Fifty-third Annual Allerton Conference
Allerton House, UIUC, Illinois, USA
September 29 - October 2, 2015

Optimum Shewhart Tests for Markovian Data

George V. Moustakides
Computer Science, CBIM, Rutgers University, New Brunswick, NJ 08854-8019, USA
and Electrical and Computer Engineering, University of Patras, 26500 Rion, Greece
Email: gm463@rutgers.edu, moustaki@upatras.gr

***Abstract*—We consider the sequential change detection problem for Markovian processes. Adopting a Lorden-like criterion where expected delays are replaced with detection probabilities we end up with a well defined constrained optimization problem which is possible to solve, exactly, in the Markovian case. The optimum scheme turns out to be a modified version of the Shewhart test (known to be optimum in the i.i.d. case) that involves two unknown functions. These functions can be identified by solving a system of two equation which is the result of applying optimal stopping theory and the need for the optimum scheme to be an equalizer rule.**

I. Introduction

Suppose $\{\xi_t\}_{t\geq 0}$ is a discrete-time observation sequence that becomes available sequentially. Define $\{\mathscr{F}_t\}_{t\geq 0}$ to be the corresponding filtration with $\mathscr{F}_t = \sigma\{\xi_0,\ldots,\xi_t\}$ the σ-algebra generated by the observations up to time t. Let $\tau \in \{0,1,\ldots\}$ denote a *changetime* and assume that the observations follow the probability measure P_∞ up to and including τ, while after τ the probability measure switches to P_0. If the change in statistics takes place at $\tau = t$ then this induces a probability measure which is denoted with P_t while we use $\mathsf{E}_t[\cdot]$ for the corresponding expectation.

We are interested in detecting the occurrence of the changetime τ using a sequential strategy, in particular, a stopping time T adapted to the filtration $\{\mathscr{F}_t\}$. T will signal the change as soon as possible avoiding, at the same time, making frequent false alarms. The detection power of the stopping time T is usually measured through the average detection delay. In [1] the generic measure $\mathcal{J}(T) = \mathsf{E}_\tau[T-\tau|T>\tau]$ is proposed which, depending on the prior knowledge we have and the model we adopt for the changetime τ, it is possible to recover all existing performance criteria encountered in the literature as: the measure proposed by Shiryaev [2]; by Pollak [3] and, finally, the criterion proposed by Lorden [4] which is the focus of our work:

$$\mathcal{J}_{\rm L}(T) = \sup_{t\geq 0} \operatorname{ess\,sup} \mathsf{E}_t[T-t|\mathscr{F}_t, T>t]. \qquad (1)$$

As we can see Lorden's measure considers the expected detection delay conditioned on the worst possible data before the change (expressed through the "ess sup") and the worst possible deterministic changetime (expressed through "$\sup_{t\geq 0}$"). In [1] it is shown that Lorden's criterion can be recovered from the generic measure $\mathcal{J}(T)$ by considering τ to be a *stopping time* (the last time instant under the nominal regime) adapted to $\{\mathscr{F}_t\}$ and, then, following a worst-case analysis with respect to τ.

A. Criteria Based on Detection Probability

From (1) it is clear that T can take upon any integer value, consequently, this quantity can become arbitrarily large. There are important applications (see [5] and references therein) where unbounded delays are meaningless (since they imply that a catastrophic event may already have occurred) and one would rather prefer to detect the change within a prespecified time window. In other words we would like to have $\tau < T \leq \tau + m$, for given $m \geq 1$. Stopping within the prescribed interval constitutes a desirable event while $T > \tau + m$ is *not* considered as successful detection.

Similarly to (1) we can now propose the following alternative Lorden-like criterion (for Shiryaev- and Pollak-like measures see [6], [7]):

$$\mathscr{J}_{\rm L}(T) = \inf_{t\geq 0} \operatorname{ess\,inf} \mathsf{P}_t(t<T\leq t+m|\mathscr{F}_t, T>t). \qquad (2)$$

Instead of the average detection delay, we focus on the worst-case detection probability. Clearly, the goal here will be to maximize the detection probability subject to a suitable false alarm constraint.

The probability maximizing idea was first adopted by Bojdecki [8] with the optimum stopping time T resulting from the maximization of the probability $\mathsf{P}(|\tau+1-T| \leq m)$. The complete solution to this problem was offered for the i.i.d. case, $m = 0$ and for the Bayesian formulation with the changetime τ following a geometric prior. The optimum stopping time turned out to be the Shewhart test introduced in [9]. We should mention that $m = 0$ corresponds to the maximization of the probability of the event $\{T = \tau+1\}$, namely that detection is achieved by using just *the first observation under the alternative regime*. A point we need to make is that Bojdecki, in his approach, did not attempt to control false alarms in any sense. Following a similar idea, in [10] this analysis was extended to cover dependent observations, but no complete solution was offered.

Regarding the Shewhart test, one can find numerous optimality properties in [6], [7] that follow the detection probability maximization idea applied to Pollak- and Lorden-like measures combined with suitable false alarm constraints. In the current work we intend to advance the analysis of [6] to dependent data. More specifically the goal is to discover the exact form of the optimum Shewhart test for Markovian observations.

B. Problem Formulation

Let us present the precise formulation of the problem of interest by explicitly defining our performance measure and the corresponding optimization problem that leads to the optimum detector. Consider (2) with $m=1$, namely

$$\mathscr{J}_{\mathrm{L}}(T)=\inf_{t\geq 0}\operatorname{ess\,inf}\mathsf{P}_t(T=t+1|\mathscr{F}_t,T>t), \quad (3)$$

and the following max-min constrained optimization problem

$$\begin{aligned}\sup_T \mathscr{J}_{\mathrm{L}}(T) = &\sup_T \inf_{t\geq 0}\operatorname{ess\,inf}\mathsf{P}_t(T=t+1|\mathscr{F}_t,T>t)\\ &\text{subject to: } \mathsf{E}_\infty[T]\geq\gamma>1.\end{aligned} \quad (4)$$

As we can see the goal is to maximize the worst-case conditional probability of detecting the change using only the first sample under the alternative regime, while preventing the average false alarm period of becoming smaller than some prescribed value $\gamma>1$.

Solving the optimization problem depicted in (4) for the i.i.d. case (before and after the change), it is shown in [6] that this leads to the Shewhart test defined as follows

$$\mathcal{S}=\inf\{t>0:\ \mathsf{L}(\xi_t)\geq\nu\}, \quad (5)$$

where $\mathsf{L}(\xi_t)=\mathsf{f}_0(\xi_t)/\mathsf{f}_\infty(\xi_t)$ denotes the likelihood ratio of the current sample ξ_t with $\mathsf{f}_\infty(\xi_t),\mathsf{f}_0(\xi_t)$ its pre- and post-change pdf respectively. Parameter $\nu>0$ is a constant threshold selected to satisfy the false alarm constraint with equality. Actually in [6] this optimality result is presented under the more general setting of independent and non-identically distributed observations, with the threshold of the test required to be a time varying sequence $\{\nu_t\}$ which is explicitly specified.

II. Markovian Observations

In this section we consider observations $\{\xi_t\}$ that are Markov with the corresponding pre- and post-change conditional pdfs denoted as $\mathsf{f}_\infty(\xi_t|\xi_{t-1})$ and $\mathsf{f}_0(\xi_t|\xi_{t-1})$. We will also assume that at the time of change τ we have a change in the *conditional* pdf. In other words the pre-change data influence the post-change observations. We should note that it is possible to experience different forms of changepoint mechanisms where, for example, post-change data are independent from their pre-change counterparts (see [1]). This case requires completely different analysis from the one we present in this work.

There is an additional detail that necessitates special attention. In order to define our optimum stopping time using a stationary rule we need to assume that we also observe ξ_0. Due to our assumption that $\tau\geq 0$ this implies that ξ_0 is necessarily under the nominal regime. Consequently this sample cannot contribute to the detection of the change when considered solely by itself. The previous assumption can be easily relaxed to cover the normal situation where observations start from ξ_1 with the first sample having two different pdfs under the two measures. This, however, leads to a stopping time with a stopping rule at time $t=1$ that differs from the rule applied for $t>1$, thus complicating the analysis. For this reason, and in order to facilitate the presentation of our results, we decided to reserve the complete analysis for a future (journal) version of this work which will be more detailed.

A. Candidate Shewhart Test

For $t\geq 1$ define the conditional likelihood ratio

$$\mathsf{L}(\xi_t,\xi_{t-1})=\frac{\mathsf{f}_0(\xi_t|\xi_{t-1})}{\mathsf{f}_\infty(\xi_t|\xi_{t-1})}. \quad (6)$$

The naive extension of the Shewhart test to the Markov case would be the definition of the corresponding stopping time similarly to (5) but with $\mathsf{L}(\xi_t)$ replaced by its conditional version defined in (6). Unfortunately, the resulting stopping time lacks exact optimality properties and even if one can show asymptotic optimality its actual performance will depend on observation ξ_0 making such optimality *nonuniform*. Instead, we propose an alternative version which we detail next and prove its *exact optimality*, provided certain functions entering in its definition are properly selected.

Let $c(\xi),\nu(\xi)$ be two functions with $c(\xi)\geq 0$ and $\nu(\xi)\geq 1$, and define the following version of the Shewhart stopping time

$$\mathcal{S}=\inf\{t>0:\ c(\xi_{t-1})\mathsf{L}(\xi_t,\xi_{t-1})\geq\nu(\xi_t)\}. \quad (7)$$

Actually this stopping time is equivalent to the original one depicted in (5) but with the threshold being a *separable* function of the current and the previous sample.

In order for (7) to correspond to a practically applicable detection scheme, the two functions $c(\xi),\nu(\xi)$ need to be explicitly specified. For this, of course, we need suitable equations. By recalling the analysis for the independent but non-identically distributed case from [6] we observe that it is necessary for the optimum test to be an *equilizer rule* over all changetime values. This means that $\operatorname{ess\,inf}\mathsf{P}_t(\mathcal{S}=t+1|\mathscr{F}_t,\mathcal{S}>t)$ must be constant over all $t\geq 0$. This requirement translates into the following equation for the stopping time defined in (7)

$$\mathsf{P}_0\big(c(\xi_0)\mathsf{L}(\xi_1,\xi_0)\geq\nu(\xi_1)|\xi_0\big)=\beta,\ \forall\xi_0, \quad (8)$$

where $\beta\in(0,1)$ some constant. Due to stationarity (8) suggests that $\mathsf{P}_t(\mathcal{S}=t+1|\mathscr{F}_t,\mathcal{S}>t)=\beta$ for every $t\geq 0$ and ξ_t. Consequently

$$\mathscr{J}_{\mathrm{L}}(\mathcal{S})=\beta. \quad (9)$$

We note that (8) defines a mapping between $\nu(\xi)$ and $c(\xi)$. Indeed, if we fix the detection probability β, and select a function $\nu(\xi)\geq 1$ then, using (8) we can specify $c(\xi)$ by solving this equation for $c(\xi_0)$ for each value of ξ_0.

We need one more equation to identify the two unknown functions. Following similar methodology as in the case of the usual Lorden measure, we recall from [6] that we upper bound $\mathscr{J}_{\mathrm{L}}(T)$ with a simpler expression (see Lemma 2) for which we can easily apply Optimal Stopping. In order for the test in (7) to solve the corresponding optimal stopping problem defined for the upper bound it is necessary the two functions to satisfy the following integral equation:

$$\nu(\xi_0)=1+\mathsf{E}_\infty[\nu(\xi_1)\mathbb{1}_{\{c(\xi_0)\mathsf{L}(\xi_1,\xi_0)<\nu(\xi_1)\}}|\xi_0]. \quad (10)$$

Actually from (10) we can immediately deduce that

$$\nu(\xi_0) = \mathsf{E}_\infty[\mathcal{S}|\xi_0]. \tag{11}$$

Combining (8) and (10) defines a pair of equations with unknowns the functions $c(\xi), \nu(\xi)$. The following theorem states that a solution always exists.

Theorem 1. *Fix $\beta \in (0,1)$ and define the sequence of functions $\{\nu_n(\xi)\}, \{c_n(\xi)\}$ by applying the recursion*

$$c_n(\xi) = \arg\big\{c(\xi) : \mathsf{P}_0\big(c(\xi)\mathsf{L}(\xi_1,\xi) \geq \nu_n(\xi_1)|\xi_0 = \xi\big) = \beta\big\}, \tag{12}$$

$$\nu_{n+1}(\xi) = 1 + \mathsf{E}_\infty[\nu_n(\xi_1)\mathbb{1}_{\{c_n(\xi)\mathsf{L}(\xi_1,\xi)<\nu_n(\xi_1)\}}|\xi_0 = \xi], \tag{13}$$

which is initialized with $\nu_0(\xi) = 1$. Then the sequences $\{\nu_n(\xi)\}, \{c_n(\xi)\}$ converge to functions $\nu(\xi), c(\xi)$ respectively that satisfy (8),(10).

Proof: The main steps of the proof are highlighted in the Appendix. ■

As we can see the iterative method during the $(n+1)$st iteration has available $\nu_n(\xi)$ and computes the function $c_n(\xi)$ from (12) to make the Shewhart test an equalizer. Then uses $c_n(\xi), \nu_n(\xi)$ to update $\nu_{n+1}(\xi)$. Theorem 1 essentially claims that, for fixed detection probability β, there exists a common solution $\nu(\xi), c(\xi)$ to the two equations (8),(10). Even though not clearly apparent, both functions $\nu(\xi), c(\xi)$ depend on the value of the parameter β. In order to complete the definition of our candidate stopping time we need to specify this last quantity as well. If $g(\xi)$ is the pdf of ξ_0 under the P_∞ measure and we force our stopping time to satisfy the false alarm constraint with equality then from (11) we obtain the following equation relating the detection probability β to the average false alarm period γ

$$\gamma = \mathsf{E}_\infty[\mathcal{S}] = \mathsf{E}_\infty\big[\mathsf{E}_\infty[\mathcal{S}|\xi_0]\big] = \int \nu(\xi)g(\xi)\,d\xi, \tag{14}$$

where we recall from Theorem 1 that $\nu(\xi)$ is a function that depends on β. Regarding the solution of (14), we have the following lemma.

Lemma 1. *The average false alarm period $\mathsf{E}_\infty[\mathcal{S}]$ in* (14) *is decreasing and continuous in $\beta \in (0,1)$ with $\lim_{\beta\to 1}\mathsf{E}_\infty[\mathcal{S}] = 1$ and $\lim_{\beta\to 0}\mathsf{E}_\infty[\mathcal{S}] = \infty$. This suggest that for any $\gamma > 1$ there exists $\beta \in (0,1)$ so that* (14) *is satisfied.*

Proof: A sketch of the proof can be found in the Appendix. ■

With Theorem 1 and Lemma 1 we have fully specified the candidate stopping time $\mathcal{S}$ since we have completely identified the two functions $c(\xi), \nu(\xi)$ and the detection probability β corresponding to the average false alarm period γ.

B. Max-Min Optimality

Our task in this section is to show that the candidate stopping time $\mathcal{S}$ defined in (7) with $c(\xi), \nu(\xi), \beta$ obtained through (8),(10),(14), is the one solving the max-min constrained optimization problem defined in (4). In order to be able to prove this fact we need to find a suitable upper bound for $\mathscr{J}_{\mathrm{L}}(T)$. The following lemma provides the necessary expression.

Lemma 2. *Let $c(\xi), \nu(\xi)$ be two functions related through* (8) *and $\mathcal{S}$ the corresponding Shewhart test. Then for any stopping time T with $\mathsf{E}_\infty[T] < \infty$ we can write*

$$\mathscr{J}_{\mathrm{L}}(T) \leq \frac{\mathsf{E}_\infty[c(\xi_{T-1})\mathsf{L}(\xi_T,\xi_{T-1})]}{\mathsf{E}_\infty\big[\sum_{t=0}^{T-1} c(\xi_t)\big]}, \tag{15}$$

with equality when $T = \mathcal{S}$.

Proof: A sketch of the main steps are presented in the Appendix. ■

The next theorem and its corollary establish the optimality of our candidate test.

Theorem 2. *Let $c(\xi), \nu(\xi), \beta$ be the two function and the detection probability corresponding to γ through the system of equations* (8),(10),(14) *and T any stopping time that satisfies the false alarm constraint $\mathsf{E}_\infty[T] \geq \gamma$. Then for each such T we have*

$$\frac{\mathsf{E}_\infty[c(\xi_{T-1})\mathsf{L}(\xi_T,\xi_{T-1})]}{\mathsf{E}_\infty\big[\sum_{t=0}^{T-1} c(\xi_t)\big]} \leq \beta,$$

with equality when $T = \mathcal{S}$.

Proof: The main steps of the proof are highlighted in the Appendix. ■

Corollary: *Let $c(\xi), \nu(\xi), \beta$ be as in the previous theorem and $\mathcal{S}$ the corresponding version of the Shewhart test depicted in* (7)*. Then $\mathcal{S}$ solves the max-min problem defined in* (4).

Proof: Combining Lemma 2, Theorem 2 and (9), the optimality of the proposed version of the Shewhart test is straightforward. Indeed we can immediately deduce that

$$\beta = \mathscr{J}_{\mathrm{L}}(\mathcal{S}) \leq \sup_T \mathscr{J}_{\mathrm{L}}(T) \leq \beta$$

with the supremum taken over all stopping times satisfying the false alarm constraint. This suggests that $\sup_T \mathscr{J}_{\mathrm{L}}(T) = \beta = \mathscr{J}_{\mathrm{L}}(\mathcal{S})$ and therefore establishes optimality of $\mathcal{S}$. ■

Remark. It is interesting to note that the Shewhart test, defined in (5) for i.i.d. observations, enjoys several optimality properties. In particular, as we mentioned, it is optimum under the Lorden-like measure defined in (3), but also under Pollak- and Shiryaev-like alternatives (see [6], [7]). When however we consider Markovian observations this general optimality result is no longer valid. The optimum test we offer in the present analysis turns out to be optimum *only* under the Lorden-like measure and *not* under the other alternative performance criteria.

III. NUMERICAL EXAMPLES

Undoubtly AR(1) can be regarded as one of the most popular Markovian models. For this reason it will be adopted for our numerical examples. Specifically we consider a nonlinear, conditionally Gaussian AR(1) process $\{\xi_t\}$ where the pre-change model is $\xi_t = w_t$ while the post-change satisfies $\xi_t = \alpha(\xi_{t-1}) + w_t$ with $\{w_t\}$ an i.i.d. zero-mean Gaussian noise sequence of unit variance and $\alpha(\xi)$ a scalar nonlinearity.

For the likelihood ratio $\mathsf{L}(\xi_t, \xi_{t-1})$ we have $\mathsf{L}(\xi_t, \xi_{t-1}) = \exp\big(-0.5\alpha^2(\xi_{t-1}) + \alpha(\xi_{t-1})\xi_t\big)$. Furthermore if we call $\tilde{c}(\xi) = \log c(\xi) - 0.5\alpha^2(\xi)$, $\tilde{\nu}(\xi) = \log \nu(\xi)$ then the proposed Shewhart test is equivalent to

$$\mathcal{S} = \inf\big\{t > 0 : \tilde{c}(\xi_{t-1}) + \alpha(\xi_{t-1})\xi_t \geq \tilde{\nu}(\xi_t)\big\}$$

while the naive version of the Shewhart test takes the form

$$\mathcal{T} = \inf\big\{t > 0 : -0.5\alpha^2(\xi_{t-1}) + \alpha(\xi_{t-1})\xi_t \geq \tilde{\nu}\big\}$$

where $\tilde{\nu} = \log \nu$ is a constant threshold. Comparing the two detection schemes, we realize that the optimum test, replaces $-0.5\alpha^2(\xi)$ with the function $\tilde{c}(\xi)$ and instead of the constant $\tilde{\nu}$ uses the function $\tilde{\nu}(\xi)$.

Let us now examine the form of the optimum functions $\tilde{c}(\xi), \tilde{\nu}(\xi)$ when we consider a linear AR(1) model with $\alpha(\xi) = 0.5\xi$ and an average false alarm period equal to $\gamma = 100$. Applying the iterations proposed in (12) and (13) and using numerical techniques similar to the ones developed in [11], [12], we can compute the functions $\tilde{c}(\xi), \tilde{\nu}(\xi)$. Their form appears in Fig. 1(a) and (b) (in red) respectively. In the same figures we can see the corresponding functions for the naive implementation of the Shewhart rule (in blue) namely $-0.5\alpha^2(\xi) = -0.125\xi^2$ and $\tilde{\nu} = 1.1$. The threshold $\tilde{\nu}$ was selected so that the naive test satisfies the false alarm constraint with equality.

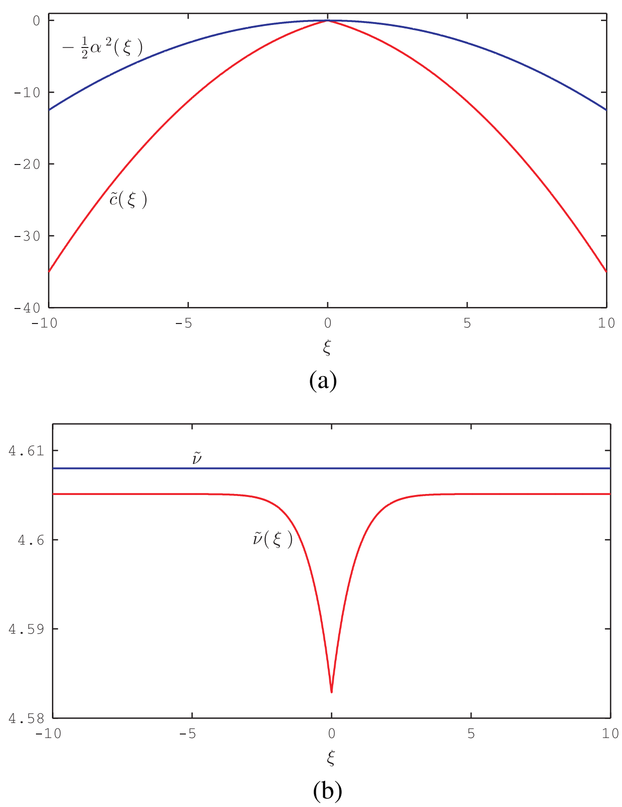

Fig. 1. For $\alpha(\xi) = 0.5\xi$. Optimum Shewhart test in red: function $\tilde{c}(\xi)$ in (a), and $\tilde{\nu}(\xi)$ in (b). Naive Shewhart test in blue: function $-\alpha^2(\xi)/2$ in (a), and threshold $\tilde{\nu}$ in (b).

From the two figures we observe that $\tilde{c}(\xi)$ diverges significantly from $-0.5\alpha^2(\xi)$. More importantly, this divergence translates into a substantial performance difference between the two tests. Specifically the (worst-case) detection probability for the optimum is $\beta = 0.022$ while in the naive it becomes 0! Indeed for the latter we observe that

$$\begin{aligned}\operatorname{ess\,inf} \mathsf{P}_t(\mathcal{T} = t+1 | \mathscr{F}_t, \mathcal{T} > t)\\ = \inf_{\xi_{t-1}} \mathsf{P}_t\left(-0.125\xi_{t-1}^2 + 0.5\xi_{t-1}\xi_t > \tilde{\nu} | \xi_{t-1}\right)\end{aligned}$$

which is equal to 0 when $\xi_{t-1} = 0$ and $\tilde{\nu} > 0$. In other words when the observation right before the change is 0, then it is impossible to detect the change just with the first sample under the alternative regime. On the other hand, with the optimum detector we have a guaranteed performance which is independent from the value of the sample before the change.

Aknowledgement

This work was supported by the US National Science Foundation under Grant CIF1513373.

Appendix

Proof of Theorem 1. By applying the recursion, in order to show existence of the limiting function $\nu(\xi)$ we observe first that $\nu_1(\xi) \geq 1 = \nu_0(\xi)$. Using induction we can then show that $\nu_n(\xi) \geq \nu_{n-1}(\xi)$. Consequently for each fixed ξ $\{\nu_n(\xi)\}$ is a monotonically increasing sequence of n. Again, using induction we can show that $\nu_n(\xi) \leq b(\xi)$, where $b(\xi)$ properly selected function independent from n. Since $\{\nu_n(\xi)\}$ is monotone and bounded, for each ξ, there is a limit $\nu(\xi)$. This also implies that $\{c_n(\xi)\}$ converges to some $c(\xi)$ which satisfies (8) and the pair $c(\xi), \nu(\xi)$ satisfies (10). ■

Proof of Lemma 1. Let us make explicit the dependence of the functions $\nu_n(\xi)$ on β by denoting them as $\nu_n(\xi, \beta)$. We can show by induction that each function $\nu_n(\xi, \beta)$, for each fixed ξ is decreasing with respect to β. Consequently, the same property passes also to the limiting function $\nu(\xi, \beta)$. Furthermore we can show, again by induction, that $\nu_n(\xi, \beta) - \nu_n(\xi, \beta + \epsilon) \leq D(\beta)\epsilon$ where $D(\beta)$ properly selected constant (dependent on β) and $\epsilon > 0$ sufficiently small quantity. This observation also implies that $\mathsf{E}_\infty[\nu(\xi_0, \beta)]$ is continuous in β. As $\beta \to 0$ we can see that the corresponding functions $\nu_n(\xi, \beta) \to \infty$ while as $\beta \to 1$ we have $\nu_n(\xi, \beta) \to 1$. These properties pass also to the limit $\nu(\xi_0, \beta)$ and, therefore, to the expectation $\mathsf{E}_\infty[\nu(\xi_0, \beta)]$. Since the latter is continuous in β taking values in the interval $[1, \infty)$ we conclude that there exists a β_* for which $\mathsf{E}_\infty[\nu(\xi_0, \beta_*)] = \gamma$. ■

Proof of Lemma 2. From its definition $\mathscr{J}_{\rm L}(T)$, for every $t \geq 0$, satisfies

$$\begin{aligned}\mathscr{J}_{\rm L}(T) &\leq \mathsf{P}_t(T = t+1 | \mathscr{F}_t, T > t)\\ &= \mathsf{E}_\infty[\mathsf{L}(\xi_{t+1}, \xi_t)\mathbb{1}_{\{T=t+1\}} | \mathscr{F}_t, T > t]\end{aligned}$$

where we have used the fact that $\{T = t+1\}$ is $\mathscr{F}_{t+1}$-measurable and applied a change of measures. Consider $c(\xi) \geq 0$, multiplying both sides with $c(\xi_t)\mathbb{1}_{\{T>t\}}$ which is nonnegative; recalling that $\{T > t\}$ is $\mathscr{F}_t$-measurable; taking expectation with respect to P_∞; and summing over all $t \geq 0$ we obtain

$$\mathscr{J}_{\mathrm{L}}(T)\sum_{t=0}^{\infty}\mathsf{E}_\infty[c(\xi_t)\mathbb{1}_{\{T>t\}}] = \mathscr{J}_{\mathrm{L}}(T)\mathsf{E}_\infty\left[\sum_{t=0}^{T-1}c(\xi_t)\right]$$
$$\leq \sum_{t=0}^{\infty}\mathsf{E}_\infty[c(\xi_t)\mathsf{L}(\xi_{t+1},\xi_t)\mathbb{1}_{\{T=t+1\}}\mathbb{1}_{\{T>t\}}]$$
$$= \sum_{t=0}^{\infty}\mathsf{E}_\infty[c(\xi_t)\mathsf{L}(\xi_{t+1},\xi_t)\mathbb{1}_{\{T=t+1\}}]$$
$$= \mathsf{E}_\infty[c(\xi_{T-1})\mathsf{L}(\xi_T,\xi_{T-1})],$$

which validates the correctness of the upper bound. When $c(\xi)$ is selected as the solution of (8) we have equality when $T = \mathcal{S}$ because $\mathcal{S}$ is an equalizer and $\mathscr{J}_{\mathrm{L}}(\mathcal{S}) = \beta = \mathsf{P}_t(\mathcal{S} = t+1|\mathscr{F}_t, \mathcal{S} > t)$ for all $t \geq 0$. ■

Proof of Theorem 2. To show the desired inequality we use similar arguments as in [6] and show that it is sufficient to limit ourselves to stopping times T that satisfy the false alarm constraint with equality. Note that for the upper bound it is sufficient to demonstrate

$$\mathsf{E}_\infty\left[c(\xi_{T-1})\mathsf{L}(\xi_T,\xi_{T-1}) - \sum_{t=0}^{T-1}\beta c(\xi_t)\right] \leq 0,$$

which, if we add the false alarm constraint in the form of equality, it becomes equivalent to

$$\mathsf{E}_\infty\left[c(\xi_{T-1})\mathsf{L}(\xi_T,\xi_{T-1}) + \sum_{t=0}^{T-1}\big(1-\beta c(\xi_t)\big)\right] \leq \gamma.$$

To prove validity of the previous inequality we can now assume that T is unconstrained. Applying optimal stopping and using ideas from [6] we can actually show that the left hand side is maximized when $T = \mathcal{S}$ which suggests that

$$\mathsf{E}_\infty\left[c(\xi_{T-1})\mathsf{L}(\xi_T,\xi_{T-1}) + \sum_{t=0}^{T-1}\big(1-\beta c(\xi_t)\big)\right]$$
$$\leq \mathsf{E}_\infty\left[c(\xi_{\mathcal{S}-1})\mathsf{L}(\xi_{\mathcal{S}},\xi_{\mathcal{S}-1}) + \sum_{t=0}^{\mathcal{S}-1}\big(1-\beta c(\xi_t)\big)\right]$$
$$= \mathsf{E}_\infty\left[c(\xi_{\mathcal{S}-1})\mathsf{L}(\xi_{\mathcal{S}},\xi_{\mathcal{S}-1}) - \beta\sum_{t=0}^{\mathcal{S}-1}c(\xi_t)\right] + \mathsf{E}_\infty[\mathcal{S}]$$
$$= \mathsf{E}_\infty[\mathcal{S}] = \gamma.$$

For the second last equality we used from Lemma 2 the fact that $\mathcal{S}$ attains the upper bound and finally that $\mathcal{S}$ was designed to satisfy the false alarm constraint with equality. ■

References

[1] G. V. Moustakides, "Sequential Change Detection Revisited", *Annals of Statistics*, vol. 36, pp. 787–807, 2008.

[2] A. N. Shiryaev, "On Optimum Methods in Quickest Detection Problems", *Theory of Probability and its Applications*, vol. 8, pp. 22–46, 1963.

[3] M. Pollak, "Optimal Detection of a Change in Distribution", *Annals of Statistics*, vol. 13, pp. 206–227, 1985.

[4] G. Lorden, "Procedures for Reacting to a Change in Distribution", *Annals of Mathematical Statistics*, vol. 42, pp. 1897–1908, 1971.

[5] B. K. Guépié, L. Fillatre and I. Nikiforov, "Sequential Detection of Transient Changes", *Sequential Analysis*, vol. 31, pp. 528–547, 2012.

[6] G. V. Moustakides, "Multiple Optimality Properties of the Shewhart Test", *Sequential Analysis*, vol. 33, pp. 318–344, 2014.

[7] M. Pollak and A. M. Krieger, "Shewhart Revisited", *Sequential Analysis*, vol. 32, pp. 230–242, 2013.

[8] T. Bojdecki, "Probability Maximizing Approach to Optimal Stopping and its Application to a Disorder Problem", *Stochastics*, vol. 3, pp. 61–71, 1979.

[9] W. A. Shewhart, *Economic Control of Quality of Manufactured Product* New York: D. Van Nostrand Company, 1931.

[10] W. Sarnowski and K. Szajowski, "Optimal Detection of Transition Probability Change in Random Sequence", *Stochastics*, vol. 86, pp. 569–581, 2011.

[11] G. V. Moustakides, A. S. Polunchenko and A. G. Tartakovsky, "A Numerical Approach to Performance Analysis of Quickest Changepoint Detection Procedures", *Statistica Sinica*, vol. 21, No. 2, pp. 571–596, 2011.

[12] A. S. Polunchenko, G. Sokolov and W. Du, "An Accurate Method for Determining the Pre-Change Run-Length Distribution of the Generalized Shiryaev-Roberts Detection Procedure", *Sequential Analysis*, vol. 33, No. 1, pp. 112–134, 2014.