Asymptotic Minimax Robust and Misspecified Lorden Quickest Change Detection For Dependent Stochastic Processes

Timothy L. Molloy and Jason J. Ford
School of Electrical Engineering and Computer Science,
Queensland University of Technology
Brisbane, QLD 4001, Australia
Email: {t.molloy, j2.ford}@qut.edu.au

***Abstract*—The quick detection of an abrupt unknown change in the conditional distribution of a dependent stochastic process has numerous applications. In this paper, we pose a minimax robust quickest change detection problem for cases where there is uncertainty about the post-change conditional distribution. Our minimax robust formulation is based on the popular Lorden criteria of optimal quickest change detection. Under a condition on the set of possible post-change distributions, we show that the widely known cumulative sum (CUSUM) rule is asymptotically minimax robust under our Lorden minimax robust formulation as a false alarm constraint becomes more strict. We also establish general asymptotic bounds on the detection delay of misspecified CUSUM rules (i.e. CUSUM rules that are designed with post-change distributions that differ from those of the observed sequence). We exploit these bounds to compare the delay performance of asymptotically minimax robust, asymptotically optimal, and other misspecified CUSUM rules. In simulation examples, we illustrate that asymptotically minimax robust CUSUM rules can provide better detection delay performance at greatly reduced computation effort compared to competing generalised likelihood ratio procedures.**

I. Introduction

Quickly detecting an abrupt unknown change in the distribution of a dependent stochastic process on the basis of sequential observations is important in many applications including manoeuvring target tracking [1]–[4], anomaly detection [2] and fault detection [5]. However, the theory of quickest change detection has only recently been extended beyond simple models where the observations before and after the unknown change-time are independent and identically distributed (i.i.d.) with known distributions [6]–[8]. Motivated by the practical importance of detecting an unknown change in a dependent stochastic process, this paper investigates minimax robust quickest change detection for general dependent processes when the post-change distribution is uncertain.

In its standard formulation, quickest change detection is a dynamic hypothesis test between a null hypothesis that a change has not yet occurred (the no-change hypothesis), and an alternative hypothesis that a change has occurred at some (unknown) previous time (the change hypothesis). Typically, performance criteria for quickest change detection consist of a measure of the detection delay (e.g. the time between when a change occurs and when the no-change hypothesis is rejected) and a constraint on the false alarm rate (e.g. the time before the no-change hypothesis is incorrectly rejected). One of the most well known formulations of the quickest change detection problem is the Lorden criterion [9] (others being the Pollak and Bayesian criteria [10]).

In the case of i.i.d. observations with known pre-change and post-change distributions, the well known cumulative sum (CUSUM) procedure was shown to be optimal under the Lorden criterion in [11]. However, there appear to be few Lorden optimality results for cases where the post-change distribution is unknown [8], [10]. Notably, only limited progress has been made in establishing the properties of the popular generalised likelihood ratio (GLR) procedure under the Lorden criterion when the post-change distribution is unknown (c.f. [6] and [7]).

Recently in the i.i.d. case, several authors have proposed minimax robust quickest change detection formulations of the Lorden criterion for detecting a change when the pre-change and post-change distributions are unknown [10], [12], [13]. These i.i.d. process Lorden minimax robust formulations seek to guarantee minimum levels of detection delay performance over uncertainty sets of possible pre-change and post-change distributions. Under the assumption of i.i.d. observations and an assumption that the uncertainty sets satisfy a stochastic boundedness condition, CUSUM was shown to be minimax robust under the Lorden minimax robust criterion [13]. The results of [13] suggest that minimax robust CUSUM rules may perform better in practice than GLR rules (partly because GLR rules are known to be computationally expensive since they are based on maximum likelihood parameter estimation [2], [6], [7]). However, the required stochastic boundedness conditions appear difficult to generalise to dependent stochastic processes.

In the dependent processes case, most previous quickest change detection work is limited to when the pre-change and post-change distributions are known [6], [14], [15]. Furthermore, the optimality of CUSUM under the Lorden criterion has only been established in an asymptotic sense as a constraint on the false alarm rate is tightened [6]. Motivated by the limited existing results for detecting a change in a dependent process when the post-change distribution is unknown, in this paper we pose a Lorden minimax robust quickest change detection problem and exploit the existing asymptotic optimality results of [6] to establish the robustness properties of CUSUM rules for general dependent processes.

The main contribution of this paper is the formulation and investigation of a Lorden minimax robust quickest change detection problem for dependent processes when the post-change distribution is unknown (the pre-change distribution is assumed to be known). Importantly, under a condition on the uncertainty set of possible post-change distributions, we establish that the popular CUSUM stopping rule asymptotically solves our Lorden minimax robust quickest change detection problem. In the i.i.d. case, our condition on the uncertainty set is a relaxation of the joint stochastic boundedness condition of [13], and hence our results also provide insight into Lorden minimax robust quickest detection when the observations are i.i.d. and the uncertainty set is not known to obey the conditions of [13]. A secondary contribution of this paper is the derivation of new bounds on the Lorden delay costs of asymptotically minimax robust CUSUM stopping rules (and more generally, rules that have been designed with a misspecified post-change distributions, i.e. misspecified rules). Finally, our simulation examples suggest that asymptotically minimax robust rules may perform better in practice than competing adaptive quickest change detection procedures that are more computationally expensive (such as GLR procedures).

This paper is organised as follows: In Section II we propose our Lorden minimax robust quickest change detection problem. In Section III we establish our delay bounds for misspecified CUSUM rules before exploiting them to develop our asymptotic Lorden minimax robust results in Section IV. In Section V we present illustrative examples and simulations to demonstrate the utility of our results. Some conclusions are given in Section VI.

II. Problem Formulation

Let $Y_k \in \mathbb{R}^n$ for $k \geq 1$ be a sequence of (possibly dependent) random variables. We will assume that the process Y_k has an unknown (possibly random) change-time $\lambda \geq 1$ in the sense that the conditional probability of Y_k given $Y_{[1,k-1]} \triangleq \{Y_1, Y_2, \dots, Y_{k-1}\}$ is described by the pre-change conditional probability distribution $\mu \in S_\mu$ for $k < \lambda$, and by the post-change conditional probability distribution $\nu \in S_\nu$ for $k \geq \lambda$. Here, we use S_μ and S_ν to denote the sets of pre-change and post-change conditional probability distributions, and $S \triangleq S_\mu \cup S_\nu$ to denote the set of all conditional probability distributions on $\mathbb{R}^n$. We will assume the existence of a probability space $(\Omega, \mathcal{F}, P_\lambda^{\mu,\nu})$ where Ω is a sample space of all infinite sequences $Y_{[1,\infty]}$, $\mathcal{F}$ is a $\sigma-$algebra of events, and where $P_\lambda^{\mu,\nu}$ is a probability measure associated with Y_k that is constructed by extending the finite-dimensional probability distributions implied by λ, μ and ν using Kolmogorov's extension theorem (see [16] for details of probability space construction). In the following, we will denote the expectation under $P_\lambda^{\mu,\nu}$ as $E_\lambda^{\mu,\nu}[\cdot]$, and use P_∞^μ and $E_\infty^\mu[\cdot]$ to denote the probability measure and expectation corresponding to a process with no change. We will let $\mathcal{F}_k \triangleq \sigma\left\{Y_{[1,k]}\right\}$ denote the filtration generated by Y_k.

In the quickest change detection problem, we observe Y_k sequentially with the aim of detecting the change (i.e. rejecting the hypothesis that a change has not occurred and stopping our observation of Y_k) as soon as possible after the change-time λ whilst avoiding false alarms. A quickest change detection procedure is therefore characterised by a stopping time $T \in S_T$ where S_T denotes the set of stopping times with respect to $\mathcal{F}_k$. Before we pose our Lorden minimax robust quickest change detection problem, we will first review the Lorden criteria for optimal quickest change detection.

A. Standard Lorden Optimal Formulation

In the standard Lorden formulation of optimal quickest change detection, the change-time $\lambda \geq 1$ is considered a deterministic unknown quantity. To introduce the Lorden formation, we will define the false alarm rate of a stopping rule $T \in S_T$ under pre-change distribution μ as

$$f_R(T) \triangleq \frac{1}{E_\infty^\mu[T]}.$$

We will denote the set of stopping times $T \in S_T$ satisfy $f_R(T) \leq \alpha$ for a given constraint $0 < \alpha < \infty$ as $S_R(\alpha)$.

Under the Lorden formulation, the worst case detection delay $D_L(T, \mu, \nu)$ of the stopping time $T \in S_T$ for a pre-change distribution μ and a post-change distribution ν is defined as

$$D_L(T, \mu, \nu) \triangleq \sup_{\lambda \geq 1} \operatorname{ess\,sup} E_\lambda^{\mu,\nu}\left[(T - \lambda + 1)^+ \middle| \mathcal{F}_{\lambda-1} \right]$$

where $x^+ \triangleq \max\{x, 0\}$. Lorden's formulation of quickest change detection is then the optimisation problem [9]

$$\inf_{T \in S_R(\alpha)} D_L(T, \mu, \nu). \tag{1}$$

B. Proposed Lorden Minimax Robust Formulation

Under the standard Lorden formulation of optimal quickest change detection, the conditional distributions μ and ν that specify the performance criteria are also assumed to describe the statistics of the sequence Y_k [10]. However, in many problems of importance the true post-change conditional distribution $\nu^* \in S_\nu$ that describes the sequence Y_k may be unknown, making it difficult to specify and solve the optimal Lorden quickest detection problem. In this paper, we will assume that the true post-change conditional distribution $\nu^* \in S_\nu$ is unknown but belongs to the (known) uncertainty set $\mathcal{Q}_1 \subset S_\nu$ (we assume that $\mu \notin \mathcal{Q}_1$). We will consider the Lorden minimax robust problem,

$$\inf_{T \in S_R(\alpha)} \sup_{\nu \in \mathcal{Q}_1} D_L(T, \mu, \nu), \tag{2}$$

where we will seek to find detection procedures that minimise the worst case delay amongst all possible post-change $\nu \in \mathcal{Q}_1$ distributions.

In many applications, even minimax robust quickest change detection problems may be difficult to specify and solve (e.g. the uncertainty set $\mathcal{Q}_1$ may be unknown). A (potentially naive) approach when minimax robust quickest change detection is difficult is to apply a misspecified detection procedure that is designed to have optimal properties under the misspecified conditional distribution $\bar{\nu}$ that differs from the true distribution ν^* of the observed sequence Y_k. In this paper, we will prove a general asymptotic bound on the misspecified detection delay performance of the popular cumulative sum (CUSUM) procedure which we introduce next.

III. Misspecified Lorden Quickest Change Detection

In this section, we consider the misspecified detection performance of the widely known cumulative sum (CUSUM) quickest detection procedure with stopping time [6]

$$T_C(\mu,\nu) \triangleq \inf\left\{k \geq 1 : \max_{1\leq n\leq k} Z_n^k(\mu,\nu) \geq h_C\right\} \quad (3)$$

where

$$Z_n^k(\mu,\nu) \triangleq \sum_{i=n}^{k} \log \frac{p^{\nu}\left(Y_k \middle| Y_{[1,k-1]}\right)}{p^{\mu}\left(Y_k \middle| Y_{[1,k-1]}\right)}$$

for all $1 \leq n \leq k$. Our results are developed under generalised versions of two key assumptions on the convergence of the log-likelihood ratio $Z_n^k(\mu,\nu)$ that were first proposed in [6]. To present our assumptions, let us introduce the non-negative function $\mathcal{K}(\cdot\|\cdot) : S \times S \mapsto [0,\infty)$ that is suitable for the following assumptions:

Assumption 1: Given $\mu \in S_\mu$ and $\nu \in S_\nu$,

$$\lim_{k\to\infty} \sup_{\lambda\geq 1} \operatorname{ess\,sup} P_\lambda^{\mu,\nu}\left(\max_{t\leq k} Z_\lambda^{\lambda+t}(\mu,\nu) \geq k(1+\delta)\mathcal{K}(\nu\|\mu) \middle| Y_{[1,\lambda-1]}\right) = 0$$

for any $\delta > 0$.

Assumption 2: Given $\mu \in S_\mu$ and $\nu,\bar{\nu} \in S_\nu$,

$$\lim_{k\to\infty} \sup_{1\leq\lambda\leq n} \operatorname{ess\,sup} P_\lambda^{\mu,\nu}\left(\frac{1}{k} Z_n^{n+k-1}(\mu,\bar{\nu}) \leq \mathcal{K}(\nu\|\mu) - \mathcal{K}(\nu\|\bar{\nu}) - \delta \middle| Y_{[1,n-1]}\right) = 0$$

for any $\delta > 0$, and for all $n \geq 1$.

Remark 3.1: We note that $\mathcal{K}(\nu\|\mu)$ can be regarded as a Kullback-Leibler information number between the joint distributions associated with ν and μ. Examples of the Kullback-Leibler information number include the Kullback-Leibler divergence (or relative entropy) in the i.i.d. case, and the Kullback-Leibler divergence rate (or relative entropy rate) in case of uniformly recurrent Markov chains (see [6]).

We will now establish the detection delay of the CUSUM procedure (3) designed with a (possibly misspecified) post-change conditional distribution $\bar{\nu} \in S_\nu$.

Lemma 3.1: Consider $\mu \in S_\mu$ and $\nu,\bar{\nu} \in S_\nu$ such that Assumption 2 holds and $\mathcal{K}(\nu\|\mu) > \mathcal{K}(\nu\|\bar{\nu})$. Then, the misspecified CUSUM rule $T_C(\mu,\bar{\nu})$ given by (3) satisfies

$$D_L(T_C(\mu,\bar{\nu}),\mu,\nu) \leq (1+o(1))\frac{h_C}{\mathcal{K}(\nu\|\mu) - \mathcal{K}(\nu\|\bar{\nu})}$$

as $h_C \to \infty$ where $o(1) \to 0$ as $h_C \to \infty$.

Proof: Please see appendix. □

Clearly, the asymptotic upper bound on the delay of the misspecified CUSUM rule established in Lemma 3.1 is only reasonable when the misspecified distribution $\bar{\nu}$ satisfies $\mathcal{K}(\nu\|\mu) > \mathcal{K}(\nu\|\bar{\nu})$. We highlight that when the misspecified distribution $\bar{\nu}$ violates this condition (i.e. when $\mathcal{K}(\nu\|\mu) \leq \mathcal{K}(\nu\|\bar{\nu})$), the misspecified CUSUM rule $T_C(\mu,\bar{\nu})$ may be better suited for detecting a change from ν to μ than μ to ν. We will next quantify the poor performance of the misspecified CUSUM rule $T_C(\mu,\bar{\nu})$ for the following class of misspecified distributions that violate the condition $\mathcal{K}(\nu\|\mu) > \mathcal{K}(\nu\|\bar{\nu})$ of Lemma 3.1.

Definition 3.1 (Conflicted Misspecified Distribution): Given $\mu \in S_\mu$ and $\nu,\bar{\nu} \in S_\nu$, we will call $\bar{\nu}$ a conflicted misspecified distribution if

$$\mathcal{K}(\nu,\mu) - \mathcal{K}(\nu,\bar{\nu}) = E_1^{\mu,\nu}\left[\log \frac{p^{\bar{\nu}}\left(Y_k \middle| Y_{[1,k-1]}\right)}{p^{\mu}\left(Y_k \middle| Y_{[1,k-1]}\right)}\right] \quad (4)$$

for all $k \geq 1$, and

$$\operatorname{ess\,sup} E_1^{\mu,\nu}\left[\left.\frac{p^{\bar{\nu}}\left(Y_k \middle| Y_{[1,k-1]}\right)}{p^{\mu}\left(Y_k \middle| Y_{[1,k-1]}\right)}\right| \mathcal{F}_{k-1}\right] \leq 1 \quad (5)$$

for all $k \geq 1$.

The following lemma establishes that conflicted misspecified distributions violate the condition $\mathcal{K}(\nu\|\mu) > \mathcal{K}(\nu\|\bar{\nu})$ of Lemma 3.1, and that the detection delay of a misspecified CUSUM rule with a conflicted misspecified distribution is lower bounded by an exponential.

Lemma 3.2: Consider $\mu \in S_\mu$ and $\nu,\bar{\nu} \in S_\nu$ and suppose that $\bar{\nu}$ is a conflicted misspecified distribution in the sense of Definition 3.1. Then $\mathcal{K}(\nu,\mu) \leq \mathcal{K}(\nu,\bar{\nu})$, and the misspecified CUSUM rule $T_C(\mu,\bar{\nu})$ satisfies

$$D_L(T_C(\mu,\bar{\nu}),\mu,\nu) \geq e^{h_C} - 1.$$

Proof: Please see appendix. □

Remark 3.2: We highlight that Lemma 3.2 implies that a misspecified CUSUM rule $T_C(\mu,\bar{\nu})$ designed with a conflicted misspecified distribution $\bar{\nu}$ (in the sense of Definition 3.1), may have a detection delay $D_L(T_C(\mu,\bar{\nu}),\mu,\nu)$ that is longer than its mean time to false alarm $E_\infty^\mu[T_C(\mu,\bar{\nu})]$ (since [6, Theorem 4] gives that $E_\infty^\mu[T_C(\mu,\bar{\nu})] \geq e^{h_C}$).

We are now in a position to investigate our Lorden minimax robust quickest change detection problem.

IV. Asymptotically Minimax Robust Lorden Quickest Change Detection

In this section, we will investigate the asymptotic solution of our Lorden minimax robust quickest change detection problem. In order to establish that an asymptotic solution to our Lorden minimax robust quickest change detection problem exists, we will require the following uncertainty set concept related to the Pythagorean property of relative entropy (c.f. [17, Theorem 1]).

Definition 4.1 (ν_L-Pythagorean): Given $\mu \in S_\mu$, we will say that the set $\mathcal{Q}_1 \subset S_\nu$ is ν_L-Pythagorean if there exists some post-change conditional distribution $\nu_L \in \mathcal{Q}_1$ such that

$$\mathcal{K}\left(\nu \,\|\mu\right) \geq \mathcal{K}\left(\nu \,\|\nu_L\right) + \mathcal{K}\left(\nu_L \,\|\mu\right) \tag{6}$$

for all $\nu \in \mathcal{Q}_1$.

Remark 4.1: Whilst we acknowledge that determining if $\mathcal{Q}_1$ is ν_L-Pythagorean is non-trivial in general (since it depends on the form of $\mathcal{K}\left(\cdot\|\cdot\right)$ and hence on the classes of processes involved), when the observations Y_k are i.i.d. before and after the change-time, several useful uncertainty sets are already known to be ν_L-Pythagorean since if $\mathcal{Q}_1$ is joint stochastically bounded by ν_L in the sense of [13, Definition 1], it is also ν_L-Pythagorean.

Remark 4.2: We note that when $\mathcal{Q}_1$ is ν_L-Pythagorean, ν_L satisfies

$$\mathcal{K}\left(\nu_L \|\mu\right) = \inf_{\nu \in \mathcal{Q}_1} \mathcal{K}\left(\nu\|\mu\right). \tag{7}$$

Thus a useful (sufficient) method for testing if $\mathcal{Q}_1$ is ν_L-Pythagorean is to identify a $\nu_L \in \mathcal{Q}_1$ satisfying (7), and to verify that

$$\inf_{\nu \in \mathcal{Q}_1} \left[\mathcal{K}\left(\nu \,\|\mu\right) - \mathcal{K}\left(\nu \,\|\nu_L\right)\right] \geq \mathcal{K}\left(\nu_L \,\|\mu\right). \tag{8}$$

The following theorem establishes that if the uncertainty set $\mathcal{Q}_1$ is ν_L-Pythagorean, the distribution ν_L is least favourable in the sense that the CUSUM rule $T_C\left(\mu, \nu_L\right)$ is an asymptotic solution to the Lorden minimax robust quickest change detection problem.

Theorem 4.1: Consider $\mu \in S_\mu$ and some uncertainty set $\mathcal{Q}_1$ such that $\mathcal{Q}_1$ is ν_L-Pythagorean. Furthermore, suppose that Assumptions 1 and 2 hold for all $\nu \in \mathcal{Q}_1$ with $\bar{\nu} = \nu_L$. Then the post-change distribution ν_L and the CUSUM rule $T_C\left(\mu, \nu_L\right)$ with threshold $h_C \sim |\log \alpha|$ (as $\alpha \to 0$) specify an asymptotic saddle point of the Lorden minimax robust problem in the sense that,

$$\begin{aligned}\inf_{T \in S_R(\alpha)} D_L\left(T, \mu, \nu_L\right) &\sim D_L\left(T_C\left(\mu, \nu_L\right), \mu, \nu_L\right) \\ &\sim \sup_{\nu \in \mathcal{Q}_1} D_L\left(T_C\left(\mu, \nu_L\right), \mu, \nu\right)\end{aligned} \tag{9}$$

as $\alpha \to 0$. Moreover, the CUSUM rule $T_C\left(\mu, \nu_L\right)$ with $h_C \sim |\log \alpha|$ is asymptotically minimax robust in the sense that

$$\inf_{T \in S_R(\alpha)} \sup_{\nu \in \mathcal{Q}_1} D_L\left(T, \mu, \nu\right) \sim D_L\left(T_C\left(\mu, \nu_L\right), \mu, \nu_L\right)$$

as $\alpha \to 0$.

Proof: The first asymptotic equality of the Lorden saddle point (9) follows from the asymptotic optimality of the CUSUM rule $T_C\left(\mu, \nu_L\right)$ with $h_C \sim |\log \alpha|$ established in [6, Theorem 1] and [6, Theorem 3] under Assumptions 1 and 2, respectively.

To prove the second asymptotic equality of the Lorden saddle point (9) we note that the definition of ν_L in (6) implies that $\mathcal{K}\left(\nu \,\|\nu_L\right) < \mathcal{K}\left(\nu \,\|\mu\right)$ for all $\nu \in \mathcal{Q}_1$. Hence, under Assumption 2, Lemma 3.1 gives that for all $\nu \in \mathcal{Q}_1$,

$$\begin{aligned}D_L\left(T_C\left(\mu, \nu_L\right), \mu, \nu\right) &\leq (1 + o(1)) \frac{|\log \alpha|}{\mathcal{K}\left(\nu \,\|\mu\right) - \mathcal{K}\left(\nu \,\|\nu_L\right)} \\ &\leq (1 + o(1)) \frac{|\log \alpha|}{\mathcal{K}\left(\nu_L \,\|\mu\right)} \\ &= D_L\left(T_C\left(\mu, \nu_L\right), \mu, \nu_L\right)\end{aligned} \tag{10}$$

as $\alpha \to 0$ where the second line follows from the definition of ν_L and the third line follows from the asymptotic optimality of CUSUM established in [6, Theorem 3]. The second asymptotic equality of (9) follows since (10) holds for all $\nu \in \mathcal{Q}_1$ and since $\nu_L \in \mathcal{Q}_1$. The point $(T_C(\mu, \nu_L), \nu_L)$ therefore specifies the asymptotic saddle point (9). The existence of the asymptotic saddle point (9) is sufficient to imply that $T_C\left(\mu, \nu_L\right)$ is asymptotically Lorden minimax robust (c.f. [18, Proposition 3.4.2]) completing the proof. □

Before we illustrate the performance of the asymptotically minimax robust CUSUM rule $T_C\left(\mu, \nu_L\right)$ in simulations, we will characterise the asymptotic cost of robustness by considering the the ratio of the detection delay of the asymptotically minimax robust CUSUM rule $T_C\left(\mu, \nu_L\right)$ to the detection delay of the asymptotically optimal CUSUM rule $T_C\left(\mu, \nu^*\right)$.

Theorem 4.2: Consider $\mu \in S_\mu$ and some uncertainty set $\mathcal{Q}_1$ such that $\mathcal{Q}_1$ is ν_L-Pythagorean. Furthermore, suppose that Assumptions 1 and 2 hold for $\nu^* \in \mathcal{Q}_1$ with $\nu = \nu^*$ and $\bar{\nu} = \nu_L$. Then the CUSUM stopping rules $T_C\left(\mu, \nu_L\right)$ and $T_C\left(\mu, \nu^*\right)$ with $h_C \sim |\log \alpha|$ satisfy,

$$\frac{D_L\left(T_C\left(\mu, \nu_L\right), \mu, \nu^*\right)}{D_L\left(T_C\left(\mu, \nu^*\right), \mu, \nu^*\right)} \leq \frac{\mathcal{K}\left(\nu^* \|\mu\right)}{\mathcal{K}\left(\nu^* \|\mu\right) - \mathcal{K}\left(\nu^* \|\nu_L\right)}$$

as $\alpha \to 0$.

Proof: The theorem result follows by dividing the upper bound established in Lemma 3.1 under Assumption 2 by the asymptotic lower bound

$$D_L\left(T_C\left(\mu, \nu^*\right), \mu, \nu^*\right) \geq (1 + o(1)) \frac{|\log \alpha|}{\mathcal{K}\left(\nu^* \|\mu\right)}$$

established in [6, Theorem 1] (also [6, Theorem 3]) for the asymptotically optimal stopping rule $T_C\left(\mu, \nu^*\right)$ with $h_C \sim |\log \alpha|$ under Assumption 1. □

V. Examples and Simulation Results

In this section, we will illustrate and examine the performance of our asymptotic minimax robust CUSUM rules in two simulation examples. In our first example, we will consider an i.i.d. process case analogous to the problem of detecting a change in the residuals generated by an observer or filter (a ubiquitous problem in the applications of fault detection

[5] and manoeuvring target tracking [3], [4]). In our second example we will consider a Markov chain case to demonstrate our results in a dependent process setting comparable to the aircraft manoeuvre detection problem presented in [1].

Importantly, in each example, we will compare our asymptotic minimax robust CUSUM rule to two alternative approaches: a nominal CUSUM approach where the CUSUM rule $T_C(\mu, \nu_N)$ is (naively) designed with a nominal distribution at the centre (in a sense we later make clear) of the uncertainty set $\mathcal{Q}_1$; and a generalised likelihood ratio (GLR) approach with stopping rule [6]

$$T_G \triangleq \inf\left\{k \geq 1 : \max_{k-w<n\leq k-1} \sup_{\nu \in S_\nu} Z_n^k(\mu, \nu) \geq h_G\right\} \quad (11)$$

where h_G is chosen so that $f_R(T_G, \mu) \leq \alpha$, and $w \geq 1$ is a window length used to trade-off performance for computational effort. In order to provide a bound on achievable detection performance, we also simulated CUSUM rules $T_C(\mu, \nu^*)$ that had ideal knowledge of the post-change distributions ν^*. We highlight that in contrast to the GLR rule T_G, in our examples, all CUSUM rules have efficient recursive implementations since the CUSUM test statistic $S_k \triangleq \max_{1\leq n\leq k} Z_n^k(\mu, \nu)$ is given by $S_k \triangleq S_{k-1}^+ + Z_k^k(\mu, \nu)$ with $S_0 = 0$.

A. *i.i.d. Process Example*

In this example, we will suppose that the observations Y_k for $1 \leq k < \lambda$ are i.i.d. with known pre-change distribution $\mu = \mathcal{N}(0, 1)$, and the observations Y_k for $\lambda \leq k$ are i.i.d. with unknown (possibly non-Gaussian) post-change distribution $\nu^* \in \mathcal{Q}_1$. Here, $\mathcal{N}(x, \sigma^2)$ denotes the Gaussian distribution with mean x and variance σ^2. To define our uncertainty set $\mathcal{Q}_1$, let us introduce the nominal post-change distribution $\nu_N = \mathcal{N}(1, 1)$. We will assume that the unknown post-change distribution ν^* resides within the relative entropy uncertainty set $\mathcal{Q}_1 = \{\nu \in S_\nu : \mathcal{K}(\nu \| \nu_N) \leq \Delta\}$ for $\Delta = 0.46$ where here $\mathcal{K}(\cdot \| \cdot)$ is the relative entropy given by

$$\mathcal{K}(\nu \| \mu) = \int_{-\infty}^{\infty} p^\nu(y) \log \frac{p^\nu(y)}{p^\mu(y)} dy.$$

We note that Assumptions 1 and 2 hold for any $\nu \in \mathcal{Q}_1$ by the weak law of large numbers (see [16]).

We highlight that uncertainty sets defined by relative entropy constraints appear new in the problem of quickest change detection. In order to establish that $\mathcal{Q}_1$ is ν_L-Pythagorean we proceed along the lines of Remark 4.2 and solve (7). Following [19, pp. 251] and applying Lagrange multiplier techniques, we have that the density $p^L(y)$ associated with the distribution ν_L solving (7) is given by

$$p^L(y) = \frac{\left(q^0(y)\right)^\delta \left(p^N(y)\right)^{1-\delta}}{\int_{-\infty}^{\infty} \left(q^0(y)\right)^\delta \left(p^N(y)\right)^{1-\delta} dy} \quad (12)$$

where $q^0(y)$ and $p^N(y)$ are the densities associated with the pre-change μ and nominal ν_N distributions respectively, and $0 \leq \delta \leq 1$ is a scalar constant such that

$$\mathcal{K}(\nu_L \| \nu_N) = \Delta. \quad (13)$$

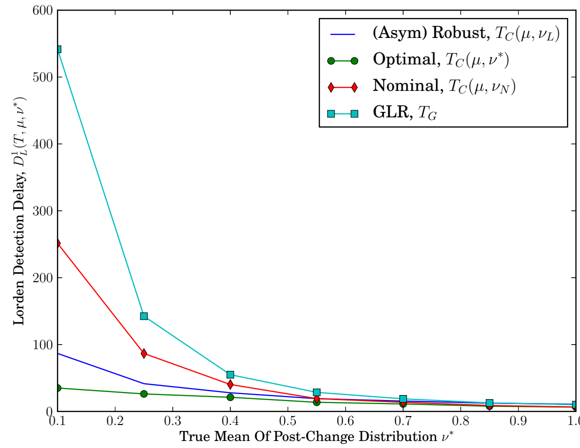

Fig. 1. I.I.D. Process Example: Detection delays $D_L(T, \mu, \nu^*)$ versus the mean x of the post-change distribution $\nu^* = \mathcal{N}(x, 0.81)$ for false alarm rate constraint $\alpha = 0.008$, nominal distribution $\nu_N = \mathcal{N}(1, 1)$ and least favourable distribution $\nu_L = \mathcal{N}(0.04, 1)$. Maximum percentage standard error of 7.7% and 7.9% for the Lorden delay and false alarm rates, respectively.

Substitution of Gaussian densities into (12) combined with algebraic manipulations give us that $\nu_L = \mathcal{N}((1-\delta), 1)$, and we solve (13) for $\delta = \sqrt{2\Delta} = 0.96$ by using the closed form expression for the relative entropy between Gaussians. Thus, $\nu_L = \mathcal{N}(0.04, 1)$ solves (7). Importantly, because $\mathcal{Q}_1$ is a convex set of probability measures (since $\mathcal{K}(\cdot \| \cdot)$ is a convex function of the probability measures) applying the Pythagorean identity of relative entropy on convex sets (c.f. [17, Theorem 1] and [17, Remark 1]) we have that $\mathcal{Q}_1$ is ν_L-Pythagorean.

To illustrate performance over a range of distributions in $\mathcal{Q}_1$, we simulated the asymptotically minimax robust CUSUM rule $T_C(\mu, \nu_L)$, the nominal CUSUM rule $T_C(\mu, \nu_N)$, and the GLR rule T_G for true post-change distributions $\nu^* = \mathcal{N}(x, 0.81)$ with means x in $[0.1, 1]$. We followed [7] and selected window lengths of $w \geq |\log \alpha|/\mathcal{K}(\nu^* \| \mu)$ with $\alpha = 0.0001$. Figure 1 shows the detection delays $D_L(T, \mu, \nu^*)$ versus the true post-change mean for a fixed false alarm rate of $\alpha = 0.008$. We note that the difference in performance between rules appears to decrease as the true mean approaches $x = 1$. However, the GLR rule performs poorly across the range of true post-change means, particularly at $x = 0.1$ (where the asymptotically robust CUSUM rule outperforms the GLR and nominal CUSUM rules). As expected, the nominal CUSUM rule performs better than the asymptotically robust CUSUM (and GLR) rule when the mean of the true post-change distribution approaches the mean of the nominal distribution ν_N (i.e. $x = 1$). Nevertheless, Figure 1 suggests that even at this modest false alarm rate, the worst delay of the asymptotically minimax robust CUSUM rule in this subset of $\mathcal{Q}_1$ is less than the worst delay of the nominal CUSUM and GLR rules.

The Lorden detection delays $D_L(T, \mu, \nu^*)$ for a fixed post-change distribution $\nu^* = \mathcal{N}(0.1, 0.81)$ are plotted against

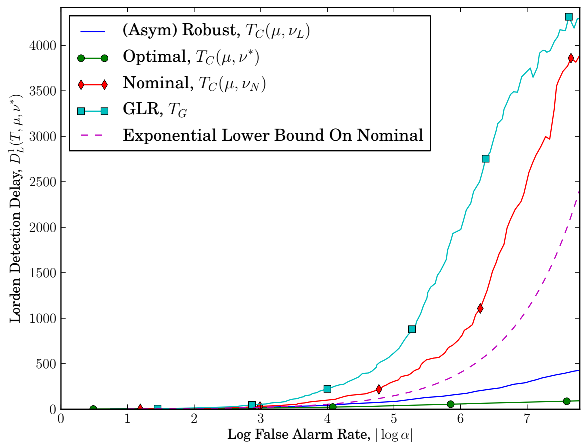

Fig. 2. I.I.D. Process Example: Detection delays $D_L(T,\mu,\nu^*)$ versus the log false alarm rate $|\log\alpha|$ for post-change distribution $\nu^* = \mathcal{N}(0.1, 0.81)$, nominal distribution $\nu_N = \mathcal{N}(1,1)$ and least favourable distribution $\nu_L = \mathcal{N}(0.04, 1)$. The exponential lower bound is computed for the nominal CUSUM $T_C(\mu,\nu_N)$ from Lemma 3.2. Maximum percentage standard error of 7.7% and 7.9% for the Lorden delay and false alarm rates, respectively.

the false alarm rate constraint in Figure 2. Figure 2 suggests that the delay of each rule increases at a different rate as $\alpha \to 0$. Indeed, for this particular true post-change distribution ν^*, the nominal distribution ν_N is a conflicted misspecified distribution in the sense of Definition 3.1 and so the delay of the nominal CUSUM rules $T_C(\mu,\nu_N)$ is under bounded by the exponential in Lemma 3.2. This lower bound is also plotted in Figure 2. We also note that the delay of the GLR rule also appears to be exponentially under bound. In contrast, Theorem 4.2 gives that the delay of the asymptotically minimax robust CUSUM rule is asymptotically over bounded by a linear function of $|\log\alpha|$ (since the delay of the optimal CUSUM rule is asymptotically linear).

B. Markov Chain Example

In this example, we will suppose that $Y_k \in S_Y$ for $k \geq 1$ is a Markov chain with state space $S_Y = \{1, 2\}$, initial state distribution $\psi_0 \in \mathbb{R}^2$ where $\psi_0^i = P(Y_1 = i)$, and one-step transition probability matrix $A \in S$ where $A^{ij} = P(Y_k = i \,|Y_{k-1} = j)$ for $1 \leq i, j \leq 2$. Here, (in a slight abuse of notation) we define $S \subset \mathbb{R}^{2\times 2}$ as the set of matrices A that satisfy $A^{ij} > 0$ and $\sum_{i=1}^{2} A^{ij} = 1$ for all $1 \leq i, j \leq 2$. We will suppose that the transition matrix $A \in S$ associated with the Markov chain Y_k changes at time $k = \lambda$ in the sense that $A = A_\mu$ for $0 \leq k < \lambda$ and $A = A_\nu$ for $k \geq \lambda$ where A_μ is a known pre-change transition matrix, and A_ν is an unknown transition matrix from the uncertainty set $\mathcal{Q}_1 \subset S$. We highlight that strict positivity of A is sufficient for Y_k to be aperiodic and irreducible under $P_\infty^{\mu,\nu}$ and $P_1^{\mu,\nu}$ for all A_μ and A_ν. Hence (similar to [6, Section IV]), the strong convergence of $k^{-1} Z_n^k(\mu,\nu)$ under $P_1^{\mu,\nu}$ for irreducible, aperiodic and absolutely continuous Markov chains (c.f. [20, Corollary 2]) implies that Assumptions 1 and 2 hold with

$$\mathcal{K}(\nu\,\|\mu) = \sum_{i=1}^{2}\sum_{j=1}^{2} \psi_\nu^i A_\nu^{ij} \log \frac{A_\nu^{ij}}{A_\mu^{ij}} \tag{14}$$

where $\psi_\nu(\cdot) : S_Y \mapsto (0, 1]$ is the stationary distribution of the chain Y_k under $P_1^{\mu,\nu}$.

For the purpose of this example, we will let the true (known) pre-change transition matrix A_μ, and the (known) nominal post-change transition matrix A_N be

$$A_\mu = \begin{bmatrix} 0.6 & 0.3 \\ 0.4 & 0.7 \end{bmatrix}, \text{ and } A_N = \begin{bmatrix} 0.8 & 0.5 \\ 0.2 & 0.5. \end{bmatrix},$$

respectively. We will assume that the unknown post-change transition matrix A_ν belongs to the uncertainty set $\mathcal{Q}_1 = \left\{ A \in S : A_N^{ij} - 0.1 \leq A^{ij} \leq A_N^{ij} + 0.1 \text{ for all } 1 \leq i, j \leq 2 \right\}$ centred on the nominal matrix A_N. In light of Remark 4.2, we used the closed form expression (14) to numerically solve the optimisation problems (7) and (8) for the transition matrix

$$A_L = \begin{bmatrix} 0.7 & 0.4 \\ 0.3 & 0.6 \end{bmatrix}$$

associated with the distribution $\nu_L \in \mathcal{Q}_1$ such that $\mathcal{Q}_1$ is ν_L-Pythagorean.

For the purpose of illustrating the performance of the asymptotic minimax robust, nominal and asymptotically optimal CUSUM rules, we simulated the rules over the set $\left\{ A \in S : 0.7 \leq A^{11} \leq 0.8 \text{ and } A^{22} = 0.5 \right\}$ of possible post-change transition matrices (we omitted the GLR rules because of computational limitations). Figure 3 shows the Lorden delays for a false alarm rate of $\alpha = 0.0005$. As in our i.i.d. example, Figure 3 suggests that asymptotically optimal performance is recovered by the nominal CUSUM rule as the true transition matrix A_ν approaches the nominal transition matrix A_N (i.e. as $A_\nu^{11} \to 0.8$). As expected, the asymptotically minimax robust CUSUM rule $T_C(\mu,\nu_L)$ has a shorter worst case detection delay than the nominal CUSUM rule $T_C(\mu,\nu_N)$.

In order to investigate the effect of the false alarm constraint on the worst detection delay ($A_\nu^{11} = 0.7$) of the asymptotically minimax robust, nominal and asymptotically optimal CUSUM rules, we plotted the Lorden delay against the log false alarm rate constraint $|\log\alpha|$ in Figure 4. We also simulated and plotted the delays of a GLR rule given by (11) with a window length of $w = 200$. As in our i.i.d. example, the delay performance of the GLR rule is poor (we believe this is partly because estimation of low probability transitions is difficult with a limited number of observations). The asymptotically minimax robust CUSUM rule therefore appears to offer improved worst case delay performance compared to the GLR and nominal CUSUM rules.

VI. Conclusion

In this paper, we have posed a minimax robust quickest change detection problem based on the popular Lorden

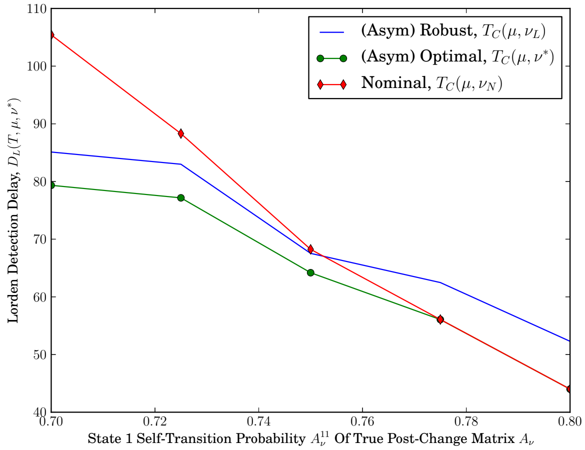

Fig. 3. Markov Chain Example: Detection delays $D_L(T,\mu,\nu^*)$ against true post-change transition probabilities $A_\nu^{11} = x$ and $A_\nu^{22} = 0.5$ for $x \in [0.7, 0.8]$ and fixed false alarm rate constraint $\alpha = 0.0005$. Maximum percentage standard error of 6.6% and 8.1% for the Lorden delay and false alarm rates, respectively.

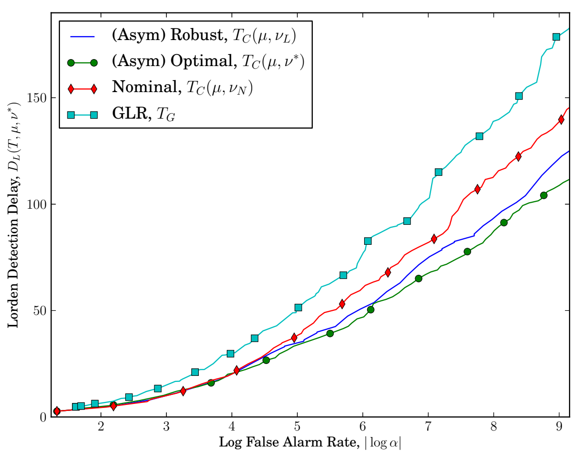

Fig. 4. Markov Chain Example: Detection delays $D_L(T,\mu,\nu^*)$ versus log false alarm rate constraint $|\log\alpha|$ for true post-change transition probabilities $A_\nu^{11} = 0.7$ and $A_\nu^{22} = 0.5$. Maximum percentage standard error of 6.6% and 8.1% for the Lorden delay and false alarm rates, respectively.

formulation of optimal quickest change detection. Under a condition on the set of possible post-change distributions, we established that CUSUM rules are asymptotic solutions to the Lorden minimax robust problem for a general class of dependent stochastic processes. Finally, we established bounds on the detection delay of misspecified CUSUM rules, and we demonstrated in two simulation examples that asymptotically minimax robust CUSUM rules can provide better detection delay performance than competing GLR procedures.

ACKNOWLEDGEMENT

Computational resources and services used in this paper were provided by the HPC and Research Support Group, Queensland University of Technology, Brisbane, Australia.

APPENDIX

Proof of Lemma 3.1

To prove the lemma result it suffices to show that for any $0 < \delta < 1$,

$$\begin{aligned} &\operatorname{ess\,sup} E_\lambda^{\mu,\nu}\left[\left(T_C(\mu,\bar\nu) - \lambda + 1\right)^+ \middle| \mathcal{F}_{\lambda-1}\right] \\ &\qquad \le (1+o(1)) \frac{h_C}{\mathcal{K}(\nu\|\mu) - \mathcal{K}(\nu\|\bar\nu)} \frac{1}{1-\delta} \end{aligned} \tag{15}$$

as $h_C \to \infty$ for all $\lambda \ge 1$.

Let k_c be the largest integer such that $0 < k_c < (1-\delta)^{-1}\left(\mathcal{K}(\nu\|\mu) - \mathcal{K}(\nu\|\bar\nu)\right)^{-1} h_C$, and let $A_j \in \mathcal{F}$ for $1 \le j \le t$ denote the event

$$Z_{\lambda+(j-1)k_c}^{\lambda+jk_c-1}(\mu,\bar\nu) < h_C.$$

Then from the CUSUM stopping rule (3), for any $\lambda \ge 1$, $t \ge 1$ and $h_C > 0$ we have that

$$\begin{aligned} &\operatorname{ess\,sup} P_\lambda^{\mu,\nu}\left(T_C(\mu,\bar\nu) - \lambda + 1 > tk_c \middle| \mathcal{F}_{\lambda-1}\right) \\ &\le \operatorname{ess\,sup} P_\lambda^{\mu,\nu}\left(A_j < h_C \text{ for all } 1 \le j \le t \middle| \mathcal{F}_{\lambda-1}\right) \\ &= \operatorname{ess\,sup} E_\lambda^{\mu,\nu}\left[I\{A_1\} \dots E_\lambda^{\mu,\nu}\left[I\{A_{t-1}\}\right.\right. \\ &\quad E_\lambda^{\mu,\nu}\left[I\{A_t\}\middle|\mathcal{F}_{\lambda+(t-1)k_c-1}\right]\left|\mathcal{F}_{\lambda+(t-2)k_c-1}\right] \dots \left|\mathcal{F}_{\lambda-1}\right] \end{aligned} \tag{16}$$

where that last line follows by the tower property of conditional expectations and since $I\{A_i\}$ is a measurable random variable with respect to $\mathcal{F}_{\lambda+(j-1)k_c-1}$ for all $i < j$. We now recall that Assumption 2 implies the convergence

$$\begin{aligned} \lim_{k\to\infty} \sup_{1\le\lambda\le t} \operatorname{ess\,sup} P_\lambda^{\mu,\nu}\bigg(\frac{1}{k} Z_t^{t+k-1}(\mu,\bar\nu) < \\ \left(\mathcal{K}(\nu\|\mu) - \mathcal{K}(\nu\|\bar\nu)\right)(1-\delta)\Big| \mathcal{F}_{t-1}\bigg) = 0 \end{aligned}$$

for any $0 < \delta < 1$. From the definition of k_c, it follows that for sufficiently large h_C,

$$\sup_{1\le\lambda\le t} \operatorname{ess\,sup} P_\lambda^{\mu,\nu}\left(Z_t^{t+k_c-1}(\mu,\bar\nu) < h_C \middle| \mathcal{F}_{t-1}\right) < \delta \tag{17}$$

for any $0 < \delta < 1$. Applying (17) separately to each of the nested conditional probabilities in (16) we have that for any $0 < \delta < 1$,

$$\operatorname{ess\,sup} P_\lambda^{\mu,\nu}\left(T_C(\mu,\bar\nu) - \lambda + 1 > tk_c \middle| \mathcal{F}_{\lambda-1}\right) \le \delta^t \tag{18}$$

for sufficiently large h_C, all $\lambda \ge 1$ and all $t \ge 1$.

Recalling our interest in the expectation associated with these bounded probabilities, for large h_C, and any $\lambda \ge 1$ we have that

$$\begin{aligned} &\operatorname{ess\,sup} E_\lambda^{\mu,\nu}\left[\left(T_C(\mu,\bar\nu) - \lambda + 1\right)^+ \middle| \mathcal{F}_{\lambda-1}\right] \\ &\quad = \operatorname{ess\,sup} \int_0^\infty P_\lambda^{\mu,\nu}\left(\left(T_C(\mu,\bar\nu) - \lambda + 1\right)^+ > y \middle| \mathcal{F}_{\lambda-1}\right) dy \\ &\quad \le k_c \sum_{t=0}^\infty \operatorname{ess\,sup} P_\lambda^{\mu,\nu}\left(\left(T_C(\mu,\bar\nu) - \lambda + 1\right)^+ > k_c t \middle| \mathcal{F}_{\lambda-1}\right) \\ &\quad \le k_c \sum_{t=0}^\infty \delta^t = k_c \frac{1}{1-\delta} \end{aligned}$$

for any $0 < \delta < 1$ where the first inequality is an over bound of the integral by the sum of rectangles, and the second inequality follows from (18) by noting that $P(X^+ > x) \leq P(X > x) I\{x > 0\} + I\{x = 0\}$. Since the definition of k_c implies that $k_c \sim (1-\delta)^{-1} \left(\mathcal{K}(\nu \| \mu) - \mathcal{K}(\nu \| \bar{\nu})\right)^{-1} h_C$, we have that (15) holds for any $0 < \delta < 1$ as $h_C \to \infty$ and the lemma result follows. □

Proof of Lemma 3.2

We first note that $\mathcal{K}(\nu, \mu) \leq \mathcal{K}(\nu, \bar{\nu})$ follows from (5) since taking the logarithm after the expectation of (5) gives

$$0 \geq \log E_1^{\mu,\nu} \left[\frac{p^{\bar{\nu}}\left(Y_k \middle| Y_{[1,k-1]}\right)}{p^{\mu}\left(Y_k \middle| Y_{[1,k-1]}\right)} \right] = \mathcal{K}(\nu, \mu) - \mathcal{K}(\nu, \bar{\nu})$$

by Jensen's inequality and (4).

We prove the second lemma result by noting that,

$$D_L\left(T_C(\mu, \bar{\nu}), \mu, \nu\right) \geq E_1^{\mu,\nu}\left[T_C(\mu, \bar{\nu})\right] - 1$$

since $P_\lambda^{\mu,\nu}\left(T_C(\mu, \bar{\nu}) \geq 1\right) = 1$. We now follow an argument similar to the proof of [6, Theorem 3] to bound the expectation $E_1^{\mu,\nu}\left[T_C(\mu, \bar{\nu})\right]$.

Let us define the sequence of zero-crossing stopping rules

$$\tau_{\ell+1} \triangleq \inf\left\{k \geq \tau_\ell + 1 : Z_{\tau_\ell+1}^k(\mu, \bar{\nu}) < 0\right\}$$

for $\ell \geq 0$ where $\tau_0 \triangleq 0$ and we follow the convention that $\inf \emptyset = \infty$ for the null set $\emptyset$. Let us now define the threshold crossing stopping rule

$$\begin{aligned} \bar{T}_h \triangleq \inf\{&\ell \geq 1 : \tau_\ell < \infty \text{ and} \\ &Z_{\tau_\ell+1}^k(\mu, \bar{\nu}) \geq h_C \text{ for some } k \geq \tau_\ell + 1\}. \end{aligned}$$

We now note that $\left\{\exp\left(Z_n^k(\mu, \bar{\nu})\right), \mathcal{F}_k, k \geq n\right\}$ is a nonnegative supermartingale under $P_1^{\mu,\nu}$ by (5). Then on events $\{\tau_\ell < \infty\}$, the maximal inequality for nonnegative supermartingales (e.g. see [21, pp. 55]) gives

$$\begin{aligned} &P_1^{\mu,\nu}\left(\max_{k > \tau_\ell} Z_{\tau_\ell+1}^k(\mu, \bar{\nu}) \geq h_C \middle| \mathcal{F}_{\tau_\ell}\right) \\ &\quad \leq e^{-h_C} E_1^{\mu,\nu}\left[\exp\left(Z_{\tau_\ell+1}^{\tau_\ell+1}(\mu, \bar{\nu})\right) | \mathcal{F}_{\tau_\ell}\right] \leq e^{-h_C} \end{aligned}$$

where we note that τ_ℓ is $\mathcal{F}_{\tau_\ell}-$measurable and (5) gives that $E_1^{\mu,\nu}\left[\exp\left(Z_{\tau_\ell+1}^{\tau_\ell+1}(\mu, \bar{\nu})\right) | \mathcal{F}_{\tau_\ell}\right] \leq 1$. Hence,

$$\begin{aligned} P_1^{\mu,\nu}\left(\bar{T}_h > \ell\right) &= E_1^{\mu,\nu}\left[P_1^{\mu,\nu}\left(\bar{T}_h \geq \ell + 1 \,|\mathcal{F}_{\tau_\ell}\right) I\{\bar{T}_h \geq \ell\}\right] \\ &\geq (1 - e^{-h_C}) P_1^{\mu,\nu}\left(\bar{T}_h > \ell - 1\right) \end{aligned}$$

and so

$$\begin{aligned} E_1^{\mu,\nu}\left[\bar{T}_h\right] &\geq \sum_{\ell=1}^{\infty} P_1^{\mu,\nu}\left(\bar{T}_h > \ell\right) \\ &\geq \sum_{\ell=1}^{\infty} (1 - e^{-h_C})^\ell = e^{h_C}. \end{aligned}$$

The lemma result follows by noting that

$$\begin{aligned} &T_C(\mu, \bar{\nu}) \\ &\quad = \inf\left\{k \geq 1 : Z_{\tau_\ell+1}^k(\mu, \bar{\nu}) \geq h_C \text{ for some } \tau_\ell < k\right\} \\ &\quad \geq \bar{T}_h. \end{aligned}$$

□

References

[1] T. Molloy and J. J. Ford, "HMM triangle relative entropy concepts in sequential change detection applied to vision-based dim target manoeuvre detection," in *Proc. 15th International Conference on Information Fusion*, Singapore, July 2012.

[2] N. Vaswani, "Additive change detection in nonlinear systems with unknown change parameters," *Signal Processing, IEEE Transactions on*, vol. 55, no. 3, pp. 859–872, 2007.

[3] J. Ru, V. Jilkov, X. Rong Li, and A. Bashi, "Detection of target maneuver onset," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 45, no. 2, pp. 536 –554, 2009.

[4] X. Rong Li and V. Jilkov, "A Survey of maneuvering target tracking—Part IV: Decision-based methods," in *Proc. 2002 SPIE Conf. on Signal and Data Processing of Small Targets*, vol. 4728. Orlando, FL., April 2002, pp. 511–534.

[5] S. Hansen and M. Blanke, "Diagnosis of airspeed measurement faults for unmanned aerial vehicles," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 50, no. 1, pp. 224–239, January 2014.

[6] T. L. Lai, "Information bounds and quick detection of parameter changes in stochastic systems," *Information Theory, IEEE Transactions on*, vol. 44, no. 7, pp. 2917 –2929, Nov. 1998.

[7] T. L. Lai and J. Shan, "Efficient recursive algorithms for detection of abrupt changes in signals and control systems," *Automatic Control, IEEE Transactions on*, vol. 44, no. 5, pp. 952 –966, May 1999.

[8] I. Nikiforov and M. Basseville, *Detection of abrupt changes: theory and application*. Englewood Cliffs, NJ: PTR Prentice Hall, 1993.

[9] G. Lorden, "Procedures for reacting to a change in distribution," *The Annals of Mathematical Statistics*, vol. 42, no. 6, pp. 1897–1908, 1971.

[10] H. Poor and O. Hadjiliadis, *Quickest detection*. Cambridge, U.K.: Cambridge University Press, 2009.

[11] G. V. Moustakides, "Optimal stopping times for detecting changes in distributions," *The Annals of Statistics*, vol. 14, no. 4, pp. 1379–1387, 1986.

[12] R. Crow and S. Schwartz, "On robust quickest detection procedures," in *Information Theory, 1994. Proceedings., 1994 IEEE International Symposium on*, 1994, pp. 258–.

[13] J. Unnikrishnan, V. Veeravalli, and S. Meyn, "Minimax robust quickest change detection," *Information Theory, IEEE Transactions on*, vol. 57, no. 3, pp. 1604 –1614, Mar. 2011.

[14] R. Bansal and P. Papantoni-Kazakos, "An algorithm for detecting a change in a stochastic process," *Information Theory, IEEE Transactions on*, vol. 32, no. 2, pp. 227 – 235, Mar. 1986.

[15] C. D. Fuh, "SPRT and CUSUM in Hidden Markov Models," *The Annals of Statistics*, vol. 31, no. 3, pp. 942–977, 2003.

[16] A. Shiryaev, *Probability*, ser. Graduate Texts in Mathematics. New York, NY: Springer-Verlag, 1996, vol. 95.

[17] I. Csiszar and F. Matúš, "Information projections revisited," *Information Theory, IEEE Transactions on*, vol. 49, no. 6, pp. 1474–1490, June 2003.

[18] R. B. Vinter, *Optimal control*. Boston, MA: Springer-Verlag, 2000.

[19] B. C. Levy, *Principles of signal detection and parameter estimation*. New York, NY: Springer-Verlag, 2008.

[20] L. Xie, V. Ugrinovskii, and I. Petersen, "Probabilistic distances between finite-state finite-alphabet hidden Markov models," *Automatic Control, IEEE Transactions on*, vol. 50, no. 4, pp. 505 – 511, 2005.

[21] D. Revuz and M. Yor, *Continuous martingales and Brownian motion*, ser. Comprehensive Studies in Mathematics. New York, NY: Springer-Verlag, 1991, vol. 293.