Taylor & Francis
Taylor & Francis Group

Multiple Optimality Properties of the Shewhart Test

George V. Moustakides
Department of Electrical and Computer Engineering, University of Patras,
Patras-Rion, Greece

Abstract: For the problem of sequential detection of changes, we adopt the probability maximizing approach in place of the classical minimization of the average detection delay and propose modified versions of the Shiryaev, Lorden, and Pollak performance measures. For these alternative formulations, we demonstrate that the optimum sequential detection scheme is the simple Shewhart rule. Interestingly, we can also solve problems that under the classical setup have been open for many years, as optimum change detection with time-varying observations or with multiple postchange probability measures. For the latter, we also offer the exact solution for Lorden's original setup involving average detection delays, for the case where the average false alarm period is within certain limits.

1. INTRODUCTION

Suppose $\{\xi_t\}_{t>0}$ is a discrete-time process that becomes available sequentially and define $\{\mathscr{F}_t\}_{t\geq0}$ to be the associated filtration with $\mathscr{F}_t = \sigma\{\xi_1, \ldots, \xi_t\}$ the σ-algebra generated by the observations up to time t. Let $\tau \in \{\ldots, -1, 0, 1, \ldots\}$ denote a *change time* 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 that we denote with P_t while $\mathsf{E}_t[\cdot]$ is reserved for the corresponding expectation. We would like to stress that, here, τ denotes the *last* time instant under the nominal regime and *not* the first under the alternative, which is the usual practice. This slight difference allows to view τ as a *stopping time* (the time the observations *stop* following the nominal statistics), a property that can be analytically very appealing (see Moustakides, 2008).

We are interested in detecting the occurrence of the change time τ with the help of a stopping time T adapted to the filtration $\{\mathscr{F}_t\}$ that will signal the change as soon as possible avoiding, at the same time, making frequent false alarms. The effectiveness of a detection scheme is commonly quantified through the average detection delay. There are, of course, various possibilities depending on the prior knowledge we have and the model we adopt for the change time. In particular, assuming τ to be random, independent from the observations, with a known prior, Shiryaev (1963) proposed the following measure:

$$\mathscr{J}_S(T) = \mathsf{E}[T - \tau | T > \tau]. \tag{1.1}$$

If we consider $\tau = t$ to be deterministic and unknown, we can then follow a worst-case analysis and consider the performance measure proposed by Lorden (1971)

$$\mathscr{J}_L(T) = \sup_{t \geq 0} \operatorname{ess\,sup} \mathsf{E}_t[T - t | \mathscr{F}_t, T > t]. \tag{1.2}$$

Finally, assuming again that $\tau = t$ is deterministic and unknown we can alternatively define

$$\mathscr{J}_P(T) = \sup_{t \geq 0} \mathsf{E}_t[T - t | T > t], \tag{1.3}$$

which is the criterion introduced by Pollak (1985). The three measures depicted in (1.1), (1.2), (1.3) are the most common criteria encountered in the literature and, as noted in Moustakides (2008), they can be recovered from a general definition that treats τ as a stopping time. An optimum stopping rule T is then specified by minimizing these performance measures subject to suitable false alarm constraints.

1.1. Criteria Based on Detection Probability

We observe from (1.1), (1.2), (1.3) that no hard limit is imposed on the detection delay; consequently, this quantity can become arbitrarily large. As reported in Guépié et al. (2012) and references therein there are several applications in practice where unbounded delays can be undesirable and one would rather detect the change within a prespecified time window, *after* the change has occurred.[1] In other words, we like to have $\tau < T \leq \tau + m$, for given $m \geq 1$. Stopping within the prescribed interval constitutes a desirable event while if $T > \tau + m$ this is *not* considered as successful detection.

Similarly to (1.1), (1.2), (1.3), we can now propose the following alternatives of the three classical performance measures:

$$\mathscr{J}_S(T) = \mathsf{P}(\tau < T \leq \tau + m | T > \tau) \tag{1.4}$$

$$\mathscr{J}_L(T) = \inf_{t \geq 0} \operatorname{ess\,inf} \mathsf{P}_t(t < T \leq t + m | \mathscr{F}_t, T > t) \tag{1.5}$$

$$\mathscr{J}_P(T) = \inf_{t \geq 0} \mathsf{P}_t(t < T \leq t + m | T > t). \tag{1.6}$$

---

[1]According to our definition, stopping before or at τ corresponds to false alarm.

As we can see, instead of focusing on the average detection delay, we now pay attention to the detection probability. Consequently, here, we need to replace the minimization of the worst-case average detection delay of the classical approach with the *maximization of the worst-case detection probability*.

Bojdecki (1979) was the first to adopt this probability maximizing idea by considering the maximization of the probability $P(|\tau + 1 - T| \leq M)$. The complete solution to this problem was offered for the case $M = 0$ and for the Bayesian formulation with the change time τ following a geometric prior. The optimum stopping time turned out to be the simple test introduced by Shewhart (1931) and will also become our main focus in the analysis that follows. 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 similar ideas, Sarnowski and Szajowski (2011) extended this result to the dependent observations case, and very recently Pollak and Krieger (2013) considered the independent and identically distributed (i.i.d.) case but with the data after the change distributed according to a parametric family of probability density functions (pdfs) and the parameters following a known prior. Pollak and Krieger (2013) also adopted a semi-Bayesian approach where the change time τ is deterministic and unknown while the postchange density, as before, is a parametric family with the parameters distributed according to a known prior. In the current work, unlike Bojdecki (1979) and Sarnowski and Szajowski (2011), we follow the common practice of the classical formulation and, as in the semi-Bayesian approach of Pollak and Krieger (2013), we impose suitable constraints for false alarm control.

Before continuing with the detailed presentation of the various formulations, it is necessary to first recall the Shewhart test. Consider observations $\{\xi_t\}$ that are independent but not necessarily identically distributed before and after the change and denote with $\{\ell_t\}$ the corresponding sequence of likelihood ratios. We are then interested in the following version of the Shewhart test[2]:

$$\mathscr{S} = \inf\{t > 0 : \ell_t \geq v_t\}. \tag{1.7}$$

The threshold sequence $\{v_t\}$ is deterministic and its exact form depends on the criterion we adopt and the statistics of the observations.

Having defined the Shewhart stopping time, we briefly recall an optimality result for this test that has already been established in Moustakides (1986). In particular, in the next subsection we discuss the fact that the Shewhart test *matches* Cumulative Sum (CUSUM) as long as the average false alarm period does not exceed a specific value.

1.2. Optimality with Respect to Lorden's Classical Criterion

In the case of i.i.d. observations before and after the change with corresponding pdfs $f_\infty(\xi)$ and $f_0(\xi)$, in Moustakides (1986) it was proved that CUSUM solves the

---

[2]To avoid unnecessary technical complications, throughout our work we are going to assume that the cumulative distribution functions (cdfs) of all likelihood ratios ℓ_t, under both probability measures, are continuous and strictly increasing functions.

following constrained optimization problem proposed by Lorden (1986):

$$\inf_{T} \sup_{t \geq 0} \text{ess sup } \mathsf{E}_t[T - t | \mathscr{F}_t, T > t]; \quad \text{over all } T : \mathsf{E}_\infty[T] \geq \gamma \geq 1. \tag{1.8}$$

The CUSUM stopping time T_C is defined in Moustakides (1986) as follows: For $t > 0$ let

$$Y_t = \max\{Y_{t-1}, 1\}\ell_t, \ Y_0 = 0; \quad T_C = \inf\{t > 0 : Y_t \geq v\}, \tag{1.9}$$

where $\{Y_t\}$ is the CUSUM statistic, while the constant threshold $v \geq 0$ is selected so that the false alarm constraint is satisfied with equality.

It is interesting to note that, customarily, the CUSUM statistic Y_t is specified in the literature slightly differently, namely, $Y_t = \max\{Y_{t-1}\ell_t, 1\}$. When $v > 1$, the two statistics give rise to exactly the same stopping time T_C. When, however, $0 \leq v \leq 1$, by adopting the classical definition we are forced to stop at $T = 1$, while (1.9) results in a nontrivial stopping time. It is in fact for these values of the threshold—that is, $v \in [0, 1]$—that CUSUM is reduced to the Shewhart test. Indeed, note from (1.9) that, as long as we do not stop at $t - 1$, we have $Y_{t-1} < v$. Consequently, when $v \leq 1$ this immediately implies that $Y_t = \max\{Y_{t-1}, 1\}\ell_t = \ell_t$, suggesting that CUSUM is reduced to the Shewhart rule (1.7) with constant threshold v.

Let us identify the range of false alarm rates for which CUSUM is equivalent to the Shewhart test. Since under each probability measure the sequence $\{\ell_t\}$ is i.i.d. we can conclude

$$\mathsf{E}_i[\mathscr{S}] = \sum_{t=0}^{\infty} \mathsf{P}_i(\mathscr{S} > t) = \sum_{t=0}^{\infty} \mathsf{P}_i(\ell_1 < v; \ell_2 < v; \cdots; \ell_t < v)$$

$$= \sum_{t=0}^{\infty} \left(\mathsf{P}_i(\ell_1 < v) \right)^t = \frac{1}{\mathsf{P}_i(\ell_1 \geq v)}, \quad i = 0, \infty. \tag{1.10}$$

From the previous equality we deduce that the largest value of the false alarm rate γ, for which $T_C = \mathscr{S}$, is achieved when $v = 1$. This implies that for $\gamma \in [1, \mathsf{P}_\infty^{-1}(\ell_1 \geq 1)]$ we can find a threshold $v \in [0, 1]$ such that CUSUM is reduced to the Shewhart rule. It is also clear that the classical definition of CUSUM cannot accommodate any false alarm rate within the same interval.

The previous range of false alarm rates can become more pronounced if we consider the exponential penalty criterion proposed by Poor (1998); that is,

$$\hat{\mathscr{J}}_L(T) = \sup_{t \geq 0} \text{ess sup } \mathsf{E}_t\left[\frac{1 - c^{T-t}}{1 - c} \Big| T > t, \mathscr{F}_t \right], \quad 0 < c, c \neq 1.$$

It is easy to see from the previous criterion that we can recover (1.2) by letting $c \to 1$. As in (1.8) we are interested in minimizing $\hat{\mathscr{J}}_L(T)$ over all stopping times that satisfy the same false alarm constraint $\mathsf{E}_\infty[T] \geq \gamma \geq 1$. The optimum stopping time (see Poor, 1998) has the following CUSUM-like form:

$$\hat{Y}_t = \max\{\hat{Y}_{t-1}, 1\}c\ell_t, \quad \hat{Y}_0 = 0; \quad \hat{T}_C = \inf\{t > 0 : \hat{Y}_t \geq v\}.$$

We can then verify that $\hat{T}_C$ is reduced to the Shewhart test when $\gamma \in [1, \mathsf{P}_\infty^{-1}(\ell_1 \geq 1/c)]$. If $0 < c < 1$, the previous interval is clearly larger than the one obtained in

the classical $c = 1$ case. The range of false alarm rates just specified can be quite significant if the two pdfs differ drastically, namely, when we have "large changes." Let us demonstrate this fact with a simple example.

Example 1.1. Consider the detection of a change in the mean of a Gaussian i.i.d. process of unit variance, from 0 to $\mu > 0$. We can then see that when

$$1 \le \gamma \le \frac{1}{\Phi(-0.5\mu)},$$

where $\Phi(x)$ is the c.d.f. of a standard Gaussian, CUSUM is reduced to Shewhart with corresponding maximal average detection delay satisfying

$$1 \le \mathcal{J}_{\mathrm{L}}(\mathscr{S}) \le \frac{1}{\Phi(0.5\mu)}.$$

If we select μ such that $\Phi(-0.5\mu) = 0.001$, resulting in $\mu = 6.1805$, this allows for average false alarm periods in the interval $1 \le \gamma \le 1000$, when the corresponding detection delay is, at worst, equal to 1.001, performance that, undoubtedly, can satisfy any exigent user.

Our previous discussion corroborates what is already known in the literature, namely, that the Shewhart test behaves extremely well when changes are "large," whereas in the case of "small" changes one needs to resort to CUSUM. Actually, it is clear that this optimal behavior of the Shewhart test is inherited from the optimality of CUSUM.

Even though the previous result concerning the Shewhart test is interesting, it is nevertheless theoretically restricted since it covers only a limited range of false alarm rates. In the next section we will demonstrate that this simple detection rule is in fact optimum according to a number of intriguing criteria. We would also like to mention that in Section 4.1 we will return to this optimality property of Shewhart and extend it to the case of multiple possibilities under the post-change regime.

2. THE PROBABILITY MAXIMIZING APPROACH

Let us now adopt the alternative performance measures introduced in Section 1.1 and analyze the special case $m = 1$. As mentioned, this corresponds to the probability of the event that detection will be achieved with the first observation under the alternative regime. Therefore, we consider

$$\mathcal{J}_{\mathrm{S}}(T) = \mathsf{P}\big(T = (\tau + 1)^+ | T > \tau\big) \tag{2.1}$$

$$\mathcal{J}_{\mathrm{L}}(T) = \inf_{t \ge 0} \mathrm{ess}\, \inf \mathsf{P}_t(T = t + 1 | \mathscr{F}_t, T > t) \tag{2.2}$$

$$\mathcal{J}_{\mathrm{P}}(T) = \inf_{t \ge 0} \mathsf{P}_t(T = t + 1 | T > t), \tag{2.3}$$

corresponding to (1.4), (1.5), (1.6), respectively, with $x^+ = \max\{x, 0\}$. In the previous measures we define the value of the conditional probability to be 1 when $\{T > \tau\}$ or $\{T > t\}$ (hence also $\{T = \tau + 1\}$ or $\{T = t + 1\}$, respectively) is the empty set. Additionally, we note that in the case of Shiryaev's modified measure (2.1), due to

the existence of the prior probability, it is possible to distinguish between the events $\tau \leq -1$ and $\tau = 0$. In the former case, the soonest we can hope to detect the change is at time 0. This is the reason why in our criterion we use $(\tau + 1)^+$ instead of $(\tau + 1)$. In the other two measures this modification is unnecessary since, due to lack of prior information, a change before 0 cannot be distinguished from a change at 0.

Regarding now the stopping time T, we need to properly enrich the σ-algebra $\mathcal{F}_0$ so randomization is permitted at time 0. In particular, at time 0, with probability ϖ we decide to stop at 0 and take no samples and with probability $(1 - \varpi)$ to employ a standard stopping time that requires sampling. Probability ϖ is selected independently from the observations. This slight modification in the definition of our stopping time is absolutely necessary for Shiryaev's formulation, whereas for the Lorden and Pollak setup it is needed only for technical reasons.

We continue our presentation by examining various optimality problems defined with the help of the previous performance measures in combination with proper false alarm constraints. We start with Shiryaev's Bayesian setup.

2.1. Modified Shiryaev Criterion

Shiryaev (1963) considered the change time τ to be random, independent from the observations, with a zero modified exponential prior of the form[3]: $\mathsf{P}(\tau \leq -1) = \pi$ and $\mathsf{P}(\tau = t) = (1 - \pi)p(1 - p)^t$, $t \geq 0$, where $\pi \in [0, 1]$ and $p \in (0, 1]$. Combining (2.1) with the classical constraint on the false alarm probability used in Baysian approaches, we propose the following constrained optimization problem:

$$\sup_T \mathcal{J}_S(T) = \sup_T \mathsf{P}\big(T = (\tau + 1)^+ | T > \tau\big); \quad \text{over all } T: \mathsf{P}(T \leq \tau) \leq \alpha, \quad (2.4)$$

where $\alpha \in (0, 1)$ is a prescribed false alarm level. The next theorem identifies the optimum scheme that solves (2.4).

Theorem 2.1. *The optimum detection rule that solves* (2.4) *is defined as follows*:

  i) $\alpha \geq (1 - \pi)$: *With probability* $\varpi = 1$ *stop at* 0 *without taking any samples*.
  ii) $(1 - \pi) > \alpha \geq (1 - \pi)\frac{(1-p)\mathsf{P}_\infty(\ell_1 \geq v^*)}{1-(1-p)\mathsf{P}_\infty(\ell_1 < v^*)}$ *where* $v^* = \frac{\pi}{1-\pi}\frac{1-p}{p}$: *With probability*

$$\varpi = \frac{\alpha}{p(1-\pi)}[1 - (1-p)\mathsf{P}_\infty(\ell_1 < v^*)] - \frac{1-p}{p}\mathsf{P}_\infty(\ell_1 \geq v^*),$$

    decide between stopping at 0 *and using Shewhart with constant threshold* v^*.
  iii) $(1 - \pi)\frac{(1-p)\mathsf{P}_\infty(\ell_1 \geq v^*)}{1-(1-p)\mathsf{P}_\infty(\ell_1 < v^*)} > \alpha$: *The optimum is the Shewhart stopping time with constant threshold* v *computed from*

$$\mathsf{P}_\infty(\ell_1 \geq v) = \frac{\alpha p}{(1 - \pi - \alpha)(1 - p)} \quad (2.5)$$

    and with randomization probability $\varpi = 0$.

---

[3]There is a slight difference between the current definition of the prior and the one encountered in the literature. This is because in our approach, τ is the last time instant under the nominal regime, whereas in the literature τ is conventionally considered as the first instant under the alternative.

Proof. The proof of Theorem 2.1 is presented in the Appendix.   ☐

The exponential prior model is theoretically very appealing because it leads to well-defined optimal stopping problems. However, one of its key weaknesses is the need to properly specify the parameter pair (π, p). If the two quantities are unknown and cannot be defined explicitly, a possible means to overcome this problem is to *adopt a worst-case analysis* with respect to these two parameters. We should point out that this idea, detailed in the next subsection, has no equivalent in the existing literature for the classical Shiryaev criterion.

2.2. Max–Min Version of the Modified Shiryaev Criterion

Consider, as before, τ to be distributed according to a zero modified exponential with unknown parameters π, p. Let us denote our performance measure as

$$\mathcal{J}_S(T, \pi, p) = \mathsf{P}\big(T = (\tau + 1)^+ | T > \tau\big),$$

making explicit its dependence on the parameter pair (π, p). Adopting a max–min approach, we are interested in the following constrained optimization problem:

$$\sup_T \inf_{\pi, p} \mathcal{J}_S(T, \pi, p) = \sup_T \inf_{\pi, p} \mathsf{P}\big(T = (\tau + 1)^+ | T > \tau\big); \quad \text{over all } T : \mathsf{E}_\infty[T] \geq \gamma \geq 1.$$

$$(2.6)$$

As we can see, we have replaced the false alarm probability constraint, used in the previous formulation, with a constraint on the average period between false alarms, commonly encountered in min-max approaches. The next theorem presents the optimum detection rule.

Theorem 2.2. *Let v be the solution of the equation*

$$\frac{\mathsf{P}_0(\ell_1 < v)}{\mathsf{P}_\infty(\ell_1 \geq v)} = \gamma, \tag{2.7}$$

then (2.6) is solved by randomizing with probability $\varpi = \mathsf{P}_0(\ell_1 \geq v)$ between stopping at 0 and using the Shewhart stopping time with constant threshold v. The resulting stopping rule is an equalizer over all parameter pairs (π, p); while the worst-case zero modified exponential prior is the degenerate uniform obtained by selecting $\pi(p) = vp/(vp + 1 - p)$ and letting $p \to 0$.

Proof. The proof of Theorem 2.2 can be found in the appendix.   ☐

It is surprising that a worst-case analysis results in an optimum stopping rule that requires non trivial randomization at 0. This is quite uncommon in min–max approaches. It is basically due to the fact that, even though we follow a worst-case approach with respect to the two parameters, the underlying setup is still Bayesian, thus accepting randomized optimum solutions, as was demonstrated in Theorem 2.1. Let us now continue our presentation with the max–min criteria introduced in (2.2), (2.3).

2.3. Modified Lorden and Pollak Criterion

We propose the following optimization problem:

$$\sup_{T} \mathcal{J}_{\mathrm{L}}(T) = \sup_{T} \inf_{t \geq 0} \operatorname{ess\,inf} \mathsf{P}_{t}(T = t + 1 | \mathcal{F}_{t}, T > t); \quad \text{over all } T : \mathsf{E}_{\infty}[T] \geq \gamma \geq 1,$$
(2.8)

where we maximize Lorden's modified measure (2.2) under the classical constraint on the average false alarm period. Similarly, for Pollak's modified criterion (2.3), we have

$$\sup_{T} \mathcal{J}_{\mathrm{P}}(T) = \sup_{T} \inf_{t \geq 0} \mathsf{P}_{t}(T = t + 1 | T > t); \quad \text{over all } T : \mathsf{E}_{\infty}[T] \geq \gamma \geq 1. \quad (2.9)$$

The following theorem offers the solution to both problems.

Theorem 2.3. *The optimum stopping time that solves the max–min problems in* (2.8) *and* (2.9) *is the Shewhart test with constant threshold v computed from the equation* $\mathsf{P}_{\infty}(\ell_{1} \geq v) = 1/\gamma$.

Proof. The proof for (2.9) (actually under a more general semi-Bayesian setting) is given in Pollak and Krieger (2013), while the one for (2.8) is detailed in the appendix. □

We note that in the case of Pollak's modified measure we have an *exact* optimality result. This should be compared with the original criterion $\mathcal{J}_{\mathrm{P}}(T)$ in (1.3) where (third-order) asymptotically optimum detection rules are available (see Pollak, 1985; Tartakovsky et al., 2012).

The simplicity of the probability maximizing approach allows for the straightforward analysis of problems that, in the classical change-point literature (involving expected delays), have been open for many years. It is worth considering two such characteristic cases in detail and develop the corresponding optimal solutions.

3. INDEPENDENT, NON-IDENTICALLY DISTRIBUTED OBSERVATIONS

Let $\{f_{\infty,t}(x)\}$, $\{f_{0,t}(x)\}$ denote two pdf sequences and consider the case where the observation process $\{\xi_{t}\}$ is independent but not identically distributed, following the first pdf sequence up to some change time τ and switching to the second after τ. We are interested in detecting the change optimally following the max–min approach proposed in (2.8) or (2.9). We recall that the likelihood ratio $\ell_{t} = f_{0,t}(\xi_{t})/f_{\infty,t}(\xi_{t})$ now has time-varying statistics. We have the following theorem that provides the optimum solution to both problems.

Theorem 3.1. *The optimum stopping time that solves* (2.8) *and* (2.9) *for the case of independent and non-identically distributed observations is the Shewhart stopping time* $\mathcal{S} = \inf\{t > 0 : \ell_{t} \geq v_{t}(\beta)\}$, *where the sequence of thresholds $\{v_{t}(\beta)\}$ is obtained by solving the equations*

$$\mathsf{P}_{0,t}\big(\ell_{t} \geq v_{t}(\beta)\big) = \beta; \quad \forall t > 0, \quad (3.1)$$

with parameter $\beta \in (0, 1)$. Assuming for each β that

$$\sup_{t>0} \mathsf{P}_{\infty,t}\big(\ell_t < v_t(\beta)\big) < 1, \tag{3.2}$$

this parameter is specified by requiring the false alarm constraint to be satisfied with equality; that is,

$$\mathsf{E}_\infty[\mathscr{S}] = 1 + \sum_{t=1}^\infty \prod_{l=1}^t \mathsf{P}_{\infty,l}\big(\ell_l < v_l(\beta)\big) = \gamma. \tag{3.3}$$

Proof. The proof of Theorem 3.1 can be found in the appendix. $\square$

Due to the time-varying statistics, the threshold sequence needs to be time-varying as well. With (3.1) we assure that the Shewhart test is an equalizer over time, a very important property for proving its optimality. This is indeed true since $\mathsf{P}_t(\mathscr{S} = t + 1|\mathscr{F}_t, \mathscr{S} > t) = \mathsf{P}_t(\mathscr{S} = t + 1|\mathscr{S} > t) = \mathsf{P}_{0,t}(\ell_t \geq v_t(\beta)) = \beta$. Of course, this condition still generates an ambiguity since β is unknown. This last parameter is then specified by forcing the Shewhart stopping time to satisfy the false alarm constraint with equality through (3.3). Condition (3.2) guarantees summability of the series in (3.3) and also simplifies, considerably, the proof of our theorem. It can be relaxed but at the expense of a far more involved analysis.

Example 3.1. Consider the case where $f_{\infty,t}(x)$ is time invariant Gaussian with mean 0 and variance 1, while $f_{0,t}(x)$ is Gaussian with mean $\mu_t > 0$ and variance 1. The sequence of thresholds then becomes

$$v_t(\beta) = e^{0.5\mu_t^2 + \mu_t s(\beta)}; \quad \text{where } s(\beta) = \Phi^{-1}(1 - \beta),$$

and $\Phi^{-1}(x)$ denotes the inverse cdf of a standard Gaussian. Assumption (3.2) is valid if the sequence of means $\{\mu_t\}$ is upper bounded by a finite constant. To find β, we observe that $\mathsf{P}_\infty(\ell_t \leq v_t) = \Phi(\mu_t + s(\beta))$. Since there is a one-to-one correspondence between $\beta \in (0, 1)$ and $s(\beta) \in \mathbb{R}$, we can instead solve (3.3) for s; that is,

$$1 + \sum_{t=1}^\infty \prod_{l=1}^t \Phi(\mu_l + s) = \gamma$$

and compute the optimum performance as $\beta = 1 - \Phi(s) = \Phi(-s)$.

4. MULTIPLE POSTCHANGE PROBABILITY MEASURES

Consider now the change detection problem with more than one postchange possibility. Our observation sequence $\{\xi_t\}$ is i.i.d. before and after the change with a common pdf $f_\infty(\xi)$ before the change and two[4] different pdf possibilities $f_0^1(\xi), f_0^2(\xi)$ after the change. Following a pure non-Bayesian approach (see Pollak and Krieger, 2013, for semi-Bayesian formulations) we extend the definition

---

[4]Extension to more than two pdfs is straightforward.

of our performance measures in order to account for the multiple postchange distributions. Define

$$\mathcal{J}_{\mathrm{L}}(T) = \min_{i=1,2} \inf_{t \geq 0} \mathrm{ess\ inf}\ \mathsf{P}_t^i(t < T \leq t + m | \mathscr{F}_t, T > t)$$

$$\mathcal{J}_{\mathrm{P}}(T) = \min_{i=1,2} \inf_{t \geq 0} \mathsf{P}_t^i(t < T \leq t + m | T > t),$$

where P_t^i is the measure induced by a change at time t with the alternative pdf being $f_0^i(\xi)$. Consequently, in our criterion we include an additional minimization over the possible alternative measures.

Limiting, again, ourselves to the special case $m = 1$, we are interested in solving the following constrained optimization problems:

$$\sup_T \mathcal{J}_{\mathrm{L}}(T) = \sup_T \min_{i=1,2} \inf_{t \geq 0} \mathrm{ess\ inf}\ \mathsf{P}_t^i(T = t + 1 | \mathscr{F}_t, T > t); \quad \text{over all } T: \ \mathsf{E}_\infty[T] \geq \gamma \geq 1,$$

$$(4.1)$$

for the Lorden and

$$\sup_T \mathcal{J}_{\mathrm{P}}(T) = \sup_T \min_{i=1,2} \inf_{t \geq 0} \mathsf{P}_t^i(T = t + 1 | T > t); \quad \text{over all } T: \mathsf{E}_\infty[T] \geq \gamma \geq 1, \quad (4.2)$$

for the Pollak criterion. We note that we have two sequences of likelihood ratios, namely, $\{\ell_t^1\}$ and $\{\ell_t^2\}$ defined as $\ell_t^i = f_0^i(\xi_t)/f_\infty(\xi_t), i = 1, 2$. For each $q \in [0, 1]$ we define a threshold $v(q) \geq 0$, so that the following version of the Shewhart test

$$\mathcal{S}(q) = \inf\{t > 0 : (1 - q)\ell_t^1 + q\ell_t^2 \geq v(q)\} \quad (4.3)$$

satisfies the equation

$$\mathsf{P}_\infty\big((1 - q)\ell_1^1 + q\ell_1^2 \geq v(q)\big) = \frac{1}{\gamma}. \quad (4.4)$$

The next theorem demonstrates that by proper selection of the parameter q, the corresponding stopping time solves both optimization problems.

Theorem 4.1. *For the solution of* (4.1) *and* (4.2) *we distinguish three cases*:

(i) *If* $\mathsf{P}_0^2(\ell_1^1 \geq v(0)) \geq \mathsf{P}_0^1(\ell_1^1 \geq v(0))$, *then the optimum test is* $\mathcal{S}(0)$.
(ii) *If* $\mathsf{P}_0^1(\ell_1^2 \geq v(1)) \geq \mathsf{P}_0^2(\ell_1^2 \geq v(1))$ *then the optimum test is* $\mathcal{S}(1)$.
(iii) *If there is* $q \in (0, 1)$ *such that*

$$\mathsf{P}_0^1\big((1 - q)\ell_1^1 + q\ell_1^2 \geq v(q)\big) = \mathsf{P}_0^2\big((1 - q)\ell_1^1 + q\ell_1^2 \geq v(q)\big), \quad (4.5)$$

then the optimum test is $\mathcal{S}(q)$. *For each* $\gamma \geq 1$, *only one of* (i), (ii) *and* (iii) *applies*.

Proof. The proof of Theorem 4.1 can be found in the appendix. □

We can use the previous outcome to find solutions for Lorden's *original* criterion involving average detection delays when there are multiple postchange probabilities. The goal is to obtain a result that is similar to the one presented in Section 1.2 for the Shewhart rule of Theorem 4.1.

4.1. Multiple Postchange Measures with Lorden's Original Criterion

Consider the Lorden criterion in (1.2) properly extended to cover multiple postchange probability distributions

$$\mathcal{J}_L(T) = \max_{i=1,2} \sup_{t \geq 0} \text{ess sup } E_t^i[T - t | \mathcal{F}_t, T > t].$$

We are then interested in the following min–max constrained optimization problem:

$$\inf_T \mathcal{J}_L(T) = \inf_T \max_{i=1,2} \sup_{t \geq 0} \text{ess sup } E_t^i[T - t | \mathcal{F}_t, T > t]; \quad \text{over all } T : E_\infty[T] \geq \gamma \geq 1.$$
$$(4.6)$$

This problem has been open for many years. Existing results typically refer to the two-sided CUSUM (2-CUSUM) and demonstrate that this rule exhibits different levels of asymptotic optimality. For example, in Hadjiliadis and Moustakides (2006) and Hadjiliadis and Poor (2009), it is proved that specially designed 2-CUSUM tests enjoy second- and third-order asymptotic optimality when detecting changes in the constant drift of a Brownian motion. Dragalin (1997) provides first-order asymptotically optimum 2-CUSUM rules for the case of single-parameter exponential families.

With the next theorem we present the analog of Section 1.2 for the case of two postchange probability measures. In particular, we demonstrate that the Shewhart test of Theorem 4.1 can be the *exact* solution to (4.6) provided threshold v (hence parameter γ) takes values within a specific range. The next theorem presents the precise form of our claim. We recall that the two likelihood ratios ℓ_t^i are known functions of the observation ξ_t.

Theorem 4.2. *With $\mathcal{S}(q)$ defined in (4.3), (4.4), we distinguish three cases that can provide partial solution to (4.6):*

  (i) *If $P_0^2(\ell_1^1 \geq v(0)) \geq P_0^1(\ell_1^1 \geq v(0))$ with $1 \geq v(0) \geq 0$, then the optimum test is $\mathcal{S}(0)$.*
 (ii) *If $P_0^1(\ell_1^2 \geq v(1)) \geq P_0^2(\ell_1^2 \geq v(1))$ with $1 \geq v(1) \geq 0$, then the optimum test is $\mathcal{S}(1)$.*
(iii) *If there is $q \in (0, 1)$ such that*

$$P_0^1((1 - q)\ell_1^1 + q\ell_1^2 \geq v(q)) = P_0^2((1 - q)\ell_1^1 + q\ell_1^2 \geq v(q)), \qquad (4.7)$$

  and

$$\min\{q + (1 - q)\inf_{\xi_1 \in \mathcal{A}_1 \cap \mathcal{A}_2^c} \ell_1^1, (1 - q) + q \inf_{\xi_1 \in \mathcal{A}_1^c \cap \mathcal{A}_2} \ell_1^2\} \geq v(q) \geq \inf_{\xi_1 \in \mathcal{A}_1 \cap \mathcal{A}_2} \{(1 - q)\ell_1^1 + q\ell_1^2\}$$
$$(4.8)$$

  where $\mathcal{A}_i = \{\xi_1 : \ell_1^i \leq 1\}$ and $\mathcal{A}_i^c$ its complement, then the optimum test is $\mathcal{S}(q)$.

Proof. The proof of Theorem 4.1 can be found in the appendix.   □

Even though the extent of this result is clearly limited, it is nonetheless the first time we have a nonasymptotic solution for Lorden's formulation when there

are multiple distributions under the alternative regime. Theorem 4.2 also establishes that 2-CUSUM is *not* strictly optimum (at least not in the sense of 4.6) despite its very strong asymptotic optimality properties. Finally, we need to mention that it is not possible to recover the same nonasymptotic result by assigning specific prior probabilities to the postchange measures (i.e., following the semi-Baysian idea in Pollak and Krieger, 2013). The extra freedom enjoyed by considering each probability measure separately is critical in demonstrating the optimality of the Shewhart test in the sense of Lorden.

Example 4.1. Consider the Gaussian case where under the nominal regime the samples are i.i.d. with mean 0 and variance 1, whereas under the alternative they can have two possible means $\pm\mu$, $\mu > 0$ with unit variance. Let us apply case (iii) of Theorem 4.2. Due to symmetry it is sufficient to select $q = 0.5$ to satisfy (4.7).

The two likelihood ratios ℓ_1^i as functions of the observation ξ_1 are equal to $e^{-0.5\mu^2 \pm \mu \xi_1}$ and the sets of interest are $\mathscr{A}_1 = \{\xi_1 : \xi_1 \leq 0.5\mu\}$ and $\mathscr{A}_2 = \{\xi_1 : \xi_1 \geq -0.5\mu\}$. We can now compute the critical range for threshold v from (4.8). Since $\inf_{\xi_1 < -0.5\mu} \ell_1^1 = \inf_{\xi_1 > 0.5\mu} \ell_1^2 = 0$ and $\inf_{-0.5\mu \leq \xi_1 \leq 0.5\mu} 0.5(\ell_1^1 + \ell_1^1) = e^{-0.5\mu^2}$, we have $0.5 \geq v \geq e^{-0.5\mu^2}$. This interval is nonempty when $\mu > \sqrt{2\log 2} = 1.1774$ and gives rise to the following range for γ:

$$1 \leq \gamma \leq \frac{1}{2\Phi(-0.5\mu + \delta)}, \quad \text{where } \delta = -\frac{1}{\mu}\log\left(\frac{1 + \sqrt{1 - 4e^{-\mu^2}}}{2}\right),$$

with the corresponding worst-case average detection delay satisfying

$$1 \leq \mathscr{J}_{\mathrm{L}}(\mathscr{S}) \leq \frac{1}{\Phi(0.5\mu + \delta) + \Phi(-1.5\mu + \delta)}.$$

Using the same numerical value we adopted in Example 1.1 for the one-sided case, namely, $\mu = 6.1805$, we obtain $1 \leq \gamma \leq 500$ while the optimum detection delay becomes, at worst, 1.001. Compared to Example 1.1, as we can see, the range of γ where the Shewhart test is optimum is reduced to half.

Remark 4.1. Because with the maximizing probability approach we focus on a single sample after the change, it turns out that Shewhart is optimum for *transient changes* as well. Specifically, the same proofs go through for any type of change provided that *it lasts at least one sample* (which is necessary for a change to exist). Clearly, this is an additional distinct optimality characteristic enjoyed by this simple detection rule. As we know, the Shiryaev, CUSUM, and Shiryaev-Roberts tests lose their optimality if the change does not last indefinitely after its occurrence.

APPENDIX: PROOFS

Proof of Theorem 2.1. We begin our analysis by writing the performance measure in a more detailed form. We have

$$\mathscr{J}_{\mathrm{S}}(T) = \mathsf{P}(T = (\tau + 1)^+ | T > \tau) = \frac{\mathsf{P}(T = (\tau + 1)^+)}{\mathsf{P}(T > \tau)}. \tag{A.1}$$

Since $T \geq 0$, for the numerator we can write

$$P(T = (\tau+1)^+) = P(\tau \leq -1)P(T=0) + \sum_{t=0}^{\infty} P(\tau=t)P_t(T=t+1)$$

$$= \pi\varpi + (1-\pi)p \sum_{t=0}^{\infty}(1-p)^t P_t(T=t+1)$$

$$= \pi\varpi + (1-\pi)p \sum_{t=0}^{\infty}(1-p)^t E_\infty[\ell_{t+1}\mathbb{1}_{\{T=t+1\}}]$$

$$= \pi\varpi + \frac{(1-\pi)p}{(1-p)} E_\infty[(1-p)^T \ell_T \mathbb{1}_{\{T>0\}}]$$

$$= \pi\varpi + \frac{(1-\pi)p}{(1-p)} E_\infty[(1-p)^T \ell_T | T > 0]P(T>0)$$

$$= \pi\varpi + \frac{(1-\pi)p}{(1-p)} E_\infty[(1-p)^T \ell_T | T > 0](1-\varpi). \qquad (A.2)$$

Similarly, for the denominator, since $\{T > t\} \in \mathscr{F}_t$ and $T \geq 0$, we have

$$P(T > \tau) = P(\tau \leq -1)P(T > -1) + \sum_{t=0}^{\infty} P(\tau=t)P_\infty(T>t)$$

$$= \pi + (1-\pi)p \sum_{t=0}^{\infty}(1-p)^t P_\infty(T>t) = \pi + (1-\pi)p E_\infty\left[\sum_{t=0}^{T-1}(1-p)^t\right]$$

$$= \pi + (1-\pi)E_\infty[1-(1-p)^T] = \pi + (1-\pi)E_\infty\left[\left(1-(1-p)^T\right)\mathbb{1}_{\{T>0\}}\right]$$

$$= \pi + (1-\pi)E_\infty[1-(1-p)^T | T > 0]P(T>0)$$

$$= \pi + (1-\pi)\{1 - E_\infty[(1-p)^T | T > 0]\}(1-\varpi), \qquad (A.3)$$

with the third last equality being true because $1 - (1-p)^0 = 0$. Combining (A.2) and (A.3) we have the following form for the modified Shiryaev measure:

$$\mathscr{J}_S(T) = \frac{\pi\varpi + \frac{(1-\pi)p}{1-p} E_\infty[(1-p)^T \ell_T | T > 0](1-\varpi)}{\pi + (1-\pi)\{1 - E_\infty[(1-p)^T | T > 0]\}(1-\varpi)}. \qquad (A.4)$$

Next, we distinguish different possibilities depending on the value of α. For case (i) where $\alpha \geq 1 - \pi$ by selecting $\varpi = 1$, in other words stopping at 0 with probability 1, as we can see from (A.4), yields $\mathscr{J}_S(T) = 1$, which is the maximum possible value for our criterion (since it is a probability). On the other hand, the denominator, which is the complement of the false alarm probability, from (A.3) is equal to π. This means that the false alarm probability is $1 - \pi$, thus satisfying the constraint.

Let now $1 - \pi > \alpha$. Since the stopping time T must satisfy the false alarm constraint, this suggests that the denominator, by being the complement of the false alarm probability, is no smaller than $1 - \alpha$. We are going to show that in order to maximize the performance measure it is sufficient to limit ourselves to stopping times that satisfy the false alarm constraint with equality. This equality will be

achieved by modifying the randomization probability ϖ in a way that will improve (increase) the value of the criterion $\mathcal{J}_S(T)$.

As we can see from (A.4), both the numerator and the denominator are linear functions of ϖ and the ratio takes its maximal value (equal to 1) for $\varpi = 1$. We can therefore conclude that the ratio is an increasing function of ϖ. If T is such that the denominator is strictly greater than $1 - \alpha$ and since we are in the case where $\pi < 1 - \alpha$, this suggests that, necessarily, we have $1 - (1 - \pi)\mathsf{E}_\infty[(1 - p)^T | T > 0] > 1 - \alpha$. Consequently, by replacing ϖ with a *larger* value $\varpi' > \varpi$ we can make the denominator exactly equal to $1 - \alpha$. Making the same change in the numerator, due to the monotonicity with respect to ϖ, this will result in an overall increase of our performance measure. Therefore, without loss of generality, we may limit ourselves to stopping times that satisfy the false alarm constraint with equality.

The previous observation suggests that we can maximize the numerator in (A.4) subject to the constraint that the denominator is equal to $(1 - \alpha)$. Using the Lagrange multiplier technique we define the following criterion $\mathcal{G}(T)$ that combines the numerator and the constraint:

$$
\begin{aligned}
\mathcal{G}(T) &= \pi\varpi + \frac{(1 - \pi)p}{1 - p}\mathsf{E}_\infty[(1 - p)^T \ell_T | T > 0](1 - \varpi) \\
&\quad + \lambda\left\{\pi\varpi + \{1 - (1 - \pi)\mathsf{E}_\infty[(1 - p)^T | T > 0]\}(1 - \varpi)\right\} \\
&= \pi(1 + \lambda)\varpi + \left\{\lambda + (1 - \pi)\mathsf{E}_\infty\left[(1 - p)^T\left(\frac{p}{1 - p}\ell_T - \lambda\right)\Big| T > 0\right]\right\}(1 - \varpi),
\end{aligned}
$$

with λ being the corresponding Lagrange multiplier. The goal, now, is first to maximize $\mathcal{G}(T)$ over $T > 0$ and then over the randomization probability $\varpi \in [0, 1]$. Fixing ϖ and maximizing over $T > 0$ means that we need to maximize the expression

$$
\hat{\mathcal{G}}(T) = \mathsf{E}_\infty\left[(1 - p)^T\left(\frac{p}{1 - p}\ell_T - \lambda\right)\Big| T > 0\right].
$$

For $v \geq 0$ consider the following specific value of the Lagrange multiplier:

$$
\lambda = \frac{v}{1 - p}\{1 - (1 - p)\mathsf{P}_\infty(\ell_1 < v)\} - \mathsf{P}_0(\ell_1 \geq v).
$$

Using standard optimal stopping theory it is then straightforward to show that $\hat{\mathcal{G}}(T)$ is maximized by the Shewhart stopping time defined in (1.7) with constant threshold v. The corresponding optimum performance can then be computed as follows:

$$
\begin{aligned}
\hat{\mathcal{G}}(\mathcal{S}) &= \sum_{t=1}^{\infty}(1 - p)^t[\mathsf{P}_\infty(\ell_1 < v)]^{t-1}\left\{\frac{p}{1 - p}\mathsf{P}_0(\ell_1 \geq v) - \lambda\mathsf{P}_\infty(\ell_1 \geq v)\right\} \\
&= \frac{p\mathsf{P}_0(\ell_1 \geq v) - \lambda(1 - p)\mathsf{P}_\infty(\ell_1 \geq v)}{1 - (1 - p)\mathsf{P}_\infty(\ell_1 < v)} = \frac{p}{1 - p}v - \lambda.
\end{aligned}
$$

The last equality can be verified by directly substituting the definition of the Lagrange multiplier λ. Using this result in the original measure, we end up with the following inequality:

$$
\mathcal{G}(T) \leq \pi(1 + \lambda)\varpi + \left[\pi(1 + \lambda) + (1 - \pi)\frac{p}{1 - p}v - \pi\right](1 - \varpi). \tag{A.5}
$$

From (A.3), we can also compute the corresponding false alarm probability, which must be set to α (we must satisfy the constraint with equality):

$$(1 - \pi)\left\{\varpi + \frac{(1 - p)\mathrm{P}_\infty(\ell_1 \geq v)}{1 - (1 - p)\mathrm{P}_\infty(\ell_1 < v)}(1 - \varpi)\right\} = \alpha. \qquad (A.6)$$

We are now left with the definition of the randomization probability ϖ. Selecting ϖ optimally amounts to maximizing the right-hand side in (A.5) over ϖ. We observe that the corresponding expression is a convex combination of the value $\pi(1 + \lambda)$, which is the gain obtained when stopping at 0, and $\pi(1 + \lambda) + (1 - \pi)\frac{p}{1-p}v - \pi$, which is the gain resulting by employing $\mathscr{S}$ for $t > 0$. Clearly, we are going to put all of the probability mass on the largest gain. Consequently, when $v > v^* = \frac{\pi}{1-\pi}\frac{1-p}{p}$ the gain provided by $\mathscr{S}$ exceeds the gain obtained by stopping at 0; therefore, in this case we select $\varpi = 0$. Of course, v must be such that the Shewhart test satisfies the false alarm constraint with equality. From (A.6), by substituting $\varpi = 0$ we can see that the constraint is satisfied when v is computed through equation (2.5). This equation always has a solution that exceeds v^* as long as α takes values in the interval specified in case (iii). When $v = v^*$, stopping at 0 provides exactly the same gain as the Shewhart test $\mathscr{S}$ with threshold v^*. Therefore, we can randomize between the two possibilities with any probability ϖ. However, since we need to satisfy the false alarm constraint with equality, from (A.6) with $v = v^*$ we can solve for ϖ and obtain the optimum ϖ depicted in case (ii). The resulting value corresponds to a legitimate probability $\varpi \in [0, 1]$ when α is within the limits prescribed for this case. This concludes our proof. $\qquad \square$

Proof of Theorem 2.2. The proof will rely on the analysis developed for the proof of Theorem 2.1. In order to solve the max–min problem defined in our theorem it is sufficient to show the existence of a triple (T^*, π^*, p^*) such that the following saddle-point relation holds

$$\mathscr{J}_\mathrm{S}(T, \pi^*, p^*) \leq \mathscr{J}_\mathrm{S}(T^*, \pi^*, p^*) \leq \mathscr{J}_\mathrm{S}(T^*, \pi, p), \qquad (A.7)$$

for all stopping times T that satisfy the false alarm constraint. It is well known that whenever a saddle-point solution exists it is also max–min optimum. Indeed, note that if T^* satisfies (A.7) then we can write

$$\inf_{\pi, p} \mathscr{J}_\mathrm{S}(T, \pi, p) \leq \mathscr{J}_\mathrm{S}(T, \pi^*, p^*) \leq \mathscr{J}_\mathrm{S}(T^*, \pi^*, p^*) = \inf_{\pi, p} \mathscr{J}_\mathrm{S}(T^*, \pi, p),$$

where the first inequality is obvious; The second corresponds to the left-hand side inequality in (A.7) and the last equality is equivalent to the right-hand side inequality in (A.7). Consequently, T^* solves the max–min problem and the parameter pair (π^*, p^*) corresponds to the worst-case (least-favorable) exponential prior.

To show (A.7), let us first define our candidate optimum stopping time T^*. Consider (2.7) and observe that for $v \to 0$ the left-hand side tends to 0, whereas for $v \to \infty$ the same expression tends to ∞. Furthermore, the ratio is a strictly increasing and continuous function of v (see Footnote 2). Due to this continuity and strict monotonicity, the equation has a unique solution v. With the help of this threshold value our candidate detection rule T^* consists in randomizing with

probability $\varpi^* = P_0(\ell_1 \geq v)$ between stopping at 0 and using the Shewhart test $\mathscr{S}$ with constant threshold v. For T^* we observe that $E_\infty[T^*] = (1 - \varpi^*)E_\infty[\mathscr{S}] = P_0(\ell_1 < v)/P_\infty(\ell_1 \geq v) = \gamma$, suggesting that it satisfies the false alarm constraint with equality.

We first demonstrate that T^* satisfies the right-hand side in (A.7). For any parameter pair (π, p), after recalling that on $\{T^* > 0\}$ we have $T^* = \mathscr{S}$, we can verify using (A.4) that

$$\mathscr{J}_S(T^*, \pi, p) = \frac{\pi\varpi^* + \frac{(1-\pi)p}{1-p}E_\infty[(1-p)^{\mathscr{S}}\ell_{\mathscr{S}}](1 - \varpi^*)}{\pi\varpi^* + \{1 - (1-\pi)E_\infty[(1-p)^{\mathscr{S}}]\}(1 - \varpi^*)} = P_0(\ell_1 \geq v).$$

The last equality is true since we can immediately compute $E_\infty[(1-p)^{\mathscr{S}}\ell_{\mathscr{S}}] = (1-p)P_0(\ell_1 \geq v)/\{1 - (1-p)P_\infty(\ell_1 < v)\}$ and $E_\infty[(1-p)^{\mathscr{S}}] = (1-p)P_\infty(\ell_1 \geq v)/\{1 - (1-p)P_\infty(\ell_1 < v)\}$. As we realize, the resulting performance of T^* is independent from (π, p); therefore, the stopping rule is an equalizer with respect to the two parameters. This, in turn, suggests that the right-hand side in (A.7) is trivially satisfied with equality.

Showing the left-hand side inequality requires more work. Note that we need to define the worst-case parameter pair (π^*, p^*). Unfortunately, this pair turns out to be a limiting case corresponding to an exponential prior that tends to a degenerate uniform. More specifically, for $p > 0$ we solve for π the following equation:

$$\frac{\pi}{1-\pi}\frac{1-p}{p} = v,$$

resulting in $\pi(p) = vp/(1 - p + vp)$. The parameter pair $(\pi(p), p)$ with $p \to 0$ yields the worst-case exponential prior we are interested in. Consequently, for the left-hand side inequality we need to prove that

$$\lim_{p \to 0} \mathscr{J}_S(T, \pi(p), p) \leq \lim_{p \to 0} \mathscr{J}_S(T^*, \pi(p), p) = P_0(\ell_1 \geq v), \tag{A.8}$$

over all T satisfying the false alarm constraint $E_\infty[T] = (1 - \varpi)E_\infty[T|T > 0] \geq \gamma$.

Fix a sufficiently small $\epsilon > 0$ so that $\varpi_\epsilon^* = P_0(\ell_1 \geq v) + \epsilon < 1$. Define the false alarm level

$$\alpha_\epsilon(p) = (1 - \pi(p))\frac{p\varpi_\epsilon^* + (1-p)P_\infty(\ell_1 \geq v)}{p + (1-p)P_\infty(\ell_1 \geq v)}$$

and the class of stopping rules

$$\mathscr{A}_\epsilon(p) = \{T: \quad P(T \leq \tau) \leq \alpha_\epsilon(p)\}. \tag{A.9}$$

It is then straightforward to verify that $\alpha_\epsilon(p) \in (0, 1)$ and for any probability p, the quantities $\pi(p), p, \alpha_\epsilon(p)$ are such that case (ii) of Theorem 2.1 applies. This suggests that when $\mathscr{J}_S(T, \pi(p), p)$ is maximized over the class $\mathscr{A}_\epsilon(p)$, the optimum stopping time is to randomize between stopping at 0 and the Shewhart test $\mathscr{S}$ with threshold v using the randomization probability ϖ_ϵ^*. The latter is a direct consequence of the specific definition of $\pi(p)$, and $\alpha_\epsilon(p)$. Call the resulting optimal stopping time T_ϵ^*. Note also that this optimality property is true for all $1 > p > 0$. Using the

definitions of $\pi(p)$, $\alpha_\epsilon(p)$ and (A.3), we can verify that the class $\mathscr{A}_\epsilon(p)$ in (A.9) can be equivalently written as

$$\mathscr{A}_\epsilon(p) = \left\{ T : \ (1-\varpi)\mathsf{E}_\infty\left[\frac{1-(1-p)^T}{p} | T > 0 \right] \geq \frac{\gamma - \frac{\epsilon}{\mathsf{P}_\infty(\ell_1 \geq v)}}{1+p\gamma} \right\}, \tag{A.10}$$

where we also used (2.7).

Fix a T that satisfies the false alarm constraint $\mathsf{E}_\infty[T] = (1-\varpi)\mathsf{E}_\infty[T|T > 0] \geq \gamma$. As we argued before, our goal is to prove (A.8). From monotone convergence we have

$$\lim_{p\to 0}(1-\varpi)\mathsf{E}_\infty\left[\frac{1-(1-p)^T}{p} | T > 0 \right] = (1-\varpi)\mathsf{E}_\infty[T|T > 0] \geq \gamma.$$

Consequently, for any $p \in (0, p_\epsilon]$, where p_ϵ sufficiently small, we can write

$$(1-\varpi)\mathsf{E}_\infty\left[\frac{1-(1-p)^T}{p} | T > 0 \right] \geq \gamma - \frac{\epsilon}{\mathsf{P}_\infty(\ell_1 \geq v)}.$$

The previous inequality, comparing with (A.10), suggests that $T \in \mathscr{A}_\epsilon(p)$ for all $0 < p \leq p_\epsilon$. A direct consequence of this fact is that $\mathscr{J}_\mathrm{S}(T, \pi(p), p) \leq \mathscr{J}_\mathrm{S}(T_\epsilon^*, \pi(p), p)$ for all $p \in (0, p_\epsilon]$. Taking the limit as $p \to 0$ and using monotone convergence, we obtain

$$\lim_{p\to 0}\mathscr{J}_\mathrm{S}(T, \pi(p), p) \leq \lim_{p\to 0}\mathscr{J}_\mathrm{S}(T_\epsilon^*, \pi(p), p) = \frac{\varpi_\epsilon^* v + (1-\varpi_\epsilon^*)\mathsf{E}_\infty[\ell_{\mathscr{S}}]}{v + (1-\varpi_\epsilon^*)\mathsf{E}_\infty[\mathscr{S}]}. \tag{A.11}$$

Since $\mathsf{E}_\infty[\mathscr{S}] = 1/\mathsf{P}_\infty(\ell_1 \geq v)$, $\mathsf{E}_\infty[\ell_{\mathscr{S}}] = \mathsf{P}_0(\ell_1 \geq v)/\mathsf{P}_\infty(\ell_1 \geq v)$ and $\varpi_\epsilon^* = \mathsf{P}_0(\ell_1 \geq v) + \epsilon$, we conclude that $\mathsf{E}_\infty[\ell_{\mathscr{S}}] \leq \varpi_\epsilon^*\mathsf{E}_\infty[\mathscr{S}]$. Substituting in (A.11) yields

$$\lim_{p\to 0}\mathscr{J}_\mathrm{S}(T, \pi(p), p) \leq \varpi_\epsilon^* = \mathsf{P}_0(\ell_1 \geq v) + \epsilon.$$

Because this inequality is true for any sufficiently small $\epsilon > 0$, we have validity of (A.8). This concludes our proof. $\square$

Proof of Theorem 2.3. If T is such that $\mathsf{E}_\infty[T] = \infty$, then we can define a sufficiently large integer M so that $\infty > \mathsf{E}_\infty[T_M] \geq \gamma$ where $T_M = \min\{T, M\}$. Since for $t < M$ we have $\{T = t+1\} = \{T_M = t+1\}$ and $\{T > t\} = \{T_M > t\}$, we conclude $\mathsf{P}_t(T_M = t+1|\mathscr{F}_t, T_M > t) = \mathsf{P}_t(T = t+1|\mathscr{F}_t, T > t)$. On the other hand, for $t \geq M$ it is true that $\{T_M = t+1\} = \{T_M > t\} = \varnothing$, suggesting that $\mathsf{P}_t(T_M = t+1|\mathscr{F}_t, T_M > t) = 1 \geq \mathsf{P}_t(T = t+1|\mathscr{F}_t, T > t)$. This means that $\mathscr{J}_\mathrm{L}(T) \leq \mathscr{J}_\mathrm{L}(T_M)$. The last inequality implies that we can limit ourselves to stopping times T that satisfy $\infty > \mathsf{E}_\infty[T] \geq \gamma$.

From Lorden's modified measure (2.8) we conclude that for all $t \geq 0$ we can write

$$\mathscr{J}_\mathrm{L}(T) \leq \mathsf{P}_t(T = t+1|\mathscr{F}_t, T > t).$$

Multiplying both sides with $\mathbb{1}_{\{T>t\}}$ and taking expectation with respect to the nominal measure yields

$$\mathscr{J}_\mathrm{L}(T)\mathsf{P}_\infty(T > t) \leq \mathsf{P}_t(T = t+1) = \mathsf{E}_\infty[\ell_{t+1}\mathbb{1}_{\{T=t+1\}}]. \tag{A.12}$$

Summing over all $t \geq 0$ we obtain

$$\mathcal{J}_L(T)\mathsf{E}_\infty[T] \leq \mathsf{E}_\infty[\ell_T],$$

where we define $\ell_0 = 0$. From the previous inequality we conclude

$$\mathcal{J}_L(T) \leq \frac{\mathsf{E}_\infty[\ell_T]}{\mathsf{E}_\infty[T]} = \frac{(1-\varpi)\mathsf{E}_\infty[\ell_T|T>0]}{(1-\varpi)\mathsf{E}_\infty[T|T>0]} = \frac{\mathsf{E}_\infty[\ell_T|T>0]}{\mathsf{E}_\infty[T|T>0]}.$$

Let us examine the ratio $\mathsf{E}_\infty[\ell_T]/\mathsf{E}_\infty[T]$ over all T that satisfy the constraint. Note that when $\infty > \mathsf{E}_\infty[T] = (1-\varpi)\mathsf{E}_\infty[T|T>0] > \gamma$ we can replace ϖ with a larger value ϖ' so that $(1-\varpi')\mathsf{E}_\infty[T|T>0] = \gamma$ without changing the value of the ratio (since it does not depend on ϖ). This in turn suggests that any value attained by this ratio can also be achieved by a stopping time that satisfies the constraint with equality. Using this observation we can write

$$\sup_{T:\mathsf{E}_\infty[T]\geq\gamma} \mathcal{J}_L(T) \leq \sup_{T:\mathsf{E}_\infty[T]=\gamma} \frac{\mathsf{E}_\infty[\ell_T]}{\mathsf{E}_\infty[T]} = \gamma^{-1} \sup_{T:\mathsf{E}_\infty[T]=\gamma} \mathsf{E}_\infty[\ell_T]. \qquad \text{(A.13)}$$

To maximize $\mathsf{E}_\infty[\ell_T]$ over all stopping times that satisfy the constraint with equality, we reduce the optimization problem into an unconstrain one using the Lagrange multiplier technique. In particular, we consider the maximization of

$$\mathcal{G}(T) = \mathsf{E}_\infty[\ell_T - \lambda T] = (1-\varpi)\mathsf{E}_\infty[\ell_T - \lambda T|T>0].$$

To find the optimum stopping time we will first optimize over $T > 0$ and then identify the optimum randomization probability ϖ. Let $v \geq 0$ be the solution of the equation $\mathsf{P}_\infty(\ell_1 \geq v) = 1/\gamma$ and define $\lambda = \mathsf{P}_0(\ell_1 \geq v) - v\mathsf{P}_\infty(\ell_1 \geq v)$. Using standard optimal stopping theory we can then conclude that $\mathcal{G}(T)$ for $T > 0$ is optimized by the Shewhart test with threshold v. Since $\mathsf{E}_\infty[\mathcal{S}] = 1/\mathsf{P}_\infty(\ell_1 \geq v) = \gamma$ and $\mathsf{E}_\infty[\ell_\mathcal{S}] = \mathsf{P}_0(\ell_1 \geq v)/\mathsf{P}_\infty(\ell_1 \geq v) = \gamma\mathsf{P}_0(\ell_1 \geq v)$, if we also use the definition of λ we conclude that

$$\mathcal{G}(T) = (1-\varpi)\mathsf{E}_\infty[\ell_T - \lambda T|T>0] \leq (1-\varpi)\mathsf{E}_\infty[\ell_\mathcal{S} - \lambda\mathcal{S}] = (1-\varpi)v \leq v.$$

The last inequality suggests that the optimum randomization probability is $\varpi = 0$. From the previous result we have that for any T satisfying the false alarm constraint with equality, we can write

$$\mathsf{E}_\infty[\ell_T] - \lambda\gamma = \mathsf{E}_\infty[\ell_T - \lambda T] \leq \mathsf{E}_\infty[\ell_\mathcal{S} - \lambda\mathcal{S}] = \mathsf{E}_\infty[\ell_\mathcal{S}] - \lambda\gamma,$$

which implies $\mathsf{E}_\infty[\ell_T] \leq \mathsf{E}_\infty[\ell_\mathcal{S}] = \gamma\mathsf{P}_0(\ell_1 \geq v)$. Observing also that for every $t \geq 0$ we have $\mathsf{P}_t(\mathcal{S} = t+1|\mathcal{F}_t, \mathcal{S} > t) = \mathsf{P}_0(\ell_1 \geq v)$, this means that Shewhart is an equalizer; consequently, $\mathcal{J}_L(\mathcal{S}) = \mathsf{P}_0(\ell_1 \geq v)$. Using these two facts in (A.13) leads to

$$\mathcal{J}_L(\mathcal{S}) \leq \sup_{T:\mathsf{E}_\infty[T]\geq\gamma} \mathcal{J}_L(T) \leq \gamma^{-1} \sup_{T:\mathsf{E}_\infty[T]=\gamma} \mathsf{E}_\infty[\ell_T] \leq \gamma^{-1}\{\gamma\mathsf{P}_0(\ell_1 \geq v)\} = \mathsf{P}_0(\ell_1 \geq v) = \mathcal{J}_L(\mathcal{S}),$$

which proves optimality for $\mathcal{S}$ and concludes the proof. Exactly the same analysis applies to (2.9). In fact, we can simply start the proof from (A.12), which is immediately satisfied by Pollak's modified measure. $\qquad \square$

Proof of Theorem 3.1. Let β and $\{v_t(\beta)\}$ be such that (3.1), (3.2), (3.3) are satisfied. If we define $\rho(\beta) = \sup_{t>0} \mathsf{P}_{\infty,t}\big(\ell_t < v_t(\beta)\big)$ then assumption (3.2) is equivalent to

$$0 \le \rho(\beta) < 1. \tag{A.14}$$

For simplicity, from now on, we drop the dependence of $v_t(\beta)$ and $\rho(\beta)$ on β. For $t \ge 0$ define the two sequences $\{\omega_t\}$, $\{c_t\}$

$$\omega_t = \mathsf{E}_\infty[\mathscr{S} - t | \mathscr{S} > t] = 1 + \sum_{n=t+1}^{\infty} \prod_{l=t+1}^{n} \mathsf{P}_{\infty,l}(\ell_l < v_l) \quad \text{and} \quad c_t = \frac{\omega_{t+1}}{v_{t+1}}.$$

Also, set $c_{-1} = 0$ and $\ell_0 = 0$. From the definition of ω_t and comparing with (3.3) we conclude that $\omega_0 = \gamma$. Note that $\{\omega_t\}$ satisfies the backward recursion

$$\omega_{t-1} = 1 + \mathsf{P}_{\infty,t}(\ell_t < v_t)\omega_t. \tag{A.15}$$

From (A.14) we have $\mathsf{P}_{\infty,t}(\ell_t < v_t) \le \rho$, suggesting that $\omega_t \le 1/(1-\rho)$. Furthermore,

$$1 - \beta = \mathsf{P}_{0,t}(\ell_t < v_t) = \mathsf{E}_\infty[\ell_t \mathbb{1}_{\{\ell_t < v_t\}}] \le v_t,$$

from which we conclude that $c_t \le 1/(1-\beta)(1-\rho)$. In other words, both sequences $\{\omega_t\}$, $\{c_t\}$ are uniformly bounded from above by some finite constant.

Consider first (3.3). The function $\phi(\beta) = 1 + \sum_{t=1}^{\infty} \prod_{l=1}^{t} \mathsf{P}_{\infty,l}\big(\ell_l < v_l(\beta)\big)$ is decreasing in β with $\phi(0) = \infty$ and $\phi(1) = 1$. From assumption (3.2) we have validity of (A.14), which allows for the use of bounded convergence to show that $\phi(\beta)$ is continuous in β. This suggests that (3.3) has a nonnegative solution.

As in the previous theorem, we can write

$$\mathscr{J}_{\mathrm{L}}(T)\mathsf{P}_\infty(T > t) \le \mathsf{E}_\infty[\ell_{t+1}\mathbb{1}_{\{T=t+1\}}].$$

Multiplying both sides with c_t, which is nonnegative and summing over $t \ge 0$, for any $T > 0$ we have

$$\mathscr{J}_{\mathrm{L}}(T) \le \frac{\mathsf{E}_\infty[\ell_T c_{T-1}]}{\mathsf{E}_\infty[\sum_{t=0}^{T-1} c_t]}.$$

Recalling that $\ell_0 = 0$ and using similar arguments as in the proof of Theorem 2.3, we can show that

$$\sup_{T:\mathsf{E}_\infty[T]\ge\gamma} \mathscr{J}_{\mathrm{L}}(T) \le \sup_{T:\mathsf{E}_\infty[T]=\gamma} \frac{\mathsf{E}_\infty[\ell_T c_{T-1}]}{\mathsf{E}_\infty[\sum_{t=0}^{T-1} c_t]}, \tag{A.16}$$

namely, to maximize the upper bound it suffices to limit ourselves to stopping times that satisfy the false alarm constraint with equality. We will show that the upper bound cannot exceed β.

Fix T with $\mathsf{E}_\infty[T] = (1 - \varpi)\mathsf{E}_\infty[T|T > 0] = \gamma$ and consider the expression

$$\mathscr{G}(T) = \mathsf{E}_\infty\left[\ell_T c_{T-1} - \beta \sum_{t=0}^{T-1} c_t + T\right] = (1-\varpi)\mathsf{E}_\infty\left[\ell_T c_{T-1} + \sum_{t=0}^{T-1}(1 - \beta c_t)|T > 0\right]. \tag{A.17}$$

Note that $\varpi = 1$ is not an acceptable value since then T cannot satisfy the false alarm constraint with equality. Therefore, $0 \le \varpi < 1$. This suggests that $E_\infty[T|T > 0] = \gamma/(1-\varpi) < \infty$. We first examine the part $T > 0$, namely, the expression

$$\hat{\mathcal{G}}(T) = E_\infty\left[\ell_T c_{T-1} + \sum_{t=0}^{T-1}(1 - \beta c_t)|T > 0\right].$$

We observe that

$$E_\infty[\ell_T|T > 0] = \sum_{t=1}^{\infty} E_\infty[\ell_t \mathbb{1}_{\{T=t\}}|T > 0] \le \sum_{t=1}^{\infty} E_\infty[\ell_t \mathbb{1}_{\{T>t-1\}}|T > 0]$$

$$= \sum_{t=1}^{\infty} E_\infty\left[E_\infty[\ell_t|\mathcal{F}_{t-1}]\mathbb{1}_{\{T>t-1\}}|T > 0\right] = E_\infty[T|T > 0] < \infty.$$

Since $\{c_t\}$ is uniformly bounded and because of the previous observation, this suggests that for every $\epsilon > 0$ we can find sufficiently large integer M so that $|\hat{\mathcal{G}}(T) - \hat{\mathcal{G}}(T_M)| \le \epsilon$, where $T_M = \min\{T,M\}$. This implies

$$\hat{\mathcal{G}}(T) \le \hat{\mathcal{G}}(T_M) + \epsilon. \tag{A.18}$$

We can now maximize $\hat{\mathcal{G}}(T_M)$ over T_M with the optimization performed over the finite time horizon $[0,M]$. From standard optimal stopping theory we can define the sequence of optimal costs with the help of the backward recursion

$$V_t(\ell_t) = \max\{\ell_t c_{t-1}, (1 - \beta c_t) + E_\infty[V_{t+1}(\ell_{t+1})]\}; t \quad = M - 1, \ldots, 0,$$

starting with $V_M(\ell_M) = \ell_M c_{M-1}$. Since $V_M(\ell_M) \le \max\{\ell_M c_{M-1}, \omega_M\}$, using induction we can show that $V_t(\ell_t) \le \max\{\ell_t c_{t-1}, \omega_t\}$ for all $t = M, M - 1, \ldots, 0$. Indeed, the inequality is true for $t = M$. Assume it is true for $t + 1 \le M$, we will then prove it for t. Note that

$$V_t(\ell_t) = \max\{\ell_t c_{t-1}, (1 - \beta c_t) + E_\infty[V_{t+1}(\ell_{t+1})]\}$$

$$\le \max\{\ell_t c_{t-1}, (1 - \beta c_t) + E_\infty[\max\{\ell_{t+1} c_t, \omega_{t+1}\}]\}$$

$$= \max\{\ell_t c_{t-1}, (1 - \beta c_t) + c_t P_{0,t+1}(\ell_{t+1} \ge v_{t+1}) + \omega_{t+1} P_{\infty,t+1}(\ell_{t+1} < v_{t+1})\}$$

$$= \max\{\ell_t c_{t-1}, 1 + \omega_{t+1} P_{\infty,t+1}(\ell_{t+1} < v_{t+1})\} = \max\{\ell_t c_{t-1}, \omega_t\}.$$

The inequality above is due to the induction assumption; furthermore, in the last three equalities we used the definition of c_t, namely, $c_t = \omega_{t+1}/v_{t+1}$; the fact that by construction of the sequence $\{v_t\}$ we have $P_{0,t+1}(\ell_{t+1} \ge v_{t+1}) = \beta$; and we also used recursion (A.15). We thus conclude that $V_t(\ell_t) \le \max\{\ell_t c_{t-1}, \omega_t\}$. Applying it for $t = 0$ yields $V_0(\ell_0) \le \max\{\ell_0 c_{-1}, \omega_0\} = \omega_0 = \gamma$, because ℓ_0 is defined to be 0 and, as we argued, $\omega_0 = \gamma$. From optimal stopping theory we have $\hat{\mathcal{G}}(T_M) \le V_0(\ell_0)$; consequently, $\hat{\mathcal{G}}(T_M) \le \gamma$. Using this in (A.18) we obtain

$$\hat{\mathcal{G}}(T) \le \hat{\mathcal{G}}(T_M) + \epsilon \le \gamma + \epsilon,$$

which implies $\hat{\mathcal{G}}(T) \le \gamma$. Substituting in (A.17) and maximizing over ϖ, we have

$$\mathcal{G}(T) \le (1 - \varpi)\gamma \le \gamma,$$

with the optimum randomization becoming $\varpi = 0$. Using the definition of $\mathcal{G}(T)$ from (A.17) and the fact that we consider T with $\mathsf{E}_\infty[T] = \gamma$ we have

$$\gamma \geq \mathcal{G}(T) = \mathsf{E}_\infty[\ell_T c_{T-1}] - \beta \mathsf{E}_\infty\left[\sum_{t=0}^{T-1} c_t\right] + \mathsf{E}_\infty[T],$$

which directly implies

$$\frac{\mathsf{E}_\infty[\ell_T c_{T-1}]}{\mathsf{E}_\infty\left[\sum_{t=0}^{T-1} c_t\right]} \leq \beta.$$

Shewhart, by construction, is an equalizer; hence, we have $\mathcal{J}_\mathrm{L}(\mathscr{S}) = \beta$. From (A.16) and the previous inequality we can then write

$$\mathcal{J}_\mathrm{L}(\mathscr{S}) \leq \sup_{T:\mathsf{E}_\infty[T]\geq\gamma} \mathcal{J}_\mathrm{L}(T) \leq \sup_{T:\mathsf{E}_\infty[T]=\gamma} \frac{\mathsf{E}_\infty[\ell_T c_{T-1}]}{\mathsf{E}_\infty\left[\sum_{t=0}^{T-1} c_t\right]} \leq \beta = \mathcal{J}_\mathrm{L}(\mathscr{S}),$$

thus proving the desired optimality for Lorden's criterion. Similar proof applies in the case of Pollak's measure. □

Proof of Theorem 4.1.   When $q = 0$ or 1 then $\mathsf{P}_\infty(\ell_1^i \geq v)$ is continuous and strictly decreasing in v (see Footnote 2). If $q \in (0, 1)$ we observe

$$\mathsf{P}_\infty\big((1 - q)\ell_1^1 + q\ell_1^2 \geq v\big) = \int_0^\infty \mathsf{P}_\infty\left(\ell_1^1 \geq \frac{v - qs}{1 - q}\right) \mathsf{P}_\infty(\ell_1^2 \in ds). \qquad (\text{A.19})$$

Consequently, if we use the continuity and strict monotonicity with respect to v of the first probability under the integral and bounded convergence we can prove continuity and strict monotonicity of $\mathsf{P}_\infty((1 - q)\ell_1^1 + q\ell_1^2 \geq v)$ as a function of v for all $q \in [0, 1]$. This probability is equal to 1 and 0 for $v = 0$ and $v \to \infty$, respectively, therefore there exists unique $v(q) \geq 0$ that satisfies the false alarm constraint (4.4) with equality.

Consider now $v(q)$ as a function of q. We like to show that this function is continuous. Fix $q_0 \in (0, 1)$ then for $q \to q_0\pm$ we will show $v(q_0\pm) = v(q_0)$. Recall that $v(q)$ is constructed so that for all $q \in [0, 1]$ we have $\mathsf{P}_\infty((1 - q)\ell_1^1 + q\ell_1^2 \geq v(q)) = 1/\gamma$. Taking the limit with respect to $q \to q_0\pm$ and using (A.19) we have

$$\frac{1}{\gamma} = \lim_{q \to q_0\pm} \int_0^\infty \mathsf{P}_\infty\left(\ell_1^1 \geq \frac{v(q) - qs}{1 - q}\right) \mathsf{P}_\infty(\ell_1^2 \in ds)$$

$$= \int_0^\infty \mathsf{P}_\infty\left(\ell_1^1 \geq \frac{v(q_0\pm) - q_0 s}{1 - q_0}\right) \mathsf{P}_\infty(\ell_1^2 \in ds) = \mathsf{P}_\infty\big((1 - q_0)\ell_1^1 + q_0\ell_1^2 \geq v(q_0\pm)\big),$$

where for the second equality we used bounded convergence and the continuity of the c.d.f. of ℓ_1^1. Since $\mathsf{P}_\infty((1 - q_0)\ell_1^1 + q_0\ell_1^2 \geq v(q_0\pm)) = 1/\gamma$, but also from the definition of $v(q_0)$ that $\mathsf{P}_\infty((1 - q_0)\ell_1^1 + q_0\ell_1^2 \geq v(q_0)) = 1/\gamma$, we can claim that $v(q_0\pm) = v(q_0)$. The claim is valid since for each q, as we argued before, the threshold that satisfies the false alarm constraint with equality is unique. Similar proof (with one-sided limits) applies for $q_0 = 0, 1$.

Let us now prove the validity of our theorem when the condition of case (i) is true. We have

$$\mathcal{J}_L(T) = \min_{i=1,2} \inf_{t \geq 0} \text{ess inf } P_t^i(T = t + 1 | \mathcal{F}_t, T > t) \leq \inf_{t \geq 0} \text{ess inf } P_t^1(T = t + 1 | \mathcal{F}_t, T > t)$$

$$\leq \inf_{t \geq 0} \text{ess inf } P_t^1(\mathcal{S}(0) = t + 1 | \mathcal{F}_t, \mathcal{S}(0) > t) = P_0^1(\ell_1 \geq v(0)), \qquad (A.20)$$

where the second inequality comes from the fact that $\mathcal{S}(0)$ is optimum when the postchange probability measure is P_0^1, and the last equality is the result of $\mathcal{S}(0)$ being an equalizer under P_0^1. Note now that

$$P_0^1(\ell_1 \geq v(0)) \leq P_0^2(\ell_1 \geq v(0)) = \inf_{t \geq 0} \text{ess inf } P_t^2(\mathcal{S}(0) = t + 1 | \mathcal{F}_t, \mathcal{S}(0) > t),$$

the inequality being the condition of case (i) and the equality that follows is the result of $\mathcal{S}(0)$ being an equalizer under P_0^2 as well. Completing what was started in (A.20), we can write

$$\mathcal{J}_L(T) \leq P_0^1(\ell_1 \geq v(0)) = \min_{i=1,2} P_0^i(\ell_1 \geq v(0))$$

$$= \min_{i=1,2} \inf_{t \geq 0} \text{ess inf } P_t^i(\mathcal{S}(0) | \mathcal{F}_t, \mathcal{S}(0) > t) = \mathcal{J}_L(\mathcal{S}(0)),$$

which proves the claim of case (i). Similar proof applies in case (ii).

Suppose now that neither the condition of case (i) nor of case (ii) is valid. This suggests that we simultaneously have $P_0^2(\ell_1^1 \geq v(0)) < P_0^1(\ell_1^1 \geq v(0))$ and $P_0^1(\ell_1^2 \geq v(1)) < P_0^2(\ell_1^2 \geq v(1))$. Define the following difference as a function of q:

$$D(q) = P_0^1((1 - q)\ell_1^1 + q\ell_1^2 \geq v(q)) - P_0^2((1 - q)\ell_1^1 + q\ell_1^2 \geq v(q)).$$

We observe that $D(0) > 0$ and $D(1) < 0$; furthermore, $D(q)$ is continuous because we can show using (A.19) and the continuity of $v(q)$ that the probabilities $P_0^i((1 - q)\ell_1^1 + q\ell_1^2 \geq v(q))$ are continuous in q. Hence, there exists $q \in (0, 1)$ so that $D(q) = 0$. For this specific q the corresponding Shewhart stopping rule $\mathcal{S}(q)$ is by construction an equalizer across time and across postchange probabilities. Furthermore, for each T and $t \geq 0$, as in (A.12), we have

$$\mathcal{J}_L(T)P_\infty(T > t) \leq P_t^i(T = t + 1); i \quad = 1, 2$$

suggesting

$$\mathcal{J}_L(T)P_\infty(T > t) \leq (1 - q)P_t^1(T = t + 1) + qP_t^2(T = t + 1)$$

$$= E_\infty[\{(1 - q)\ell_{t+1}^1 + q\ell_{t+1}^2\} \mathbb{1}_{\{T = t+1\}}].$$

Summing over $t \geq 0$ we obtain the following upper bound:

$$\mathcal{J}_L(T) \leq \frac{E_\infty[(1 - q)\ell_T^1 + q\ell_T^2]}{E_\infty[T]}.$$

The proof continues along the same lines of the proof of Theorem 2.3. Basically we show that the upper bound is optimized by $\mathcal{S}(q)$; furthermore, this optimal value

is also attained by $\mathscr{J}_{\mathrm{L}}(\mathscr{S}(q))$ because $\mathscr{S}(q)$ is an equalizer across time and across postchange probabilities. This establishes the desired optimality for $\mathscr{S}(q)$.

What is now left to demonstrate is that for each γ, only one of the three cases can be valid. Call

$$\mathscr{J}_{\mathrm{L}}^i(T) = \inf_{t \geq 0} \operatorname{ess\,inf} \mathsf{P}_t^i(T = t + 1 | \mathscr{F}_t, T > t); i \quad = 1, 2,$$

then we know that $\mathscr{J}_{\mathrm{L}}^1(T)$ is maximized by $\mathscr{S}(0)$ and $\mathscr{J}_{\mathrm{L}}^2(T)$ by $\mathscr{S}(1)$. In fact no other stopping time can attain the same optimal value unless it is equal, with probability 1, to the corresponding Shewhart test. If case (i) applies then we will show that it is not possible the condition of case (ii) to be true. Indeed, if both conditions were valid simultaneously, then we could write

$$\mathscr{J}_{\mathrm{L}}^1\big(\mathscr{S}(0)\big) = \mathsf{P}_0^1\big(\ell_1^1 \geq v(0)\big) \leq \mathsf{P}_0^2\big(\ell_1^1 \geq v(0)\big) = \mathscr{J}_{\mathrm{L}}^2\big(\mathscr{S}(0)\big)$$
$$\leq \mathscr{J}_{\mathrm{L}}^2\big(\mathscr{S}(1)\big) = \mathsf{P}_0^2\big(\ell_2^1 \geq v(1)\big) \leq \mathsf{P}_0^1\big(\ell_2^1 \geq v(1)\big) = \mathscr{J}_{\mathrm{L}}^1\big(\mathscr{S}(1)\big),$$

where the first inequality comes from case (i), the second inequality from the fact that $\mathscr{S}(1)$ optimizes $\mathscr{J}_{\mathrm{L}}^2(T)$, and the third inequality is the condition of case (ii). From the above we conclude that $\mathscr{S}(1)$ has a better $\mathscr{J}_{\mathrm{L}}^1(\cdot)$ performance than $\mathscr{S}(0)$, which optimizes $\mathscr{J}_{\mathrm{L}}^1(\cdot)$, leading to contradiction. Actually, since $\mathscr{S}(1)$ is not equal to $\mathscr{S}(0)$ with probability 1, its corresponding performance is strictly smaller than the optimum.

Similarly, it is not possible to have the conditions of case (i) and case (iii) satisfied at the same time. Again, if this were true then

$$\mathscr{J}_{\mathrm{L}}^1\big(\mathscr{S}(0)\big) = \mathsf{P}_0^1\big(\ell_1^1 \geq v(0)\big) \leq (1 - q)\mathsf{P}_0^1\big(\ell_1^1 \geq v(0)\big) + q\mathsf{P}_0^2\big(\ell_1^1 \geq v(0)\big)$$
$$= (1 - q)\mathscr{J}_{\mathrm{L}}^1\big(\mathscr{S}(0)\big) + q\mathscr{J}_{\mathrm{L}}^2\big(\mathscr{S}(0)\big)$$
$$\leq (1 - q)\mathscr{J}_{\mathrm{L}}^1\big(\mathscr{S}(q)\big) + q\mathscr{J}_{\mathrm{L}}^2\big(\mathscr{S}(q)\big) = \mathscr{J}_{\mathrm{L}}^1\big(\mathscr{S}(q)\big),$$

with the first inequality due to case (i) and the second due to the fact that the convex combination of the two measures is maximized by $\mathscr{S}(q)$. Finally, the last equality is true because of case (iii), namely, that the stopping time $\mathscr{S}(q)$ is an equalizer for the two postchange measures. Again this is a contradiction since $\mathscr{S}(q)$ has larger $\mathscr{J}_{\mathrm{L}}^1(\cdot)$ measure than $\mathscr{S}(0)$, which is the optimum. Therefore, case (i) and case (iii) cannot be valid at the same time. Similarly, we can show that case (ii) and case (iii) are incompatible.

Since we have shown that when neither case (i) nor case (ii) is valid, we necessarily have case (iii) being true; this suggests that, for each value of γ, exactly one of the three cases applies. This concludes the proof for Lorden's criterion. Similar proof applies in the case of Pollak's measure. □

Proof of Theorem 4.2. Let case (i) be true, then we can write

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

$$\geq \sup_{t \geq 0} \operatorname{ess\,sup} \mathsf{E}_t^1[\mathscr{S}(0) - t | \mathscr{F}_t, \mathscr{S}(0) > t] = \mathsf{E}_0^1[\mathscr{S}(0)] = \frac{1}{\mathsf{P}_0^1(\ell_1 \geq v(0))},$$

$$(\mathrm{A.21})$$

where the first inequality is obvious and the second comes from the fact that if $v(0) \leq 1$ then the Shewhart stopping time $\mathscr{S}(0)$, according to Section 1.2, optimizes Lorden's original criterion for the postchange probability measure P_0^1. The second last equality comes from the fact that Shewhart, exactly as CUSUM, is an equalizer and the last equality is true due to (1.10). We also have

$$\sup_{t \geq 0} \operatorname{ess\,sup} \mathsf{E}_t^2[\mathscr{S}(0) - t | \mathscr{F}_t, \mathscr{S}(0) > t] = \mathsf{E}_0^2[\mathscr{S}(0)] = \frac{1}{P_0^2(\ell_1 \geq v(0))},$$

because $\mathscr{S}(0)$ is an equalizer under P_0^2 as well. Since by assumption, $P_0^1(\ell_1 \geq v(0)) \leq P_0^2(\ell_1 \geq v(0))$ this suggests that $\mathscr{J}_L(\mathscr{S}(0)) = \max_{i=1,2} 1/P_0^i(\ell_1 \geq v(0)) = 1/P_0^1(\ell_1 \geq v(0))$. Using this last observation in (A.21) we conclude that $\mathscr{J}_L(T) \geq \mathscr{J}_L(\mathscr{S}(0))$, thus proving optimality of $\mathscr{S}(0)$. In a similar way, we can prove optimality for $\mathscr{S}(1)$ under the condition of case (ii).

Assume now that we are in case (iii), then

$$\mathscr{J}_L(T) \geq \mathsf{E}_t^i[T - t | \mathscr{F}_t, T > t] = \mathsf{E}_t^i\left[\sum_{n=t}^{\infty} \mathbb{1}_{\{T>n\}} | \mathscr{F}_t, T > t\right] = \sum_{n=t}^{\infty} \mathsf{E}_t^i[\mathbb{1}_{\{T>n\}} | \mathscr{F}_t, T > t]$$

$$= \sum_{n=t}^{\infty} \mathsf{E}_{\infty}\left[\mathbb{1}_{\{T>n\}} \prod_{m=t+1}^{n} \ell_m^i | \mathscr{F}_t, T > t\right] = \mathsf{E}_{\infty}\left[\sum_{n=t}^{T-1} \prod_{m=t+1}^{n} \ell_m^i | \mathscr{F}_t, T > t\right],$$

where we applied a change of measures and used the fact that $\{T > n\}$ is $\mathscr{F}_n$-measurable. We also define $\prod_a^b = 1$ and $\sum_a^b = 0$ when $b < a$ while we recall that ℓ_0^i is defined to be 0. Multiplying both sides of the previous inequality with $\mathbb{1}_{\{T>t\}}(1 - \ell_t^i)^+$, which is nonnegative and $\mathscr{F}_t$-measurable and taking expectation with respect to the nominal measure, we obtain

$$\mathscr{J}_L(T)\mathsf{E}_{\infty}[\mathbb{1}_{\{T>t\}}(1 - \ell_t^i)^+] \geq \mathsf{E}_{\infty}\left[\sum_{n=t}^{T-1} \mathbb{1}_{\{T>t\}} \prod_{m=t+1}^{n} \ell_m^i(1 - \ell_t^i)^+\right]$$

$$\geq \mathsf{E}_{\infty}\left[\sum_{n=t}^{T-1} \mathbb{1}_{\{T>t\}} \prod_{m=t+1}^{n} \ell_m^i(1 - \ell_t^i)\right]$$

$$= \mathsf{E}_{\infty}\left[\sum_{n=t}^{T-1} \mathbb{1}_{\{T>t\}} \left(\prod_{m=t+1}^{n} \ell_m^i - \prod_{m=t}^{n} \ell_m^i\right)\right].$$

Summing over all $t \geq 0$ and recalling that $\ell_0^i = 0$, $\prod_{n+1}^n = 1$, yields

$$\mathscr{J}_L(T)\mathsf{E}_{\infty}\left[\sum_{t=0}^{T-1}(1 - \ell_t^i)^+\right] \geq \mathsf{E}_{\infty}\left[\sum_{t=0}^{T-1}\sum_{n=t}^{T-1} \left(\prod_{m=t+1}^{n} \ell_m^i - \prod_{m=t}^{n} \ell_m^i\right)\right]$$

$$= \mathsf{E}_{\infty}\left[\sum_{n=0}^{T-1}\sum_{t=0}^{n} \left(\prod_{m=t+1}^{n} \ell_m^i - \prod_{m=t}^{n} \ell_m^i\right)\right] = \mathsf{E}_{\infty}\left[\sum_{n=0}^{T-1} 1\right] = \mathsf{E}_{\infty}[T].$$

Finally, multiplying the previous inequality for $i = 1$ with $(1 - q)$ and the one for $i = 2$ with q and adding the resulting expressions we obtain the following lower bound:

$$\mathscr{J}_L(T) \geq \frac{\mathsf{E}_{\infty}[T]}{\mathsf{E}_{\infty}\left[\sum_{t=0}^{T-1}(1 - q)(1 - \ell_t^1)^+ + q(1 - \ell_t^2)^+\right]}.$$

Following the usual methodology we have adopted in the previous proofs, in order to minimize the lower bound, with the help of the randomization probability ϖ we can show that we can limit ourselves to stopping times that satisfy the false alarm constraint with equality. Consequently,

$$\inf_{T:\mathsf{E}_\infty[T]\geq\gamma} \mathcal{J}_\mathrm{L}(T) \geq \inf_{T:\mathsf{E}_\infty[T]=\gamma} \frac{\mathsf{E}_\infty[T]}{\mathsf{E}_\infty\left[\sum_{t=0}^{T-1}(1-q)(1-\ell_t^1)^+ + q(1-\ell_t^2)^+\right]}$$

$$= \frac{\gamma}{\sup_{T:\mathsf{E}_\infty[T]=\gamma}\mathsf{E}_\infty\left[\sum_{t=0}^{T-1}(1-q)(1-\ell_t^1)^+ + q(1-\ell_t^2)^+\right]}. \quad (A.22)$$

For simplicity denote $z_t = (1-q)(1-\ell_t^1)^+ + q(1-\ell_t^2)^+$, then maximizing the denominator subject to the equality constraint is straightforward. Using a Lagrange multiplier with value

$$\lambda = (1-v)\mathsf{P}_\infty(z_1 < 1-v) + \mathsf{E}_\infty[z_1 \mathbb{1}_{\{z_1 \geq 1-v\}}]$$

and applying standard optimal stopping theory, we can conclude that the optimum stopping time is

$$\mathcal{T} = \inf\{t > 0 : z_t \leq 1-v\}.$$

For $v = v(q)$ we will show that $\mathcal{T}$ is in fact equivalent to $\mathcal{S}(q)$ under condition (4.8). Indeed, notice that when $\mathcal{T}$ stops we have

$$1 - v(q) \geq z_{\mathcal{T}} = (1-q)(1-\ell_{\mathcal{T}}^1)^+ + q(1-\ell_{\mathcal{T}}^2)^+ \geq (1-q)(1-\ell_{\mathcal{T}}^1) + q(1-\ell_{\mathcal{T}}^2),$$

which implies

$$(1-q)\ell_{\mathcal{T}}^1 + q\ell_{\mathcal{T}}^2 \geq v(q),$$

suggesting $\mathcal{S}(q) \leq \mathcal{T}$ (because $\mathcal{S}(q)$ is the *first* time instant the above inequality is true). For any $t < \mathcal{T}$ we have

$$1 - v(q) < z_t = (1-q)(1-\ell_t^1)^+ + q(1-\ell_t^2)^+. \quad (A.23)$$

Because $v(q)$ satisfies (4.8) we will show that the previous inequality can be true only when $\xi_t \in \mathcal{A}_1 \cap \mathcal{A}_2$; that is, when the likelihood ratios ℓ_t^1 and ℓ_t^2 are simultaneously no larger than 1. Indeed, from (4.8) we have that the upper bound of $v(q)$ is no larger than 1; consequently, in (A.23) the two likelihood ratios cannot be larger than 1 simultaneously. Let $\ell_t^1 \leq 1$ and $\ell_t^2 > 1$ then (A.23) becomes $1 - v(q) < (1-q)(1-\ell_t^1)$ or $q + (1-q)\ell_t^1 < v(q)$. But the latter is again not possible because of the left-hand side inequality of (4.8). The same is true when $\ell_t^2 \leq 1$ and $\ell_t^1 > 1$. Hence, (A.23) can be valid only when both likelihood ratios are smaller than 1. This means that when $t < \mathcal{T}$, (A.23) is equivalent to

$$(1-q)\ell_t^1 + q\ell_t^2 < v(q).$$

This observation suggests that $t < \mathcal{T}$ combined with (4.8) implies $t < \mathcal{S}(q)$; therefore, $\mathcal{T} - 1 < \mathcal{S}(q)$ or $\mathcal{S}(q) \geq \mathcal{T}$. Consequently, $\mathcal{S}(q) = \mathcal{T}$, which means that $\mathcal{S}(q)$ optimizes the lower bound in (A.22).

To compute the optimum value of the lower bound, since before stopping, as we pointed out, both likelihood ratios are no larger than 1, we note

$$\mathsf{E}_\infty\left[\sum_{t=0}^{\mathscr{S}(q)-1}(1-q)(1-\ell_t^1)^+ + q(1-\ell_t^2)^+\right]$$

$$= \gamma - \mathsf{E}_\infty\left[\sum_{t=0}^{\mathscr{S}(q)-1}(1-q)\ell_t^1 + q\ell_t^2\right]$$

$$= \gamma\{(1-q)\mathsf{P}_0^1((1-q)\ell_t^1 + q\ell_t^2 \geq v(q)) + q\mathsf{P}_0^2((1-q)\ell_t^1 + q\ell_t^2 \geq v(q))\}$$

$$= \gamma\mathsf{P}_0^1((1-q)\ell_t^1 + q\ell_t^2 \geq v(q)) = \gamma\mathsf{P}_0^2((1-q)\ell_t^1 + q\ell_t^2 \geq v(q)),$$

where we used the fact that $\gamma = \mathsf{E}_\infty[\mathscr{S}(q)] = 1/\mathsf{P}_\infty((1-q)\ell_t^1 + q\ell_t^2 \geq v(q))$ and that we are in case (iii) with condition (4.5) being valid. Consequently,

$$\inf_{T:\mathsf{E}_\infty[T]\geq\gamma} \mathscr{J}_{\mathrm{L}}(T) \geq \frac{1}{\mathsf{P}_0^i\big((1-q)\ell_t^1 + q\ell_t^2 \geq v(q)\big)}.$$

Now it is straightforward to verify that the lower bound is attainable by the Lorden measure of the Shewhart stopping time $\mathscr{S}(q)$. This is clearly due to the fact that $\mathscr{S}(q)$ is an equalizer across time and across postchange measures. This concludes our proof. $\qquad\square$

REFERENCES

Bojdecki, T. (1979). Probability Maximizing Approach to Optimal Stopping and Its Application to a Disorder Problem, *Stochastics* 3: 61–71.

Dragalin, V. (1997). The Design and Analysis of 2-CUSUM Procedure, *Communications in Statistics-Simulation and Computation* 26: 67–81.

Guépié, B. K., Fillatre, L., and Nikiforov, I. (2012). Sequential Detection of Transient Changes, *Sequential Analysis* 31: 528–547.

Hadjiliadis, O. and Moustakides, G. V. (2006). Optimal and Asymptotically Optimal CUSUM Rules for Change Point Detection in the Brownian Motion Model with Multiple Alternatives, *Theory of Probability and Its Applications* 50: 75–85.

Hadjiliadis, O. and Poor, H. V. (2009). On the Best 2-CUSUM Stopping Rule for Quickest Detection of Two-Sided Alternatives in a Brownian Motion Model, *Theory of Probability and Its Applications* 53: 537–547.

Lorden, G. (1971). Procedures for Reacting to a Change in Distribution, *Annals of Mathematical Statistics* 42: 1897–1908.

Moustakides, G. V. (1986). Optimal Stopping Times for Detecting Changes in Distributions, *Annals of Statistics* 14: 1379–1387.

Moustakides, G. V. (2008). Sequential Change Detection Revisited, *Annals of Statistics* 36: 787–807.

Pollak, M. (1985). Optimal Detection of a Change in Distribution, *Annals of Statistics* 13: 206–227.

Pollak, M. and Krieger, A. M. (2013). Shewhart Revisited, *Sequential Analysis* 32: 230–242.

Poor, H. V. (1998). Quickest Detection with Exponential Penalty for Delay, *Annals of Statistics* 26: 2179–2205.

Sarnowski, W. and Szajowski, K. (2011). Optimal Detection of Transition Probability Change in Random Sequence, *Stochastics* 86: 569–581.

Shewhart, W. A. (1931). *Economic Control of Quality of Manufactured Product*, New York: Van Nostrand.

Shiryaev, A. N. (1963). On Optimum Methods in Quickest Detection Problems, *Theory of Probability and Its Applications* 8: 22–46.

Tartakovsky, A. G., Pollak M., and Polunchenko A. S. (2012). Third-Order Asymptotic Optimality of the Generalized Shiryaev-Roberts Changepoint Detection Procedures, *Theory of Probability and Its Applications* 56: 457–484.