \min\{c^*/(\gamma - 1), b_0, 1\}$ and, therefore, $b^* \le \rho_1 \le 1 - \alpha^*$.

Let $T^* = T_\gamma(c^*, b^*)$. It follows from (26), (27) and (31) that

$$\begin{aligned} &\mathbf{E}(T - \tau + 1 | T \ge \tau)\mathbf{P}(T \ge \tau) + (1 + c^*)\mathbf{P}(T < \tau) - b\mathbf{E}_\infty(T) \\ &\quad \ge \mathbf{E}(T^* - \tau + 1 | T^* \ge \tau)\mathbf{P}(T^* \ge \tau) + (c^* + 1)\mathbf{P}(T^* < \tau) - b\gamma \end{aligned}$$

and, therefore,

$$
\begin{aligned}
&\mathbf{E}(T-\tau+1|T\geq\tau)\mathbf{P}(T\geq\tau)+\bigl(1+c^*\bigr)\mathbf{P}(T<\tau)\\
&\qquad\geq\mathbf{E}\bigl(T^*-\tau+1|T^*\geq\tau\bigr)\mathbf{P}\bigl(T^*\geq\tau\bigr)+\bigl(c^*+1\bigr)\mathbf{P}\bigl(T^*<\tau\bigr)
\end{aligned}\tag{32}
$$

since $\mathbf{E}_\infty(T)\geq\gamma$.

Furthermore,

$$
\begin{aligned}
\mathbf{E}\bigl(T^*-\tau+1|T^*\geq\tau\bigr)&=1+\frac{\sum_{k=1}^\infty\mathbf{E}_\infty(Y_kI(T^*\geq k+1))}{\sum_{k=1}^\infty\rho_k\mathbf{P}_\infty(T^*\geq k)}\\
&\leq 1+\frac{\sum_{k=1}^\infty(c^*\rho_{k+1}+b^*)\mathbf{P}_\infty(T^*\geq k+1)}{\sum_{k=1}^\infty\rho_k\mathbf{P}_\infty(T^*\geq k)}\\
&=1+c^*-\frac{c^*\rho_1-b^*(\gamma-1)}{\sum_{k=1}^\infty\rho_k\mathbf{P}_\infty(T^*\geq k)}\\
&\leq 1+c^*-\bigl(c^*\rho_1-b^*(\gamma-1)\bigr).
\end{aligned}
$$

This is (30). If $\mathbf{E}(T-\tau+1|T\geq\tau)>1+c^*$, then

$$\mathbf{E}(T-\tau+1|T\geq\tau)>1+c^*\geq\mathbf{E}\bigl(T^*-\tau+1|T^*\geq\tau\bigr).$$

Let $\mathbf{E}(T-\tau+1|T\geq\tau)\leq 1+c^*$. Since

$$\mathbf{E}(T-\tau+1|T\geq\tau)(1-x)+\bigl(1+c^*\bigr)x$$

is monotonically increasing in x and $\mathbf{P}(T<\tau)\leq\mathbf{P}(T^*<\tau)$, it follows from (32) that

$$
\begin{aligned}
&\mathbf{E}(T-\tau+1|T\geq\tau)\mathbf{P}\bigl(T^*\geq\tau\bigr)+\bigl(1+c^*\bigr)\mathbf{P}(T^*<\tau)\\
&\qquad\geq\mathbf{E}(T-\tau+1|T\geq\tau)\mathbf{P}(T\geq\tau)+\bigl(1+c^*\bigr)\mathbf{P}(T<\tau)\\
&\qquad\geq\mathbf{E}\bigl(T^*-\tau+1|T^*\geq\tau\bigr)\mathbf{P}\bigl(T^*\geq\tau\bigr)+\bigl(c^*+1\bigr)\mathbf{P}\bigl(T^*<\tau\bigr)
\end{aligned}
$$

and, therefore,

$$\mathbf{E}(T-\tau+1|T\geq\tau)\geq\mathbf{E}\bigl(T^*-\tau+1|T^*\geq\tau\bigr).$$

Thus, (29) is true. □

By using Theorem 2, we can obtain the following corollary.

COROLLARY 1. *Let $\beta=\{\beta_k,k\geq1\}$ satisfying $0<\beta_{k+1}\leq\beta_k, k\sqrt{\beta_{k+1}}\geq(k-1)\sqrt{\beta_k}$ for $k\geq1$ and $\sum_{k=1}^\infty\beta_k=\infty$. Then the following stopping time*:

$$T_\gamma(\beta)=\min\left\{n\geq1:\sum_{k=1}^n\frac{\beta_k}{\beta_{n+1}}\prod_{i=k}^n\frac{p_{v_1}(X_{i-1},X_i)}{p_{v_0}(X_{i-1},X_i)}\geq c_\beta(\gamma)\right\}$$

is optimal in the sense that

$$\frac{\sum_{k=1}^{\infty}\beta_k\mathbf{E}_k(T_\gamma(\beta)-k+1)^+}{\sum_{k=1}^{\infty}\beta_k\mathbf{P}_k(T_\gamma(\beta)\geq k)}=\inf_{T\in\Delta_\gamma}\frac{\sum_{k=1}^{\infty}\beta_k\mathbf{E}_k(T-k+1)^+}{\sum_{k=1}^{\infty}\beta_k\mathbf{P}_k(T\geq k)}$$

or equivalently,

$$\frac{\sum_{k=1}^{\infty}\beta_k\mathbf{E}_k(T_\gamma(\beta)-k)^+}{\sum_{k=1}^{\infty}\beta_k\mathbf{P}_k(T_\gamma(\beta)\geq k)}=\inf_{T\in\Delta_\gamma}\frac{\sum_{k=1}^{\infty}\beta_k\mathbf{E}_k(T-k)^+}{\sum_{k=1}^{\infty}\beta_k\mathbf{P}_k(T\geq k)},$$

where $c_\beta(\gamma)$ *is such that* $\mathbf{E}_\infty(T_\gamma(\beta))=\gamma$ *and* $\Delta_\gamma=\{T:\mathbf{E}_\infty(T)\geq\gamma\}$.

PROOF. Let $\rho_k=\rho(a)(a(k-1)+\beta_k^{-1/2})^{-2}$ for $k\geq 1$, where $\rho(a)=[\sum_{k=1}^{\infty}(a(k-1)+\beta_k^{-1/2})^{-2}]^{-1}$ and $a>0$. Note that $\beta_k\geq\beta_{k+1}>0$ for $k\geq 1$ and $\sum_{k=1}^{\infty}\beta_k=\infty$. It follows that $\rho_k\geq\rho_{k+1}>0$ for $k\geq 1$, $\lim_{a\to 0}\rho(a)=0$ and, therefore, $\lim_{a\to 0}\rho_k=0$ for every $k\geq 1$.

Taking $b^*=0$ in Theorem 2, we have

$$\inf_{T\in D_\gamma}\mathcal{J}_S(T)=\mathcal{J}_S(T_\gamma(c^*,0)),$$

where $\mathbf{E}_\infty(T_\gamma(c^*,0))=\gamma$ and $c^*=c_a$ depends on a. Note that

$$Y_n/\rho_{n+1}=\sum_{t=1}^{n}\frac{\rho_t}{\rho_{n+1}}\prod_{k=t}^{n}\Lambda_k$$

for every $n\geq 1$, where $\Lambda_k=p_{v_1}(X_{k-1},X_k)/p_{v_0}(X_{k-1},X_k)$, and $\rho_t/\rho_{n+1}\to\beta_t/\beta_{n+1}$ as $a\to 0$. It follows that

$$\lim_{\alpha\to 0}Y_n/\rho_{n+1}=\sum_{t=1}^{n}\frac{\beta_t}{\beta_{n+1}}\prod_{k=t}^{n}\Lambda_k.$$

Moreover,

$$\frac{\rho_t}{\rho_{n+1}}=\frac{(an+\beta_{n+1}^{-1/2})^2}{(a(t-1)+\beta_t^{-1/2})^2}$$

is monotonically increasing in a for $t<n+1$ since $k\sqrt{\beta_{k+1}}\geq(k-1)\sqrt{\beta_k}$ for $k\geq 1$ and, therefore, $c^*=c_a$ is also monotonically increasing in a for $\mathbf{E}_\infty(T_\gamma(c^*,0))=\gamma$. Thus, $\lim_{a\to 0}T_\gamma(c^*,0)=T_\gamma(\beta)$ and $\lim_{a\to 0}c_a=c_\beta(\gamma)$.

On the other hand, from $\mathbf{P}(T<\tau)\leq\alpha^*$ it follows that

$$\begin{aligned}\mathbf{P}(T\geq\tau)&=\sum_{k=1}^{\infty}\rho_k\mathbf{P}_\infty(T\geq k)\\&\geq 1-\alpha^*=\sum_{k=1}^{\infty}\rho_k\mathbf{P}_\infty(T_\gamma(c^*,0)\geq k).\end{aligned}$$

Note that $\rho_k \to 0$ and $\rho_k/\rho(a) \to \beta_k$ as $a \to 0$. Hence, $\alpha^* \to 1$ as $a \to 0$,

$$\lim_{a\to 0} D_\gamma = \lim_{a\to 0}\{T \geq 1 : \mathbf{E}_\infty(T) \geq \gamma; \mathbf{P}(T < \tau) \leq \alpha^*\} = \Delta_\gamma$$

and

$$\begin{aligned}\lim_{a\to 0} \mathcal{J}_S(T) &= \lim_{a\to 0} \frac{\sum_{k=1}^\infty \rho_k \mathbf{E}_k(T-k+1)^+}{\sum_{k=1}^\infty \rho_k \mathbf{P}_\infty(T \geq k)} \\ &= \frac{\sum_{k=1}^\infty \beta_k \mathbf{E}_k(T-k+1)^+}{\sum_{k=1}^\infty \beta_k \mathbf{P}_k(T \geq k)}.\end{aligned}$$

Thus, we have

$$\begin{aligned}&\frac{\sum_{k=1}^\infty \beta_k \mathbf{E}_k(T_\gamma(\beta)-k+1)^+}{\sum_{k=1}^\infty \beta_k \mathbf{P}_k(T_\gamma(\beta) \geq k)} \\ &\qquad = \lim_{a\to 0} \mathcal{J}_S(T_\gamma(c^*, 0)) = \lim_{a\to 0} \inf_{T\in D_\gamma} \mathcal{J}_S(T) \\ &\qquad = \inf_{T\in\Delta_\gamma} \frac{\sum_{k=1}^\infty \beta_k \mathbf{E}_k(T-k+1)^+}{\sum_{k=1}^\infty \beta_k \mathbf{P}_k(T \geq k)}.\end{aligned}$$

Furthermore, by $\sum_{k=1}^\infty \beta_k \mathbf{E}_k(T-k+1)^+ - \sum_{k=1}^\infty \beta_k \mathbf{E}_k(T-k)^+ = \sum_{k=1}^\infty \beta_k \mathbf{P}_k(T \geq k)$, we have also

$$\frac{\sum_{k=1}^\infty \beta_k \mathbf{E}_k(T_\gamma(\beta)-k)^+}{\sum_{k=1}^\infty \beta_k \mathbf{P}_k(T_\gamma(\beta) \geq k)} = \inf_{T\in\Delta_\gamma} \frac{\sum_{k=1}^\infty \beta_k \mathbf{E}_k(T-k)^+}{\sum_{k=1}^\infty \beta_k \mathbf{P}_k(T \geq k)}. \qquad \square$$

REMARK 2. Note that $\sum_{k=1}^\infty \mathbf{P}_k(T \geq k) = \mathbf{E}_\infty(T)$. When the observations $X_i, i \geq 0$, are mutually independent and $\beta_k \equiv 1$ for $k \geq 1$, we have $T_\gamma(\beta) = T_\gamma$ and

$$\begin{aligned}\frac{\sum_{k=1}^\infty \mathbf{E}_k(T_\gamma(\beta)-k)^+}{\gamma} &= \inf_{T\in\Delta_\gamma} \frac{\sum_{k=1}^\infty \mathbf{E}_k(T-k)^+}{\mathbf{E}_\infty(T)} \\ &\leq \inf_{T\in\Delta_\gamma} \frac{\sum_{k=1}^\infty \mathbf{E}_k(T-k)^+}{\gamma} \\ &\leq \frac{\sum_{k=1}^\infty \mathbf{E}_k(T_\gamma(\beta)-k)^+}{\gamma}\end{aligned}$$

and, therefore,

$$\sum_{k=1}^\infty \mathbf{E}_k(T_\gamma(\beta)-k)^+ = \inf_{T\in\Delta_\gamma} \sum_{k=1}^\infty \mathbf{E}_k(T-k)^+.$$

Thus, the result in [10] is only a special case of the corollary. If $\beta_1 = 1 + r$, $\beta_k \equiv 1$ for $k \geq 2$ and the observations $X_i, i \geq 0$, are mutually independent, then, by the corollary, $T_\gamma(\beta) = T_{\mathrm{SR}}(c_\gamma)$ and

$$\frac{r\mathbf{E}_\infty(T_\gamma(\beta)) + \sum_{k=1}^\infty \mathbf{E}_k(T_\gamma(\beta) - k + 1)^+}{r + \mathbf{E}_\infty(T_\gamma(\beta))} = \inf_{T\in\Delta_\gamma} \frac{r\mathbf{E}_\infty(T) + \sum_{k=1}^\infty \mathbf{E}_k(T - k + 1)^+}{r + \mathbf{E}_\infty(T)},$$

where $T_{\mathrm{SR}}(c_\gamma)$ is the Shiryaev–Roberts test defined in (8) and $r \ge 0$. This is just the result of Lemma 1 proved by Polunchenko and Tartakovsky [9].

Next, we will consider the optimal test for the case "$\mathbf{P}(T < \tau) \ge \alpha^*$" which is just the opposite of "$\mathbf{P}(T < \tau) \le \alpha^*$" in D_γ.

Consider the following loss random variables of detection:

$$\eta_n = \sum_{k=1}^n Y_{k-1} + 1 - cn, \tag{33}$$

where $c > 0$ is a constant. Note that η_n may be negative.

By Lemma 1, we have

$$\begin{aligned} \mathbf{E}_\infty(\eta_T) &= \mathbf{E}_\infty\left(\sum_{k=1}^T (Y_{k-1} + \rho_k) + 1\right) - \mathbf{E}_\infty\left(\sum_{k=1}^T \rho_k\right) - c\mathbf{E}_\infty(T) \\ &= \mathbf{E}(T - \tau + 1 | T \ge \tau)\mathbf{P}(T \ge \tau) + \mathbf{P}(T < \tau) - c\mathbf{E}_\infty(T). \end{aligned} \tag{34}$$

Let

$$T_\eta(c) = \min\{n \ge 1 : Y_n \ge c_n(c)\} \tag{35}$$

for $c > 0$, where $\{c_n(c), n \ge 1\}$ is a series of positive random variables satisfying $c_n(c) \in \mathfrak{F}_n$.

THEOREM 3. *Let $\gamma > 1$ and $\mathbf{E}_\infty(T_\eta(c_0) < \infty$ for some number $c_0 > 0$. There exists a positive number c^* and a series of positive random variables $c_n(c^*)$ such that $\mathbf{E}_\infty(T_\eta(c^*)) = \gamma$ and*

$$c_n(c^*) = c^*/\rho_{n+1} + \mathbf{E}_\infty\left(\sum_{m=n+1}^\infty \frac{\rho_{m+1}}{\rho_{n+1}} I(B_{m,n+1})[c^*/\rho_{m+1} - \tilde{Y}_m]\Big|\mathfrak{F}_n\right),$$

$$\inf_{T\in D'_\gamma} \mathcal{J}_S(T) = \mathcal{J}_S(T_\eta(c^*)), \tag{36}$$

where

$$D'_\gamma = \{T \ge 1 : \mathbf{E}_\infty(T) \ge \gamma; \mathbf{P}(T < \tau) \ge \alpha^*\}$$

and $\alpha^* = \mathbf{P}(T_\eta(c^*) < \tau)$.

PROOF. It needs only to be proved that (36) holds for the following subset of D'_α:

$$\hat{D}''_\gamma = \{T \ge 1 : \mathbf{E}_\infty(T) \ge \gamma; \mathbf{P}(T < \tau) \ge \alpha^*; \mathbf{E}_\infty(|\eta_T|) < \infty\}.$$

Let $\tilde{\eta}_n = -\tilde{\eta}_n$ for $n \ge 1$. By the same way of proving Theorem 1, we can prove $\mathbf{E}_\infty(\eta_T) \ge \mathbf{E}_\infty(\eta_{T_\eta})$ for all T satisfying $\mathbf{E}_\infty(|\eta_T|) < \infty$.

On the other hand, we know that $\mathbf{E}_\infty(T_\eta(c))$ is continuous and strictly increasing in c with $\mathbf{E}_\infty(T_\eta(0)) = 1$ and $\mathbf{E}_\infty(T_\eta(c)) \nearrow \infty$ as $c \nearrow \infty$ since $\mathbf{E}_\infty(T_\eta(c_0) < \infty$ for some number $c_0 > 0$. This means that there exists a unique positive number c^* such that $\mathbf{E}_\infty(T_\eta(c^*)) = \gamma$. Note that $\mathbf{E}_\infty(\eta_T) \ge \mathbf{E}_\infty(\eta_{T_\eta(c)})$. It follows from (34) that

$$\begin{aligned}
&\mathbf{E}(T-\tau+1|T \ge \tau)\mathbf{P}(T \ge \tau) + \mathbf{P}(T < \tau) - c^*\mathbf{E}_\infty(T)\\
&\quad \ge \mathbf{E}(T_\eta(c^*) - \tau + 1|T_\eta(c^*) \ge \tau)\mathbf{P}(T_\eta(c^*) \ge \tau) + \mathbf{P}(T_\eta(c^*) < \tau) - c^*\gamma.
\end{aligned}$$

Moreover, $\mathbf{E}(T-\tau+1|T \ge \tau) \ge 1$ and, therefore,

$$\begin{aligned}
&(\mathbf{E}(T-\tau+1|T \ge \tau) - 1)\mathbf{P}(T_\eta(c^*) \ge \tau) + 1 - c^*\mathbf{E}_\infty(T)\\
&\quad \ge (\mathbf{E}(T-\tau+1|T \ge \tau) - 1)\mathbf{P}(T \ge \tau) + 1 - c^*\mathbf{E}_\infty(T)\\
&\quad = \mathbf{E}(T-\tau+1|T \ge \tau)\mathbf{P}(T \ge \tau) + \mathbf{P}(T < \tau) - c^*\mathbf{E}_\infty(T)\\
&\quad \ge \mathbf{E}(T_\eta(c^*) - \tau + 1|T_\eta(c^*) \ge \tau)\mathbf{P}(T_\eta(c^*) \ge \tau) + \mathbf{P}(T_\eta(c^*) < \tau) - c^*\gamma\\
&\quad = (\mathbf{E}(T_\eta(c^*) - \tau + 1|T_\eta(c^*) \ge \tau) - 1)\mathbf{P}(T_\eta(c^*) \ge \tau) + 1 - c^*\gamma
\end{aligned}$$

for $\mathbf{P}(T_\eta(c^*) \ge \tau) \ge \mathbf{P}(T \ge \tau)$. Thus,

$$\mathbf{E}(T-\tau+1|T \ge \tau) \ge \mathbf{E}(T_\eta(c^*) - \tau + 1|T_\eta(c^*) \ge \tau)$$

for all $T \in D''_\gamma$. □

3. Optimal tests for Bayesian change-point detection with finite possible change-points. In this section, we assume that the change-point τ is at most finite, that is, $\tau \le N < \infty$. Its distribution $\{\rho_k, k \ge 1\}$ satisfies $\sum_{k=1}^N \rho_k = 1$ and $\rho_k = 0$ for $k \ge N+1$. We can think of the distribution of the change-point τ as being finite. Here, the integer N can be sufficiently large. For example, $N = 10^8$ is large enough in practical application.

Consider a test T which is unbounded, that is, $\mathbf{P}(T > M) > 0$ for any large $M > 0$.

Since $p_{v_1}(x, \cdot)/p_{v_0}(x, \cdot)$ has no atoms with respect to $\mathbf{P}_\infty$ for all $x \in E$, it follows that

$$\mathbf{E}_\infty(I(\Lambda_1 < 1)|X_0 = x) < 1, \qquad \mathbf{E}_\infty(I(\Lambda_1 < 1)|X_0 = x) < 1$$

for all $x \in E$.

For example, let $p_{v_1}(x,y) = (\sqrt{2\pi}\sigma)^{-1}e^{-(y-x-\mu_1)^2/2\sigma^2}$ and $p_{v_0}(x,y) = (\sqrt{2\pi}\sigma)^{-1}e^{-(y-x)^2/2\sigma^2}$. Then $\Lambda_1 = e^{\mu_1(X_1-X_0-\mu_1/2)/\sigma^2}$, where $\mu_1 > 0$. Then

$$\begin{aligned}
\mathbf{E}_\infty\big(I(\Lambda_1 < 1)|X_0 = x\big) &= \mathbf{E}_\infty\big(I(X_1 - x - \mu_1/2 < 0)|X_0 = x\big)\\
&= \int_{-\infty}^{\mu_1/2\sigma} (\sqrt{2\pi})^{-1} e^{-u^2/2}\,du < 1,\\
\mathbf{E}_\infty\big(I(\Lambda_1 \ge 1)|X_0 = x\big) &= \int_{\mu_1/2\sigma}^{+\infty} (\sqrt{2\pi})^{-1} e^{-u^2/2}\,du < 1
\end{aligned}$$

for all x. Let

$$A = \sup_{x\in E} \mathbf{E}_\infty\big(I(\Lambda_1 < 1)|X_0 = x\big), \qquad B = \sup_{x\in E} \mathbf{E}_\infty\big(I(\Lambda_1 \ge 1)|X_0 = x\big) < 1$$

$$C = \sup_{x\in E} \mathbf{E}_\infty\big([\ln \Lambda_1]^2|X_0 = x\big).$$

Under the conditions that $A, B < 1$ and $C < \infty$. Theorem 4 below shows that if only the restrictive condition $\mathbf{E}_\infty(T) \ge \gamma$ exists, then

$$\inf_{T:\mathbf{E}_\infty(T)\ge\gamma} \mathcal{J}_S(T) = 1$$

for every finite prior distribution $\{\rho_k, 1 \le k \le N\}$ of the change-point τ, where the number $\gamma > 1$ and $N < \infty$.

THEOREM 4. *Let $\rho_1 > 0$, $\sum_{k=1}^N \rho_k = 1$, $A, B < 1$ and $C < \infty$. Then there is a series of tests (stopping times) $\{T_N(c), c > 0\}$ such that $\mathbf{E}_\infty(T_N(c)) \ge \gamma$,*

$$\lim_{c\to 0} \mathcal{J}_S\big(T_N(c)\big) = 1, \qquad \lim_{c\to 0} \mathbf{E}_\infty\big(T_N^2(c)\big) = \infty, \tag{37}$$

and, therefore, $\inf_{T:\mathbf{E}_\infty(T)\ge\gamma} \mathcal{J}_S(T) = 1$ for every finite prior distribution $\{\rho_k, 1 \le k \le N\}$ of change-point τ.

PROOF. The proof of Theorem 4 is shown in supplementary materials [3]. □

Now we consider the probability of false alarm $\mathbf{P}_\infty(T < \tau)$. Since

$$\mathbf{P}_\infty(T < \tau) = \sum_{k=1}^N \rho_k \mathbf{P}_\infty(T < k) = 1 - \rho_1 - \sum_{k=2}^N \rho_k \mathbf{P}_\infty(T \ge k) \le 1 - \rho_1,$$

it follows that $\mathbf{P}_\infty(T < \tau) < 1 - \rho_1$ if and only if $\mathbf{E}_\infty(T) = 1 + \sum_{k=2}^N \mathbf{P}_\infty(T \ge k) > 1$ for $\rho_k > 0, 2 \le k \le N$. Thus, we have the following corollary.

COROLLARY 2. *Assume that the conditions of Theorem 4 hold. Then*

$$\inf_{\{T:\mathbf{P}_\infty(T<\tau)<1-\rho_1\}} \mathcal{J}_S(T) = 1$$

for every finite positive prior distribution $\{\rho_k, 1 \le k \le N\}$ of the change-point τ.

TABLE 1
$\mathcal{J}_S(T)$ *with* $\mathrm{ARL}_0 \approx 1000$

Shift in μ	**0.0**	**0.1**	**0.25**	**0.5**	**0.75**	**1.0**	**1.25**	**1.5**	**2.00**	**3.00**
$r = 0.000166$	1006.95	4.46	2.72	1.95	1.63	1.44	1.31	1.22	1.09	1.01
$c' = -0.6$	(164,713.92)	(21.40)	(5.69)	(2.59)	(1.83)	(1.49)	(1.32)	(1.20)	(0.97)	(0.83)
$r = 0.000344$	1002.15	6.47	3.48	2.39	1.92	1.64	1.45	1.34	1.18	1.03
$c' = -0.2$	(82,057.39)	(26.64)	(6.89)	(2.82)	(1.84)	(1.50)	(1.30)	(1.12)	(0.90)	(0.74)
$r = 0.000463$	995.04	7.65	3.91	2.63	2.09	1.77	1.55	1.40	1.22	1.06
$c' = 0.0$	(62,621.72)	(28.80)	(7.38)	(3.06)	(1.93)	(1.49)	(1.25)	(1.10)	(0.88)	(0.65)
$r = 0.000744$	1003.42	10.37	4.94	3.06	2.44	2.04	1.77	1.58	1.33	1.11
$c' = 0.4$	(40,856.77)	(32.86)	(8.42)	(3.23)	(2.06)	(1.52)	(1.25)	(1.08)	(0.85)	(0.58)
$r = 0.001078$	1002.66	13.34	6.09	3.60	2.74	2.32	2.01	1.78	1.47	1.17
$c' = 0.8$	(28,285.51)	(36.53)	(9.47)	(3.58)	(2.17)	(1.60)	(1.25)	(1.09)	(0.83)	(0.56)
$r = 0.001813$	1003.58	19.82	8.72	4.87	3.55	2.87	2.46	2.20	1.81	1.36
$c' = 1.6$	(17,225.21)	(43.10)	(11.39)	(4.20)	(2.43)	(1.67)	(1.33)	(1.09)	(0.83)	(0.57)
$r = 0.003347$	1004.85	33.96	14.74	7.80	5.50	4.34	3.60	3.10	2.51	1.90
$c' = 3.2$	(9312.11)	(54.93)	(14.56)	(5.32)	(2.98)	(2.06)	(1.54)	(1.25)	(0.87)	(0.60)
$r = 0.0065$	996.64	63.01	27.12	14.13	9.74	7.36	6.04	5.21	4.08	2.89
$c' = 6.4$	(4786.32)	(73.22)	(19.71)	(7.19)	(4.02)	(2.68)	(1.99)	(1.57)	(1.09)	(0.68)
$r = 0.012773$	998.26	117.17	51.03	26.45	18.16	13.56	11.03	9.44	6.97	4.99
$c' = 12.80000$	(2406.36)	(94.81)	(26.64)	(9.85)	(5.48)	(3.61)	(2.65)	(2.07)	(1.43)	(0.84)
$r = 0.025781$	1000.47	208.35	95.08	49.54	34.82	25.54	20.77	18.14	13.03	9.34
$c' = 25.60000$	(1201.08)	(113.21)	(34.89)	(13.29)	(7.51)	(4.98)	(3.61)	(2.83)	(1.90)	(1.08)
$r = 0.032120$	999.09	242.32	113.01	60.29	41.22	32.16	24.80	20.92	16.99	10.72
$c' = 31.50000$	(980.29)	(116.60)	(37.35)	(14.48)	(8.17)	(5.42)	(4.06)	(3.09)	(2.14)	(1.19)
CUSUM	1001.83	437.25	147.34	38.21	16.93	10.14	7.21	5.60	3.90	2.51
$c = 5.075$, $\delta = 1.0$	(988.56)	(434.01)	(141.95)	(31.89)	(11.28)	(5.49)	(3.34)	(2.27)	(1.32)	(0.68)
EWMA	998.35	349.33	104.92	30.94	15.61	10.22	7.58	6.04	4.35	2.86
$c = 0.644$, $\lambda = 0.1$	(984.87)	(339.44)	(95.54)	(22.50)	(8.96)	(4.87)	(3.13)	(2.24)	(1.39)	(0.78)

4. Numerical simulations. As an application of Theorem 4, we consider the test $T_5(c)$ with $\rho_k = 1/5$, $1 \le k \le 5$. It follows from Theorem 4 that

$$\mathcal{J}_S\big(T_5(c)\big) = 1 + o(1)$$

and $\mathbf{E}_\infty(T_5^2(c))$ is sufficiently large for a small number c.

In the following example, we compare the detection performance of the three tests: $T_5(c)$, the CUSUM test $T_C(c)$ and EWMA test $T_E(c)$. Table 1 corresponding to the example shows the simulation results of $\mathcal{J}_S(\cdot)$ of the tests $T_5(c)$, $T_C(c)$ and $T_E(c)$ with $\mathrm{ARL}_0 \approx \mathbf{1000}$ for mean shifts from $\mu = 0$ to $\mu = 0.1$, 0.25, 0.5, 0.75, 1, 1.25, 1.5, 2 and 3, where $\mathrm{ARL}_0(T) = \mathbf{E}_\infty(T)$,

$$\mathcal{J}_S(T) = \frac{\sum_{k=1}^5 \mathbf{E}_{k,\mu}(T-k+1)^+}{\sum_{k=1}^5 \mathbf{P}_\infty(T \ge k)}$$

and $\mathbf{E}_{k,\mu}(T-k+1)^+$ denotes the out-control average run length (ARL) from 0 to μ at the change-point k. The numerical simulation results of ARL_0 and $\mathcal{J}_S(\cdot)$ in the table were obtained based on $\mathbf{10}^7$ repetitions. The paper [2] compares the ARLs of the four tests: CUSUM, EWMA, GEWMA and GLR in detecting mean shifts when $\tau = 1$ or equally, $\rho_1 = 1$.

EXAMPLE 1. Let the observation processes $\{X_n, n \geq 0\}$ be mutually independent. Let $X_0 \sim N(0,1)$ and after the change point τ, $X_k \sim N(\mu, 1)$ for $k \geq \tau$. Hence, $Z_k = \mu(X_k - \mu/2)$. In order to compare the detection performance of $T_5(c)$, $T_C(c)$ and $T_E(c)$, we write their definitions in the following:

$$T_5(c) = \begin{cases} \min\{1 \leq n \leq 5 : Z'_n \geq c' + \ln 5\}, \\ \min\left\{n \geq 6 : Z'_5 + \sum_{k=6}^{n}(Z_k - \mu_0) \geq c' + \ln 5 - (n-5)r(c)\right\}, \end{cases}$$

where $c = e^{c'}$ and

$$\begin{aligned}
&Z'_1 = Z_1 - \mu_0, \qquad Z'_2 = Z_2 - \mu_0 + \ln(1 + e^{Z_1}), \\
&Z'_3 = Z_3 - \mu_0 + \ln(1 + e^{Z_2} + e^{Z_2 + Z_1}), \\
&Z'_4 = Z_4 - \mu_0 + \ln(1 + e^{Z_3} + e^{Z_3 + Z_2} + e^{Z_3 + Z_2 + Z_1}), \\
&Z'_5 = Z_5 - \mu_0 + \ln(1 + e^{Z_4} + e^{Z_4 + Z_3} + e^{Z_4 + Z_3 + Z_2} + e^{Z_4 + Z_3 + Z_2 + Z_1}).
\end{aligned}$$

The popular upward-sided CUSUM test, T_C, is defined as

$$T_C(c) = \inf\left\{n : \max_{1 \leq k \leq n}\left[\sum_{i=n-k+1}^{n} \delta(X_i - \delta/2)\right] \geq c\right\},$$

where $c > 0$ is a control limit and $\delta/2 > 0$ is the reference value. Here, we take $\delta = 1$.

The EWMA is also a popularly used test which can be written by

$$T_E(c) = \inf\{n \geq 1 : E_n(\lambda) \geq c\},$$

where λ is a weighting parameter ($0 < \lambda \leq 1$), $c > 0$ is the control limit and $E_n(\lambda) = \lambda X_n + (1-\lambda)E_{n-1}(\lambda)$ with $E_0(\lambda) = 0$. Here, we take $\lambda = 0.1$.

The simulation results of $\mathcal{J}_S(T_5(c))$, $\mathcal{J}_S(T_C(c))$ and $\mathcal{J}_S(T_E(c))$ are shown in rows 1–11, the twelfth row and the last row, respectively. The values in parentheses are the standard deviations of the simulation results of detection delay. We can see from Table 1 that: (1) The larger the standard deviations of the in-control run length (false alarm), the smaller the average detection delay $\mathcal{J}_S(T_5(c))$ before detecting the mean shifts. For example, the standard deviation in the first row is **164713.92** with $\mathrm{ARL}_0 =$ **1006.95**, but the average detection delay is only **4.46** for a small mean shift $\mu = \mathbf{0.1}$. When the standard deviation is **980.55** with $\mathrm{ARL}_0 = \mathbf{999.09}$ (see thirteenth row), the average detection delay becomes **242.32** for $\mu = \mathbf{0.1}$. (2) $T_5(c)$ performs better than T_C and T_E for a small mean shift $\mu = \mathbf{0.1}$ even if they have the same standard deviations for $\mu = 0$ (see the last three rows). (3) $T_5(c)$ performs better than T_C and T_E for all mean shifts μ when the standard deviations of T_5 for $\mu = 0$ is greater than **9312.11** (see rows 1–7).

5. Conclusions. By introducing suitable loss random variables of detection, we have derived four strictly optimal tests $T^*(c)$, $T'(c)$, $T_\gamma(c, b)$ and $T_\eta(c)$ for Bayesian change-point detection not only for a general prior distribution of the change-point but also for observations $X_0, X_1, X_2, \dots$ that form a Markov process under the restrictive conditions that $\{\mathbf{P}_\infty(T < \tau) \le \alpha\}$, $\{\mathbf{E}_\infty(T) \ge \gamma, \mathbf{P}(T < \tau) \le \alpha^*\}$ and $\{\mathbf{E}_\infty(T) \ge \gamma, \mathbf{P}(T < \tau) \ge \alpha^*\}$.

When the number of change-points τ is finite ($N < \infty$), that is, the prior distribution $\{\rho_k, 1 \le k \le N\}$ of τ satisfies $\rho_k = 0$ for all $k > N$, we have constructed a series of tests $\{T_N(c) : c > 0\}$ in Theorem 4 and proved that $\lim_{c\to 0} \mathcal{J}_S(T_N(c)) = 1$ and $\lim_{c\to 0} \mathbf{E}_\infty(T_N^2(c)) = \infty$ with $\mathbf{E}_\infty(T_N(c)) \ge \gamma > 1$. Since $\mathcal{J}_S(T) \ge 1$ for every test T, $T_N(c)$ can be considered as an asymptotically optimal test when $c \to 0$. This implies that

$$\inf_{T:\mathbf{E}_\infty(T)\ge\gamma} \mathcal{J}_S(T) = 1$$

for a finite number of change-points τ.

It follows from Theorem 4 and the numerical simulations in Section 4 that the smaller the value of $\mathcal{J}_S(T_N(c))$, the larger the variance $\mathbf{Var}_\infty(T_N(c))$ for a given finite false alarm rate $\mathbf{E}_\infty(T_N(c)) \ge \gamma$. This also means that the larger the variance, the greater the risk of false alarm. In order to reduce the risk or the variance, we must consider a stronger restrictive condition $\{\mathbf{E}_\infty(T) \ge \gamma, \mathbf{Var}_\infty(T) \le \sigma^2\}$. An interesting problem would be to find a test $T_N(c^*)$ that is optimal in the sense that

$$\inf_{\{T:\mathbf{E}_\infty(T)\ge\gamma, \mathrm{Var}_\infty(T)\le\sigma^2\}} \mathcal{J}_S(T) = \mathcal{J}_S(T_N(c^*))$$

with $\mathbf{E}_\infty(T_N(c^*)) = \gamma$ and $\mathbf{Var}_\infty(T_N(c^*)) = \sigma^2$ when the number of change-points τ is finite. Other problems such as how to calculate $\mathcal{J}_S(T_N(c^*))$ and how to estimate c^* are also worth studying in the future.

Acknowledgments. The authors would like to thank the editor and anonymous referees for their many valuable comments that have resulted in signicant improvements in the article.

SUPPLEMENTARY MATERIAL

Supplement A: Proofs of Theorem 4 of the paper "On the optimality of Bayesian change-point detection" (DOI: 10.1214/16-AOS1479SUPP; .pdf). We prove in the supplementary material that the optimal (minimal) average detection delay is equal to 1 for any (possibly large) average run length to false alarm if the number of possible change-points is finite for observations being a Markov process.

REFERENCES

[1] CHOW, Y. S., ROBBINS, H. and SIEGMUND, D. (1991). *The Theory of Optimal Stopping*. Dover, New York. MR1105325

[2] HAN, D. and TSUNG, F. (2004). A generalized EWMA control chart and its comparison with the optimal EWMA, CUSUM and GLR schemes. *Ann. Statist.* **32** 316–339. MR2051010

[3] HAN, D., TSUNG, F. and XIAN, J. (2017). Supplement to "On the optimality of Bayesian change-point detection." DOI:10.1214/16-AOS1479SUPP.

[4] LAWLER, G. F. and LIMIC, V. (2010). *Random Walk: A Modern Introduction. Cambridge Studies in Advanced Mathematics* **123**. Cambridge Univ. Press, Cambridge. MR2677157

[5] LORDEN, G. (1971). Procedures for reacting to a change in distribution. *Ann. Math. Stat.* **42** 1897–1908. MR0309251

[6] MOUSTAKIDES, G. V. (2008). Sequential change detection revisited. *Ann. Statist.* **36** 787–807. MR2396815

[7] POLLAK, M. (1985). Optimal detection of a change in distribution. *Ann. Statist.* **13** 206–227. MR0773162

[8] POLLAK, M. and TARTAKOVSKY, A. G. (2009). Optimality properties of the Shiryaev–Roberts procedure. *Statist. Sinica* **19** 1729–1739. MR2589207

[9] POLUNCHENKO, A. S. and TARTAKOVSKY, A. G. (2010). On optimality of the Shiryaev–Roberts procedure for detecting a change in distribution. *Ann. Statist.* **38** 3445–3457. MR2766858

[10] POLUNCHENKO, A. S. and TARTAKOVSKY, A. G. (2012). State-of-the-art in sequential change-point detection. *Methodol. Comput. Appl. Probab.* **14** 649–684. MR2966314

[11] POOR, H. V. and HADJILIADIS, O. (2009). *Quickest Detection*. Cambridge Univ. Press, Cambridge. MR2482527

[12] ROBERTS, W. S. (1966). A comparison of some control chart procedures. *Technometrics* **8** 411–430. MR0196887

[13] SHIRYAEV, A. N. (1963). On optimum methods in quickest detection problems. *Theory Probab. Appl.* **13** 22–46.

[14] SHIRYAEV, A. N. and ZRYUMOV, P. Y. (2009). On the linear and nonlinear generalized Bayesian disorder problem (discrete time case). In *Optimality and Risk—Modern Trends in Mathematical Finance* 227–235. Springer, Berlin. MR2648606

[15] SHIRYAYEV, A. N. (1978). *Optimal Stopping Rules*. Springer, New York. MR0468067

[16] TARTAKOVSKIĬ, A. G. (2008). Asymptotic optimality in Bayesian quickest detection problems under a global false alarm probability constraint. *Theory Probab. Appl.* **53** 443–466.

[17] TARTAKOVSKY, A., NIKIFOROV, I. and BASSEVILLE, M. (2015). *Sequential Analysis: Hypothesis Testing and Changepoint Detection. Monographs on Statistics and Applied Probability* **136**. CRC Press, Boca Raton, FL. MR3241619

[18] TARTAKOVSKY, A. G., POLLAK, M. and POLUNCHENKO, A. S. (2012). Third-order asymptotic optimality of the generalized Shiryaev–Roberts changepoint detection procedures. *Theory Probab. Appl.* **56** 457–484. MR3136464

[19] TARTAKOVSKY, A. G. and VEERAVALLI, V. V. (2005). General asymptotic Bayesian theory of quickest change detection. *Theory Probab. Appl.* **49** 458–497. MR2144868

[20] YAKIR, B. (1994). Optimal detection of a change in distribution when the observations form a Markov chain with a finite state space. In *Change-Point Problems* (*South Hadley*, *MA*, 1992) (E. Carlstein, H. Muller and D. Siegmund, eds.). *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **23** 346–358. IMS, Hayward, CA. MR1477935

D. HAN
J. XIAN
DEPARTMENT OF MATHEMATICS
SHANGHAI JIAO TONG UNIVERSITY
SHANGHAI 200030
P.R. CHINA
E-MAIL: donghan@sjtu.edu.cn
jgxian@sjtu.edu.cn

F. TSUNG
DEPARTMENT OF INDUSTRIAL ENGINEERING
AND LOGISTICS MANAGEMENT
HONG KONG UNIVERSITY
OF SCIENCE AND TECHNOLOGY
HONG KONG
P.R. CHINA
E-MAIL: season@ust.hk