Least Favorable Distributions for Robust Quickest Change Detection

Jayakrishnan Unnikrishnan, Venugopal V. Veeravalli, Sean Meyn
Department of Electrical and Computer Engineering, and Coordinated Science Laboratory
University of Illinois at Urbana-Champaign
Email: {junnikr2, vvv, meyn}@illinois.edu

***Abstract*—We study the problem of robust quickest change detection where the pre-change and post-change distributions are not known exactly but belong to known uncertainty classes of distributions. Both Bayesian and minimax versions of the quickest change detection problem are considered. When the uncertainty classes satisfy some specific conditions, we identify least favorable distributions (LFD's) from the uncertainty classes, and show that the detection rule designed for the LFD's is optimal in a minimax sense. The condition is similar to that required for the existence of LFD's for the robust hypothesis testing problem studied by Huber.**

I. Introduction

A dynamic hypothesis testing problem with a rich set of applications is that of detecting an abrupt change in a system based on observations. Change-point detection was first studied by Page nearly fifty years ago in the context of quality control [1]. In its standard formulation, there is a sequence of observations whose distribution changes at some unknown point in time. The goal is to detect this change as soon as possible, subject to false alarm constraints. Some applications of change detection are: intrusion detection in computer networks and security systems, detecting faults in infrastructure of various kinds, and spectrum monitoring for opportunistic access.

Most of the past work in the area of change detection has been restricted to the setting where the distributions of the observations prior to the change and after the change are known exactly. (see, e.g., [2], [3], [4], [5]; For an overview of the work in this area, see [6], [7] and [8].) The two most popular formulations for optimizing the tradeoff between the detection delay and the false alarms are Lorden's criterion [3], in which the change point is a deterministic quantity, and Shiryaev's Bayesian formulation [9], in which the change-point is modeled as a random variable with a known prior distribution. In this paper, we study both these versions of change detection, under the setting where the pre-change and post-change distributions are not known exactly but belong to known uncertainty classes. We pose a minimax robust version of the standard quickest change detection problem wherein the objective is to identify the change detection rule that minimizes the maximum delay over all possible distributions. This minimization should be done while meeting the false alarm constraint for all possible values of the unknown distributions. We obtain a solution to this problem when the uncertainty classes satisfy some specific conditions. Under these conditions we can identify Least Favorable Distributions (LFD's) from the uncertainty classes and the optimal robust change detection rule is the optimal non-robust change detection rule when the underlying distributions are the LFD's.

Although there has been some prior work on robust change detection, the approaches are distinctly different from ours. The maximin approach of [10] is similar in that they also identify LFD's for the robust problem. However, their result is restricted to asymptotic optimality (as the false alarm constraint goes to zero) under the Lorden criterion. A similar formulation is also discussed in [11, sec. 7.3.1]. Other approaches (e.g. [12]) are used to develop algorithms for quickest change-detection with unknown distributions, whose provable optimality properties are again restricted to the asymptotic setting.

The rest of the paper is organized as follows. We describe the exact problem that we are studying in Section II and the solution in Section III. We discuss some examples in Section IV and conclude in Section V.

II. Problem Statement

In the online quickest change detection problem, we are given observations from a sequence $\{X_n : i = 1, 2, \ldots\}$ taking values in a set $\mathcal{X}$. There are two known distributions $\nu_0, \nu_1 \in \mathcal{P}(\mathcal{X})$ where $\mathcal{P}(\mathcal{X})$ is the set of probability distributions on $\mathcal{X}$. Initially, the observations are drawn i.i.d. under distribution ν_0. Their distribution switches abruptly to ν_1 at some unknown time λ so that $X_n \sim \nu_0$ for $n \leq \lambda - 1$ and $X_n \sim \nu_1$ for $n \geq \lambda$. It is assumed that the observations are stochastically independent conditioned on the change-point. The objective is to identify the occurrence of change with minimum delay subject to false alarm constraints. We use E_m^{ν} to denote the expectation operator and P_m^{ν} to denote the probability operator when the change happens at m and the pre-change and post-change distributions are ν_0 and ν_1 respectively. The symbols are replaced with E_∞^{ν} and P_∞^{ν} when the change does not happen. Similarly, if the pre-change and post-change distributions are some μ and γ, respectively, and the change happens at m, we use $\mathsf{E}_m^{\mu,\gamma}$ to denote the expectation operator and $\mathsf{P}_m^{\mu,\gamma}$. We further use $\mathcal{F}_m$ to denote the sigma algebra generated by $(X_1, X_2, \ldots, X_m)$.

A sequential change-point detection procedure is characterized by a stopping time τ with respect to the observation sequence. The design of the quickest change-point detection procedure involves optimizing the tradeoff between two performance measures: detection delay, and frequency of false alarms. There are two standard mathematical formulations for the optimal tradeoff. In the minimax formulation of Lorden [3] the change-point is assumed to be an unknown deterministic quantity. The worst-case detection delay is defined as,

$$\text{WDD}(\tau) = \sup_{\lambda \geq 1} \text{ ess sup } \mathsf{E}_\lambda^\nu[(\tau - \lambda + 1)^+ | \mathcal{F}_{\lambda-1}]$$

where $x^+ = \max(x, 0)$, and the false alarm rate is defined as,

$$\text{FAR}(\tau) = \frac{1}{\mathsf{E}_\infty^\nu[\tau]}.$$

Here $\mathsf{E}_\infty^\nu[\tau]$ can be interpreted as the mean time between false alarms. Under the Lorden criterion, the objective is to find the stopping rule that minimizes the worst-case delay subject to an upper bound on the false alarm rate:

$$\text{Minimize WDD}(\tau) \text{ subject to FAR}(\tau) \leq \alpha \tag{1}$$

It was shown by Moustakides in [2] that the optimal solution to (1) is given by the cumulative sum (CUSUM) test proposed by Page [1]. We describe this test later in the paper.

The other approach to change detection is the Bayesian formulation of Shiryaev [9], [5]. Here the change point is modeled as a random variable Λ with prior probability distribution, $\pi_k = \mathsf{P}(\Lambda = k), k = 1, 2, \ldots$. The performance measures are the average detection delay (ADD) and probability of false alarm (PFA) defined by:

$$\text{ADD}(\tau) = \mathsf{E}^\nu[(\tau - \Lambda)^+], \qquad \text{PFA}(\tau) = \mathsf{P}^\nu(\tau < \Lambda)$$

when the pre-change and post-change distributions are ν_0 and ν_1 respectively. For a given $\alpha \in (0, 1)$, the optimization problem under the Bayesian criterion is:

$$\text{Minimize ADD}(\tau) \text{ subject to PFA}(\tau) \leq \alpha \tag{2}$$

When the prior distribution on the change-point follows a geometric distribution, the optimal solution to the above problem is given by the Shiryaev test [9].

The robust versions of (1) and (2) are relevant when one or both of the distributions ν_0 and ν_1 are not known exactly, but are known to belong to uncertainty classes of distributions, $\mathcal{P}_0, \mathcal{P}_1 \subset \mathcal{P}(\mathcal{X})$. The objective is to minimize the worst-case delay amongst all possible values of the unknown distributions, while satisfying the false-alarm constraint for all possible values of the unknown distributions. Thus the robust version of the Lorden criterion is to identify the stopping rule that solves the following optimization problem:

$$\begin{aligned} \min \quad & \sup_{\nu_0 \in \mathcal{P}_0, \nu_1 \in \mathcal{P}_1} \text{WDD}(\tau) \\ \text{s.t.} \quad & \sup_{\nu_0 \in \mathcal{P}_0} \text{FAR}(\tau) \leq \alpha. \end{aligned} \tag{3}$$

Similarly, the robust version of the Bayesian criterion is,

$$\begin{aligned} \min \quad & \sup_{\nu_0 \in \mathcal{P}_0, \nu_1 \in \mathcal{P}_1} \text{ADD}(\tau) \\ \text{s.t.} \quad & \sup_{\nu_0 \in \mathcal{P}_0} \text{PFA}(\tau) \leq \alpha \end{aligned} \tag{4}$$

The optimal stopping rule τ under either robust criteria described above has the following minimax interpretation. For any other stopping rule τ' that guarantees the false alarm constraint for all values of unknown distributions from the uncertainty classes, there is at least one pair of distributions such that the delay obtained under τ' will be at least as high as the maximum delay obtained with τ over all pairs of distributions from the uncertainty classes. In the rest of this paper, we provide a solution to the robust problems (3) and (4) when the uncertainty classes satisfy some specific conditions.

III. Robust change detection

A. Least Favorable Distributions

The solution to the robust problem is simplified greatly if we can identify least favorable distributions (LFD's) from the uncertainty classes such that the solution to the robust problem is given by the solution to the non-robust problem designed with respect to the LFD's. LFD's were first identified for a simpler problem - the robust hypothesis testing problem - by Huber et al. in [13] and [14]. It was later shown in [15] that the sufficient conditions that the uncertainty classes should satisfy for the existence of LFD's is a joint stochastic boundedness condition. Before we introduce this condition, we need the following notation. If X and X' are two real-valued random variables defined on a probability space $(\Omega, \mathcal{F}, \mathsf{P})$ such that,

$$\mathsf{P}(X \geq t) \geq \mathsf{P}(X' \geq t), \text{ for all } t \in \mathbb{R},$$

then we say that the random variable X is *stochastically larger than* the random variable X'. We denote this relation via the following notation $X \succ X'$. Equivalently if $X \sim \mu$ and $X' \sim \mu'$, we also denote $\mu \succ \mu'$.

Definition 1 (Joint Stochastic Boundedness) [15]: Consider the pair $(\mathcal{P}_0, \mathcal{P}_1)$ of classes of distributions defined on a measurable space $(\mathcal{X}, \mathcal{F})$. Let $(\overline{\nu}_0, \underline{\nu}_1) \in \mathcal{P}_0 \times \mathcal{P}_1$ be some pair of distributions from this pair of classes. Let L^* denote the log-likelihood ratio between $\underline{\nu}_1$ and $\overline{\nu}_0$ defined as the logarithm of the Radon-Nikodym derivative $\log \frac{d\underline{\nu}_1}{d\overline{\nu}_0}$. Corresponding to each $\nu_j \in \mathcal{P}_j$, we use μ_j to denote the distribution of $L^*(X)$ when $X \sim \nu_j, j = 0, 1$. Similarly we use $\overline{\mu}_0$ (respectively $\underline{\mu}_1$) to denote the distribution of $L^*(X)$ when $X \sim \overline{\nu}_0$ (respectively $\underline{\nu}_1$). The pair $(\mathcal{P}_0, \mathcal{P}_1)$ is said to be jointly stochastically bounded by $(\overline{\nu}_0, \underline{\nu}_1)$ if for all $(\nu_0, \nu_1) \in \mathcal{P}_0 \times \mathcal{P}_1$,

$$\overline{\mu}_0 \succ \mu_0 \text{ and } \mu_1 \succ \underline{\mu}_1 \qquad \blacksquare$$

Examples of uncertainty classes that satisfy the joint stochastic boundedness condition can be found in [15], [14], and [16].

We show that under certain assumptions on $\mathcal{P}_0$ and $\mathcal{P}_1$, $\overline{\nu}_0$ and $\underline{\nu}_1$ are LFD's for the robust change detection problem in (3) and (4) and thus the optimal stopping rules designed assuming known pre-change and post-change distributions of $\overline{\nu}_0$ and $\underline{\nu}_1$, respectively, are optimal for the robust problems (3) and (4). We use E_m^* to denote the expectation operator and P_m^* to denote the probability operator when the change happens at m and the pre-change and post-change distributions are $\overline{\nu}_0$ and $\underline{\nu}_1$, respectively.

We need the following basic lemma.

Lemma 3.1: Suppose $\{U_i : 1 \leq i \leq n\}$ is a set of mutually independent random variables, and $\{V_i : 1 \leq i \leq n\}$ is another set of mutually independent random variables such that $U_i \succ V_i, 1 \leq i \leq n$. Now let $h : \mathbb{R}^n \mapsto \mathbb{R}$ be a continuous real-valued function defined on $\mathbb{R}^n$ that satisfies,

$$\begin{aligned} &h(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n) \\ &\qquad \leq \quad h(x_1, \ldots, x_{i-1}, a, x_{i+1}, \ldots, x_n), \end{aligned}$$

for all $x_1^n \in \mathbb{R}^n, a > x_i$, and $i \in \{1, \ldots, n\}$. Then we have,

$$h(U_1, U_2, \ldots, U_n) \succ h(V_1, V_2, \ldots, V_n)$$

Proof: We prove this claim by induction. For $n = 1$, the claim holds because if $h : \mathbb{R} \mapsto \mathbb{R}$ is a non-decreasing continuous function we have,

$$\begin{aligned} \mathsf{P}(h(U_1) \geq t) &= \mathsf{P}(U_1 \geq \sup\{x : h(x) \leq t\} \\ &\geq \mathsf{P}(V_1 \geq \sup\{x : h(x) \leq t\} \\ &= \mathsf{P}(h(V_1) \geq t). \end{aligned}$$

Assume the claim is true for $n = N$ and now consider $n = N + 1$. For any fixed $x_1^N \in \mathbb{R}^N$, define $g : \mathbb{R} \mapsto \mathbb{R}$ by $g(y) = h(x_1, x_2, \ldots, x_N, y)$. We know by the property of function h that function g is non-decreasing. Now, we know by the proof for $n = 1$ that,

$$g(U_{N+1}) \succ g(V_{N+1}). \tag{5}$$

We further have,

$$\begin{aligned} &\mathsf{P}(h(U_1, U_2, \ldots, U_{N+1}) \geq t) \\ &= \int f_{U_1^N}(x_1^N) \mathsf{P}(g(U_{N+1}) \geq t) dx_1^N \\ &\geq \int f_{U_1^N}(x_1^N) \mathsf{P}(g(V_{N+1}) \geq t) dx_1^N \qquad (6) \\ &= \mathsf{P}(h(\tilde{U}_1, \tilde{U}_2, \ldots, \tilde{U}_N, V_{N+1}) \geq t) \qquad (7) \\ &= \int f_{V_{N+1}}(y) \mathsf{P}(h(\tilde{U}_1, \tilde{U}_2, \ldots, \tilde{U}_N, y) \geq t) dy \\ &\geq \int f_{V_{N+1}}(y) \mathsf{P}(h(V_1, V_2, \ldots, V_N, y) \geq t) dy \qquad (8) \\ &= \mathsf{P}(h(V_1, V_2, \ldots, V_{N+1}) \geq t). \end{aligned}$$

where (6) is obtained via (5). The variables $\tilde{U}_i$ appearing in (7) are random variables with exact same statistics as U_i and independent of V_i's. The inequality of (8) is obtained by using the induction hypothesis for $n = N$. Thus we have shown that,

$$h(U_1, U_2, \ldots, U_{N+1}) \succ h(V_1, V_2, \ldots, V_{N+1})$$

which proves the lemma by the principle of mathematical induction. ■

B. Lorden criterion

When the distributions ν_0 and ν_1 are known, the solution to (1) is given by the CUSUM test [2]. The optimal stopping time is given by,

$$\tau_{\mathrm{C}} = \inf\{n \geq 1 : \max_{1 \leq k \leq n} \sum_{i=k}^{n} L^{\nu}(X_i) \geq \eta\} \tag{9}$$

where L^{ν} is the log-likelihood ratio between ν_1 and ν_0, and threshold η is chosen so that, $\mathsf{E}_{\infty}^{\nu}(\tau_{\mathrm{C}}) = \frac{1}{\alpha}$.

Theorem 3.2: Suppose the following conditions hold:

(i) Uncertainty classes $\mathcal{P}_0, \mathcal{P}_1$ are jointly stochastically bounded by $(\overline{\nu}_0, \underline{\nu}_1)$.

(ii) All distributions $\nu_0 \in \mathcal{P}_0$ are absolutely continuous with respect to $\overline{\nu}_0$. i.e.,

$$\nu_0 \ll \overline{\nu}_0, \quad \nu_0 \in \mathcal{P}_0. \tag{10}$$

(iii) Function $L^*(.)$, representing the log-likelihood ratio between $\underline{\nu}_1$ and $\overline{\nu}_0$ is continuous over the support of $\overline{\nu}_0$.

Then the optimal stopping rule that solves (3) is given by the following CUSUM test:

$$\tau_{\mathrm{C}}^* = \inf\left\{n \geq 1 : \max_{1 \leq k \leq n} \sum_{i=k}^{n} L^*(X_i) \geq \eta\right\} \tag{11}$$

where threshold η is chosen so that, $\mathsf{E}_{\infty}^{*}(\tau_{\mathrm{C}}^*) = \frac{1}{\alpha}$.

Proof: Suppose $\mathcal{P}_0$ and $\mathcal{P}_1$ satisfy the conditions of the theorem. Since the CUSUM test is optimal for known distributions, it is clear that the test given in (11) is optimal when the pre- and post-change distributions are $\overline{\nu}_0$ and $\underline{\nu}_1$, respectively. Hence, it suffices to show that the values of WDD(τ_{C}^*) and FAR(τ_{C}^*) obtained under any $\nu_0 \in \mathcal{P}_0$ and any $\nu_1 \in \mathcal{P}_1$, are no higher than their respective values when the pre- and post-change distributions are $\overline{\nu}_0$ and $\underline{\nu}_1$. We use Y_i^* to denote the random variable $L^*(X_i)$ when the pre-change and post-change distributions of the observations from the sequence $\{X_i : i = 1, 2, \ldots\}$ are $\overline{\nu}_0$ and $\underline{\nu}_1$, respectively, and Y_i^{ν} to denote the random variable $L^*(X_i)$ when the pre- and post-change distributions are ν_0 and ν_1, respectively. We first prove the theorem for a special case.

Case 1: $\mathcal{P}_0$ is a singleton given by $\mathcal{P}_0 = \{\nu_0\}$.

Clearly, in this case $\overline{\nu}_0 = \nu_0$ and (10) is met trivially. Furthermore, in this case, the false alarm constraint is also met trivially since the false alarm rate obtained by using the stopping rule τ_{C}^* is independent of the true value of the post-change distribution. Fix the change-point to be λ. Now, to complete the proof for the scenario where $\mathcal{P}_0$ is a singleton, we will show that for all $\lambda \geq 1$,

$$\mathsf{E}_{\lambda}^{*}[(\tau_{\mathrm{C}}^* - \lambda + 1)^+ | \mathcal{F}_{\lambda-1}] \succ \mathsf{E}_{\lambda}^{\nu}[(\tau_{\mathrm{C}}^* - \lambda + 1)^+ | \mathcal{F}_{\lambda-1}] \tag{12}$$

which would establish that the value of WDD(τ_{C}^*) obtained under any $\nu_1 \in \mathcal{P}_1$, is no higher than the value when the true post-change distribution is $\underline{\nu}_1$.

Since we now have $\overline{\nu}_0 = \nu_0$, we assume without loss of generality that for all $i < \lambda$, $Y_i^* = Y_i^{\nu}$ with probability one. Under this assumption, we will show that for all integers $N \geq 0$, the following relation holds with probability one,

$$\begin{aligned} &\mathsf{P}_{\lambda}^{*}((\tau_{\mathrm{C}}^* - \lambda + 1)^+ \leq N | \mathcal{F}_{\lambda-1}) \\ &\qquad \leq \quad \mathsf{P}_{\lambda}^{\nu}((\tau_{\mathrm{C}}^* - \lambda + 1)^+ \leq N | \mathcal{F}_{\lambda-1}), \end{aligned} \tag{13}$$

which would then establish (12). Since τ_{C}^* is a stopping time, the event $\{(\tau_{\mathrm{C}}^* - \lambda + 1)^+ \leq 0\}$ is $\mathcal{F}_{\lambda-1}$-measurable. Hence, with probability one, (13) holds with equality for $N = 0$. Now it suffices to verify (13) for $N \geq 1$. We know by the stochastic ordering condition on $\mathcal{P}_1$ that,

$$Y_i^\nu \succ Y_i^*, \text{ for all } i \geq \lambda \quad (14)$$

Now we have the following equivalence between two events:

$$\begin{aligned}\{\tau_{\mathrm{C}}^* \leq N\} &= \left\{\max_{1\leq n\leq N}\max_{1\leq k\leq n}\sum_{i=k}^{n} L^*(X_i) \geq \eta\right\} \\ &= \left\{\max_{1\leq k\leq n\leq N}\sum_{i=k}^{n} L^*(X_i) \geq \eta\right\}.\end{aligned}$$

It is easy to see that the function,

$$f(x_1, \ldots, x_N) \triangleq \max_{1\leq k\leq n\leq N}\sum_{i=k}^{n} x_i$$

is continuous and non-decreasing in each of its components as required by Lemma 3.1. Hence for $N \geq 1$, the following hold with probability one:

$$\begin{aligned}&\mathsf{P}_\lambda^*((\tau_{\mathrm{C}}^* - \lambda + 1)^+ \leq N|\mathcal{F}_{\lambda-1}) \\ &\quad= \mathsf{P}_\lambda^*(\tau_{\mathrm{C}}^* \leq N + \lambda - 1|\mathcal{F}_{\lambda-1}) \\ &\quad= \mathsf{P}_\lambda(f(Y_1^*, \ldots, Y_{N+\lambda-1}^*) \geq \eta|\mathcal{F}_{\lambda-1}) \\ &\quad\leq \mathsf{P}_\lambda(f(Y_1^\nu, \ldots, Y_{N+\lambda-1}^\nu) \geq \eta|\mathcal{F}_{\lambda-1}) \\ &\quad= \mathsf{P}_\lambda^\nu(\tau_{\mathrm{C}}^* \leq N|\mathcal{F}_{\lambda-1}) \\ &\quad= \mathsf{P}_\lambda^\nu((\tau_{\mathrm{C}}^* - \lambda + 1)^+ \leq N|\mathcal{F}_{\lambda-1})\end{aligned}$$

where the inequality follows from Lemma 3.1 and (14), using the fact that f is a non-decreasing function with respect to its last N arguments and the fact that $Y_i^\nu = Y_i^*$ for $i < \lambda$. Thus, for all integers $N \geq 0$, (13) holds with probability one and hence (12) is satisfied. This proves the result for the case where $\mathcal{P}_0$ is a singleton.

Case 2: $\mathcal{P}_0$ is any class of distributions satisfying (10).

Suppose that the change does not occur. Then we know by the stochastic ordering condition on $\mathcal{P}_0$ that, $Y_i^* \succ Y_i^\nu$ for all i. It follows by Lemma 3.1 that,

$$\begin{aligned}\mathsf{P}_\infty^*(\tau_{\mathrm{C}}^* \leq N) &= \mathsf{P}_\infty(f(Y_1^*, \ldots, Y_N^*) \geq \eta) \\ &\geq \mathsf{P}_\infty(f(Y_1^\nu, \ldots, Y_N^\nu) \geq \eta) \\ &= \mathsf{P}_\infty^\nu(\tau_{\mathrm{C}}^* \leq N)\end{aligned}$$

Since the above relation holds for all $N \geq 1$, we have

$$\mathsf{E}_\infty^\nu(\tau_{\mathrm{C}}^*) \geq \mathsf{E}_\infty^*(\tau_{\mathrm{C}}^*) = \frac{1}{\alpha}$$

and hence the value of $\mathrm{FAR}(\tau_{\mathrm{C}}^*)$ is no higher than α for all values of $\nu_0 \in \mathcal{P}_0$ and $\nu_1 \in \mathcal{P}_1$.

Now suppose the change-point is fixed at λ. A useful observation is that for any given stopping rule τ and fixed post-change distribution ν_1, the random variable $\mathsf{E}_\lambda^{\nu_0,\nu_1}[(\tau - \lambda+1)^+|\mathcal{F}_{\lambda-1}]$ is a fixed deterministic function of the random observations $(X_1, \ldots, X_{\lambda-1})$, irrespective of the distribution ν_0. Thus the essential supremum of this random variable depends only on the support of ν_0. Applying this observation to the stopping rule τ_{C}^* and using the relation (10) we have for all $\nu_0 \in \mathcal{P}_0, \nu_1 \in \mathcal{P}_1$,

$$\begin{aligned}&\text{ess sup } \mathsf{E}_\lambda^{\nu_0,\nu_1}[(\tau_{\mathrm{C}}^* - \lambda + 1)^+|\mathcal{F}_{\lambda-1}] \\ &\qquad\leq \quad \text{ess sup } \mathsf{E}_\lambda^{\overline{\nu}_0,\nu_1}[(\tau_{\mathrm{C}}^* - \lambda + 1)^+|\mathcal{F}_{\lambda-1}].\end{aligned}$$

We also know from *Case 1* above that for all $\nu_1 \in \mathcal{P}_1$,

$$\begin{aligned}&\text{ess sup } \mathsf{E}_\lambda^{\overline{\nu}_0,\nu_1}[(\tau_{\mathrm{C}}^* - \lambda + 1)^+|\mathcal{F}_{\lambda-1}] \\ &\qquad\leq \quad \text{ess sup } \mathsf{E}_\lambda^*[(\tau_{\mathrm{C}}^* - \lambda + 1)^+|\mathcal{F}_{\lambda-1}].\end{aligned}$$

Taking supremum over $\lambda \geq 1$, it follows from the above two relations that the value of $\mathrm{WDD}(\tau_{\mathrm{C}}^*)$ under any pair of distributions $(\nu_0, \nu_1) \in \mathcal{P}_0 \times \mathcal{P}_1$ is no larger than that under $(\overline{\nu}_0, \underline{\nu}_1)$. Thus τ_{C}^* solves the robust problem (3). ∎

Remark 3.1: The discussion in [11, p. 198] suggests that when LFD's exist under our formulation, they also solve the asymptotic problem, as expected.

C. Bayesian criterion

When the distributions ν_0 and ν_1 are known and the prior distribution of the change point is geometric, the solution to (2) is given by the Shiryaev test [9]. Denoting the parameter of the geometric distribution by ρ, we have,

$$\pi_k = \rho(1-\rho)^{k-1}, \quad k \geq 1$$

The optimal stopping time is given by,

$$\tau_{\mathrm{S}} = \inf\left\{n \geq 1 : \log(\sum_{k=1}^{n} \pi_k \exp(\sum_{i=k}^{n} L^\nu(X_i))) \geq \eta\right\} \quad (15)$$

where threshold η is chosen such that, $\mathrm{PFA}(\tau_{\mathrm{S}}) = \mathsf{P}^\nu(\tau_{\mathrm{S}} < \Lambda) = \alpha$.

Theorem 3.3: Suppose the following conditions hold:

(i) Uncertainty class $\mathcal{P}_0$ is a singleton $\mathcal{P}_0 = \{\nu_0\}$ and the pair $(\mathcal{P}_0, \mathcal{P}_1)$ is jointly stochastically bounded by $(\nu_0, \underline{\nu}_1)$.
(ii) Prior distribution of the change point is a geometric distribution.
(iii) Function $L^*(.)$, representing the log-likelihood ratio between $\underline{\nu}_1$ and ν_0 is continuous over the support of ν_0.

Then the optimal stopping rule that solves (4) is given by the following Shiryaev test:

$$\tau_{\mathrm{S}}^* = \inf\left\{n \geq 1 : \log(\sum_{k=1}^{n} \pi_k \exp(\sum_{i=k}^{n} L^*(X_i))) \geq \eta\right\} \quad (16)$$

where threshold η is chosen so that, $\mathsf{P}^*(\tau_{\mathrm{S}}^* < \Lambda) = \alpha$.

Proof: The proof is very similar to that of *Case 1* in Theorem 3.2. Since the Shiryaev test is optimal for known distributions, it is clear that the test given in (16) is optimal under the Bayesian criterion when the post-change distribution is $\underline{\nu}_1$. Also from the definition of $\mathrm{PFA}(\tau_{\mathrm{S}}^*)$ it is clear that the probability of false alarm depends only on the pre-change distribution and hence the constraint in (4) is met by the stopping time τ_{S}^*. Hence, it suffices to show that the value of $\mathrm{ADD}(\tau_{\mathrm{S}}^*)$ obtained under any $\nu_1 \in \mathcal{P}_1$, is no higher than the value when the true post-change distribution is $\underline{\nu}_1$.

Let us first fix $\Lambda = \lambda$. We know by the stochastic ordering condition that conditioned on $\Lambda = \lambda$, for all $i \geq \lambda$, we have

$Y_i^\nu \succ Y_i^*$ where Y_i^* and Y_i^ν are as defined in the proof of Theorem 3.2. As before, the function,

$$f'(x_1, \ldots, x_N) \triangleq \max_{1 \le n \le N} \log \left(\sum_{k=1}^{n} \pi_k \exp(\sum_{i=k}^{n} x_i) \right)$$

is continuous and non-decreasing in each of its components as required by Lemma 3.1. Using these facts, we can show the following by proceeding exactly as in the proof of Theorem 3.2: Conditioned on $\Lambda = \lambda$,

$$\mathsf{E}_\lambda^*((\tau_s^* - \lambda)^+ | \mathcal{F}_{\lambda-1}) \succ \mathsf{E}_\lambda^\nu((\tau_s^* - \lambda)^+ | \mathcal{F}_{\lambda-1}).$$

Thus, we have $\mathsf{E}_\lambda^*((\tau_s^* - \lambda)^+) \ge \mathsf{E}_\lambda^\nu((\tau_s^* - \lambda)^+)$ and by averaging over λ, we get,

$$\mathsf{E}^*((\tau_s^* - \Lambda)^+) \ge \mathsf{E}^\nu((\tau_s^* - \Lambda)^+)$$

∎

Remark 3.2: Our result under the Bayesian criterion is weaker than that under the Lorden criterion since our solution does not hold for the case where $\mathcal{P}_0$ is not a singleton.

IV. An Example

A. Gaussian mean shift

Here we consider a simple example to illustrate the result. Assume ν_0 is known to be a standard Gaussian distribution with mean zero and unit variance, so that $\mathcal{P}_0$ is a singleton. Let $\mathcal{P}_1$ be the collection of Gaussian distributions with means from the interval $[0.1, 3]$ and unit variance.

$$\begin{aligned} \mathcal{P}_0 &= \{\mathcal{N}(0,1)\} \\ \mathcal{P}_1 &= \{\mathcal{N}(\theta,1) : \theta \in [0.1, 3]\} \end{aligned}$$

It is easily verified that $(\mathcal{P}_0, \mathcal{P}_1)$ is jointly stochastically bounded by $(\overline{\nu}_0, \underline{\nu}_1)$ given by

$$\overline{\nu}_0 \sim \mathcal{N}(0,1), \qquad \underline{\nu}_1 \sim \mathcal{N}(0.1, 1).$$

We simulated the Bayesian and robust Bayesian change-detection tests for this problem assuming a geometric prior distribution for the change-point with parameter 0.1 and a false alarm constraint of $\alpha = 0.001$. From the performance curves plotted in Figure 1, we can see that the robust Shiryaev test gives the exact same performance as the optimal Shiryaev test at $\underline{\nu}_1$. For all other values of $\nu_1 \in \mathcal{P}_1$, the performance is strictly better than the performance at $\underline{\nu}_1$ and hence the robust Shiryaev test is indeed minimax optimal. We can also see in Figure 1 that the average delays obtained with the robust test are much higher than those obtained with the optimal test especially at high values of the mean θ.

V. Conclusion

We have shown that for uncertainty classes that satisfy some specific conditions, the optimal change detectors designed for the least favorable distributions are optimal in a minimax sense. This is shown for both the Lorden criterion and Shiryaev's Bayesian criterion. However, robustness comes with a potential cost. The optimal stopping rule designed for the LFD's may perform quite sub-optimally for other

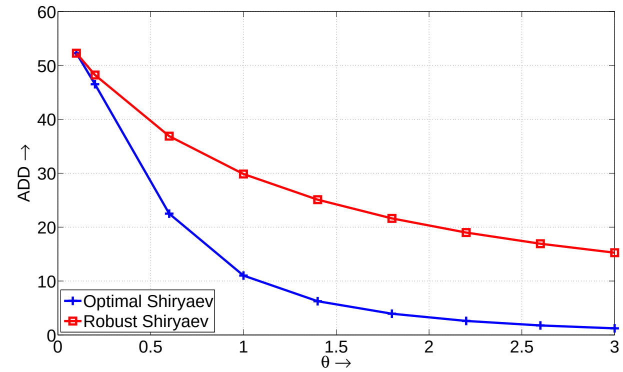

Fig. 1. Comparison of robust and non-robust Shiryaev tests for $\alpha = 0.001$

distributions from the uncertainty class, when compared with the optimal performance that can be obtained under the same distributions in the case where the distributions are known exactly.

Acknowledgment

This work was supported by NSF Grants CCF 07-29031 and CCF 08-30169, through the University of Illinois.

References

[1] E. S. Page, "Continuous Inspection Schemes," Biometrika, vol. 41, pp. 100-115, 1954.

[2] G. V. Moustakides, "Optimal Stopping Times for Detecting Changes in Distributions," *Ann. Statist.*, vol. 14, no. 4, pp. 1379-1387, Dec. 1986.

[3] G. Lorden, "Procedures for reacting to a change in distribution," *Ann. Math. Statist.*, vol. 42, no. 6, pp. 1897-1908, 1971.

[4] M. Pollak, "Optimal detection of a change in distribution," *Ann. Statist.*, vol. 13, no. 1, pp. 206-227, 1985.

[5] A. N. Shiryaev, *Optimal Stopping Rules*, New York: Springer-Verlag, 1978.

[6] A. G. Tartakovsky, *Sequential Methods in the Theory of Information Systems*, Radio i Svyaz', Moscow, 1991.

[7] M. Basseville and I. Nikiforov, *Detection of Abrupt Changes: Theory and Applications*, Prentice-Hall, Englewood Cliffs, NJ, 1993.

[8] B. E. Brodsky and B.S. Darkhovsky, *Nonparametric Methods in Change-Point Problems*, Kluwer, Dordrecht, 1993.

[9] A. N. Shiryaev, "On optimum methods in quickest detection problems," *Theory of Prob. and App.*, vol. 8, pp. 22-46, 1963.

[10] R. Crow and S. Schwartz, "On robust quickest change detection procedures," *Proc., IEEE International Symposium on Information Theory, Trondheim, Norway*, pp. 258, 1994

[11] H. V. Poor and O. Hadjiliadis, *Quickest Detection*, Cambridge University Press, Cambridge, UK, 2009.

[12] A. G. Tartakovsky and A.S. Polunchenko, "Quickest Changepoint Detection in Distributed Multisensor Systems under Unknown Parameters," *Proc., 11th International Conference on Information Fusion*, Cologne, Germany, July 2008.

[13] P. J. Huber, "A robust version of the probability ratio test," *Ann. Math. Statist.*, vol. 36, pp. 1753-1758, 1965.

[14] P. J. Huber and V. Strassen, "Minimax tests and the Neyman-Pearson lemma for capacities," *Ann. Statist.*, vol. 1, no. 2, pp. 251-263, 1973.

[15] V. V. Veeravalli, T. Başar, and H. V. Poor, "Minimax Robust Decentralized Detection," in *IEEE Trans. Inform. Theory*, vol. 40, no. 1, pp. 35-40, Jan. 1994.

[16] B. C. Levy, *Principles of Signal Detection and Parameter Estimation*, Springer, 2008.