ON THE OPTIMALITY OF BAYESIAN CHANGE-POINT DETECTION*

BY DONG HAN† FUGE  TEUNG‡ AND JINGUO XIAN†

Shanghai Jiao Tong University †
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 $X_0, X_1, X_2, \ldots$, 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 $X_0, X_1, X_2, \ldots, X_n, \ldots$ is known and is $P_{v_0}$ for $n < \tau$, and that after the change-point $\tau$, the distribution of $X_\tau, X_{\tau+1}, \ldots$ 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 $P_\tau$ and $E_\tau$ be the probability distribution and the expectation of $\{X_\tau, X_{\tau+1}, \ldots\}$ respectively if a change occurs at time $\tau$, where $\tau = 1, 2, \ldots$. When $\tau = \infty$, i.e., a change never occurs, the probability distribution and the expectation are denoted by $P_\infty$ and $E_\infty$ respectively for all

---

*Supported by RGC Competitive Earmarked Research Grants, National Natural Science Foundation of China (11171216) and National Basic Research Program of China (973 Program, 2011CB808000)

MSC 2010 subject classifications: Primary 62L10; secondary 62L15

Keywords and phrases: Optimal test, Bayesian change-point detection, Markov Process
observations \(X_0, X_1, X_2, ..., X_n, \ldots\).

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 \(\{\mathcal{F}_t\}_{t \geq 0}\), where \(\mathcal{F}_t = \sigma\{X_k, 0 \leq k \leq t\}\) denotes the smallest \(\sigma\)-algebra with respect to which all of the random variables \(X_0, \ldots, X_t\) are measurable. We usually assume that \(X_0 \equiv x_0\) (a constant) and \(\mathcal{F}_0 = \{\phi, \Omega\}\).

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 \(\mathbb{P}_\infty(T < \tau)\) 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 \(\gamma > 1\), i.e., \(\mathbb{E}_\infty(T) \geq \gamma\).

Three major optimal performance measures exist for detecting change-points. The first one is Shiryaev([12, 13]) procedure which can be written as

\[
\inf_{T: \mathbb{P}_\infty(T < \tau) \leq \alpha} J_S(T)
\]

or

\[
\inf_{T: \mathbb{E}_\infty(T) \geq \gamma} J_S(T),
\]

where

\begin{align}
J_S(T) &= \mathbb{E}((T - \tau + 1)^+ | T \geq \tau) = \frac{\sum_{k=1}^{\infty} \rho_k \mathbb{E}_k(T - k + 1)^+}{\sum_{k=1}^{\infty} \rho_k \mathbb{P}_\infty(T \geq k)}, \\
\end{align}

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

The second is Lorden’s procedure ([4]):

\[
\inf_{T: \mathbb{E}_\infty(T) \geq \gamma} J_L(T),
\]

where \(J_L(T)\) is the worst average delay, i.e.,

\[
J_L(T) = \sup_{k \geq 1} \mathbb{E}_k((T - k + 1)^+ | \mathcal{F}_{k-1}).
\]

The third one, proposed by Pollak [6], also considers the worst average delay \(J_P(T) = \sup_{t \geq 1} \mathbb{E}_t(T - t + 1 | T \geq t)\), and the optimal stopping time can be written as

\[
\inf_{T: \mathbb{E}_\infty(T) \geq \gamma} J_P(T).
\]
Note that \( \tau \) 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 [5] introduced a general framework for capturing and better understanding the above three optimization criteria and showed that \( J_S(T) \leq J_P(T) \leq J_L(T) \).

Since this paper focuses 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 [10], [9], [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, i.e. \( \rho_1 = \rho_0 + (1 - \rho_0) \rho \), \( \rho_k = (1 - \rho_0) \rho (1 - \rho)^{k-1} \) (\( k \geq 2 \)), where \( 0 \leq \rho_0 < 1 \) and \( 0 < \rho < 1 \), Shiryaev [12, 13] first proved that

\[
S_{\rho} = \min \{ n : P(\tau \leq n | \mathcal{F}_n) \geq a(\rho, \alpha) \}
\]

is optimal, i.e., it minimizes \( J_S(T) \) for an appropriately chosen threshold \( a(\rho, \alpha) \in (0, 1) \) such that \( \alpha = P_\infty(S_{\rho} < \tau) \).

It follows from the Bayes rule that (8)

\[
P(\tau \leq n | \mathcal{F}_n) = \frac{R_n(\rho)}{R_n(\rho) + 1/\rho}
\]

where

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

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

\[
T_S(\rho) = \min \{ n : R_n(\rho) \geq c \},
\]

where \( c \) is a positive constant.

For a general prior distribution of the change-point \( \tau \), Chow, Robbins and Siegmund [1] have shown that the following test

\[
T_{CRS} = \min \{ n : X_n \geq E(\zeta_{n+1} | \mathcal{F}_n) \}
\]

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

\[
\psi_n = c(1 - \pi_n) + \sum_{k=0}^{n-1} (n - i)n_i^n,
\]
\[ \pi_n = \mathbf{P}(\tau \leq n | \mathcal{F}_n), \quad \pi_k^n = \mathbf{P}(\tau = k | \mathcal{F}_n) \quad \text{and} \quad \zeta_n = \text{ess sup}_{T \geq n} \{ \mathbb{E}(\psi_T | \mathcal{F}_n) \}. \]

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 \( \tau \) is geometric, Yakir [19] generalized Shiryaev’s result for the Markov chain case. Tartakovsky and Veeravalli [18] proved that the Shiryaev procedure is asymptotically optimal in the general non-i.i.d. case when the false alarm probability approaches zero. Tartakovsky [15] proposed an asymptotically optimal Bayesian change point detection procedure for general non-i.i.d. models when the global false alarm probability \( P_\infty(T < \infty) \) approaches zero.

Note that for \( \rho_0 = r \rho \), as \( \rho \to 0 \), where \( r \geq 0 \), Shiryaev’s stopping time \( T_{SR}(\rho) \) converges to the following Shiryaev-Roberts ([12], [11]) test:

\begin{equation}
T_{SR}(c) = \min\{ n : r \prod_{k=1}^{n} \Lambda_k + \sum_{t=1}^{n} \prod_{k=t}^{n} \Lambda_k \geq c \} \tag{8}
\end{equation}

for a positive threshold \( c > 0 \) and

\begin{equation}
\hat{J}_{SR}(T) = \lim_{\rho \to 0} \frac{r \rho \mathbb{E}_1(T) + (1 - r \rho) \sum_{k=1}^{\infty} \rho(1 - \rho)^{k-1} \mathbb{E}_k(T - k + 1)^+}{r \rho + (1 - r \rho) \sum_{k=1}^{\infty} \rho(1 - \rho)^{k-1} \mathbb{P}_\infty(T \geq k)}
\end{equation}

\begin{align}
&= \frac{r \mathbb{E}_1(T) + \sum_{k=1}^{\infty} \mathbb{E}_k(T - k + 1)^+}{r + \mathbb{E}_\infty(T)} \tag{9}
\end{align}

\( T_{SR}(c) \) is often called the generalized Bayesian change-point detection. Pollak and Tartakovsky [7] proved that the Shiryaev-Roberts test is optimal in the measure (9), that is, \( T_{SR}(c) \) minimizes \( \hat{J}_{SR}(T) \) over all stopping times \( T \) satisfying \( \mathbb{E}_\infty(T) \geq \gamma \), where \( \mathbb{E}_\infty(T_{SR}(c)) = \gamma \). Furthermore, Tartakovsky, Pollak and Polunchenko [16] 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, \( S_{\rho} \) and \( T_{SR}(c) \) have been proved to be strictly (non-asymptotically) 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 \( \tau \).

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

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

$$\inf_{T, E_\infty(T) \geq \gamma} J_S(T) = 1$$

for every finite ($N < \infty$) prior distribution $\{\rho_k, 1 \leq k \leq N\}$ 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 $\pi_n = P(\tau \leq n | \mathcal{F}_n), n \geq 1$, is not a Markov process when the prior distribution is not geometric and the observation sequence $\{X_n, n \geq 0\}$ 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 $\tau$. In Section 3, for any finite possible change-points ($\rho_k = P(\tau = k) = 0$ for all $k \geq N + 1$, $N < \infty$), we prove that the optimal (minimal) average detection delay is 1, that is, $\inf_{T, E_\infty(T) \geq \gamma} J_S(T) = 1$. Section 4 illustrates simulation results of comparing the detection performance of the three tests: $T_\delta(c)$, the CUSUM test $T_C(c)$ and EWMA test $T_E(c)$ for monitoring mean shifts in the special cases of $\tau$: $\rho_k = 1/5, 1 \leq k \leq 5$. 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 $\tau$ will take infinite possible non-negative integer values, that is, $P(\tau_k = k) = \rho_k > 0$ for all $k \geq 1$.

Let the observations $X_0, X_1, X_2, \ldots$ be a time homogeneous Markov process taking values in some real number space $E \subseteq (-\infty, +\infty)$ with a known transition probability density $p_{0u}(\ldots)$ before the change-point $\tau \geq 1$, and with the possible post-change transition probability density $p_{u}(\ldots)$ 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 \( p_v(\ldots) \) lies in a closed domain \( \Gamma(V) \), where \( \Gamma(V) = \{ p_v(\ldots) : v \in V \} \), \( V \) 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 \( v = (\mu, \sigma) \), \( \mu \) and \( \sigma \) denote the mean and standard deviation respectively. We may take the set \( V \) satisfying \( V = \{ (\mu, \sigma) : |\mu| \leq \mu_1, 0 < \sigma \leq \sigma_1 \} \), where \( \mu_1 \) and \( \sigma_1 \) are known; that is, the domain \( V \) and the set of possible post-change distributions, i.e., \( \Gamma(V) \) and the uniform distribution \( Q \) of \( V \) are known.

Let

\[
p_v(\ldots) = \int_V p_v(\ldots)dQ(v).
\]

The function \( p_v(\ldots) \) can be considered as a special, known post-change transition probability density of \( E \).

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

\[
F(x_1, \ldots, x_n) = \sum_{k=1}^{n} \rho_k \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_{k-1}} \int_{-\infty}^{x_n} \prod_{j=1}^{k-1} p_v(y_{j-1}, y_j) \prod_{j=k}^{n} p_v(y_{j-1}, y_j) dy_1 \cdots dy_{k-1} \cdots dy_n + (1 - \sum_{k=1}^{n} \rho_k) F_\infty(x_1, \ldots, x_n),
\]

where

\[
F_\infty(x_1, \ldots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} \prod_{j=1}^{n} p_v(y_{j-1}, y_j) dy_1 \cdots dy_n
\]

and \( y_0 = x_0 \).

When the observations \( X_1, \ldots, X_n \ldots \) are independent with the probability density \( p_v \) before the change-point and the probability density \( p_v(\ldots) \) after the
change-point, the joint probability \( P(X_1 \leq x_1, \ldots, X_n \leq x_n) = F(x_1, \ldots, x_n) \) can be written as

\[
F(x_1, \ldots, x_n) = \sum_{k=1}^{n} \rho_k F_\infty(x_1, \ldots, x_{k-1}) F_{v_1}(x_k, \ldots, x_n) + (1 - \sum_{k=1}^{n} \rho_k) F_\infty(x_1, \ldots, x_n),
\]

where \( F_\infty(x_1, \ldots, x_k) = \prod_{j=1}^{k} F_\infty(x_k), \) \( F_{v_1}(x_1, \ldots, x_k) = \prod_{j=1}^{k} F_{v_1}(x_k) \) 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 \( X_0 \equiv x_0 \in E, p_{v_1}(x,.)/p_{v_0}(x,.) \) has no atoms with respect to \( P_\infty \) 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

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

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 \( T \), i.e. \( P_\infty(T < \infty) = 1, \)

\[
\sum_{k=1}^{\infty} \rho_k E_{k}(T - k + 1) = E_{\infty}\left( \sum_{k=1}^{T} [Y_{k-1} + \rho_k] \right)
\]

\[
= \sum_{k=1}^{\infty} E_{\infty}(Y_k I(T \geq k + 1)) + \sum_{k=1}^{\infty} \rho_k P_\infty(T \geq k),
\]

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

\[
Y_k = (Y_{k-1} + \rho_k) \Lambda_k, \quad Y_0 = 0
\]

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

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

\[
(T - k + 1) = \sum_{m=1}^{\infty} I(T - k \geq m) = \sum_{m=k}^{\infty} I(T \geq m),
\]
where \( T \geq m \in \mathbb{F}_{m-1} \) and \( m \geq k \), it follows that

\[
\sum_{m=k}^{\infty} E_k(I(T \geq m)) = \sum_{m=k}^{\infty} E_k\left(I(T \geq m) \prod_{j=1}^{k-1} p_{v_0}(X_{j-1}, X_j) \prod_{j=k}^{m-1} p_{v_1}(X_{j-1}, X_j)\right)
\]

\[
= \sum_{m=k}^{\infty} E_{\infty}\left(I(T \geq m) \frac{\prod_{j=k}^{k-1} p_{v_0}(X_{j-1}, X_j) \prod_{j=k}^{m-1} p_{v_1}(X_{j-1}, X_j)}{\prod_{j=1}^{m-1} p_{v_0}(X_{j-1}, X_j)}\right)
\]

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

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

Thus

\[
\sum_{k=1}^{\infty} \rho_k E_k(T - k + 1)^+ = E_{\infty}\left(\sum_{k=1}^{\infty} \rho_k \sum_{m=k}^{\infty} I(T \geq m) \frac{L_{m-1}}{L_{k-1}}\right)
\]

\[
= E_{\infty}\left(\sum_{k=1}^{\infty} \rho_k \sum_{m=k}^{T} \frac{L_{m-1}}{L_{k-1}}\right)
\]

\[
= E_{\infty}\left(\sum_{m=1}^{T} \sum_{k=1}^{m} \rho_k \frac{L_{m-1}}{L_{k-1}}\right) = E_{\infty}\left(\sum_{m=1}^{T} [Y_{m-1} + \rho_m]\right).
\]

Furthermore, we have

\[
E_{\infty}\left(\sum_{m=1}^{T} [Y_{m-1} + \rho_m]\right) = E_{\infty}\left(\sum_{k=1}^{\infty} I(T = k)\left(\sum_{m=1}^{k} [Y_{m-1} + \rho_m]\right)\right)
\]

\[
= \sum_{k=1}^{\infty} E_{\infty}(Y_k I(T \geq k + 1)) + \sum_{k=1}^{\infty} \rho_k P_{\infty}(T \geq k).
\]

It follows from (1) and (10) that

\[
(12) \quad \mathcal{J}_S(T) = 1 + \tilde{\mathcal{J}}_S(T)
\]
for any test (stopping time) $T$, where

$$
\tilde{J}_S(T) = \frac{\sum_{k=1}^{\infty} E_\infty(Y_k I(T \geq k + 1))}{\sum_{k=1}^{\infty} \rho_k P_\infty(T \geq k)}.
$$

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

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

It follows that

$$
E_\infty(Y_n - \sum_{k=1}^{n} \rho_k | \mathcal{F}_{n-1}) = Y_{n-1} - \sum_{k=1}^{n-1} \rho_k,
$$

that is, $\{Y_n - \sum_{k=1}^{n} \rho_k, \ n \geq 0\}$ is a martingales with $E_\infty(Y_n) = \sum_{k=1}^{n} \rho_k$. By

the optional sampling theorem of martingale (see [3] p.333) we can obtain

the following lemma.

**Lemma 2.** For any finite stopping time $T$, if

$$
E_\infty(|Y_T - \sum_{k=1}^{T} \rho_k|) < \infty, \lim_{n \to \infty} E_\infty(|Y_n - \sum_{k=1}^{n} \rho_k|; T > n) = 0,
$$

then

$$
E_\infty(Y_T) = E_\infty(\sum_{k=1}^{T} \rho_k) = \sum_{k=1}^{\infty} \rho_k P_\infty(T \geq k) = P(T \geq \tau).
$$

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

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

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

$$
E_\infty(\xi_T) = E_\infty(\sum_{k=1}^{T} [Y_{k-1} + \rho_k] + (c + 1)[1 - E_\infty(\sum_{k=1}^{T} \rho_k)]
$$

$$
= \sum_{k=1}^{\infty} \rho_k E_k(T - k + 1)^+ + (c + 1)[1 - \sum_{k=1}^{\infty} \rho_k P_\infty(T \geq k)]
$$

$$
= E(T - \tau + 1 | T \geq \tau) P(T \geq \tau) + (c + 1)P(T < \tau).
$$

(16)
The last equality follows from (1) and (15). If $c \leq 0$, the value of $\xi_n$ will be smallest when $n = 1$, that is, $E_\infty(\xi_T)$ 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 $\mathcal{F}_n$, 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 $P_{T}(T^*(c^*) < \tau) = \alpha$,

$$c_n(c^*) = c^* + E_\infty\left(\sum_{m=n+1}^{\infty} \frac{\rho_{m+1}}{\rho_{n+1}} I(B_{m,n+1})[c^* - \tilde{Y}_m]|\mathcal{F}_n\right).$$

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

$$\inf_{T \in D_\alpha} J_S(T) = J_S(T^*(c^*))$$

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

$$J_S(T^*(c^*)) \leq 1 + c^*(1 - \rho_1).$$

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

$$\frac{1 - \alpha}{E_\infty(\rho_{T^*})} - 1 < c_1 \leq \frac{1 - \alpha}{E_\infty(\rho_{T^*+1})}$$

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

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

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

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

$$E_\infty(\xi_T) \geq E_\infty(\xi_{T^*}) \quad \text{or} \quad E_\infty(\tilde{\xi}_{T^*}) \geq E_\infty(\tilde{\xi}_T)$$
for any finite test $T$ with $\mathbb{E}_\infty(\xi_T) < \infty$.

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

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

and

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

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

$$\mathbb{E}_\infty(\tilde{\xi}_{T^*}) = \int_{\{T^* \leq T\}} \tilde{\xi}_{T^*} + \int_{\{T^* > T\}} \tilde{\xi}_{T^*}$$

$$= \sum_{n=1}^{\infty} \int_{\{T^* = n \leq T\}} \tilde{\xi}_{T^*} + \sum_{n=1}^{\infty} \int_{\{T = n < T^*\}} \mathbb{E}_\infty(\tilde{\xi}_{T^*}|\tilde{\mathfrak{F}}_n)$$

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

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.

$$c_{N,N} = c$$

$$c_{n,N} = c + \frac{\rho_{n+2}}{\rho_{n+1}} \mathbb{E}_\infty(I(B_{n+1,n+1}(N))[c_{n+1,N} - \bar{Y}_{n+1}])|\tilde{\mathfrak{F}}_n)$$

$$c_{n,N} = c + \mathbb{E}_\infty\left( \sum_{m=n+1}^{N} \frac{\rho_{m+1}}{\rho_{n+1}} I(B_{m,n+1}(N))[c - \bar{Y}_m]|\tilde{\mathfrak{F}}_n \right)$$

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

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

$$c_{n-1,N} = c + \frac{\rho_{n+1}}{\rho_{n}} \mathbb{E}_\infty\left( I(B_{n,n}(N)) \mathbb{E}_\infty\left( \left[ c - \bar{Y}_n + \sum_{m=n+1}^{N} \frac{\rho_{m+1}}{\rho_{n+1}} I(B_{m,n+1}(N))[c - \bar{Y}_m] \right]|\tilde{\mathfrak{F}}_n \right) \right)$$

$$= c + \frac{\rho_{n+1}}{\rho_{n}} \mathbb{E}_\infty\left( I(B_{n,n}(N))[c_{n,N} - \bar{Y}_n]|\tilde{\mathfrak{F}}_{n-1} \right)$$

and $I(B_{n,n}(N))[c_{n,N} - \bar{Y}_n] \geq 0$, 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, \( c_{n,N} \leq c(1 + \rho_n^{-1} \sum_{m=n+1}^{\infty} \rho_m + 1) \) 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

\[
\begin{align*}
    c_n &= c + \rho_n^{-2} \mathbb{E} \left( I(\tilde{B}_{n+1,n+1}[c_{n+1} - \tilde{Y}_{n+1}]|\mathcal{F}_n) \right) \\
    &= c + \mathbb{E} \left( \sum_{m=n+1}^{\infty} \frac{\rho_{m+1}}{\rho_n} I(\tilde{B}_{m,n+1}[c - \tilde{Y}_m]|\mathcal{F}_n) \right).
\end{align*}
\]

This is (17).

Hence,

\[
I(T_* > n)\mathbb{E}_\infty(\tilde{\xi}_{T_*} - \tilde{\xi}_n)|\mathcal{F}_n)
= I(T_* > n) \sum_{m=n}^{\infty} \mathbb{E}_\infty(I(T_*>m)[\tilde{\xi}_{m+1} - \tilde{\xi}_m]|\mathcal{F}_n)
= -\rho_n I(T_*>n)[\tilde{Y}_n - c + \sum_{m=n+1}^{\infty} \frac{\rho_{m+1}}{\rho_n} I(\tilde{B}_{m,n+1}[\tilde{Y}_m - c]|\mathcal{F}_n)]
= -\rho_n I(T_*>n)(\tilde{Y}_n - c_n) > 0.
\]

That is, (21) holds.

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

\[
(24) \quad \mathbb{E} \left( \sum_{m=n+1}^{N} \frac{\rho_{m+1}}{\rho_n} I(T_N > m)[c - \tilde{Y}_m]|\mathcal{F}_n \right) \leq (c_{n,N} - c) I(T_N > n)
\]

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

\[
\begin{align*}
    \frac{\rho_{N+1}}{\rho_N} \mathbb{E} \left( I(T_N > N)[c - \tilde{Y}_N]|\mathcal{F}_{N-1} \right) \\
    \leq \frac{\rho_{N+1}}{\rho_N} \mathbb{E} \left( I(T_N > N)I(B_{N,N}(N))|c - \tilde{Y}_N]|\mathcal{F}_{N-1} \right) \\
    \leq I(T_N > N-1) \frac{\rho_{N+1}}{\rho_N} \mathbb{E} \left( I(B_{N,N}(N))[c - \tilde{Y}_N]|\mathcal{F}_{N-1} \right) \\
    = (c_{N-1,N} - c) I(T_N > N-1).
\end{align*}
\]
Assume that (24) holds for \( n \leq N - 1 \). It follows that

\[
\begin{align*}
\mathbb{E}_\infty \left( \sum_{m=n}^{N} \frac{\rho_{m+1}}{\rho_n} I(T_N > m)[c - \tilde{Y}_m]|\mathfrak{F}_{n-1} \right) \\
eq \frac{\rho_{n+1}}{\rho_n} \mathbb{E}_\infty (I(T_N > n)[c - \tilde{Y}_n]|\mathfrak{F}_{n-1}) \\
+ \frac{\rho_{n+1}}{\rho_n} \mathbb{E}_\infty \left( \mathbb{E}_\infty \left[ \sum_{m=n+1}^{N} \frac{\rho_{m+1}}{\rho_{n+1}} I(T_N > m)(c - \tilde{Y}_m)|\mathfrak{F}_n \right]|\mathfrak{F}_{n-1} \right) \\
\leq \frac{\rho_{n+1}}{\rho_n} \mathbb{E}_\infty (I(T_N > n)[c - \tilde{Y}_n]|\mathfrak{F}_{n-1}) \\
+ \frac{\rho_{n+1}}{\rho_n} \mathbb{E}_\infty \left( [c_{n,N} - c] I(T_N > n)|\mathfrak{F}_{n-1} \right) \\
= \frac{\rho_{n+1}}{\rho_n} \mathbb{E}_\infty \left( [c_{n,N} - \tilde{Y}_n] I(T_N > n)|\mathfrak{F}_{n-1} \right) \\
\leq \frac{\rho_{n+1}}{\rho_n} \mathbb{E}_\infty \left( I(B_{n,n}(N))[c_{n,N} - \tilde{Y}_n] I(T_N > n)|\mathfrak{F}_{n-1} \right) \\
\leq I(T_N > n - 1)(c_{n-1,N} - c).
\end{align*}
\]

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

\[
\begin{align*}
\mathbb{E}_\infty \left( \sum_{m=n+1}^{\infty} \frac{\rho_{m+1}}{\rho_{n+1}} I(T > m)[c - \tilde{Y}_m]|\mathfrak{F}_n \right) \leq (c_n - c) I(T > n)
\end{align*}
\]

for all \( n \geq 1 \). Thus

\[
\begin{align*}
I(T^*_n)I(T > n) \mathbb{E}_\infty (\tilde{\xi}_T - \tilde{\xi}_n)|\mathfrak{F}_n) \\
= I(T^*_n) = \sum_{m=n}^{\infty} \mathbb{E}_\infty (I(T > m)[\tilde{\xi}_{m+1} - \tilde{\xi}_m])|\mathfrak{F}_n) \\
= -\rho_{n+1}I(T^*_n)I(T > n)(\tilde{Y}_n - c) + \sum_{m=n+1}^{\infty} \mathbb{E}_\infty \left( \frac{\rho_{m+1}}{\rho_{n+1}} I(T > m)(\tilde{Y}_m - c)|\mathfrak{F}_n \right) \\
= \rho_{n+1}I(T^*_n)I[T^*_n = n][-I(T > n)\tilde{Y}_n + I(T > n)c + \sum_{m=n+1}^{\infty} \mathbb{E}_\infty \left( \frac{\rho_{m+1}}{\rho_{n+1}} I(T > m)(c - \tilde{Y}_m)|\mathfrak{F}_n \right)] \\
\leq \rho_{n+1}I(T^*_n)I[T^*_n = n][-I(T > n)\tilde{Y}_n + I(T > n)c_n] = \rho_{n+1}I(T^*_n = n)I(T > n)[c_n - \tilde{Y}_n] \leq 0
\end{align*}
\]

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

From (21) and (22) it follows that \( \mathbb{E}_\infty (\tilde{\xi}_T) \geq \mathbb{E}_\infty (\tilde{\xi}_{T^*_n}) \) for any finite test \( T \) with \( \mathbb{E}_\infty (\tilde{\xi}_T) < \infty \).
Since the probability
\[ P(T^*(c) \geq \tau) = \rho_1 + \sum_{k=2}^{\infty} \rho_k P_\infty(T^*(c) \geq 2) \]
is continuous and strictly increasing in \( c \), \( P(T^*(0) \geq \tau) = \rho_1 < 1 - \alpha \) for \( c = 0 \) and \( P(T^*(c) \geq \tau) > 1 - \alpha \) for a large \( c \), it follows that there exists a positive number \( c^* \) such that \( P(T^*(c^*) \geq \tau) = 1 - \alpha \) or \( P(T^*(c^*) < \tau) = \alpha \).

It follows from (16) and \( E_\infty(\xi_T) \geq E_\infty(\xi_T^*(c^*)) \) that
\[ E((T - \tau + 1)|T \geq \tau)P(T \geq \tau) + (1 + c^*)P(T < \tau) \geq (1 - \alpha)E((T^*(c^*) - \tau + 1)|T^*(c^*) \geq \tau) + (1 + c^*)\alpha. \]

Furthermore, by (12), (13) and (21) we have
\[
\mathcal{J}_S(T^*(c^*)) = E((T^*(c^*) - \tau + 1)|T^*(c^*) \geq \tau) = 1 + \frac{\sum_{k=1}^{\infty} E_\infty(Y_k I(T^*(c^*) \geq k + 1))}{\sum_{k=1}^{\infty} \rho_k P_\infty(T^*(c^*) \geq k)}
\[
\leq 1 + \frac{\sum_{k=1}^{\infty} c^* \rho_{k+1} P_\infty(T^*(c^*) \geq k + 1)}{\sum_{k=1}^{\infty} \rho_k P_\infty(T^*(c^*) \geq k)} = 1 + c^*(1 - \rho_1). \]

This is (19). Thus, if \( E((T - \tau + 1)|T \geq \tau) > 1 + c^* \), then
\[ E((T - \tau + 1)|T \geq \tau) > 1 + c^* \geq E((T^*(c^*) - \tau + 1)|T^*(c^*) \geq \tau). \]

If \( E((T - \tau + 1)|T \geq \tau) < 1 + c^* \), then \( E((T - \tau + 1)|T \geq \tau)(1 - x) + (1 + c^*)x \) is increasing in \( x = P(T < \tau) \), and therefore
\[ (1 - \alpha)E((T - \tau + 1)|T \geq \tau) + (1 + c^*)\alpha \geq E((T - \tau + 1)|T \geq \tau)(1 - \rho_1 P(T < \tau)) + (1 + c^*)P(T < \tau) \geq (1 - \alpha)E((T^*(c^*) - \tau + 1)|T^*(c^*) \geq \tau) + (1 + c^*)\alpha \]
since \( x = P(T < \tau) \leq \alpha \). This means that
\[ E((T - \tau + 1)|T \geq \tau) \geq E((T^*(c^*) - \tau + 1)|T^*(c^*) \geq \tau). \]

Thus, (18) holds for \( \hat{D}_\alpha \).
In order to prove that (18) also holds for $D_\alpha$, we define a set of stopping times as:

$$D_\alpha(M) = \{ T_M = T \wedge M : P_r(T < \tau) \leq \alpha \}.$$  

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

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

$$E_\infty(|Y_{T^*} - \sum_{k=1}^{T^*} \rho_k|) \leq E_\infty(\sum_{n=1}^{T^*} |Y_n - Y_{n-1} - \rho_n|)$$

$$\leq E_\infty(\sum_{n=1}^{T^*} (\Lambda_n + 1)(Y_{n-1} + \rho_n)) = \sum_{n=1}^{\infty} E_\infty((\Lambda_n + 1)(Y_{n-1} + \rho_n)I(T^* \geq n))$$

$$= E_\infty(\sum_{n=1}^{\infty} I(T^* \geq n)E_\infty(\Lambda_n + 1)(Y_{n-1} + \rho_n)|\mathcal{F}_{n-1}))$$

$$= 2E_\infty(\sum_{n=1}^{\infty} (Y_{n-1} + \rho_n)I(T^* \geq n))$$

$$\leq 2(c_1 + 1)\sum_{n=1}^{\infty} \rho_n P_\infty(T^* \geq n) < \infty$$

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

$$\lim_{n \to \infty} E_\infty(|Y_n - \sum_{k=1}^{n} \rho_k|, T^* > n)$$

$$\leq \lim_{n \to \infty} E_\infty(\sum_{k=1}^{T^*} (\Lambda_k + 1)(Y_{k-1} + \rho_k), T^* > n) = 0.$$  

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

$$\frac{1 - \alpha}{E_\infty(\rho_{T^*}) - 1} = \frac{E_\infty(Y_{T^* - 1})}{E_\infty(\rho_{T^*})} < c_1 \leq \frac{E_\infty(Y_{T^*})}{E_\infty(\rho_{T^* + 1})} = \frac{1 - \alpha}{E_\infty(\rho_{T^* + 1})}.$$
\textbf{Remark 1.} When the observation processes \( \{X_n, n \geq 0\} \) are mutually independent, we know that \( \{\tilde{Y}_m, m \geq 0\} \) is a Markov process. In this case, the positive random variables \( \{c_n\} \) become a series of positive numbers which can be written by

\[
c_n = c + \frac{\rho_{n+2}}{\rho_{n+1}} E_\infty(I(B_{n+1,n+1}[c_{n+1} - \tilde{Y}_{n+1}]) \tilde{Y}_n = c_n)
\]

\[
= c + E_\infty\left( \sum_{m=n+1}^{\infty} \frac{\rho_{m+1}}{\rho_{n+1}} I(B_{m,n+1})[c - \tilde{Y}_{m}]|\tilde{Y}_n = c_n) \right).
\]

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

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

\[
c_1 = c_n = c + \rho E_\infty(I(B_{n+1,n+1}[c_{n+1} - \tilde{Y}_{n+1}]) \tilde{Y}_n = c_n)
\]

\[
= c + E_\infty\left( \sum_{m=1}^{\infty} \rho^m I(B_{m+1,2})[c - \tilde{Y}_{m+1}]|\tilde{Y}_1 = c_1) \right)
\]

for all \( n \geq 1 \). Note that \( 1 - \alpha = 1 - E_\infty((1 - \rho)^T^*) \) and

\[
E_\infty(\rho T^{*+1}) = \sum_{k=1}^{\infty} \rho(1 - \rho)^k P_\infty(T^* = k) = \rho \alpha, \quad E_\infty(\rho T^*) = \frac{\rho \alpha}{1 - \rho}.
\]

It follows from Theorem 1 that

\[
\frac{1 - \alpha}{E_\infty(\rho T^*)} - 1 = \frac{1 - \alpha - \rho}{\rho \alpha} < c_1 \leq \frac{1 - \alpha}{\rho \alpha}.
\]

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 $c'''_1$ in the following respectively.

$$
\frac{1 - 0.05 - 0.7}{0.05 \times 0.7} \approx 7.14 < c'_1 < 27.14 \approx \frac{1 - 0.05}{0.05 \times 0.7}
$$

$$
\frac{1 - 0.1 - 0.6}{0.1 \times 0.6} = 5 < c''_1 < 15 = \frac{1 - 0.1}{0.1 \times 0.6}
$$

$$
\frac{1 - 0.2 - 0.5}{0.2 \times 0.5} = 3 < c'''_1 < 8 = \frac{1 - 0.2}{0.2 \times 0.5}.
$$

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 $E_\infty(T) \geq \gamma > 1$.

Consider the loss random variable of detection, $\varpi_n$, at time $n$ in the following:

$$
\varpi_n = \sum_{k=1}^{n} Y_{k-1} + 1 + c(1 - \sum_{k=1}^{n} \rho_k) - bn
$$

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

$$(27) \quad E_\infty(\varpi_T)$$

$$
= E_\infty(\sum_{k=1}^{T}[Y_{k-1} + \rho_k]) + (c + 1)(1 - E_\infty(\sum_{k=1}^{T} \rho_k)) - bE_\infty(T)
$$

$$
= \sum_{k=1}^{\infty} \rho_k E_k(T - k + 1)^+ + (c + 1)(1 - \sum_{k=1}^{\infty} \rho_k P_\infty(T \geq k)) - bE_\infty(T)
$$

$$
= E(T - \tau + 1|T \geq \tau)P(T \geq \tau) + (c + 1)P(T < \tau) - bE_\infty(T).
$$

We present a test:

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

where $\{c_n(c, b), n \geq 1\}$ 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 $k \geq 1$ and $E_\infty(T_\gamma(c, b_0)) < \infty$ for $c > 0$ and some number $b_0 \geq 0$. Then there exist two numbers $b^*, c^*$ and a
series of positive random variables \( c_n(c^*, b^*) \) such that \( E_\infty(T_\gamma(c^*, b^*)) = \gamma \), \( b^* \leq \rho_1 \min\{c^*/(\gamma - 1), b_0, 1\} \),

\[
c_n(c^*, b^*) = c^* + b^*/\rho_{n+1} + E_\infty\left( \sum_{m=n+1}^{\infty} \frac{\rho_{m+1}}{\rho_{n+1}} I(B_{m,n+1})[c^* + b^*/\rho_{m+1} - \bar{Y}_m] \right).
\]

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

\[
\inf_{T \in \hat{D}_\gamma} J_S(T) = J_S(T_\gamma(c^*, b^*))
\]

where

\[
D_\gamma = \{ T \geq 1 : E_\infty(T) \geq \gamma; \ P(T < \tau) \leq \alpha^* \},
\]

\[
B_{m,n+1} = \{ \bar{Y}_k \leq c_k(c^*, b^*) + b^*/\rho_{n+1}, n+1 \leq k \leq m \} \text{ and } \alpha^* = P(T_\gamma(c^*, b^*) < \tau).
\]

Moreover

\[
J_S(T_\gamma(c^*, b^*)) \leq 1 + c^*(1 - \rho_1) + b^*(\gamma - 1)).
\]

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

\[
\hat{D}_\gamma = \{ T \geq 1 : E_\infty(T) \geq \gamma; \ P(T < \tau) \leq \alpha^*; \ E_\infty(|\varpi_T|) < \infty \}.
\]

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

\[
E_\infty(|\varpi_T|) \geq E_\infty(|\varpi_{T_\gamma(c, b)}|)
\]

for any finite test \( T \) and \( E_\infty(|\varpi_T|) < \infty \).

On the other hand, we know that \( E_\infty(T_\gamma(c, b)) \) is continuous and strictly increasing in \( c \) and \( b \) with \( E_\infty(T_\gamma(0, 0)) = 1 \) and

\[
E_\infty(T_\gamma(c, b)) \nearrow \infty
\]

as \( c \nearrow \infty \) for every fixed \( b \geq 0 \). This means that for every small \( b^* \leq b_0 \), there exists a unique positive number \( c^* > 0 \) such that

\[
E_\infty(T_\gamma(c^*, b^*)) = \gamma.
\]

Obviously, the number \( c^* = c^*(b^*) \) is monotonically decreasing in \( b^* \). Hence, we can take a small \( b^* \) such that \( b^* \leq \rho_1 \min\{c^*/(\gamma - 1), b_0, 1\} \) and therefore, \( b^* \leq \rho_1 \leq 1 - \alpha^* \).
Let $T^* = T_\gamma(c^*, b^*)$. It follows from (26), (27) and (31) that

$$E(T - \tau + 1 | T \geq \tau) P(T \geq \tau) + (1 + c^*) P(T < \tau) - b E_\infty(T)$$

$$\geq E(T^* - \tau + 1 | T^* \geq \tau) P(T^* \geq \tau) + (c^* + 1) P(T^* < \tau) - b \gamma,$$

and therefore

$$E(T - \tau + 1 | T \geq \tau) P(T \geq \tau) + (1 + c^*) P(T < \tau)$$

$$\geq E(T^* - \tau + 1 | T^* \geq \tau) P(T^* \geq \tau) + (c^* + 1) P(T^* < \tau)$$

since $E_\infty(T) \geq \gamma$.

Furthermore,

$$E(T^* - \tau + 1 | T^* \geq \tau) = 1 + \frac{\sum_{k=1}^{\infty} E_\infty(Y_k I(T^* \geq k + 1))}{\sum_{k=1}^{\infty} \rho_k P_\infty(T^* \geq k)}$$

$$\leq 1 + \frac{\sum_{k=1}^{\infty} (c^* \rho_k + b^*) P_\infty(T^* \geq k + 1)}{\sum_{k=1}^{\infty} \rho_k P_\infty(T^* \geq k)}$$

$$= 1 + c^* - \frac{c^* \rho_1 - b^*(\gamma - 1)}{\sum_{k=1}^{\infty} \rho_k P_\infty(T^* \geq k)}$$

$$\leq 1 + c^* - (c^* \rho_1 - b^*(\gamma - 1)).$$

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

$$E(T - \tau + 1 | T \geq \tau) > 1 + c^* \geq E(T^* - \tau + 1 | T^* \geq \tau).$$

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

$$E(T - \tau + 1 | T \geq \tau) (1 - x) + (1 + c^*) x$$

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

$$E(T - \tau + 1 | T \geq \tau) P(T \geq \tau) + (1 + c^*) P(T^* < \tau)$$

$$\geq E(T - \tau + 1 | T \geq \tau) P(T \geq \tau) + (1 + c^*) P(T < \tau)$$

$$\geq E(T^* - \tau + 1 | T^* \geq \tau) P(T^* \geq \tau) + (c^* + 1) P(T^* < \tau).$$

and therefore

$$E(T - \tau + 1 | T \geq \tau) \geq E(T^* - \tau + 1 | T^* \geq \tau).$$

Thus, (29) is true. \qed

By using Theorem 2 we can obtain the following corollary.
COROLLARY 1. Let $\beta = \{\beta_k, k \geq 1\}$ satisfying $0 < \beta_{k+1} \leq \beta_k, k\sqrt{\beta_k} \geq (k-1)\sqrt{\beta_k}$ for $k \geq 1$ and $\sum_{k=1}^{\infty} \beta_k = \infty$. Then, the following stopping time

$$T_\gamma(\beta) = \min \left\{ n \geq 1 : \sum_{k=1}^{n} \frac{\beta_k}{\beta_{n+1}} \prod_{i=k}^{n} \frac{p_{\nu_1}(X_{i-1}, X_i)}{p_{\nu_0}(X_{i-1}, X_i)} \geq c_\beta(\gamma) \right\}$$

is optimal in the sense that

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

or equivalently,

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

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

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

$$\inf_{T \in \Delta_\gamma} J_S(T) = J_S(T_\gamma(c^*, 0)),$$

where $E^\infty(T_\gamma(c^*, 0)) = \gamma$ 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 $\Lambda_k = \rho_{\nu_1}(X_{k-1}, X_k)/\rho_{\nu_0}(X_{k-1}, X_k)$, and $\rho_t/\rho_{n+1} \rightarrow \beta_t/\beta_{n+1}$ as $a \to 0$. It follows that

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

Moreover,

$$\frac{\rho_t}{\rho_{n+1}} = \frac{(a(t-1) + \beta_t^{-1/2})^2}{(an + \beta_{n+1}^{-1/2})^2}$$
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 $E_\infty(T_\gamma(c^*, 0)) = \gamma$. 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 $P(T < \tau) \leq \alpha^*$ it follows that

$$P(T \geq \tau) = \sum_{k=1}^\infty \rho_k P_\infty(T \geq k)$$

$$\geq 1 - \alpha^* = \sum_{k=1}^\infty \rho_k P_\infty(T_\gamma(c^*, 0) \geq k).$$

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

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

and

$$\lim_{a \to 0} J_S(T) = \lim_{a \to 0} \frac{\sum_{k=1}^\infty \rho_k E_k(T - k + 1)^+}{\sum_{k=1}^\infty \rho_k P_\infty(T \geq k)} = \frac{\sum_{k=1}^\infty \beta_k E_k(T - k + 1)^+}{\sum_{k=1}^\infty \beta_k P_k(T \geq k)}.$$

Thus, we have

$$\sum_{k=1}^\infty \beta_k E_k(T_\gamma(\beta) - k + 1)^+ - \sum_{k=1}^\infty \beta_k E_k(T - k)^+ = \sum_{k=1}^\infty \beta_k P_k(T \geq k),$$

Furthermore, by $\sum_{k=1}^\infty \beta_k E_k(T - k + 1)^+ - \sum_{k=1}^\infty \beta_k E_k(T - k)^+ = \sum_{k=1}^\infty \beta_k P_k(T \geq k)$, we have also

$$\sum_{k=1}^\infty \beta_k E_k(T_\gamma(\beta) - k)^+ - \sum_{k=1}^\infty \beta_k E_k(T - k)^+ = \inf_{T \in \Delta_\gamma} \frac{\sum_{k=1}^\infty \beta_k E_k(T - k)^+}{\sum_{k=1}^\infty \beta_k P_k(T \geq k)}.$$

\[ \square \]

Remark 2. \textit{Note that $\sum_{k=1}^\infty P_k(T \geq k) = E_\infty(T)$. 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}

$$\frac{\sum_{k=1}^\infty E_k(T_\gamma(\beta) - k)^+}{\gamma} = \inf_{T \in \Delta_\gamma} \frac{\sum_{k=1}^\infty E_k(T - k)^+}{E_\infty(T)}$$

$$\leq \inf_{T \in \Delta_\gamma} \frac{\sum_{k=1}^\infty E_k(T - k)^+}{\gamma}$$

$$\leq \frac{\sum_{k=1}^\infty E_k(T_\gamma(\beta) - k)^+}{\gamma}.$$
and therefore,
\[ \sum_{k=1}^{\infty} E_k(T_\gamma(\beta) - k)^+ = \inf_{T \in \Delta} \sum_{k=1}^{\infty} E_k(T - k)^+ \]

Thus, the result in (8) 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, \( T_\gamma(\beta) = T_{SR}(c_\gamma) \)

Next we will consider the optimal test for the case "\( P(T < \tau) \geq \alpha^* \)" which is just the opposite of "\( P(T < \tau) \leq \alpha^* \)" in \( D_\gamma \).

Consider the following loss random variables of detection:

\[ \eta_n = \sum_{k=1}^{n} Y_{k-1} + 1 - cn \]

where \( c > 0 \) is a constant. Note that \( \eta_n \) may be negative.

By Lemma 1 we have

\[ E_\infty(\eta_T) = E_\infty(\sum_{k=1}^{T} (Y_{k-1} + \rho_k) + 1 - E_\infty(\sum_{k=1}^{T} \rho_k) - cE_\infty(T) \]

\[ = E(T - \tau + 1|T \geq \tau)P(T \geq \tau) + P(T < \tau) - cE_\infty(T). \]

Let

\[ T_\eta(c) = \min\{n \geq 1 : Y_n \geq c_n(c)\} \]

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

**Theorem 3.** Let \( \gamma > 1 \) and \( E_\infty(T_\eta(c_0) < \infty \) for some number \( c_0 > 0 \). There exists a positive number \( c^* \) and a series of positive random variables \( c_n(c^*) \) such that \( E_\infty(T_\eta(c^*)) = \gamma \) and

\[ c_n(c^*) = c^*/\rho_{n+1} + E_\infty\left( \sum_{m=n+1}^{\infty} \frac{\rho_{m+1}}{\rho_n} I(B_{m,n+1})[c^*/\rho_{m+1} - \tilde{Y}_m]|\mathcal{F}_n \right). \]
(36) \[ \inf_{T \in D'_\gamma} \mathcal{J}_S(T) = \mathcal{J}_S(T_\eta(c^*)) , \]

where

\[ D'_\gamma = \{ T \geq 1 : E_\infty(T) \geq \gamma ; \quad P(T < \tau) \geq \alpha^* \} \]

and \( \alpha^* = P(T_\eta(c^*) < \tau) \).

\[ \text{Proof.} \quad \text{It needs only to be proved that (36) holds for the following subset of } D'_\alpha : \]

\[ \hat{D}'_\gamma = \{ T \geq 1 : E_\infty(T) \geq \gamma ; \quad P(T < \tau) \geq \alpha^* ; \quad E_\infty(|\eta_T|) < \infty \} . \]

Let \( \tilde{\eta}_n = -\tilde{\eta}_n \) for \( n \geq 1 \). By the same way of proving Theorem 1 we can prove \( E_\infty(\eta_T) \geq E_\infty(\eta_{\tilde{\eta}_n}) \) for all \( T \) satisfying \( E_\infty(|\eta_T|) < \infty \).

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

\[ E(T - \tau + 1|T \geq \tau)P(T \geq \tau) + P(T < \tau) - c^*E_\infty(T) \]

\[ \geq E(T_\eta(c^*) - \tau + 1|T_\eta(c^*) \geq \tau)P(T_\eta(c^*) \geq \tau) + P(T_\eta(c^*) < \tau) - c^*\gamma. \]

Moreover, \( E(T - \tau + 1|T \geq \tau) \geq 1 \) and therefore

\[ (E(T - \tau + 1|T \geq \tau) - 1)P(T_\eta(c^*) \geq \tau) + 1 - c^*E_\infty(T) \]

\[ \geq (E(T - \tau + 1|T \geq \tau) - 1)P(T \geq \tau) + 1 - c^*E_\infty(T) \]

\[ = E(T - \tau + 1|T \geq \tau)P(T \geq \tau) + P(T < \tau) - c^*E_\infty(T) \]

\[ \geq E(T_\eta(c^*) - \tau + 1|T_\eta(c^*) \geq \tau)P(T_\eta(c^*) \geq \tau) + P(T_\eta(c^*) < \tau) - c^*\gamma \]

\[ = (E(T_\eta(c^*) - \tau + 1|T_\eta(c^*) \geq \tau) - 1)P(T_\eta(c^*) \geq \tau) + 1 - c^*\gamma \]

for \( P(T_\eta(c^*) \geq \tau) \geq P(T \geq \tau) \). Thus

\[ E(T - \tau + 1|T \geq \tau) \geq E(T_\eta(c^*) - \tau + 1|T_\eta(c^*) \geq \tau) \]

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 \( \tau \) is at most finite, i.e., \( \tau \leq N < \infty \). Its distribution \( \{ \rho_k, k \geq 1 \} \) satisfies \( \sum_{k=1}^{N} \rho_k = 1 \) and \( \rho_k = 0 \) for \( k \geq N + 1 \). We can think of the distribution
of the change-point $\tau$ 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, $P(T > M) > 0$ for any
large $M > 0$.
Since $p_{v_1}(x, . )/p_{v_0}(x, . )$ has no atoms with respect to $P_\infty$ for all $x \in E$, it
follows that
$$
E_\infty(I(\Lambda_1 < 1)|X_0 = x) < 1, \quad E_\infty(I(\Lambda_1 < 1)|X_0 = x) < 1
$$
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, $\Lambda_1 = e^{\mu_1(X_1-X_0-\mu_1)/\sigma^2}$, where $\mu_1 > 0$. Then
$$
E_\infty(I(\Lambda_1 < 1)|X_0 = x) = E_\infty(I(X_1 - x - \mu_1/2 < 0)|X_0 = x) = \int_{-\infty}^{\mu_1/2\sigma}(\sqrt{2\pi})^{-1}e^{-u^2/2}du < 1
$$
$$
E_\infty(I(\Lambda_1 \geq 1)|X_0 = x) = \int_{\mu_1/2\sigma}^{\infty}(\sqrt{2\pi})^{-1}e^{-u^2/2}du < 1
$$
for all $x$. Let
$$
A = \sup_{x \in E} E_\infty(I(\Lambda_1 < 1)|X_0 = x), \quad B = \sup_{x \in E} E_\infty(I(\Lambda_1 \geq 1)|X_0 = x) < 1
$$
$$
C = \sup_{x \in E} E_\infty(|\ln \Lambda_1|^2|X_0 = x).
$$
Under the conditions that $A, B < 1$ and $C < \infty$. Theorem ?? below shows
that if only the restrictive condition $E_\infty(T) \geq \gamma$ exists, then
$$
\inf_{T: E_\infty(T) \geq \gamma} J_S(T) = 1
$$
for every finite prior distribution $\{\rho_k, 1 \leq k \leq N\}$ of the change-point $\tau$,
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 $E_\infty(T_N(c)) \geq \gamma$,

$$
\lim_{c \to 0} J_S(T_N(c)) = 1, \quad \lim_{c \to 0} E_\infty(T_N^2(c)) = \infty,
$$
and therefore, $\inf_{T: E_\infty(T) \geq \gamma} J_S(T) = 1$ for every finite prior distribution
$\{\rho_k, 1 \leq k \leq N\}$ of change-point $\tau$. 
Proof. The proof of Theorem 4 in shown in "Supplementary Materials for AOS"

Now we consider the probability of false alarm \( P_\infty(T < \tau) \). Since

\[
P_\infty(T < \tau) = \sum_{k=1}^{N} \rho_k P_\infty(T < k) = 1 - \rho_1 - \sum_{k=2}^{N} \rho_k P_\infty(T \geq k) \leq 1 - \rho_1,
\]

it follows that \( P_\infty(T < \tau) < 1 - \rho_1 \) if and only if \( E_\infty(T) = 1 + \sum_{k=2}^{N} P_\infty(T \geq k) > 1 \) for \( \rho_k > 0, 2 \leq k \leq N \). Thus we have the following corollary.

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

\[
\inf_{\{T: P_\infty(T < \tau) < 1 - \rho_1\}} J_S(T) = 1
\]

for every finite positive prior distribution \( \{\rho_k, 1 \leq k \leq N\} \) of the change-point \( \tau \)

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

\[
J_S(T_5(c)) = 1 + o(1)
\]

and \( E_\infty(T_5^2(c)) \) 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 \( J_S(.) \) of the tests \( T_5(c), T_C(c) \) and \( T_E(c) \) with \( ARL_0 \approx 1000 \) 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 \( ARL_0(T) = E_\infty(T) \),

\[
J_S(T) = \frac{\sum_{k=1}^{5} E_{k,\mu}(T - k + 1)^+}{\sum_{k=1}^{5} P_\infty(T \geq k)}
\]

and \( E_{k,\mu}(T - k + 1)^+ \) denotes the out-control average run length (ARL) from from 0 to \( \mu \) at the change-point \( k \). The numerical simulation results of \( ARL_0 \) and \( J_S(.) \) in the table were obtained based on \( 10^7 \) 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 \( \tau \), \( X_k \sim N(\mu, 1) \) 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:

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

where \( c = e^{c'} \) and

\[
\begin{align*}
Z'_1 &= Z_1 - \mu_0, \quad Z'_2 = Z_2 - \mu_0 + \ln(1 + e^{Z_1}) \\
Z'_3 &= Z_3 - \mu_0 + \ln(1 + e^{Z_2 + Z_1}), \\
Z'_4 &= Z_4 - \mu_0 + \ln(1 + e^{Z_3 + Z_2 + e^{Z_1+Z_2}}) \\
Z'_5 &= Z_5 - \mu_0 + \ln(1 + e^{Z_4 + Z_3 + e^{Z_1+Z_2+Z_1}}).
\end{align*}
\]

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

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

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 \( \lambda \) 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 \( \lambda = 0.1 \).
ON THE OPTIMALITY OF BAYSIAN CHANGE-POINT DETECTION

Table 1

\( J_S(T) \) with \( ARL_0 \approx 1000 \)

<table>
<thead>
<tr>
<th>shift in ( \mu )</th>
<th>0.0</th>
<th>0.1</th>
<th>0.25</th>
<th>0.5</th>
<th>0.75</th>
<th>1.0</th>
<th>1.25</th>
<th>1.5</th>
<th>2.00</th>
<th>3.00</th>
</tr>
</thead>
<tbody>
<tr>
<td>( r_1 = 0.00166 )</td>
<td>1006.95</td>
<td>4.46</td>
<td>2.72</td>
<td>1.95</td>
<td>1.63</td>
<td>1.44</td>
<td>1.31</td>
<td>1.22</td>
<td>1.09</td>
<td>1.01</td>
</tr>
<tr>
<td>( c' = -0.6 )</td>
<td>(164713.92)</td>
<td>(21.40)</td>
<td>(5.69)</td>
<td>(2.59)</td>
<td>(1.83)</td>
<td>(1.49)</td>
<td>(1.32)</td>
<td>(1.20)</td>
<td>(0.97)</td>
<td>(0.83)</td>
</tr>
<tr>
<td>( r_1 = 0.00344 )</td>
<td>1002.15</td>
<td>6.47</td>
<td>3.48</td>
<td>2.39</td>
<td>1.92</td>
<td>1.64</td>
<td>1.45</td>
<td>1.34</td>
<td>1.18</td>
<td>1.03</td>
</tr>
<tr>
<td>( c' = -0.2 )</td>
<td>(82057.39)</td>
<td>(26.64)</td>
<td>(6.89)</td>
<td>(2.82)</td>
<td>(1.84)</td>
<td>(1.50)</td>
<td>(1.30)</td>
<td>(1.12)</td>
<td>(0.90)</td>
<td>(0.74)</td>
</tr>
<tr>
<td>( r_1 = 0.00463 )</td>
<td>995.04</td>
<td>7.65</td>
<td>3.91</td>
<td>2.63</td>
<td>2.09</td>
<td>1.77</td>
<td>1.55</td>
<td>1.40</td>
<td>1.22</td>
<td>1.06</td>
</tr>
<tr>
<td>( c' = 0.0 )</td>
<td>(62621.72)</td>
<td>(28.80)</td>
<td>(7.38)</td>
<td>(3.06)</td>
<td>(1.93)</td>
<td>(1.49)</td>
<td>(1.25)</td>
<td>(1.10)</td>
<td>(0.88)</td>
<td>(0.65)</td>
</tr>
<tr>
<td>( r_1 = 0.00744 )</td>
<td>1003.42</td>
<td>10.37</td>
<td>4.94</td>
<td>3.06</td>
<td>2.44</td>
<td>2.04</td>
<td>1.77</td>
<td>1.58</td>
<td>1.40</td>
<td>1.11</td>
</tr>
<tr>
<td>( c' = -0.2 )</td>
<td>(40856.77)</td>
<td>(32.86)</td>
<td>(8.42)</td>
<td>(3.23)</td>
<td>(2.06)</td>
<td>(1.52)</td>
<td>(1.25)</td>
<td>(1.08)</td>
<td>(0.85)</td>
<td>(0.58)</td>
</tr>
<tr>
<td>( r_1 = 0.01078 )</td>
<td>1002.66</td>
<td>13.34</td>
<td>6.09</td>
<td>3.60</td>
<td>2.74</td>
<td>2.32</td>
<td>2.01</td>
<td>1.78</td>
<td>1.47</td>
<td>1.17</td>
</tr>
<tr>
<td>( c' = 0.4 )</td>
<td>(82057.39)</td>
<td>(26.64)</td>
<td>(6.89)</td>
<td>(2.82)</td>
<td>(1.84)</td>
<td>(1.50)</td>
<td>(1.30)</td>
<td>(1.12)</td>
<td>(0.90)</td>
<td>(0.74)</td>
</tr>
<tr>
<td>( r_1 = 0.01813 )</td>
<td>1003.58</td>
<td>19.82</td>
<td>8.72</td>
<td>4.87</td>
<td>3.55</td>
<td>2.87</td>
<td>2.46</td>
<td>2.20</td>
<td>1.81</td>
<td>1.36</td>
</tr>
<tr>
<td>( c' = 1.6 )</td>
<td>(17225.21)</td>
<td>(43.10)</td>
<td>(11.39)</td>
<td>(4.02)</td>
<td>(2.98)</td>
<td>(2.17)</td>
<td>(1.60)</td>
<td>(1.25)</td>
<td>(1.09)</td>
<td>(0.83)</td>
</tr>
<tr>
<td>( r_1 = 0.03347 )</td>
<td>1004.85</td>
<td>33.96</td>
<td>14.74</td>
<td>7.80</td>
<td>5.50</td>
<td>4.34</td>
<td>3.60</td>
<td>3.10</td>
<td>2.51</td>
<td>1.90</td>
</tr>
<tr>
<td>( c' = 3.2 )</td>
<td>(988.56)</td>
<td>(43.10)</td>
<td>(11.39)</td>
<td>(4.02)</td>
<td>(2.98)</td>
<td>(2.17)</td>
<td>(1.60)</td>
<td>(1.25)</td>
<td>(1.09)</td>
<td>(0.83)</td>
</tr>
<tr>
<td>( r_1 = 0.0665 )</td>
<td>1000.47</td>
<td>63.01</td>
<td>27.12</td>
<td>14.13</td>
<td>9.74</td>
<td>7.36</td>
<td>6.04</td>
<td>5.21</td>
<td>4.08</td>
<td>2.89</td>
</tr>
<tr>
<td>( c' = 6.4 )</td>
<td>(912.11)</td>
<td>(54.93)</td>
<td>(14.56)</td>
<td>(5.32)</td>
<td>(2.98)</td>
<td>(2.43)</td>
<td>(1.60)</td>
<td>(1.25)</td>
<td>(1.25)</td>
<td>(0.87)</td>
</tr>
<tr>
<td>( r_1 = 0.12773 )</td>
<td>1000.47</td>
<td>117.17</td>
<td>51.03</td>
<td>26.45</td>
<td>18.16</td>
<td>13.56</td>
<td>11.03</td>
<td>9.44</td>
<td>6.97</td>
<td>4.99</td>
</tr>
<tr>
<td>( c' = 12.80000 )</td>
<td>(2406.36)</td>
<td>(73.22)</td>
<td>(19.71)</td>
<td>(7.19)</td>
<td>(4.02)</td>
<td>(2.68)</td>
<td>(1.99)</td>
<td>(1.57)</td>
<td>(1.09)</td>
<td>(0.68)</td>
</tr>
<tr>
<td>( r_1 = 0.25781 )</td>
<td>1000.47</td>
<td>208.35</td>
<td>95.08</td>
<td>49.54</td>
<td>34.82</td>
<td>25.54</td>
<td>20.77</td>
<td>18.14</td>
<td>13.03</td>
<td>9.34</td>
</tr>
<tr>
<td>( c' = 25.60000 )</td>
<td>(1201.08)</td>
<td>(34.89)</td>
<td>(9.85)</td>
<td>(4.58)</td>
<td>(3.61)</td>
<td>(2.65)</td>
<td>(2.07)</td>
<td>(1.43)</td>
<td>(0.84)</td>
<td>(0.68)</td>
</tr>
<tr>
<td>( r_1 = 0.52120 )</td>
<td>1000.47</td>
<td>242.32</td>
<td>113.01</td>
<td>60.29</td>
<td>41.22</td>
<td>32.16</td>
<td>24.80</td>
<td>20.92</td>
<td>16.99</td>
<td>10.72</td>
</tr>
<tr>
<td>( c' = 31.50000 )</td>
<td>(980.29)</td>
<td>(116.60)</td>
<td>(37.35)</td>
<td>(14.48)</td>
<td>(8.17)</td>
<td>(5.42)</td>
<td>(4.06)</td>
<td>(3.09)</td>
<td>(2.14)</td>
<td>(1.19)</td>
</tr>
</tbody>
</table>

The simulation results of \( J_S(T_5(c)) \), \( J_S(T_C(c)) \) and \( J_S(T_E(c)) \) 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 \( J_S(T_5(c)) \) before detecting the mean shifts. For example, the standard deviation in the first row is 164713.92 with \( ARL_0 = 1006.95 \), but the average detection delay is only 4.46 for a small mean shift \( \mu = 0.1 \). When the standard deviation is 980.55 with \( ARL_0 = 999.09 \) (see thirteenth row), the average detection delay becomes 242.32 for \( \mu = 0.1 \). (2) \( T_5(c) \) performs better than \( T_C \) and \( T_E \) for a small mean shift \( \mu = 0.1 \) 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 \( \mu \) 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 $X_0, X_1, X_2, \ldots$ that form a Markov process under the restrictive conditions that \{\(\mathbb{P}_\infty(T < \tau) \leq \alpha\)\}, \{\(\mathbb{E}_\infty(T) \geq \gamma\), \(\mathbb{P}(T < \tau) \leq \alpha^*\)\} and \{\(\mathbb{E}_\infty(T) \geq \gamma\), \(\mathbb{P}(T < \tau) \geq \alpha^*\)\}.

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

$$\inf_{T: \mathbb{E}_\infty(T) \geq \gamma} J_S(T) = 1$$

for a finite number of change-points $\tau$.

It follows from Theorem 4 and the numerical simulations in Section 4 that the smaller the value of $J_S(T_N(c))$, the larger the variance $\mathbb{Var}_\infty(T_N(c))$ for a given finite false alarm rate $\mathbb{E}_\infty(T_N(c)) \geq \gamma$. 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 \{\(\mathbb{E}_\infty(T) \geq \gamma\), \(\mathbb{Var}_\infty(T) \leq \sigma^2\)\}. An interesting problem would be to find a test $T_N(c^*)$ that is optimal in the sense that

$$\inf_{\{T: \mathbb{E}_\infty(T) \geq \gamma, \mathbb{Var}_\infty(T) \leq \sigma^2\}} J_S(T) = J_S(T_N(c^*))$$

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

References.


**SUPPLEMENTARY MATERIAL**

**Supplement A: PROOFS OF THEOREM 4 OF THE PAPER**

**ON THE OPTIMALITY OF BAYESIAN CHANGE-POINT DETECTION**

( ; .pdf).

**Department of Mathematics**
**Department of Industrial Engineering and Logistics Management**
Shanghai Jiao Tong University
Shanghai 200030, P. R. China
Hong Kong University of Science and Technology
E-mail: dounghan@sjtu.edu.cn
E-mail: season@ust.hk
jgxian@sjtu.edu.cn