# Minimax Robust Quickest Change Detection

Jayakrishnan Unnikrishnan, *Member, IEEE*, Venugopal V. Veeravalli, *Fellow, IEEE*, and Sean P. Meyn, *Fellow, IEEE*

*Abstract*—The popular criteria of optimality for quickest change detection procedures are the Lorden criterion, the Pollak criterion, and the Bayesian criterion. In this paper, a robust version of these quickest change detection problems is considered when the pre-change and post-change distributions are not known exactly but belong to known uncertainty classes of distributions. For uncertainty classes that satisfy a specific condition, it is shown that one can identify least favorable distributions (LFDs) from the uncertainty classes, such that the detection rule designed for the LFDs is optimal for the robust problem in a minimax sense. The condition is similar to that required for the identification of LFDs for the robust hypothesis testing problem originally studied by Huber. An upper bound on the delay incurred by the robust test is also obtained in the asymptotic setting under the Lorden criterion of optimality. This bound quantifies the delay penalty incurred to guarantee robustness. When the LFDs can be identified, the proposed test is easier to implement than the CUSUM test based on the Generalized Likelihood Ratio (GLR) statistic which is a popular approach for such robust change detection problems. The proposed test is also shown to give better performance than the GLR test in simulations for some parameter values.

*Index Terms*—CUSUM test, least favorable distributions, minimax robustness, quickest change detection, Shiryaev test.

## I. INTRODUCTION

**T**HE problem of detecting an abrupt change in a system based on observations is a dynamic hypothesis testing problem with a rich set of applications. Such problems of change detection were first studied by Page over 50 years ago in the context of quality control [2]. In its standard formulation, there is a sequence of observations whose distribution changes at some unknown point in time, referred to as the 'change-point'. The goal is to detect this change as soon as possible, subject to a false alarm constraint. 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 to wireless networks.

J. Unnikrishnan was with the Department of Electrical and Computer Engineering and the Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, Urbana , IL 61801 USA. He is now with the Audiovisual Communications Laboratory, Swiss Federal Institute of Technology Lausanne (EPFL), Lausanne 1015, Switzerland (e-mail: jay.unnikrishnan@epfl.ch).

V. V. Veeravalli and S. Meyn are with the Department of Electrical and Computer Engineering and the Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, Urbana , IL 61801 USA (e-mail: vvv@illinois.edu; meyn@illinois.edu).

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., [3]–[6]; for an overview of the work in this area, see [7]–[9].). The three most popular criteria for optimizing the tradeoff between detection delay and false alarm rate are the Lorden criterion [4] and the Pollak criterion, in which the change-point is a deterministic quantity, and Shiryaev's Bayesian formulation [10], in which the change-point is modeled as a random variable with a known prior distribution. In this paper, we study all these three 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 performed 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 (LFDs) from the uncertainty classes, and the optimal robust change detection rule is then the optimal (nonrobust) change detection rule for the LFDs. These conditions are similar to those given by Huber [11] for robust hypothesis testing problems. We also discuss related results on robust sequential detection [11], [12] later in the paper.

Although there has been some prior work on robust change detection, these approaches are distinctly different from ours. The maximin approach of [13] is similar in that they also identify LFDs 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 [14, Sec.7.3.1]. Some other approaches to this problem (e.g., [15], [16]) are aimed at developing algorithms for quickest change detection with unknown distributions. These works study the asymptotic performance of the proposed tests under different distributions but do not seek to guarantee minimax robustness over a given class of distributions.

A closely related problem is the composite quickest change detection problem. In general, these problems also address the setting where the pre-change and post-change distributions are unknown. However, unlike the robust problem, in composite problems one seeks to identify a change detection procedure that is simultaneously optimal under all possible values of the unknown distributions. Exact solutions to these problems are often intractable and hence most results are restricted to asymptotic optimality. One such solution to a composite change detection problem is discussed in [4] when only the post-change distribution is unknown. In [4], a test is given that is asymptotically optimal under the Lorden criterion for all possible values of the

unknown post-change distribution in a one-dimensional exponential family of distributions. This test is also referred to as the Generalized Likelihood Ratio Test (GLR Test), and was also studied in [17] and [18]. An alternate asymptotically optimal solution for the setting in which both pre-change and post-change distributions are unknown was studied in [19].

We provide a performance comparison of our proposed robust test with the GLR test. Although the GLR test asymptotically performs as well as the optimal test with known distributions, we show via simulations that our robust test can give improved performance over the GLR test for moderate values of the false alarm constraint. The GLR test is also often prohibitively complex to implement in practice, while the proposed robust CUSUM test admits a simple recursive implementation.

For the asymptotic version of the problem, we also provide an analytical upper bound on the delay incurred by our robust test and use it to provide an upper bound on the drop in performance of our test relative to the optimal nonrobust test.

The rest of the paper is organized as follows. We first state the problem that we are studying in Section II. In Section III we describe the robust solution and present some analysis. 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 $\nu_0$. Their distribution switches abruptly to $\nu_1$ at some unknown time $\lambda$ so that $X_n \sim \nu_0$ for $n \leq \lambda - 1$ and $X_n \sim \nu_1$ for $n \geq \lambda$. 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 $\mathsf{E}_m^\nu$ to denote the expectation operator and $\mathsf{P}_m^\nu$ to denote the probability law when the change happens at $m$ and the pre-change and post-change distributions are $\nu_0$ and $\nu_1$ respectively. The symbols are replaced with $\mathsf{E}_\infty^\nu$ and $\mathsf{P}_\infty^\nu$ when the change does not happen. Similarly, if the pre-change and post-change distributions are some $\mu$ and $\gamma$, respectively, and the change happens at time $m$, we use $\mathsf{E}_m^{\mu;\gamma}$ to denote the expectation operator and $\mathsf{P}_m^{\mu;\gamma}$ the probability law. We further use $\mathcal{F}_m$ to denote the sigma algebra generated by $(X_1, X_2, \ldots, X_m)$.

A sequential change detection procedure is characterized by a stopping time $\tau$ with respect to the observation sequence. The design of the quickest change detection procedure involves optimizing the tradeoff between two performance measures: detection delay and frequency of false alarms. There are various standard mathematical formulations for the optimal tradeoff. In the minimax formulation of [4] 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 $x^+ = \max(x, 0)$. This quantity captures the worst-case value of the expected detection delay over all possible locations of the change-point and all possible realizations

of the pre-change observations. 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 to false alarm. 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

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

It was shown by Moustakides [3] that the optimal solution to (1) is given by the cumulative sum (CUSUM) test proposed by Page [2]. We describe this test later in the paper.

An alternate formulation of the change detection problem was studied by Pollak [5]. Even here the change point is modeled as a deterministic quantity. But the delay to be minimized is no longer the worst-case delay but a worst-case average delay (also referred to as supremum average detection delay by some authors) defined by

$$J_{\text{SRP}}(\tau) = \sup_{\lambda \geq 1} \mathsf{E}_\lambda^\nu[\tau - \lambda | \tau \geq \lambda].$$

The Pollak criterion of optimality of a stopping rule $\tau$ for change detection is given by

$$\text{Minimize } J_{\text{SRP}}(\tau) \text{ subject to FAR}(\tau) \leq \alpha \quad (2)$$

where the minimization is over all stopping times $\tau$ such that $J_{\text{SRP}}(\tau)$ is well-defined. Pollak [5] established the asymptotic optimality of the Shiryaev-Roberts-Pollak (SRP) stopping rule for (2).

Another approach to change detection is the Bayesian formulation of [6], [10]. Here, the change-point is modeled as a random variable $\Lambda$ 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

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

where $\mathsf{E}^\nu$ represents the expectation operator and $\mathsf{P}^\nu$ the probability law when the pre-change and post-change distributions are $\nu_0$ and $\nu_1$ respectively. For a given $\alpha \in (0, 1)$, the optimization problem under the Bayesian criterion is

$$\text{Minimize ADD}(\tau) \text{ subject to PFA}(\tau) \leq \alpha. \quad (3)$$

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 [10].

The robust versions of (1), (2) and (3) are relevant when one or both of the distributions $\nu_0$ and $\nu_1$ 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:

$$\min_{\nu_0 \in \mathcal{P}_0, \nu_1 \in \mathcal{P}_1} \sup \quad \mathrm{WDD}(\tau)$$
$$\text{s.t.} \quad \sup_{\nu_0 \in \mathcal{P}_0} \mathrm{FAR}(\tau) \leq \alpha. \tag{4}$$

Similarly, the robust version of the Pollak criterion is

$$\min_{\nu_0 \in \mathcal{P}_0, \nu_1 \in \mathcal{P}_1} \sup \quad J_{\mathrm{SRP}}(\tau)$$
$$\text{s.t.} \quad \sup_{\nu_0 \in \mathcal{P}_0} \mathrm{FAR}(\tau) \leq \alpha. \tag{5}$$

and the robust version of the Bayesian criterion is

$$\min_{\nu_0 \in \mathcal{P}_0, \nu_1 \in \mathcal{P}_1} \sup \quad \mathrm{ADD}(\tau)$$
$$\text{s.t.} \quad \sup_{\nu_0 \in \mathcal{P}_0} \mathrm{PFA}(\tau) \leq \alpha. \tag{6}$$

The optimal stopping rule $\tau$ under each of the robust criteria described above has the following minimax interpretation. For any other stopping rule $\tau'$ 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 $\tau'$ will be at least as high as the maximum delay obtained with $\tau$ over all pairs of distributions from the uncertainty classes. In the rest of this paper we provide solutions to the robust problems (4), (5) and (6) 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 (LFDs) from the uncertainty classes such that the solution to the robust problem is given by the solution to the nonrobust problem designed with respect to the LFDs. LFDs were first identified for a simpler problem—the robust hypothesis testing problem—by Huber and Strassen in [11] and [20]. It was later shown in [21] that if the uncertainty classes satisfy a joint stochastic boundedness condition, one can identify these LFDs. 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 [21] the random variable $X'$. We denote this relation via the notation $X \succ X'$. Equivalently, if $X \sim \nu$ and $X' \sim \nu'$, we also denote $\nu \succ \nu'$.

*Definition 1 (Joint Stochastic Boundedness) [21]:* 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 such that $\underline{\nu}_1$ is absolutely continuous with respect to $\overline{\nu}_0$. 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 $\mu_j$ 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 $(\nu_0, \nu_1) \in \mathcal{P}_0 \times \mathcal{P}_1$,

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

Loosely speaking, the LFD from one uncertainty class is the distribution that is *nearest* to the other uncertainty class. This notion can be made rigorous in terms of Kullback-Leibler divergence and other Ali-Silvey distances between distributions in the uncertainty classes, as shown in [22, Corollary 1].

Huber and Strassen [20] have established a procedure to obtain robust solutions to the Neyman-Pearson hypothesis testing problem provided the uncertainty classes can be described in terms of 2-alternating capacities. As pointed out in [21], any pair of uncertainty classes that can be described in terms of 2-alternating capacities also satisfy the joint stochastic boundedness (JSB) condition (see [20, Theorem 4.1]). This observation suggests that we can identify examples of uncertainty classes which satisfy the joint stochastic boundedness condition using the results in [20], [21], and [23]. These include $\epsilon$-contamination classes, total variation neighborhoods, Prohorov distance neighborhoods, band classes, and p-point classes. In general it is difficult to identify the distributions $\overline{\nu}_0$ and $\underline{\nu}_1$. However, for $\epsilon$-contamination classes, total variation neighborhoods, and Lévy metric neighborhoods, the method suggested in [23, pp. 241–248] can be used to identify these distributions.

We show that under certain assumptions on $\mathcal{P}_0$ and $\mathcal{P}_1$, the pair of distributions $(\overline{\nu}_0, \underline{\nu}_1)$ are LFDs for the robust change detection problem in (4), (5) and (6). Thus, the optimal stopping rules designed *assuming* known pre-change and post-change distributions $\overline{\nu}_0$ and $\underline{\nu}_1$, respectively, are optimal for the robust problems (4), (5) and (6). We use $\mathsf{E}_m^*$ to denote the expectation operator and $\mathsf{P}_m^*$ to denote the probability law 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 straightforward result. For completeness, we provide a proof in the Appendix.

*Lemma III.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

$$h(x_1, \ldots, x_{i-1}, a, x_{i+1}, \ldots, x_n)$$
$$\geq h(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n)$$

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

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

## B. Lorden Criterion

When the distributions $\nu_0$ and $\nu_1$ are known, the solution to (1) is given by the CUSUM test [3]. The optimal stopping time is given by,

$$\tau_C = \inf \left\{ n \geq 1 : \max_{1 \leq k \leq n} \sum_{i=k}^{n} L^\nu(X_i) \geq \eta \right\} \qquad (7)$$

where $L^\nu$ is the log-likelihood ratio between $\nu_1$ and $\nu_0$, and the threshold $\eta$ is chosen so that, $\mathsf{E}^\nu_\infty(\tau_C) = \frac{1}{\alpha}$. The following theorem provides a solution to the robust Lorden problem when the distributions are unknown.

*Theorem III.2:* Suppose the following conditions hold:
(i) The 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.,

$$\nu_0 \ll \overline{\nu}_0, \quad \nu_0 \in \mathcal{P}_0. \qquad (8)$$

(iii) The function $L^*(\cdot)$, 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 (4) is given by the following CUSUM test

$$\tau_C^* = \inf \left\{ n \geq 1 : \max_{1 \leq k \leq n} \sum_{i=k}^{n} L^*(X_i) \geq \eta \right\} \qquad (9)$$

where the threshold $\eta$ is chosen so that, $\mathsf{E}^*_\infty(\tau_C^*) = \frac{1}{\alpha}$. $\qquad \square$

We prove the theorem in the Appendix. Two brief remarks are in order. Firstly, the discussion in [14, p. 198] suggests that when LFDs exist under our formulation, they also solve the asymptotic problem, as expected. Secondly, the robust CUSUM test admits a simple recursive implementation similar to the ordinary CUSUM test. Clearly,

$$S_{n+1} = S_n^+ + L^*(X_{n+1}) \qquad (10)$$

where $S_n = \max_{1 \leq k \leq n} \sum_{i=k}^{n} L^*(X_i)$ is the test statistic appearing in (9). Thus, it is easy to compute the test statistic recursively.

*1) Asymptotic Analysis of the Robust CUSUM Test:* In general, for any pair of pre-change and post-change distributions $(\nu_0, \nu_1)$ from the uncertainty classes, we expect the performance of the robust CUSUM test to be poorer than that of the optimal CUSUM test designed with respect to the correct distributions. The drop in performance can be interpreted as the cost of robustness. Although it is not easy to characterize this cost in general, some insight can be obtained by performing an asymptotic analysis in the setting where the false alarm constraint $\alpha$ goes to zero. Our analysis uses the result of [4, Theorem 2] (also see [14, Theorem 6.16]). We use $\mathrm{WDD}^\nu(\tau_C^*)$ to denote the worst-case delay obtained by employing the stopping rule $\tau_C^*$ when the pre-change and post-change distributions are given by $\nu_0$ and $\nu_1$. Similarly, $\mathrm{WDD}^*(\tau_C^*)$ is used to denote the same quantity when the pre-change and post-change distributions are the LFDs.

As mentioned in the remark following [4, Theorem 2], we can interpret the robust CUSUM test as a repeated one-sided

sequential probability ratio test (SPRT) between $\underline{\nu}_1$ and $\overline{\nu}_0$. Let $\tau_{\mathrm{SPRT}}$ denote the stopping rule of the SPRT. We apply [4, Theorem 2] to $\tau_{\mathrm{SPRT}}$ when the true distributions are the LFDs. It follows that

$$\mathsf{E}^*_\infty(\tau_C^*) \geq \frac{1}{\alpha}$$

where $B = \frac{1}{\alpha}$ is used as the upper threshold in the SPRT given by $\tau_{\mathrm{SPRT}}$. From (30), we know that

$$\mathsf{E}^\nu_\infty(\tau_C^*) \geq \mathsf{E}^*_\infty(\tau_C^*) \geq \frac{1}{\alpha}.$$

We again apply the theorem to $\tau_{\mathrm{SPRT}}$, but with the true distributions given by any $\nu_0 \in \mathcal{P}_0$ and $\nu_1 \in \mathcal{P}_1$. We now have

$$\mathrm{WDD}^\nu(\tau_C^*) \leq \mathsf{E}(\tau_{\mathrm{SPRT}})$$

where the expression on the right hand side denotes the expected stopping time of the SPRT when the observations follow distribution $\nu_1$. Now, by applying the well-known Wald's identity [24] as suggested in the remark following [4, Theorem 2], we obtain

$$\mathsf{E}(\tau_{\mathrm{SPRT}}) = \frac{|\log \alpha|}{I_{\nu_1}}(1 + o(1)), \quad \text{as } \alpha \to 0$$

where $o(1) \to 0$ as $\alpha \to 0$ and

$$I_{\nu_1} = \int L^*(x) d\nu_1(x) = D(\nu_1 \| \overline{\nu}_0) - D(\nu_1 \| \underline{\nu}_1).$$

Thus

$$\mathrm{WDD}^\nu(\tau_C^*) \leq \frac{|\log(\alpha)|(1 + o(1))}{D(\nu_1 \| \overline{\nu}_0) - D(\nu_1 \| \underline{\nu}_1)}.$$

It is also known from [4, Theorem 3] that any stopping rule $\tau$ that satisfies the false alarm constraint $\mathrm{FAR}(\tau) \leq \alpha$ must satisfy the lower bound

$$\mathrm{WDD}^\nu(\tau) \geq \frac{|\log(\alpha)|(1 + o(1))}{D(\nu_1 \| \nu_0)}$$

and that this lower bound is achieved by the optimal CUSUM test between $\nu_1$ and $\nu_0$. Thus, the worst-case delay of the robust test is asymptotically larger by a factor no more than

$$\frac{D(\nu_1 \| \nu_0)}{D(\nu_1 \| \overline{\nu}_0) - D(\nu_1 \| \underline{\nu}_1)}$$

when compared with the delay incurred by the optimal test. This factor is thus an upper bound on the asymptotic cost of robustness.

## C. Pollak Criterion

The SRP stopping rule is asymptotically optimal for (2). Let $R_0^\nu$ be a random variable with distribution $\psi$ supported on $\mathbb{R}_+$ and define,

$$R_n^\nu = L^\nu(X_n)\left(1 + R_{n-1}^\nu\right), \quad n \geq 1. \qquad (11)$$

When the distributions $\nu_0$ and $\nu_1$ are known the SRP stopping rule is given by

$$\tau_{\mathrm{SRP}}^{\nu,\eta,\psi} = \inf \left\{ n \geq 0 : R_n^\nu \geq \eta \right\}. \qquad (12)$$

*Asymptotic Optimality Property:* The SRP test of (12) is asymptotically optimal for (2) in the following sense [5]: For every $0 < \alpha < 1$ there exists threshold $\eta$ and probability measure $\psi_\eta$ such that the stopping rule $\tau_{\mathrm{SRP}} := \tau_{\mathrm{SRP}}^{\nu, \eta, \psi_\eta}$ satisfies $\mathrm{FAR}(\tau_{\mathrm{SRP}}) = \alpha$ and for any other stopping rule $\tau$ that satisfies the false alarm constraint $\mathrm{FAR}(\tau) \leq \alpha$, we have

$$J_{\mathrm{SRP}}(\tau) \geq J_{\mathrm{SRP}}(\tau_{\mathrm{SRP}}) + o(1) \qquad (13)$$

where $o(1) \to 0$ as $\alpha \to 0$.

The following theorem identifies a stopping rule that extends the above asymptotic optimality property to the setting where the post-change distribution is unknown.

*Theorem III.3:* Suppose the following conditions hold:
(i) The 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) The function $L^*(\cdot)$, representing the log-likelihood ratio between $\underline{\nu}_1$ and $\nu_0$ is continuous over the support of $\nu_0$.
Let $\tau_{\mathrm{SRP}}^* := \tau_{\mathrm{SRP}}^{\underline{\nu}^*, \eta, \psi_\eta}$ denote the SRP stopping rule defined with respect to the LFDs $(\nu_0, \underline{\nu}_1)$, with parameters $\eta$ and $\psi_\eta$ chosen such that the asymptotic optimality property of (13) is satisfied. Then the stopping rule $\tau_{\mathrm{SRP}}^*$ is also asymptotically optimal for (5) in the following sense: For every $0 < \alpha < 1$ and for any stopping rule $\tau$ that satisfies the false alarm constraint $\mathrm{FAR}(\tau) \leq \alpha$, we have

$$\sup_{\nu^1 \in \mathcal{P}_1} J_{\mathrm{SRP}}^\nu(\tau) \geq \sup_{\nu^1 \in \mathcal{P}_1} J_{\mathrm{SRP}}^\nu(\tau_{\mathrm{SRP}}^*) + o(1) \qquad (14)$$

where $o(1) \to 0$ as $\alpha \to 0$.                                          $\square$

The result of (14) can be interpreted as follows: The difference between the worst-case values of the delays incurred by the stopping rule $\tau_{\mathrm{SRP}}^*$ and any other stopping rule $\tau$ approaches zero as the false alarm constraint $\alpha$ approaches zero.

Our proof, provided in the Appendix, is useful only when $\mathcal{P}_0$ is a singleton. It is possible that the asymptotic optimality result may still hold even for general $\mathcal{P}_0$, although the current proof is not applicable. We elaborate on this further in the discussion in Section IV on the Bayesian criterion, and also in the Appendix following the proof of the theorem.

We also note that in some cases our proof can be adapted to obtain tests that are exactly optimal for the robust Pollak criterion of (5). Polunchenko and Tartakovsky [25] study the Shiryaev-Roberts procedure (SR-$r$) which is identical to the SRP procedure described earlier, except for the fact that $R_0$ is not random but fixed at some constant $r$. In [25, Theorem 2], the exact nonasymptotic optimality of the SR-$r$ procedure for detecting a change in distribution from $\mathrm{Exp}(1)$ to $\mathrm{Exp}(2)$ is shown, where $\mathrm{Exp}(\theta)$ refers to an exponential distribution with mean $\theta^{-1}$. Using that result, the proof of Theorem III.3 can be adapted to obtain the exact robust solution to the optimization problem in (5). In particular it can be shown that the SR-$r$ procedure for detecting change from $\mathrm{Exp}(1)$ to $\mathrm{Exp}(2)$ given in [25, Theorem 2] is also optimal for (5) when $\mathcal{P}_0 = \{\mathrm{Exp}(1)\}$ and $\mathcal{P}_1 = \{\mathrm{Exp}(\theta) : \theta \geq 2\}$.

### D. Bayesian Criterion

When the distributions $\nu_0$ and $\nu_1$ are known and the prior distribution of the change-point is geometric, the solution to (3) is given by the Shiryaev test [10]. Denoting the parameter of the geometric distribution by $\rho$, we have

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

The Shiryaev stopping rule is based on comparing the posterior probability of change to a threshold $\eta'$

$$\tau_S = \inf \{n \geq 1 : \mathsf{P}^\nu(\Lambda \leq n | \mathcal{F}_n) \geq \eta'\}.$$

It can be equivalently expressed as

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

where the threshold $\eta$ is chosen such that $\mathrm{PFA}(\tau_S) = \mathsf{P}^\nu(\tau_S < \Lambda) = \alpha$. The following theorem, proved in the Appendix, identifies a solution to the robust Shiryaev problem (6).

*Theorem III.4:* Suppose the following conditions hold:
(i) The 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) The prior distribution of the change-point is a geometric distribution.
(iii) The function $L^*(\cdot)$, representing the log-likelihood ratio between $\underline{\nu}_1$ and $\nu_0$, is continuous over the support of $\nu_0$.
Then the optimal stopping rule that solves (6) is given by the following Shiryaev test:

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

where the threshold $\eta$ is chosen so that $\mathsf{P}^*(\tau_S^* < \Lambda) = \alpha$.     $\square$

We note that our results under the Bayesian and Pollak criteria are applicable only when the pre-change distribution is known exactly and hence these results are weaker than our result under the Lorden criterion. Suppose $\mathcal{P}_0$ is not a singleton and $(\mathcal{P}_0, \mathcal{P}_1)$ is jointly stochastically bounded by $(\overline{\nu}_0, \underline{\nu}_1)$. In this case, the stopping rule $\tau_S^*$ defined with respect to $(\overline{\nu}_0, \underline{\nu}_1)$ is not optimal for the robust Bayesian criterion (6). In particular, when the pre-change distribution is $\nu_0 \neq \overline{\nu}_0$ and the post-change distribution is $\nu_1 = \underline{\nu}_1$, it can be shown that the average detection delay $\mathrm{ADD}^\nu(\tau_S^*)$ of the stopping rule $\tau_S^*$ is in general higher than the average detection delay $\mathrm{ADD}^*(\tau_S^*)$ when the pre-change and post-change distributions are $(\overline{\nu}_0, \underline{\nu}_1)$. This is because the likelihood ratios of the pre-change observations appearing in (16) are stochastically larger under $\overline{\nu}_0$ than under $\nu_0$. This leads to a stopping time that is stochastically smaller under $(\overline{\nu}_0, \underline{\nu}_1)$ than under $(\nu_0, \underline{\nu}_1)$. Hence, there is no reason to believe that $\tau_S^*$ solves the robust problem (6).

Even in the case of the Pollak criterion studied in Section III-C, our robust result holds only when $\mathcal{P}_0$ is a singleton and the JSB condition holds. However, unlike in the Bayesian case, we do not have a simple explanation for why the result cannot be extended to the setting where the pre-change

distribution is not known exactly. It is possible that for some specific choices of the uncertainty classes, the stopping rule designed with respect to $(\overline{\nu}_0, \underline{\nu}_1)$ may be asymptotically optimal for the robust problem of (5), although we do not expect this to be true in general.

However, such a problem does not arise for the robust CUSUM test we studied in Section III-B, since the worst-case detection delay $\text{WDD}^{\nu}(\tau_C^*)$ of the robust CUSUM depends only on the support of the pre-change distribution when post-change distribution is kept fixed at $\nu_1 = \underline{\nu}_1$.

*Comparison With Robust Sequential Detection:* It is interesting to compare our results with some known results on robust sequential detection. We have shown that provided the JSB condition and other regularity conditions hold, change detection tests designed with respect to the LFDs exactly solve the minimax robust change detection problem under the Lorden and Bayesian criteria. However, the known minimax optimality results in robust sequential detection are all for the asymptotic settings—as error probabilities go to zero [11] or as the size of the uncertainty classes diminishes [12]. Huber [11] showed that an exact minimax result does not hold for the robust sequential detection problem in general. He provided examples where the expected stopping times of the SPRT designed with respect to the LFDs are not least favorable under the LFDs. This is similar to the reason why the robust Shiryaev test is not optimal for the Bayesian problem when $\mathcal{P}_0$ is not a singleton as explained above.

## IV. SOME EXAMPLES AND SIMULATION RESULTS

### A. Gaussian Mean Shift

Here we consider a simple example to illustrate the results. Assume $\nu_0$ 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.

$$\mathcal{P}_0 = \{\mathcal{N}(0, 1)\}$$
$$\mathcal{P}_1 = \{\mathcal{N}(\theta, 1) : \theta \in [0.1, 3]\}. \tag{17}$$

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

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

*1) Bayesian Criterion:* 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 Fig. 1, we can see that the robust Shiryaev test gives the same average detection delay (ADD) as the optimal Shiryaev test at $\underline{\nu}_1$ which corresponds to $\theta = 0.1$ in the figure. This is expected since the robust test is identical to the optimal test at $\underline{\nu}_1$. For all other values of $\nu_1 \in \mathcal{P}_1$, the performance of the robust test is strictly better than the performance at $\underline{\nu}_1$ and hence this test is indeed minimax optimal. We also see in Fig. 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 $\theta$. The probability of false
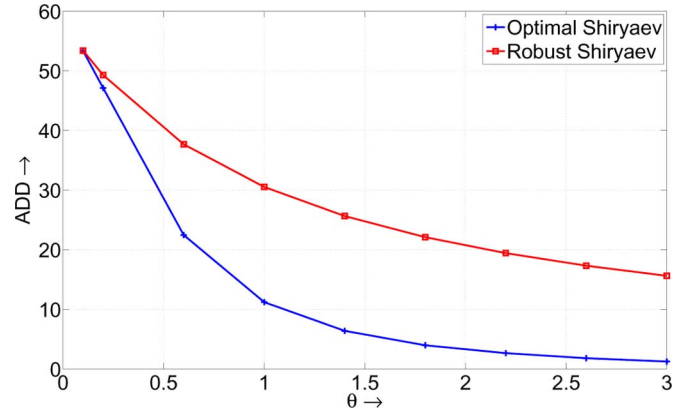


Fig. 1. Comparison of robust and nonrobust Shiryaev tests for $\alpha = 0.001$ for the Gaussian mean shift example.
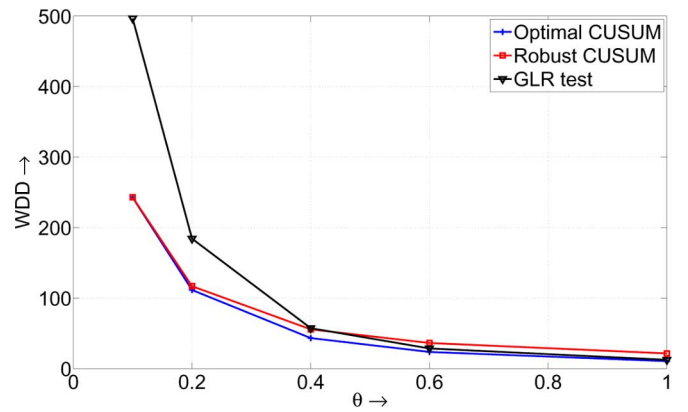


Fig. 2. Comparison of various tests for false alarm rate of $\alpha = 0.001$ for the Gaussian mean shift example.

TABLE I
DELAYS OBTAINED USING VARIOUS TESTS UNDER THE LORDEN CRITERION
FOR A FALSE ALARM RATE OF $\alpha = 0.001$

| $\theta$ | Optimal CUSUM | Robust CUSUM | GLR test |
|---|---|---|---|
| 0.1 | 242.7 | 242.7 | 496 |
| 0.2 | 111.5 | 116.8 | 184 |
| 0.4 | 43.2 | 55.6 | 57.2 |
| 0.6 | 23.5 | 36.3 | 28.6 |
| 1.0 | 10.5 | 21.5 | 12.35 |

alarm and average detection were estimated via Monte-Carlo simulations with a standard deviation of 0.1% for the estimates.

*2) Lorden Criterion and Comparison With GLR Test:* Under the Lorden criterion, we compared the performances of three tests—the optimal CUSUM test with known $\theta$, the robust CUSUM test designed with respect to the LFDs, and the CUSUM test based on the GLR Test suggested in [4]. The stopping time under the GLR test is given by

$$\tau_{\text{GLR}} = \inf \left\{ n \geq 1 : \max_{1 \leq k \leq n} \sup_{\nu_1 \in \mathcal{P}_1} \sum_{i=k}^{n} L^{\nu}(X_i) \geq \eta \right\} \tag{18}$$

where $\eta$ is chosen so that the false alarm constraint is met with equality. The GLR test does not require knowledge of $\theta$ but still achieves the same asymptotic performance as the optimal

TABLE II
DELAYS OBTAINED USING VARIOUS TESTS UNDER THE LORDEN CRITERION FOR $\epsilon$-UNCERTAINTY CLASSES WITH $\alpha = 0.001$ AND $\sigma_0 = 1$

| | $\epsilon = 0.05$ | | $\epsilon = 0.005$ | |
|---|---|---|---|---|
| $\sigma_1$ | Robust CUSUM | Optimal CUSUM | Robust CUSUM | Optimal CUSUM |
| 0.1 | 14.77 | 9.17 | 11.27 | 10.38 |
| 0.5 | 14.86 | 9.12 | 11.27 | 10.39 |
| 1 | 15.09 | 9.08 | 11.27 | 10.35 |
| 5 | 15.52 | 8.78 | 11.29 | 10.33 |
| 10 | 15.59 | 8.65 | 11.29 | 10.34 |

CUSUM test with known $\theta$ when the false alarm constraint goes to zero for some choices of the uncertainty classes including the example considered above.

Fig. 2 and Table I shows estimates of the worst-case detection delay (WDD) obtained under the these tests designed for a false alarm constraint of $\alpha = 0.001$, for various values of $\theta$. These values are estimated using Monte-Carlo simulations. The delay values have a standard deviation lower than 1% and the false alarm value has a standard deviation lower than 3%.

From the performance curves in Fig. 2 and the values in Table I we see that the GLR test gives better performance than our robust solution at higher values of $\theta$, and is close to optimal at these high values of $\theta$. However, the robust test gives much better performance than the GLR test at the low values of $\theta$. This is expected since the robust solution is minimax optimal and hence is expected to perform better at the unfavorable values of $\theta$.

An important difference between the two solutions is that although the robust CUSUM test based on the LFDs admits a simple recursive implementation like we described in (10), the GLR test is in general very complex to implement. This is because the supremum in (18) may be achieved at different values of $\nu_1$ for different $n$. Furthermore, the optimization in (18) may not be easy to solve for general uncertainty classes—particularly nonparametric classes like the $\epsilon$-uncertainty classes considered next.

### B. $\epsilon$-Contamination Classes

We now discuss an example in which the uncertainty class $\mathcal{P}_0$ is no longer a singleton. For some scalar $\epsilon \in (0, 1)$, consider the following $\epsilon$-contamination classes:

$$\mathcal{P}_0 = \{\nu_0 : \nu_0 = (1 - \epsilon)\mathcal{N}(0, 1) + \epsilon H_0, \quad H_0 \in \mathcal{P}(\mathbb{R})\} \tag{19}$$

$$\mathcal{P}_1 = \{\nu_1 : \nu_1 = (1 - \epsilon)\mathcal{N}(1, 1) + \epsilon H_1, \quad H_1 \in \mathcal{P}(\mathbb{R})\} \tag{20}$$

where $\mathcal{P}(\mathbb{R})$ is the collection of all probability measures on $\mathbb{R}$ and $\mathcal{N}(\mu, \sigma)$ denotes the probability measure corresponding to a Gaussian random variable with mean $\mu$ and variance $\sigma^2$. In other words, the distributions in uncertainty class $\mathcal{P}_i$ are mixtures of a Gaussian distribution with mean $i$ and unit variance, and an arbitrary probability distribution on $\mathbb{R}$ with weights given by $1 - \epsilon$ and $\epsilon$, respectively.

Following the method outlined in [11], we identified LFDs for these uncertainty classes and evaluated the performance of the robust test. Let $p_i$ denote the density function of a $\mathcal{N}(\mu, 1)$

random variable and let $q_i$ denote the density function of the least favorable distribution from $\mathcal{P}_i$. It is established in [11] that the densities of the LFDs have the following structure:

$$q_0(x) = \begin{cases} (1 - \epsilon)p_0(x) & \text{if } L(x) \leq b \\ \frac{1-\epsilon}{b}p_1(x) & \text{if } L(x) > b \end{cases} \tag{21}$$

$$q_1(x) = \begin{cases} (1 - \epsilon)p_1(x) & \text{if } L(x) > a \\ a(1 - \epsilon)p_0(x) & \text{if } L(x) \leq a \end{cases} \tag{22}$$

where $L(x) = \frac{p_1(x)}{p_0(x)}$. The scalars $a$ and $b$ are identified by the following relation:

$$(1 - \epsilon) \int_{\{x:L(x) \leq b\}} p_0(x)dx$$
$$+ \frac{1 - \epsilon}{b} \int_{\{x:L(x) > b\}} p_1(x)dx = 1$$
$$(1 - \epsilon) \int_{\{x:L(x) > a\}} p_1(x)dx$$
$$+ a(1 - \epsilon) \int_{\{x:L(x) \leq a\}} p_0(x)dx = 1.$$

In order to compare the performance of the robust test with that of the optimal test we chose the following distributions for $H_0$ and $H_1$:

$$H_0 = \mathcal{N}(0, \sigma_0), \sigma_0 \in [0.1, 10]$$
$$H_1 = \mathcal{N}(1, \sigma_1), \sigma_1 \in [0.1, 10].$$

Table II shows the values of the worst-case delay (WDD) obtained when $\sigma_0$ is kept fixed at $\sigma_0 = 1$ and $\sigma_1$ is varied. Shown are the results obtained using the robust CUSUM test as well as the optimal CUSUM test for $\epsilon = 0.05$ and for $\epsilon = 0.005$. We notice that the difference in performance between the robust test and the optimal test is larger for larger values of $\epsilon$. This matches the intuition that the cost of robustness would be higher for a larger uncertainty class of distributions. The delay values and false alarm rates were estimated to have standard deviations lower than 0.1% and 1%, respectively.

Table III shows the values of worst-case delay obtained under the optimal CUSUM tests when $\sigma_1$ is kept fixed at $\sigma_1 = 1$ and $\sigma_0$ is varied. The delay values and false alarm rates were estimated to have standard deviations lower than 0.1% and 1%, respectively. We have not included the delays obtained under the robust test, since the delay of the robust test is invariant with $\sigma_0$. The delay obtained under the robust test for $\epsilon = 0.05$ and $\epsilon = 0.005$ are respectively 15.09 and 11.27 as shown in the third row of Table II corresponding to $\sigma_1 = 1$.

TABLE III
DELAYS OBTAINED USING THE OPTIMAL CUSUM TEST FOR $\epsilon$-UNCERTAINTY
CLASSES WITH $\alpha = 0.001$ AND $\sigma_1 = 1$

| $\sigma_0$ | Optimal CUSUM for $\epsilon = 0.05$ | Optimal CUSUM for $\epsilon = 0.005$ |
|---|---|---|
| 0.1 | 10.56 | 10.55 |
| 0.5 | 10.50 | 10.52 |
| 1 | 10.44 | 10.56 |
| 5 | 10.02 | 10.58 |
| 10 | 9.85 | 10.59 |

## 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 the Lorden criterion, the Pollak criterion, and Shiryaev's Bayesian criterion. However, robustness comes at a potential cost. The optimal stopping rule designed for the LFDs may perform quite suboptimally for other distributions from the uncertainty class when compared with the optimal performance that can be obtained in the case where these distributions are known exactly. Using an asymptotic analysis, we have also obtained an analytic upper bound on this cost of robustness for the robust solution under the Lorden criterion. Nevertheless for some parameter ranges our robust test obtains significant performance improvement over the CUSUM test designed for the Generalized Likelihood Ratio statistic, which is a benchmark for the composite quickest change detection problem. Our robust solution also has the added advantage that it can be implemented in a simple recursive manner, while the GLR test does not admit a recursive solution in general, and may require the solution to a complex nonconvex optimization problem at every time instant.

## APPENDIX

*1) Proof of Lemma III.1:* We prove this claim by induction. For $n = 1$, the claim holds because if $h : \mathbb{R} \mapsto \mathbb{R}$ is a nondecreasing continuous function we have

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

Assume the claim is true for $n = N$ and now consider $n = N + 1$. For any fixed $x_1^N \in \mathbb{R}^N$, since the function $h$ is nondecreasing in each of its components, it follows by the proof for $n = 1$ that

$$h(x_1, x_2, \ldots, x_N, U_{N+1}) \succ h(x_1, x_2, \ldots, x_N, V_{N+1}). \quad (23)$$

We further have

$$\begin{aligned}
&\mathsf{P}(h(U_1, U_2, \ldots, U_{N+1}) \geq t) \\
&= \int f_{U_1^N}\left(x_1^N\right) \mathsf{P}(h(x_1, x_2, \ldots, x_N, U_{N+1}) \geq t) dx_1^N \\
&\geq \int f_{U_1^N}\left(x_1^N\right) \mathsf{P}(h(x_1, x_2, \ldots, x_N, V_{N+1}) \geq t) dx_1^N
\end{aligned}$$
$$(24)$$

$$= \mathsf{P}(h(\tilde{U}_1, \tilde{U}_2, \ldots, \tilde{U}_N, V_{N+1}) \geq t) \quad (25)$$

$$= \int f_{V_{N+1}}(y) \mathsf{P}(h(\tilde{U}_1, \tilde{U}_2, \ldots, \tilde{U}_N, y) \geq t) dy$$

$$\geq \int f_{V_{N+1}}(y) \mathsf{P}(h(V_1, V_2, \ldots, V_N, y) \geq t) dy$$

$$= \mathsf{P}(h(V_1, V_2, \ldots, V_{N+1}) \geq t) \quad (26)$$

where (24) is obtained via (23). The variables $\tilde{U}_i$ appearing in (25) are random variables with exact same statistics as $U_i$ and independent of $V_i$'s. The inequality of (26) 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. $\square$

*2) Proof of Theorem III.2:*

*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 (9) 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 $\mathrm{WDD}\left(\tau_C^*\right)$ and $\mathrm{FAR}\left(\tau_C^*\right)$ 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 $Y_i^\nu$ to denote the random variable $L^*(X_i)$ when the pre-change and post-change distributions are $\nu_0$ and $\nu_1$, 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 (8) 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 $\tau_C^*$ is independent of the true value of the post-change distribution. Fix the change-point to be $\lambda$. Now, to complete the proof for the scenario where $\mathcal{P}_0$ is a singleton, we will show that for all $\lambda \geq 1$

$$\mathsf{E}_\lambda^*\left[(\tau_C^* - \lambda + 1)^+ | \mathcal{F}_{\lambda-1}\right] \succ \mathsf{E}_\lambda^\nu\left[(\tau_C^* - \lambda + 1)^+ | \mathcal{F}_{\lambda-1}\right]$$
$$(27)$$

which will establish that the value of $\mathrm{WDD}\left(\tau_C^*\right)$, 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 $\overline{\nu}_0 = \nu_0$, both $Y_i^*$ and $Y_i^\nu$ have the same distributions for $i < \lambda$ and hence 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:

$$\mathsf{P}_\lambda^*\left((\tau_C^* - \lambda + 1)^+ \leq N | \mathcal{F}_{\lambda-1}\right)$$
$$\leq \mathsf{P}_\lambda^\nu\left((\tau_C^* - \lambda + 1)^+ \leq N | \mathcal{F}_{\lambda-1}\right) \quad (28)$$

which will then establish (27). Since $\tau_C^*$ is a stopping time, the event $\left\{(\tau_C^* - \lambda + 1)^+ \leq 0\right\}$ is $\mathcal{F}_{\lambda-1}-$ measurable. Hence, with probability one, (28) holds with equality for $N = 0$. Now

it suffices to verify (28) for $N \geq 1$. We know by the stochastic ordering condition on $\mathcal{P}_1$ that

$$Y_i^\nu \succ Y_i^*, \text{ for all } i \geq \lambda. \tag{29}$$

Now we have the following equivalence between two events:

$$\{\tau_C^* \leq N\} = \left\{ \max_{1 \leq n \leq N} \max_{1 \leq k \leq n} \sum_{i=k}^{n} L^*(X_i) \geq \eta \right\}$$
$$= \left\{ \max_{1 \leq k \leq n \leq N} \sum_{i=k}^{n} L^*(X_i) \geq \eta \right\}.$$

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 nondecreasing in each of its components as required by Lemma III.1. Hence, for $N \geq 1$, the following hold with probability one:

$$\mathsf{P}_\lambda^* \left( (\tau_C^* - \lambda + 1)^+ \leq N | \mathcal{F}_{\lambda-1} \right)$$
$$= \mathsf{P}_\lambda^* (\tau_C^* \leq N + \lambda - 1 | \mathcal{F}_{\lambda-1})$$
$$= \mathsf{P}_\lambda \left( f\left( Y_1^*, \ldots, Y_{N+\lambda-1}^* \right) \geq \eta | \mathcal{F}_{\lambda-1} \right)$$
$$\leq \mathsf{P}_\lambda \left( f\left( Y_1^\nu, \ldots, Y_{N+\lambda-1}^\nu \right) \geq \eta | \mathcal{F}_{\lambda-1} \right)$$
$$= \mathsf{P}_\lambda^\nu (\tau_C^* \leq N | \mathcal{F}_{\lambda-1})$$
$$= \mathsf{P}_\lambda^\nu \left( (\tau_C^* - \lambda + 1)^+ \leq N | \mathcal{F}_{\lambda-1} \right)$$

where the inequality follows from Lemma III.1 and (29), using the fact that $f$ is a nondecreasing function with respect to its last $N$ arguments and the fact that $Y_i^\nu = Y_i^*$ for $i < \lambda$. Thus, for all integers $N \geq 0$, (28) holds with probability one and hence (27) 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 (8).

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 III.1 that

$$\mathsf{P}_\infty^* (\tau_C^* \leq N) = \mathsf{P}_\infty \left( f\left( Y_1^*, \ldots, Y_N^* \right) \geq \eta \right)$$
$$\geq \mathsf{P}_\infty \left( f\left( Y_1^\nu, \ldots, Y_N^\nu \right) \geq \eta \right)$$
$$= \mathsf{P}_\infty^\nu (\tau_C^* \leq N).$$

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

$$\mathsf{E}_\infty^\nu (\tau_C^*) \geq \mathsf{E}_\infty^* (\tau_C^*) = \frac{1}{\alpha} \tag{30}$$

and hence the value of $\mathrm{FAR}(\tau_C^*)$ is no higher than $\alpha$ 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 $\lambda$. A useful observation is that for any given stopping rule $\tau$ and fixed post-change distribution $\nu_1$, the random variable $\mathsf{E}_\lambda^{\nu_0, \nu_1} \left[ (\tau - \lambda + 1)^+ | \mathcal{F}_{\lambda-1} \right]$ is a fixed deterministic function of the random observations $(X_1, \ldots, X_{\lambda-1})$, irrespective of the distribution $\nu_0$. Thus, the essential supremum of this random variable depends only on the support of $\nu_0$. Applying this observation to the stopping rule $\tau_C^*$, and using the relation (8), we have for all $\nu_0 \in \mathcal{P}_0$, $\nu_1 \in \mathcal{P}_1$

$$\text{ess sup } \mathsf{E}_\lambda^{\nu_0, \nu_1} \left[ (\tau_C^* - \lambda + 1)^+ | \mathcal{F}_{\lambda-1} \right]$$
$$\leq \text{ess sup } \mathsf{E}_\lambda^{\overline{\nu}_0, \nu_1} \left[ (\tau_C^* - \lambda + 1)^+ | \mathcal{F}_{\lambda-1} \right].$$

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

$$\text{ess sup } \mathsf{E}_\lambda^{\overline{\nu}_0, \nu_1} \left[ (\tau_C^* - \lambda + 1)^+ | \mathcal{F}_{\lambda-1} \right]$$
$$\leq \text{ess sup } \mathsf{E}_\lambda^* \left[ (\tau_C^* - \lambda + 1)^+ | \mathcal{F}_{\lambda-1} \right].$$

Taking the supremum over $\lambda \geq 1$, it follows from the above two relations that the value of $\mathrm{WDD}(\tau_C^*)$ 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, $\tau_C^*$ solves the robust problem (4). ∎

*3) Proof of Theorem III.3:*

*Proof:* Let $\tau_{\mathrm{SRP}}^* := \tau_{\mathrm{SRP}}^{\nu^*, \eta, \psi_\eta}$ denote the SRP stopping rule defined with respect to the LFDs $(\nu_0, \underline{\nu}_1)$ satisfying the asymptotic optimality property of (13) as mentioned in the statement of the theorem. It is easy to see that for any integers $\lambda \geq 1$ and $N \geq 1$, we have

$$\mathsf{P}_\lambda^\nu (\tau_{\mathrm{SRP}}^* - \lambda \leq N | \tau_{\mathrm{SRP}}^* \geq \lambda, R_0^* = r)$$
$$= \frac{\mathsf{P}_\lambda^\nu (\{\tau_{\mathrm{SRP}}^* - \lambda \leq N\} \cap \{\tau_{\mathrm{SRP}}^* \geq \lambda\} | R_0^* = r)}{\mathsf{P}_\lambda^\nu (\tau_{\mathrm{SRP}}^* \geq \lambda | R_0^* = r)}$$

where $R_0^*$ denotes the random variable with distribution $\psi_\eta$ used for initializing the iteration in (11). We follow the same steps as in the proof of Theorem III.2. Let $Y_i^\nu$ denote the random variable $L^*(X_i)$ when the pre-change and post-change distributions are $\nu_0$ and $\nu_1$ respectively. Since $\tau_{\mathrm{SRP}}^*$ is a stopping time the event $\{\tau_{\mathrm{SRP}}^* \geq \lambda\}$ is measurable with respect to the pre-change observations and hence we can represent this event as

$$\{\tau_{\mathrm{SRP}}^* \geq \lambda\} = \left\{ \left( Y_1^\nu, Y_2^\nu, \ldots, Y_{\lambda-1}^\nu \right) \in T \right\}$$

where $T$ is the set of pre-change trajectories corresponding to the event $\{\tau_{\mathrm{SRP}}^* \geq \lambda\}$. Hence, for any $r \in [0, \eta)$, we can express the conditional probability as shown in (31), at the bottom of the page, such that for all $t \in T$, the function $f_t : \mathbb{R}^{N-\lambda+1} \mapsto \mathbb{R}$

$$\mathsf{P}_\lambda^\nu (\tau_{\mathrm{SRP}}^* - \lambda \leq N | \tau_{\mathrm{SRP}}^* \geq \lambda, R_0^* = r)$$
$$= \frac{\int_T \mathsf{P}_\lambda \left( f_t \left( Y_\lambda^\nu, Y_{\lambda+1}^\nu, \ldots, Y_{\lambda+N}^\nu \right) \geq g(t, N, \eta) | R_0^* = r \right) \mathsf{P}_\lambda \left( \left( Y_1^\nu, Y_2^\nu, \ldots, Y_{\lambda-1}^\nu \right) \in dt | R_0^* = r \right)}{\int_T \mathsf{P}_\lambda \left( \left( Y_1^\nu, Y_2^\nu, \ldots, Y_{\lambda-1}^\nu \right) \in dt | R_0^* = r \right)} \tag{31}$$

satisfies the requirements of Lemma III.1, and $g(t, N, \eta)$ is some real-valued function. The exact form of function $f_t$ can be obtained from the iterations of (11) used to define the SRP stopping rule of (12). We note that in (31) the post-change distribution $\nu_1$ affects only the first term under the integral in the numerator. Thus, it follows by applying Lemma III.1 that

$$
\begin{aligned}
\mathsf{P}^*_\lambda \left(\tau^*_{\mathrm{SRP}} - \lambda \leq N | \tau^*_{\mathrm{SRP}} \geq \lambda, R^*_0 = r\right) \\
\leq \mathsf{P}^\nu_\lambda \left(\tau^*_{\mathrm{SRP}} - \lambda \leq N | \tau^*_{\mathrm{SRP}} \geq \lambda, R^*_0 = r\right)
\end{aligned}
\tag{32}
$$

for all $\nu_1 \in \mathcal{P}_1$. Hence, it further follows that

$$
\sup_{\nu_1 \in \mathcal{P}_1} \mathsf{E}^\nu_\lambda \left(\tau^*_{\mathrm{SRP}} - \lambda | \tau^*_{\mathrm{SRP}} \geq \lambda\right) = \mathsf{E}^*_\lambda \left(\tau^*_{\mathrm{SRP}} - \lambda | \tau^*_{\mathrm{SRP}} \geq \lambda\right).
\tag{33}
$$

We also observe that for any stopping rule $\tau$ that satisfies the false alarm constraint $\mathrm{FAR}(\tau) \leq \alpha$, we have

$$
\begin{aligned}
\sup_{\nu_1 \in \mathcal{P}_1} \sup_{\lambda \geq 1} & \mathsf{E}^\nu_\lambda (\tau - \lambda | \tau \geq \lambda) \\
& \geq \sup_{\lambda \geq 1} \mathsf{E}^*_\lambda (\tau - \lambda | \tau \geq \lambda) \\
& \geq \sup_{\lambda \geq 1} \mathsf{E}^*_\lambda \left(\tau^*_{\mathrm{SRP}} - \lambda | \tau^*_{\mathrm{SRP}} \geq \lambda\right) + o(1) \\
& = \sup_{\nu_1 \in \mathcal{P}_1} \sup_{\lambda \geq 1} \mathsf{E}^\nu_\lambda \left(\tau^*_{\mathrm{SRP}} - \lambda | \tau^*_{\mathrm{SRP}} \geq \lambda\right) + o(1)
\end{aligned}
$$

where the second relation follows from the fact that $\tau^*_{\mathrm{SRP}}$ satisfies the asymptotic optimality of (13) when the true post-change distribution is $\underline{\nu}_1$, and the last equality follows from (33). This completes the proof of the theorem. ∎

We note that if the robust SRP stopping rule $\tau^*_{\mathrm{SRP}}$ is used when $\mathcal{P}_0$ is not a singleton, the crucial step of (32) does not hold for $\nu_0 \neq \overline{\nu}_0$ and $\nu_1 = \underline{\nu}_1$. Thus, our proof of optimality of the robust SRP stopping rule does not hold when the pre-change distribution is unknown.

*4) Proof of Theorem III.4:*

*Proof:* The proof is very similar to that of *Case 1* in Theorem III.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 $\mathrm{PFA}\left(\tau^*_S\right)$, it is clear that the probability of false alarm depends only on the pre-change distribution and hence the constraint in (6) is met by the stopping time $\tau^*_S$. Hence, it suffices to show that the value of $\mathrm{ADD}\left(\tau^*_S\right)$ 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 $Y^\nu_i \succ Y^*_i$ where $Y^*_i$ and $Y^\nu_i$ are as defined in the proof of Theorem III.2. As before, the function

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

is continuous and nondecreasing in each of its components as required by Lemma III.1. Using these facts, we can show the fol-

lowing by proceeding exactly as in the proof of Theorem III.2: Conditioned on $\Lambda = \lambda$

$$
\mathsf{E}^*_\lambda \left((\tau^*_S - \lambda)^+ | \mathcal{F}_{\lambda-1}\right) \succ \mathsf{E}^\nu_\lambda \left((\tau^*_S - \lambda)^+ | \mathcal{F}_{\lambda-1}\right).
$$

Thus, we have $\mathsf{E}^*_\lambda \left((\tau^*_S - \lambda)^+\right) \geq \mathsf{E}^\nu_\lambda \left((\tau^*_S - \lambda)^+\right)$ and by averaging over $\lambda$, we get

$$
\mathsf{E}^* \left((\tau^*_S - \Lambda)^+\right) \geq \mathsf{E}^\nu \left((\tau^*_S - \Lambda)^+\right).
$$

∎

## REFERENCES

[1] J. Unnikrishnan, V. V. Venugopal, and S. Meyn, "Least favorable distributions for robust quickest change detection," in *Proc. 2009 IEEE Int. Symp. Information Theory*, Seoul, Korea, 2009.

[2] E. S. Page, "Continuous inspection schemes," *Biometrika*, vol. 41, pp. 100–115, 1954.

[3] G. V. Moustakides, "Optimal stopping times for detecting changes in distributions," *Ann. Statist.*, vol. 14, no. 4, pp. 1379–1387, Dec. 1986.

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

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

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

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

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

[9] B. E. Brodsky and B. Darkhovsky, *Nonparametric Methods in Change-Point Problems*. Dordrecht, The Netherlands: Kluwer, 1993.

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

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

[12] P. X. Quang, "Robust sequential testing," *Ann. Statist.*, vol. 13, no. 2, pp. 638–649, Jun. 1985.

[13] R. Crow and S. Schwartz, "On robust quickest change detection procedures," in *Proc., IEEE Int. Symp. Information Theory*, Trondheim, Norway, 1994, pp. 258–258.

[14] H. V. Poor and O. Hadjiliadis, *Quickest Detection*. Cambridge, U.K.: Cambridge Univ. Press, 2009.

[15] A. G. Tartakovsky and A. Polunchenko, "Quickest changepoint detection in distributed multisensor systems under unknown parameters," in *Proc. 11th Int. Conf. Information Fusion*, Cologne, Germany, Jul. 2008.

[16] L. Gordon and M. Pollak, "A robust surveillance scheme for stochastically ordered alternatives," *Ann. Statist.*, vol. 23, no. 4, pp. 1350–1375, 1995.

[17] T. Lai, "Sequential analysis: Some classical problems and new challenges," *Stat. Sin.*, vol. 11, pp. 303–408, 2001.

[18] D. Siegmund and E. S. Venkatraman, "Using the generalized likelihood ratio statistic for sequential detection of a change-point," *Ann. Statist.*, vol. 23, no. 1, pp. 255–271, 1995.

[19] B. Brodsky and B. Darkhovsky, "Asymptotically optimal sequential change-point detection under composite hypotheses," in *Proc. 44th Conf. Decision and Control/European Control Conf. (CDC-ECC )'05*, Dec. 2005, pp. 7347–7351.

[20] 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.

[21] V. V. Veeravalli, T. Başar, and H. V. Poor, "Minimax robust decentralized detection," *IEEE Trans. Inf. Theory*, vol. 40, no. 1, pp. 35–40, Jan. 1994.

[22] H. Poor, "Robust decision design using a distance criterion," *IEEE Trans. Inf. Theory*, vol. 26, no. 5, pp. 575–587, Sep. 1994.

[23] B. C. Levy, *Principles of Signal Detection and Parameter Estimation*. New York: Springer, 2008.

[24] A. Wald, *Sequential Analysis*. New York: Wiley, 1947.

[25] A. S. Polunchenko and A. G. Tartakovsky, "On optimality of the Shiryaev-Roberts procedure for detecting a change in distribution," *Ann. Statist.*, vol. 38, no. 6, pp. 3445–3457, 2010.

**Jayakrishnan Unnikrishnan** (S'06–M'10) received the B.Tech. degree in electrical engineering from the Indian Institute of Technology Madras, Chennai, in 2005 and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Illinois at Urbana-Champaign in 2007 and 2010, respectively.

He is a postdoctoral Researcher at the School of Computer and Communication Sciences, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland. His current research interests include detection and estimation theory, wireless communications, and information theory.

Dr. Unnikrishnan is a recipient of the Vodafone Graduate Fellowship Award from the University of Illinois at Urbana-Champaign for 2007–2008 and the E. A. Reid Fellowship Award from the ECE department at the University of Illinois at Urbana-Champaign for 2010–2011.

**Venugopal V. Veeravalli** (S'86–M'92–SM'98–F'06) received the Ph.D. degree in 1992 from the University of Illinois at Urbana-Champaign, the M.S. degree in 1987 from Carnegie-Mellon University, Pittsburgh, PA, and the B.Tech. degree (Silver Medal Honors) in 1985 from the Indian Institute of Technology, Bombay, all in electrical engineering.

He joined the University of Illinois at Urbana-Champaign in 2000, where he is currently a Professor in the Department of Electrical and Computer Engineering, and a Research Professor in the Coordinated Science Laboratory. He served as a Program Director for Communications Research at the U.S. National Science Foundation, Arlington, VA, from 2003 to 2005. He has previously held academic positions at Harvard University, Rice University, and Cornell University, and has been on sabbatical at MIT, IISc Bangalore, and Qualcomm, Inc. His research interests include distributed sensor systems and networks, wireless communications, detection and estimation theory, and information theory.

Dr. Veeravalli is a Distinguished Lecturer for the IEEE Signal Processing Society for 2010–2011. He has been on the Board of Governors of the IEEE Information Theory Society. He has been an Associate Editor for Detection and Estimation for the IEEE TRANSACTIONS ON INFORMATION THEORY and for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS. Among the awards he has received for research are the IEEE Browder J. Thompson Best Paper Award, the National Science Foundation CAREER Award, and the Presidential Early Career Award for Scientists and Engineers (PECASE).

**Sean P. Meyn** (S'85–M'87–SM'95–F'02) received the B.A. degree in mathematics from the University of California at Los Angeles in 1982 and the Ph.D. degree in electrical engineering from McGill University, Montreal, QC, Canada, in 1987.

He is a Professor in the Department of Electrical and Computer Engineering, a Research Professor in the Coordinated Science Laboratory, and Director of the Decision and Control Laboratory, all at the University of Illinois at Urbana-Champaign. His research interests include stochastic processes, control and optimization, complex networks, and information theory. Current funding is provided by the National Science Foundation, the Department of Energy, AFOSR, and DARPA. He is author of the book *Control Techniques for Complex Networks* (Cambridge, U.K.: Cambridge University Press, 2007) and coauthor (with R. Tweedie) of the monograph *Markov Chains and Stochastic Stability* (London, U.K.: Springer-Verlag, 1993) (the 2009 edition is published in the Cambridge Mathematical Library).

Dr. Meyn received jointly (with R. Tweedie) the 1994 ORSA/TIMS Best Publication In Applied Probability Award.