Methodology and Computing in Applied Probability

# MODEL MISSPECIFICATION IN DISCRETE TIME BAYESIAN ONLINE CHANGE DETECTION

--Manuscript Draft--

| | |
|---|---|
| **Manuscript Number:** | MCAP-D-22-00119 |
| **Full Title:** | MODEL MISSPECIFICATION IN DISCRETE TIME BAYESIAN ONLINE CHANGE DETECTION |
| **Article Type:** | Manuscript |
| **Keywords:** | Sequential analysis; Bayesian quickest detection; optimal stopping |
| **Corresponding Author:** | Semih Onur Sezer<br>Sabanci University: Sabanci Universitesi<br>Istanbul, TURKEY |
| **Corresponding Author Secondary Information:** | |
| **Corresponding Author's Institution:** | Sabanci University: Sabanci Universitesi |
| **Corresponding Author's Secondary Institution:** | |
| **First Author:** | Savas Dayanik |
| **First Author Secondary Information:** | |
| **Order of Authors:** | Savas Dayanik |
| | Semih Onur Sezer |
| **Order of Authors Secondary Information:** | |
| **Funding Information:** | |

# MODEL MISSPECIFICATION IN DISCRETE TIME BAYESIAN ONLINE CHANGE DETECTION

SAVAS DAYANIK[1] and SEMIH ONUR SEZER[2]

**Abstract.** We revisit the classical formulation of the discrete time Bayesian online change detection problem in which the common distribution of an observed sequence of random variables changes at an unknown point in time. The objective is to detect the change with a stopping time of the observations and minimize a given Bayes risk. When the change time has a zero-modified geometric prior distribution, the first crossing time of the odds-ratio process over a threshold is known to be an optimal solution. In the current paper, we consider a modeler who misspecifies some of the elements of this formulation. Because of this misspecification, the modeler computes a wrong stopping threshold and follows an incorrect odds-ratio process in implementation. To find her actual Bayes risk, which is different from the value function evaluated with the wrong choices, one needs to compute the expected costs accumulated by the true odds-ratio process until modeler's odd-ratio process crosses this wrong boundary. In the paper, we carry out these computations in the extended state space of both processes, and we illustrate these computations on examples. In those examples, we construct tolerance regions for the parameters to be estimated by the modeler. For a given choice by the modeler, the tolerance region is the set of true values for which her relative loss is less than or equal to a predetermined level.

*Keywords:* Sequential analysis; Bayesian quickest detection; optimal stopping.
*2010 Mathematics Subject Classification:* Primary 62L10; Secondary 62L15, 62C10, 60G40.

**Declarations for the journal:**
*Data availability statement:* Current manuscript does not use any dataset.
*Conflict of interest statement:* Authors do not have any conflict of interest.
*Grant/funding information:* No funding has been received for the current work.

## 1. Introduction

Detecting an unobservable change in the probability law of a random sequence is one of the fundamental problems in sequential analysis. Parallel to its rich and diverse application areas, there has been a well-developed literature studying various models with both Bayesian and non-Bayesian aprooaches. We invite the reader to refer to Shiryaev [40, 43], Lorden [25], Zacks [53], Moustakides [29], Basseville and Nikiforov [2], Poor and Hadjiliadis [36], Tartakovsky [46, 48], Dayanik et al. [12], Tartakovsky and Moustakides [50], Polunchenko and Tartakovsky [34], Tartakovsky et al. [47] and their references for a comprehensive treatment of these models and their applications. For some of the latest contributions, Moustakides [30], Fuh and Tartakovsky [19], Tartakovsky [49], Zou et al. [55], Dayanik and Kazutoshi [15], Xu et al. [51] and the references therein can be consulted.

As noted in some of the references above, the main focus in the existing literature is either on finding optimal solutions for relatively simpler formulations or bringing easily implementable solutions with certain desirable (asymptotic) properties for more realistic and difficult formulations. Even though model misspecification issues are of utmost importance from an implementation point of view, they have been discussed in only a small number of papers so far. The exceptions in the non-Bayesian framework include Srivastava and Wu [45], Pollak and Siegmund [33], Molloy and Timothy [27], and Du et al. [16] which

[1]Bilkent University, Industrial Engineering Department, 06800 Bilkent, Ankara, Turkey
E-mail: sdayanik@bilkent.edu.tr

[2](Corresponding author) Sabancı University, Faculty of Engineering and Natural Sciences, 34956 Tuzla, Istanbul, Turkey
E-mail: sezer@sabanciuniv.edu

bring theoretical and numerical results for the performance of a modeler/decision maker who misspecifies especially the post-change probability law of the observations. On the Bayesian side, Ekström and Vaicenavicius [17, Section 3] and Blanchet-Scalliet et al. [7] (see also Blanchet-Scalliet et al. [6]) discuss certain misspecification issues in continuous time settings with Brownian observations. Ekström and Vaicenavicius [17, Section 3] brings certain error bounds for the detection problem under some ordering assumptions, and Blanchet-Scalliet et al. [7] numerically compares in a utility maximization problem (with misspecified parameters) the performance of a detection based investment strategy against those of others. The exact computation of modeler's Bayes risk and her error have never been investigated before in any Bayesian model. This is the contribution of the current paper in the well-known discrete time setting studied in Shiryaev [42, 39]. Similar analyses could also be provided for the analogous continuous time formulations with Brownian motions and (compound) Poisson processes. These analyses require different techniques, and they will be presented in subsequent papers.

The notion of model misspecification generally refers to setting up certain parameters or distributions incorrectly. It does not exclude the existence of parameters or distributions unknown to a modeler/practitioner. Such formulations are given in the literature by means of nuisance parameters or variables; see Hadjiliadis [22], Hadjiliadis and Moustakides [23], Bayraktar et al. [4], Dayanik and Goulding [11], Banerjee and Veeravalli [1], Muravlev and Shiryaev [31], Yang et al. [52] for some examples. In Bayesian formulations, nuisance variables have their own prior distributions, and these distributions may be determined incorrectly. In non-Bayesian settings, on the other hand, unknown parameters may be replaced with their (maximum likelihood) estimates, and this may be based on wrong assumptions on the distributions or uncertainy sets. Overall, misspecifications may exist in any type of model irrespective of how simple or complex it may be.

Studying the consequences of a misspecification in a detection problem requires more than carrying out a sensitivity analysis. In a sensitivity analysis, we investigate how the optimal solution is affected by a change in the parameters/distributions. This does not tell us the magnitude of a loss because of selecting them incorrectly. Even though we still need to solve the problem with modeler's choices to derive her detection policy, additional work is needed to compute the resulting loss.

In the current paper, we address this issue in the classical Bayesian model studied in Shiryaev [42, 39] (see Chapter 4.3 in the latter). As our analysis illustrates, dealing with a misspecification is non-trivial even in this relatively simple setting. The model assumes that we observe a sequence of random variables $X_1, X_2, \dots$ having a common distribution $\nu_0$. At some unobservable random time $\Theta$ having a zero-modified geometric prior

$$\mathbb{P}\{\Theta = 0\} = \pi \quad \text{and} \quad \mathbb{P}\{\Theta = n\} = (1-\pi)(1-p)^{n-1}p, \qquad n \geq 1, \tag{1}$$

for some $\pi \in [0, 1)$ and $p \in (0, 1)$, the common distribution changes to some other distribution $\nu_1$. The objective is to find a stopping time of our observations which will detect the change as accurately as possible by minimizing a Bayes risk. When the Bayes risk is a linear combination of the expected delay in detection and the probability of a false alarm, the analyses in Shiryaev [42, 39] show that the first crossing time of the *posterior probability process* $\Pi$ over some critical threshold $\pi^*$ is an optimal alarm time. This stopping time can equivalently be expressed as the first crossing time of the *odds-ratio process* $\Phi = \Pi/(1-\Pi)$ over the threshold $\phi^* = \pi^*/(1-\pi^*)$. In our main analysis in the paper, we use the latter representation since we work with an auxiliary probability measure under which the detection problem translates into an optimal stopping problem for the odds-ratio process.

When it comes to implementation, a modeler has to estimate/set the parameters $\pi$, $p$ and the distributions $\nu_0$, $\nu_1$ in the formulation above. To our best knowledge, there is no single approach that applies to all possibly different applications, and some ad-hoc methods may be relied upon depending on the specific situation under consideration. Estimation based on past data (if available), asking experts' opinions, running candidate detection algorithms (for each possible parameter set) in parallel, and even some simple trial and error are among the available options. For example, as in Shiryaev et al. [41, 44], Zhitlukhin and Ziemba [54], and Buonaguidi et al. [8] (studying different models), if the modeler has a clear idea on when the window for a disorder opens, then she may simply set $\pi = 0$. We also make this choice in our numerical examples in Section 4. Buonaguidi et al. [8] studies a credit card fraud detection problem using a continuous time model with compound Poisson arrivals. Authors set the arrival rate of the fraudulent

attacks by expert opinion in a way to match an average for the number of attacks per year (per cardholder). In Shiryaev et al. [41], a financial application (to close a market position) is considered using a discrete time adaptation of a continuous time formulation. Authors consider different values for pre- and post-disorder drift parameters and compare the performances of these choices on the available data. This approach of trying different combinations is also used in Shiryaev et al. [44] and Zhitlukhin and Ziemba [54] for similar applications.

All these methods above and others are naturally prone to errors, and it is important to compute the resulting Bayes risks in order to understand how severe these mistakes can be, should the true parameters/distributions be different. A modeler with incorrect parameters and/or distributions obtains a wrong stopping threshold $\widetilde{\phi}^*$ from her model, and she evaluates a different odds-ratio process $\widetilde{\Phi}$. Therefore, she applies a sub-optimal stopping rule in practice. Finding her Bayes risk requires computing the expected costs incurred in the true model by the true odds-ratio process until this incorrect stopping time. In the paper, we show how the resulting Bayes risk can be computed in the extended state space of both processes after noting that they are jointly Markov and therefore standard arguments of dynamic programming apply. Using the observation that the true odds-ratio $\Phi$ is linearly dependent on its initial value, we reduce these computations to a pair of one dimensional function iterations. We give the details in Sections 2 and 3, and we illustrate the computations in Section 4 on some examples.

In practice, a modeler usually does not know all parameters and distributions exactly. Our analysis here does not provide a method for selecting them either. In fact, there can be multiple methods as indicated above. Regardless of the method(s) she prefers, the modeler can use the results of this paper to compute her potential error against alternative candidates. In particular, she can construct tolerance ranges/intervals for the parameters to be estimated. Given modeler's estimates, the $\alpha\%$ tolerance region denotes the set of true values for which her percentage loss

$$\frac{\text{Modeler's Bayes risk} - \text{True minimal Bayes risk}}{\text{True minimal Bayes risk}} \cdot 100 \tag{2}$$

is less than or equal to a given value $\alpha$. If this region is relatively large or includes other potential candidates, this gives the modeler some confidence for implementation. In that respect, the computation of the modeler's Bayes risk and the construction of such tolerance regions are the contributions of the current paper from a practical point of view. In Section 4, we construct such regions and intervals in some of our numerical examples.

Clearly, some or all of the modeler's choices may be incorrect in implementation. We do not make any specific assumption regarding which of her parameters/distributions match the correct ones (except that at least one of her choices is incorrect). We assume, however, that her distributions $\widetilde{\nu}_0$ and $\widetilde{\nu}_1$ are respectively equivalent (as measures) to true distributions $\nu_0$ and $\nu_1$. That is, we exclude the problematic and rather unexpected cases in practice where at each observation the modeler may realize with a positive probability that her assumptions are wrong. Dealing with such a model requires further assumptions on the actions taken by the modeler upon such an observation. In particular, one should specify how the distributions $\widetilde{\nu}_0$ and $\widetilde{\nu}_1$ would be revised in this case. Posterior probabilities should be re-calculated accordingly, and a penalty term may be added to the Bayes risk for changing/updating the modeling assumptions. We leave the study of the effects of various updating rules in different models/applications as future research.

As for the true distributions $\nu_0$ and $\nu_1$, in Sections 2 and 3, we give our analysis for the case where $\nu_1$ is absolutely continuous with respect to $\nu_0$. Many interesting applications in practice satisfy this condition; see also our examples in Section 4. The absolute continuity of the post-change distribution allows us to restate the problem as an optimal stopping problem for the odds-ratio process under an auxiliary probability measure as mentioned above. This approach of changing the measure and working with the odds-ratio has the major advantages that i) the computations are relatively simpler since all observations have the distribution $\nu_0$ under this measure, ii) joint Markov property of the pair $(\Phi, \widetilde{\Phi})$ under the new measure and with the observation filtration follows naturally, and iii) linear dependence of the modeler's Bayes risk on the initial point of the process $\Phi$ is an immediate consequence of the linear dependence of this process on this point. Therefore, our main analysis in the paper is based on the convenient and reasonably realistic assumption of absolute continuity of $\nu_1$ with respect to $\nu_0$. In Appendix B at the end, we complete the picture, and we give a short analysis under the physical measure without this assumption. Under the

physical measure, we work with the correct conditional probability process $\Pi$ and the incorrect one $\widetilde{\Pi}$ used by the modeler. The process $\Pi$ does not depend on its starting point linearly, and the linearity of the Bayes risk on this point is proved inductively instead.

The rest of the paper is organized as follows. In Section 2, we remind the reader the formulation of the detection problem, and we summarize its solution under our auxiliary probability measure. Results in Section 2 are already known; most of them are simply analogous to the corresponding results given in Shiryaev [42, 39] under the physical measure. We include that section for completeness and convenience of the reader as the section sets the notation for the rest of the paper. Section 3 gives our main contribution. In that section, we assume that the modeler selects some of the parameters and/or distributions incorrectly, and we compute her Bayes risk. In Section 4, we illustrate the results of Sections 2 and 3 numerically in three settings with different pre- and post-disorder distributions. In Section 5, we conclude our discussion with final remarks and potential extensions. Appendices at the end include supplementary analyses and proofs.

## 2. Detection problem with correct parameters/distributions

In the formulation below, the observations take values on some general measurable space. In Section 4, we give examples with real-valued observations.

### 2.1. **Formulation of the problem.**

Consider a probability space $(\Omega, \mathcal{H}, \mathbb{P})$ hosting a sequence of observation random variables $X_1, X_2, \dots$ taking values on some measurable space $(E, \mathcal{E})$, and also a non-negative integer valued random variable $\Theta$ representing our change time. The random variable $\Theta$ has the zero-modified geometric prior distribution given in (1) above. At time $\Theta$, common distribution of random variables $X_1, X_2, \dots$ changes from $\nu_0$ to $\nu_1$. This means that, conditioned on $\Theta = n$, the observations $X_1, \dots, X_{n-1}$ are independent and identically distributed according to $\nu_0$, and $X_n, X_{n+1}, \dots$ are independent with distribution $\nu_1$. The post-disorder probability distribution $\nu_1$ is absolutely continuous with respect to the pre-change distribution $\nu_0$.

Let $\mathbb{F} = (\mathcal{F}_n)_{n\in\mathbb{N}}$ be the filtration generated by the sequence $X_1, X_2, \dots$ over time. Our objective is to find an $\mathbb{F}$-stopping time (i.e., an alarm time) $\tau$ which will detect the change optimally in the sense that it will minimize the Bayes risk

$$R_\tau(\pi) := c\,\mathbb{E}(\tau - \Theta)^+ + \mathbb{P}(\tau < \Theta) \tag{3}$$

consisting of the expected delay cost and the probability of a false alarm. This probability has a unit coefficient in (3) without loss of generality, and the parameter $c > 0$ gives the relative importance of the terms. The Bayes risk in (3) is a common choice in the literature. It can also be considered as the Lagrangian form of a constrained optimization problem where the delay is minimized subject to an upper bound on the false alarm probability. As illustrated in Davis [10] and Bayraktar et al. [3], other alternative Bayes risks can be analyzed similarly once re-written under an auxiliary probability measure in terms of the odds-ratio introduced below. See also Poor [35] and Beibel [5] for a slightly different Bayes risk with an exponential penalty for delay and its generalized odds-ratio.

Let us introduce the conditional probability and odds-ratio processes respectively as

$$\Pi_k := \mathbb{P}(\Theta \le k|\mathcal{F}_k) \quad \text{and} \quad \Phi_k := \frac{\Pi_k}{1-\Pi_k}, \qquad k \in \mathbb{N}, \tag{4}$$

with $\Pi_0 = \pi$ and $\Phi_0 = \frac{\pi}{1-\pi}$. In terms of the process $\Phi$, using standard arguments (see for example Bayraktar et al. [3]), the expected delay and early false alarm probability of an $\mathbb{F}$-stopping time can be re-written under an auxiliary probability measure $\mathbb{P}_\infty$ as

$$\mathbb{E}(\tau-\Theta)^+ = (1-\pi)\mathbb{E}_\infty \sum_{k=0}^{\tau-1}(1-p)^k\Phi_k \qquad \text{and} \qquad \mathbb{P}(\tau<\Theta) = (1-\pi) - (1-\pi)\mathbb{E}_\infty \sum_{k=0}^{\tau-1} p(1-p)^k, \tag{5}$$

where $\mathbb{E}_\infty$ above denotes the expectation under this new measure. Our Bayes risk then becomes

$$R_\tau(\pi) = (1-\pi) + c(1-\pi)\,\mathbb{E}_\infty\Big[\sum_{k=0}^{\tau-1}(1-p)^k\Big(\Phi_k - \frac{p}{c}\Big)\Big]. \tag{6}$$

The auxiliary measure $\mathbb{P}_\infty$ and the physical measure $\mathbb{P}$ are related locally as

$$\frac{d\mathbb{P}}{d\mathbb{P}_\infty}\Big|_{\mathcal{F}_k\vee\sigma\Theta} = 1_{\{\Theta=0\}}Z_k + \sum_{i=1}^{k} 1_{\{\Theta=i\}}\frac{Z_k}{Z_{i-1}} + 1_{\{\Theta>k\}}, \qquad k\in\mathbb{N},$$

with the likelihood ratio process $Z$ given by

$$Z_0 = 1 \quad \text{and} \quad Z_k := \prod_{i=1}^{k} f(X_i), \quad \text{for } k\geq 1, \text{ where } \quad f(x) := \frac{\nu_1(dx)}{\nu_0(dx)} \tag{7}$$

is the Radon-Nykodym derivative of $\nu_1$ with respect to $\nu_0$. $\mathbb{P}_\infty$ has the property that no change occurs in the distribution of our observations (or a change occurs at $\infty$). That is, our observations $X_1, X_2, \ldots$ are independent with the distribution $\nu_0$, and furthermore they are independent of $\Theta$. Hence, computations can be carried out more conveniently under $\mathbb{P}_\infty$.

Thanks to the new representation in (6), the minimal Bayes risk can be expressed as

$$U(\pi) := \inf_\tau R_\tau(\pi) = (1-\pi) + c(1-\pi)V\big(\frac{\pi}{1-\pi}\big) \tag{8}$$

in terms of the value function of the optimal stopping problem

$$V(\phi) := \inf_\tau \mathbb{E}_\infty^\phi\Big[\sum_{k=0}^{\tau-1}(1-p)^k g(\Phi_k)\Big] \quad \text{with the running cost function } g(\phi) := \phi - \frac{p}{c}, \tag{9}$$

where $\mathbb{E}_\infty^\phi$ denotes the expectation under $\mathbb{P}_\infty$ for which the underlying process $\Phi$ starts from the point $\phi\in\mathbb{R}_+$ with probability one.

As in, for example, Shiryaev [40, Chapter 2.3], it can be shown that the odds-ratio process has the explicit form

$$\Phi_k = \frac{Z_k}{(1-p)^k}\Big(\phi + p\sum_{i=1}^{k}(1-p)^{i-1}\frac{1}{Z_{i-1}}\Big), \quad k\geq 1, \quad \text{with } \phi = \Phi_0, \tag{10}$$

and it is easy to verify that $\Phi$ satisfies the recursive relation

$$\Phi_k = \frac{f(X_k)}{1-p}\left(\Phi_{k-1} + p\right), \qquad k\geq 1, \tag{11}$$

which is useful for simulating its sample path realizations. The relation (11) also shows that $\Phi$ is a $(\mathbb{P}_\infty, \mathbb{F})$-Markov process (starting from any point $\phi\in\mathbb{R}_+$).

2.2. **Solution of the detection problem.** It follows by the linear dependence of $\Phi$ in (4) on the initial point $\Phi_0 = \phi$ and the non-negativity of the coefficient of $\phi$ that the function $\phi\mapsto V(\phi)$ is non-decreasing and concave. We have the bounds

$$0 \geq V(\phi) \geq -\frac{p}{c}\sum_{k=0}^{\infty}(1-p)^k = -\frac{1}{c},$$

where the upper bound follows by taking $\tau = 0$, and the lower bound is due to the non-negativity of $\Phi$. More importantly, it can be shown that there exists a point $\phi^*\in[\frac{p}{c}, \frac{1}{c}]$ such that $V(\phi) < 0$ for $\phi < \phi^*$ and $V(\phi) = 0$ for $\phi\geq\phi^*$. Using standard arguments of optimal stopping theory (see Shiryaev [39]) one can verify that

$$\tau^* = \inf\{k\in\mathbb{N} : V(\Phi_k) = 0\} = \inf\{k\in\mathbb{N} : \Phi_k \geq \phi^*\}$$

is an optimal stopping rule. Due to one-to-one relation between processes $\Pi$ and $\Phi$, the optimal stopping time can equivalently be expressed as the first time the process $\Pi$ exceeds the threshold $\pi^* = \phi^*/(1+\phi^*)$. In plain words, we stop and announce the new regime when its posterior probability or its odds-ratio is sufficiently large. Since $\nu_0\{x : f(x)\geq 1\} > 0$ and $X_i$s are i.i.d. under $\mathbb{P}_\infty$, the arguments in Appendix A.1. (given for modeler's odds-ratio process introduced in Section 3 below) can be modified easily to show that $\tau^*$ has (over all starting points) uniformly bounded first moment under $\mathbb{P}_\infty$.

From a computational point of view, the function $V$ is the limit of a non-increasing sequence of functions $V_n$s defined as

$$V_n(\phi) := \inf_{\tau\in\mathbb{F}} \mathbb{E}^{\phi}_{\infty}\Big[\sum_{k=0}^{(\tau\wedge n)-1}(1-p)^k g(\Phi_k)\Big], \qquad \phi\in\mathbb{R}_+, \tag{12}$$

with the notation $a\wedge b=\min\{a,b\}$. The convergence is uniform on $\mathbb{R}_+$ with error bounds

$$0\le V_n(\phi)-V(\phi)\le (1-p)^n/c \quad \text{for all } \phi\in\mathbb{R}_+ \text{ and } n\in\mathbb{N}, \tag{13}$$

and consecutive members of the sequence are computed via the recursive relation

$$V_n(\phi) := \min\Big\{0\ ,\ g(\phi)+(1-p)\ \mathbb{E}^{\phi}_{\infty}[V_{n-1}(\Phi_1)]\Big\}, \qquad n\ge 1, \tag{14}$$

starting from $V_0(\phi)=0$. The function $V$ and therefore the point $\phi^*$ can be computed/approximated iterating (14) for a large value of $n$. We can select $n$ so that the convergence error in (13) is negligible.

2.3. **Computing the delay and false alarm components.** Using the optimal time $\tau^*$, we define

$$C(\phi) := \mathbb{E}^{\phi}_{\infty}\sum_{k=0}^{\tau^*-1}(1-p)^k \quad \text{and} \quad C_n(\phi) := \mathbb{E}^{\phi}_{\infty}\sum_{k=0}^{\tau^*\wedge n-1}(1-p)^k, \quad n\in\mathbb{N},$$

for which, we have

$$C_n(\phi) = 1_{[0,\phi^*)}(\phi)\ \Big\{1+(1-p)\mathbb{E}^{\phi}_{\infty}\Big[C_{n-1}(\Phi_1)\Big]\Big\}, \qquad n\ge 1.$$

Here, $1_{[0,\phi^*)}(\phi)$ is the indicator function which is equal to 1 when $\phi<\phi^*$ and zero otherwise. The functions $C_n$s are non-decreasing and they converge to $C$ uniformly with the error bounds

$$0\le C(\phi)-C_n(\phi)\le (1-p)^n/p, \qquad \phi\in\mathbb{R}_+,\ n\in\mathbb{N},$$

and clearly all functions are zero for $\phi\ge\phi^*$.

Once $C$ is computed/approximated, the false alarm probability is obtained thanks to (5) as

$$\mathbb{P}(\tau^*<\Theta) = (1-\pi)-(1-\pi)p\ C\big(\frac{\pi}{1-\pi}\big),$$

and the expected delay associated with the optimal alarm rule is then given by

$$\mathbb{E}(\tau^*-\Theta)^+ = \frac{1}{c}[U(\pi)-\mathbb{P}(\tau^*<\Theta)] = (1-\pi)\Big[V\big(\frac{\pi}{1-\pi}\big)+\frac{p}{c}\ C\big(\frac{\pi}{1-\pi}\big)\Big].$$

## 3. A modeler using incorrect parameters/distributions

Consider now a modeler/decision maker who selects some of the parameters or distributions in the formulation above incorrectly. Following the notational convention in the introduction, we use the tilde symbol $\widetilde{\ }$ throughout the rest of the paper to denote the parameters, distributions she chooses and also the resulting expressions she computes. For example, we let $\widetilde{\pi}$ and $\widetilde{p}$ be the parameters she selects for the prior distribution of the change time, and we define $\widetilde{\phi}=\widetilde{\pi}/(1-\widetilde{\pi})$. Similarly, we let $\widetilde{f}$ be the Radon-Nykodym derivative of her pre- and post-disorder distributions $\widetilde{\nu}_0$ and $\widetilde{\nu}_1$, and $\widetilde{Z}_n$ denotes the corresponding likelihood ratio process in (7). As noted in the introduction above, we make the natural assumption that $\widetilde{\nu}_0$ and $\widetilde{\nu}_1$ are respectively equivalent to $\nu_0$ and $\nu_1$.

In line with the solution given in Section 2.2 above, the modeler computes the odds-ratio process in (10) with her parameters/distributions, and she stops at the first time $\widetilde{\tau}^*$ this process $\widetilde{\Phi}$ exceeds the incorrectly computed threshold $\widetilde{\phi}^*$. That is, her alarm time is

$$\widetilde{\tau}^* = \inf\Big\{k\ge 0 : \widetilde{\Phi}_k\ge\widetilde{\phi}^*\Big\} = \inf\left\{k\ge 0 : \frac{\widetilde{Z}_k}{(1-\widetilde{p})^k}\Big(\widetilde{\phi}+\widetilde{p}\sum_{i=1}^{k}(1-\widetilde{p})^{i-1}\frac{1}{\widetilde{Z}_{i-1}}\Big)\ge\widetilde{\phi}^*\right\}, \tag{15}$$

and the boundary $\widetilde{\phi}^*$ is the smallest point at which her value function $\widetilde{V}$ is equal to zero. In Appendix A.1., we show that the exit time of $\widetilde{\Phi}$ from an interval of the form $[0,\alpha)$, for $\alpha>0$, has uniformly bounded first moment over all starting points $\widetilde{\phi}\in\mathbb{R}_+$.

Recall that the expressions in (5) hold for any choice of $\mathbb{F}$-stopping time, and it is obvious that $\widetilde{\tau}^* \in \mathbb{F}$. Therefore, the resulting Bayes risk is

$$R_{\widetilde{\tau}^*}(\pi) = (1-\pi) + c(1-\pi)\mathbb{E}_\infty\Big[\sum_{k=0}^{\widetilde{\tau}^*-1}(1-p)^k g(\Phi_k)\Big] \tag{16}$$

computed, however, with correct parameters $\pi$, $p$, and true odds-ratio process $\Phi$. In order to find modeler's Bayes risk and her error, we need to compute the expectation in (16) above.

Observe that the processes $\Phi$ and $\widetilde{\Phi}$ are jointly driven by our observations as

$$\begin{bmatrix}\Phi_k \\ \widetilde{\Phi}_k\end{bmatrix} = \begin{bmatrix} f(X_k)/(1-p) & 0 \\ 0 & \widetilde{f}(X_k)/(1-\widetilde{p})\end{bmatrix}\begin{bmatrix}\Phi_{k-1} \\ \widetilde{\Phi}_{k-1}\end{bmatrix} + \begin{bmatrix} p/(1-p) \\ \widetilde{p}/(1-\widetilde{p})\end{bmatrix}. \tag{17}$$

Since our observations are i.i.d. under $\mathbb{P}_\infty$, it follows easily from (17) that $(\Phi, \widetilde{\Phi})$ is a Markov pair under this measure (and with the filtration $\mathbb{F}$). Hence, we can express the expectation in (16) as a function of their initial values $\phi$ and $\widetilde{\phi}$ respectively as

$$W(\phi,\widetilde{\phi}) := \mathbb{E}_\infty^{(\phi,\widetilde{\phi})}\Big[\sum_{k=0}^{\widetilde{\tau}^*-1}(1-p)^k g(\Phi_k)\Big], \qquad (\phi,\widetilde{\phi}) \in \mathbb{R}_+^2, \tag{18}$$

where, similar to the notation $\mathbb{E}_\infty^\phi$ in the previous section, $\mathbb{E}_\infty^{(\phi,\widetilde{\phi})}$ denotes the expectation under the measure $\mathbb{P}_\infty$ for which $\Phi_0 = \phi$ and $\widetilde{\Phi}_0 = \widetilde{\phi}$ with probability one. In the sequel, we omit the first argument in the superscript if the random variable whose expectation we compute depends only on the process $\widetilde{\Phi}$ and its initial point $\widetilde{\phi}$. We also adopt the same convention for the measure $\mathbb{P}_0$ introduced in Lemma 1 below.

Inside the expectation in (18), $\widetilde{\tau}^*$ depends on $\widetilde{\phi}$ and it is free of $\phi$; see (15). Clearly, $\Phi_k$ is a function of $\phi$ and it is free of $\widetilde{\phi}$. Using now the explicit form of $\Phi$ in (10), it is easy to verify that we can write $W$ as

$$W(\phi,\widetilde{\phi}) = \phi A(\widetilde{\phi}) + B(\widetilde{\phi}) \tag{19}$$

with

$$A(\widetilde{\phi}) := \mathbb{E}_\infty^{(\cdot,\widetilde{\phi})}\Big[\sum_{k=0}^{\widetilde{\tau}^*-1} Z_k\Big] \quad \text{and} \quad B(\widetilde{\phi}) := \mathbb{E}_\infty^{(\cdot,\widetilde{\phi})}\left[\sum_{k=0}^{\widetilde{\tau}^*-1}(1-p)^k\Big(\frac{Z_k}{(1-p)^k}\sum_{i=1}^{k}\frac{p(1-p)^{i-1}}{Z_i} - \frac{p}{c}\Big)\right]. \tag{20}$$

Hence, the computation of the function $W$ on $\mathbb{R}_+^2$ boils down to the computation of the one dimensional functions $A$ and $B$ on $\mathbb{R}_+$.

**Lemma 1.** *The function $\widetilde{\phi} \mapsto A(\widetilde{\phi})$ is non-negative and non-increasing. Let $\mathbb{P}_0$ denote a different probability measure under which the observations all have the distribution $\nu_1$ (as if the change happens at time 0) and $\mathbb{E}_0^{(\phi,\widetilde{\phi})}$ be the corresponding expectation operator with $(\Phi_0, \widetilde{\Phi}_0) = (\phi, \widetilde{\phi})$ with probability one. Then we have the identity*

$$A(\widetilde{\phi}) = \mathbb{E}_0^{(\cdot,\widetilde{\phi})}\left[\widetilde{\tau}^*\right], \qquad \widetilde{\phi} \in \mathbb{R}_+, \tag{21}$$

*with uniform upper bound*

$$\mathbb{E}_0^{(\cdot,\widetilde{\phi})}\left[\widetilde{\tau}^*\right] \le k_{\widetilde{\phi}^*}/\gamma^{k_{\widetilde{\phi}^*}} < +\infty \quad \textit{where} \quad k_{\widetilde{\phi}^*} := \lceil -\ln(1+\widetilde{\phi}^*)/\ln(1-\widetilde{p})\rceil \textit{ and } \gamma := \mathbb{P}_0(\widetilde{f}(X_i) \ge 1) > 0. \tag{22}$$

*The other function $\widetilde{\phi} \mapsto B(\widetilde{\phi})$ in (20) is uniformly bounded as $-\frac{1}{c} \le B(\cdot) \le A(0)$, and we have*

$$B(\widetilde{\phi}) = \mathbb{E}_\infty^{(\cdot,\widetilde{\phi})}\left[\sum_{k=0}^{\widetilde{\tau}^*-1}(1-p)^k\Big(p[A(\widetilde{\Phi}_k)-1] - \frac{p}{c}\Big)\right], \qquad \widetilde{\phi} \in \mathbb{R}_+. \tag{23}$$

The measure $\mathbb{P}_0$ in Lemma 1 and the auxiliary measure $\mathbb{P}_\infty$ are related locally via their Radon-Nykodym process

$$\frac{d\mathbb{P}_0}{d\mathbb{P}_\infty}\Big|_{\mathcal{F}_n \vee \sigma\Theta} = Z_n, \qquad n \in \mathbb{N}. \tag{24}$$

The identity in (21) is then a natural consequence of (24) and the explicit form of A in (20). We use this identity to establish the uniform upper bound for the function $A$ given in (22); see Appendix A at the end.

3.1. **A recursive approach.** Using a truncation idea as in (12), let us introduce

$$W_n(\phi,\widetilde{\phi}) := \mathbb{E}_\infty\Big[\sum_{k=0}^{(\widetilde{\tau}^*\wedge n)-1}(1-p)^k g(\Phi_k)\Big], \qquad (\phi,\widetilde{\phi})\in\mathbb{R}^2_+,\ n\in\mathbb{N}. \tag{25}$$

As $n\nearrow\infty$, we have by monotone convergence theorem

$$\mathbb{E}_\infty\Big[\sum_{k=0}^{(\widetilde{\tau}^*\wedge n)-1}(1-p)^k\Phi_k\Big]\nearrow\mathbb{E}_\infty\Big[\sum_{k=0}^{\widetilde{\tau}^*-1}(1-p)^k\Phi_k\Big] \quad\text{and}$$
$$\mathbb{E}_\infty\Big[\sum_{k=0}^{(\widetilde{\tau}^*\wedge n)-1}(1-p)^k\frac{p}{c}\Big]\nearrow\mathbb{E}_\infty\Big[\sum_{k=0}^{\widetilde{\tau}^*-1}(1-p)^k\frac{p}{c}\Big]\le\frac{1}{c},$$

therefore $W_n(\phi,\widetilde{\phi})\to W(\phi,\widetilde{\phi})$ pointwise. Markov property of the pair $(\Phi,\widetilde{\Phi})$ gives us the recursive relation

$$W_n(\phi,\widetilde{\phi}) = 1_{[0,\widetilde{\phi}^*)}(\widetilde{\phi})\cdot\Big[g(\phi)+(1-p)\ \mathbb{E}_\infty\big(W_{n-1}(\Phi_1,\widetilde{\Phi}_1)\big)\Big], \qquad n\ge 1, \tag{26}$$

with $W_0(\phi,\widetilde{\phi})=0$. Similar to (19), the functions $W_n$s also depend linearly on $\phi$. That is, we have

$$W_n(\phi,\widetilde{\phi}) = \phi A_n(\widetilde{\phi}) + B_n(\widetilde{\phi}) \tag{27}$$

where $A_n$ and $B_n$ are respectively similar to $A$ and $B$ in (20) with $\widetilde{\tau}^*$ is replaced by $\widetilde{\tau}^*\wedge n$. Plugging this linear form into (27), then using (17), and finally matching the coefficients of $\phi^0$ and $\phi^1$, we obtain

$$\begin{aligned} A_n(\widetilde{\phi}) &= 1_{[0,\widetilde{\phi}^*)}(\widetilde{\phi})\ \Big\{1+\mathbb{E}_0\Big[A_{n-1}(\widetilde{\Phi}_1)\Big]\Big\}\\ B_n(\widetilde{\phi}) &= 1_{[0,\widetilde{\phi}^*)}(\widetilde{\phi})\ \Big\{-\frac{p}{c}+p[A_n(\widetilde{\phi})-1]+(1-p)\mathbb{E}_\infty\Big[B_{n-1}(\widetilde{\Phi}_1)\Big]\Big\}. \end{aligned} \tag{28}$$

These recursions can be established alternatively after showing

$$A_n(\widetilde{\phi}) = \mathbb{E}_0^{(\cdot,\widetilde{\phi})}\left[\widetilde{\tau}^*\wedge n\right] \quad\text{and}\quad B_n(\widetilde{\phi}) = \mathbb{E}_\infty^{(\cdot,\widetilde{\phi})}\left[\sum_{k=0}^{\widetilde{\tau}^*\wedge n-1}(1-p)^k\Big(p[A_{n-k}(\widetilde{\Phi}_k)-1]-\frac{p}{c}\Big)\right] \tag{29}$$

by modifying the corresponding arguments in the proof of Lemma 1 given for the identities (21-23).

Starting with $A_0(\widetilde{\phi})=B_0(\widetilde{\phi})=0$, we successively iterate (28) to compute/approximate $A$ and $B$ and obtain $W$ by (19). The true Bayes risk of the modeler is then given by

$$R_{\widetilde{\tau}^*}(\pi) = (1-\pi)+c(1-\pi)\left[\frac{\pi}{1-\pi}A\big(\frac{\widetilde{\pi}}{1-\widetilde{\pi}}\big)+B\big(\frac{\widetilde{\pi}}{1-\widetilde{\pi}}\big)\right] \tag{30}$$

thanks to (16). This gives us the percentage error in (2) as

$$\frac{R_{\widetilde{\tau}^*}(\pi)-U(\pi)}{U(\pi)}\cdot 100\ \% = \frac{c(1-\pi)\Big[\frac{\pi}{1-\pi}A\big(\frac{\widetilde{\pi}}{1-\widetilde{\pi}}\big)+B\big(\frac{\widetilde{\pi}}{1-\widetilde{\pi}}\big)-V\big(\frac{\pi}{1-\pi}\big)\Big]}{(1-\pi)+c(1-\pi)V\big(\frac{\pi}{1-\pi}\big)}\cdot 100\ \% \tag{31}$$

Number of iterations can be determined using Proposition 2 below. An alternative approach would be to stop the iterations when the supnorm of the difference of consecutive functions becomes negligible.

The representation in (30) implies that the Bayes risk is linear in $\pi$ on the interval $[0,1]$. We verify this result for the general case without the absolute continuity of $\nu_1$ with respect to $\nu_0$ in Appendix B by induction.

**Proposition 2.** *The collection of functions $A_n$s and $B_n$s converge to $A$ and $B$, respectively, with the uniform error bounds*

$$0\le A(\widetilde{\phi})-A_n(\widetilde{\phi})\le\mathbb{E}_0^{(\cdot,\widetilde{\phi})}[1_{\{\widetilde{\tau}^*>n\}}\widetilde{\tau}^*]\le\mathbb{E}_0^{(\cdot,0)}[1_{\{\widetilde{\tau}^*>n\}}\widetilde{\tau}^*] \tag{32}$$

*and*

$$-\frac{1}{c}(1-p)^n \le B(\widetilde{\phi}) - B_n(\widetilde{\phi}) \le A(0)(1-p)^n + \sum_{k=0}^{n}(1-p)^k p \|A - A_{n-k}\| \tag{33}$$

*where* $\|A - A_{n-k}\| := \sup_{\widetilde{\phi}\in[0,\widetilde{\phi}^*]}(A(\widetilde{\phi}) - A_{n-k}(\widetilde{\phi}))$. *We have* $\sum_{k=0}^{n}(1-p)^k p\|A - A_{n-k}\| \to 0$ *as* $n \to \infty$.

3.2. **Modeler's expected delay and false alarm probability.** The expected delay and the false alarm probability associated with modeler's stopping time can be computed by replicating similar computations as in Section 2.3, with careful modifications using $\widetilde{\Phi}$ and its hitting time $\widetilde{\tau}^*$. If we define

$$\widetilde{C}(\widetilde{\phi}) := \mathbb{E}_\infty^{(\cdot,\widetilde{\phi})}\sum_{k=0}^{\widetilde{\tau}^*-1}(1-p)^k \quad \text{and} \quad \widetilde{C}_n(\widetilde{\phi}) := \mathbb{E}_\infty^{(\cdot,\widetilde{\phi})}\sum_{k=0}^{\widetilde{\tau}^*\wedge n-1}(1-p)^k, \quad n\in\mathbb{N},$$

we have the iterations

$$\widetilde{C}_n(\widetilde{\phi}) = 1_{[0,\widetilde{\phi}^*)}(\widetilde{\phi})\,\Big\{1 + (1-p)\mathbb{E}_\infty^{\phi}\Big[\widetilde{C}_{n-1}(\widetilde{\Phi}_1)\Big]\Big\}, \qquad n \ge 1, \tag{34}$$

with convergence bounds $0 \le \widetilde{C}(\cdot) - \widetilde{C}_n(\cdot) \le (1-p)^n/p$, for $n\in\mathbb{N}$. In terms of the function $\widetilde{C}$, modeler's false alarm probability is $\mathbb{P}(\widetilde{\tau}^* < \Theta) = (1-\pi) - (1-\pi)p\,\widetilde{C}\big(\frac{\widetilde{\pi}}{1-\widetilde{\pi}}\big)$, and her expected delay is obtained as

$$\mathbb{E}(\widetilde{\tau}^* - \Theta)^+ = \frac{1}{c}[R_{\widetilde{\tau}^*}(\pi) - \mathbb{P}(\widetilde{\tau}^* < \Theta)] = (1-\pi)\Big[\frac{\pi}{1-\pi}A\big(\frac{\widetilde{\pi}}{1-\widetilde{\pi}}\big) + B\big(\frac{\widetilde{\pi}}{1-\widetilde{\pi}}\big) + \frac{p}{c}\,C\big(\frac{\widetilde{\pi}}{1-\widetilde{\pi}}\big)\Big]. \tag{35}$$

## 4. Examples

In this section, the computations and results of Section 3 are illustrated in three different settings.

4.1. **Example I: Bernoulli observations with misspecified post-disorder success probability.** Consider a model where the obervations $X_1, X_2, \dots$ are $\{0,1\}$-valued Bernoulli random variables whose common success probability changes from $q_0$ to $q_1$ at the disorder time. In this setup, we have

$$f(x) = \Big(\frac{q_1}{q_0}\Big)^x\Big(\frac{1-q_1}{1-q_0}\Big)^{1-x} \text{ for } x\in\{0,1\}, \text{ and } Z_k = \Big(\frac{q_1}{q_0}\Big)^{S_k}\Big(\frac{1-q_1}{1-q_0}\Big)^{k-S_k} \text{ with } S_k := \sum_{i=1}^{k} X_i. \tag{36}$$

The odds-ratio process has the form

$$\Phi_k = \frac{1}{(1-p)^k}\Big(\frac{q_1}{q_0}\Big)^{S_k}\Big(\frac{1-q_1}{1-q_0}\Big)^{k-S_k}\left[\phi + p\sum_{i=1}^{k}(1-p)^{i-1}\Big(\frac{q_0}{q_1}\Big)^{S_{i-1}}\Big(\frac{1-q_0}{1-q_1}\Big)^{(i-1)-S_{i-1}}\right], \tag{37}$$

and the value function $V$ in (9) and the approximating sequence $V_n$s in (12) are obtained via the iterations

$$V_n(\phi) = \min\left\{0\ ,\ g(\phi) + (1-p)\left[q_0\cdot V_{n-1}\Big(\frac{q_1}{q_0}\cdot\frac{\phi+p}{1-p}\Big) + (1-q_0)\cdot V_{n-1}\Big(\frac{1-q_1}{1-q_0}\cdot\frac{\phi+p}{1-p}\Big)\right]\right\}. \tag{38}$$

Suppose that the modeler chooses a wrong value $\widetilde{q}_1$ instead of the true value $q_1$, and all other parameters are selected correctly. She evaluates the expressions in (37) with this wrong choice $\widetilde{q}_1$ and obtains a different process $\widetilde{\Phi}$. Also, she computes a different value function $\widetilde{V}$ from (38), and this yields a different stopping boundary $\widetilde{\phi}^*$. Note however that both $\Phi$ and $\widetilde{\Phi}$ start from $\phi = \widetilde{\phi} = \pi/(1-\pi)$ since the modeler uses the correct value of $\pi$ by assumption.

In Figure 1, we provide a numerical example in which Bernoulli experiments are fair initially; that is, $q_0 = 0.5$. The parameter $p$ of the prior distribution in (1) is set as 0.2. Consistent with the choices in some of the numerical examples in Dayanik and Sezer [14], Bayraktar et al [4], Dayanik et al. [13], Sezer [37], Buonaguidi et al. [8], we take the delay cost as $c = 0.25$. We set the true post-disorder probability as $q_1 = 0.3$ whereas the modeler's choice is $\widetilde{q}_1 = 0.4$. The correct value function $V$ and the one computed by the modeler $\widetilde{V}$ are shown in Panel (a) of the figure. The dashed lines represent the thresholds $\phi^*$ and $\widetilde{\phi}^*$. Compared to $\widetilde{q}_1$, $q_1$ is further away from the pre-disorder probability $q_0$. Hence, the orderings $V \le \widetilde{V}$ and $\phi^* \ge \widetilde{\phi}^*$ are natural.

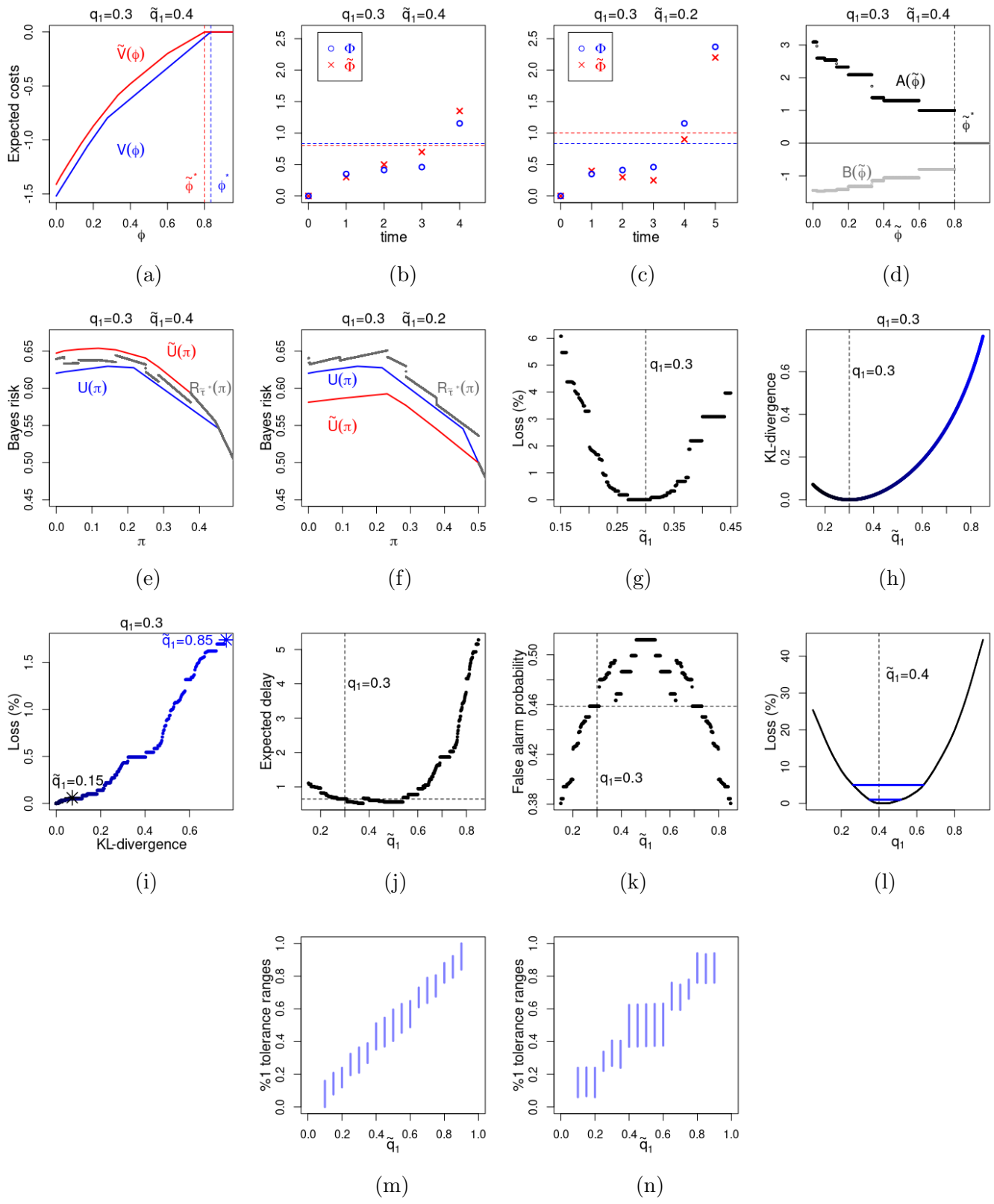


FIGURE 1. Plots for the examples where pre- and post-disorder observations are Bernoulli random variables whose success probability changes from $q_0$ to $q_1$. In all plots, we have $q_0 = 0.5$ and $c = 0.25$.

Panel (b) is an illustration of the simulated sample paths of $\Phi$ and $\widetilde{\Phi}$ when the sequential observations are $(X_1, X_2, \ldots) = (0, 1, 1, 0, 0, \ldots)$. Both processes start from the point $\phi = \widetilde{\phi} = 0$ corresponding to the value $\pi = 0$. The gap between $\phi^*$ and $\widetilde{\phi}^*$ is not large, and $\Phi$ and $\widetilde{\Phi}$ stop at the same time here (though with different accumulated costs). Depending on the setup, this gap may be larger and these two processes may stop at different times. For example, when modeler's choice is $\widetilde{q}_1$ is 0.2 (instead of 0.4), same sequence of observations causes $\widetilde{\Phi}$ to stop at time 5 whereas $\Phi$ stops at time 4; see Panel (c).

To evaluate the Bayes risk of the modeler, we need to compute the function $W$ in (18), or equivalently the functions $A$ and $B$ in (19). Once $W$ is obtained, we compute $W(\phi, \widetilde{\phi})$ with $\phi = \widetilde{\phi} = \pi/(1-\pi)$, and we use this expression in (16) to obtain the Bayes risk. Since

$$\widetilde{\Phi}_1 = \Big(\frac{\widetilde{q}_1}{\widetilde{q}_0}\Big)^{X_1}\Big(\frac{1-\widetilde{q}_1}{1-\widetilde{q}_0}\Big)^{1-X_1}\Big(\frac{\widetilde{\phi}+\widetilde{p}}{1-\widetilde{p}}\Big),$$

the functions $A$ and $B$ in (18-19) are computed iteratively using (28) as

$$\begin{aligned} A_n(\widetilde{\phi}) &= 1 + q_1 \cdot A_{n-1}\Big(\frac{\widetilde{q}_1}{\widetilde{q}_0}\cdot\frac{\widetilde{\phi}+\widetilde{p}}{1-\widetilde{p}}\Big) + (1-q_1)\cdot A_{n-1}\Big(\frac{1-\widetilde{q}_1}{1-\widetilde{q}_0}\cdot\frac{\widetilde{\phi}+\widetilde{p}}{1-\widetilde{p}}\Big) \\ B_n(\widetilde{\phi}) &= -\frac{p}{c} + p[A_n(\widetilde{\phi})-1] + (1-p)\Big\{q_0\cdot B_{n-1}\Big(\frac{\widetilde{q}_1}{\widetilde{q}_0}\cdot\frac{\widetilde{\phi}+\widetilde{p}}{1-\widetilde{p}}\Big) + (1-q_0)\cdot B_{n-1}\Big(\frac{1-\widetilde{q}_1}{1-\widetilde{q}_0}\cdot\frac{\widetilde{\phi}+\widetilde{p}}{1-\widetilde{p}}\Big)\Big\} \end{aligned} \tag{39}$$

for $\widetilde{\phi} < \widetilde{\phi}^*$. Clearly, these functions are all zero for $\widetilde{\phi} \geq \widetilde{\phi}^*$.

In the last plot of the first row of Figure 1, we show the functions $A$ and $B$ when $q_1 = 0.3$ and $\widetilde{q}_1 = 0.4$ again. First plot of the second row gives the true minimal Bayes risk $U$, the erroneous Bayes risk computed by the modeler $\widetilde{U}$, and the actual Bayes risk of the modeler $R_{\widetilde{\tau}^*}$ in (30) as functions of $\pi$. The irregular structure of the function $R_{\widetilde{\tau}^*}$ is inherited from the functions $A$ and $B$. The discontinuities of $A$ and $B$ follow by the binary jump behavior of $\widetilde{\Phi}$ reflected in (39).

Observe that $R_{\widetilde{\tau}^*}$ is lower than $\widetilde{U}$. In the true model, the post-disorder probability $q_1 = 0.3$ is further away from the pre-disorder probability $q_0 = 0.5$ compared to her choice $\widetilde{q}_1 = 0.4$. As a result, the observations signal the change better compared to what her model expects. Hence, her actual Bayes risk is lower than what she thinks she incurs. We observe the opposite ordering when $\widetilde{q}_1 = 0.2$ (and $q_1 = 0.3$). This is illustrated in Panel (f). This time, the post-disorder observations are less vocal and we have $R_{\widetilde{\tau}^*} \geq \widetilde{U}$.

Panel (g) gives the percentage loss in (30) with $\pi = 0$ as modeler's choice $\widetilde{q}_1$ varies. The plot does not show a definite uniform advantage of overestimating $q_1$ compared to an underestimation by the same amount on this range in terms of the percentage losses. Panel (h) and (i) looks at the losses from the perceptive of the KL-divergence. Panel (h) shows the Kullback-Leibler (KL) divergence of the Bernoulli distributions as $\widetilde{q}_1$ deviates from $q_1$

$$D_{KL}(q_1||\widetilde{q}_1) = q_1 \ln\big(\frac{q_1}{\widetilde{q}_1}\big) + (1-q_1)\ln\big(\frac{1-q_1}{1-\widetilde{q}_1}\big)$$

as $\widetilde{q}_1$ varies. Clearly, the divergence is an increasing function of the modeler's error $|\widetilde{q}_1 - q_1|$ (in Panel (h)). Panel (i) shows, as $\widetilde{q}_1$ varies on the range $[0.15, 0.85]$, the monotone dependence of the percentage loss in the KL divergence as expected.

In Panels (j-k), we give the expected delay and the false alarm probability components of modeler's Bayes risk (again as functions of her choice of $\widetilde{q}_1$ and for $\pi = 0$). Horizontal dashed lines show the values in the true model, and the vertical dashed lines indicate the location of the true probability $q_1 = 0.3$. Intuitively speaking, when $\widetilde{q}_1$ is lower than $q_1 = 0.3$, the modeler expects to observe 0s even more frequently with the post-change regime. However, such outcomes are relatively less likely in the true post-disorder regime, and she would therefore potentially stop later. Furthermore, as $\widetilde{q}_1$ gets lower, there is a greater mismatch between her anticipations and what the actual post-disorder environments offers, and this should cause higher expected delays. The plots confirm this intuition; we observe that the expected delay is decreasing, and the false probability is increasing on this range indeed. By a similar intuition, we would expect higher (compared to the value in the true model) and increasing expected delays on the other side of the $\widetilde{q}_1$ range; this time, modeler anticipates more 1s after the change. This is indeed what we observe in the plot of expected delays. Since $q_1 < 0.5$, higher expected delays for large $\widetilde{q}_1$ values are natural. For the false alarm

probabilities, on the other hand, we observe a symmetry around $\widetilde{q}_1 = 0.5$. It is easy to verify that, when $\widetilde{q}_0 = q_0 = 0.5$, the probability law of the process $\widetilde{\Phi}$ under $\mathbb{P}_\infty$ remains the same if $\widetilde{q}_1$ is replaced by $1 - \widetilde{q}_1$. We have the same value function iterations for $\widetilde{V}$, we obtain the same threshold $\widetilde{\phi}^*$, and the iterations in (34) for the false alarm probabilities are again the same. This gives us the symmetry in the plot. Note that we don't have such a symmetry in the computation of modeler's expected delay. Recall that the value of $q_1$ is fixed in both plots and the expected delay is given by the expectation $(1-\pi)\mathbb{E}_\infty[\sum_{k=0}^{\widetilde{\tau}^*-1}(1-p)^k\Phi_k]$; see (5). When we change $\widetilde{q}_1$ only, we change the joint law of $\Phi$ and $\widetilde{\Phi}$, and this gives us a different expected value. We can also check that the recursions in (39) are not the same. Hence, the modeler's Bayes risk and her expected delay are different.

In panel (l), we fix the modeler's choice $\widetilde{q}_1 = 0.4$, and we plot the relative loss in (31) as a function of true value of $q_1$ (again with $\pi = 0$). Horizontal line segments show 1% and 5% tolerance intervals. Clearly, the error is zero when $q_1 = 0.4$, and it increases as $q_1$ deviates from this value. The plot indicates that the error becomes more severe as the deviation from the true value increases. Nevertheless, 1% tolerance interval is, relatively speaking, non-negligible, and 5% tolerance interval is considerably larger. This gives a certain level of confidence to the modeler in implementation. In Panel (m), where the y-axis is for $q_1$ values, we show 1% tolerance intervals at various choices of $\widetilde{q}_1$ by the modeler. We observe that the lengths of these intervals do not vary significantly. However, this is not always the case as the last panel illustrates where we change the value of $p$ as 0.5. In this case, we have a greater fluctuation depending on the choice of $\widetilde{q}_1$.

Clearly, we can always re-label the success/failure (or 0/1) outcomes in our Bernoulli experiments, in which case the success and failure probabilities are interchanged in both true and incorrect models. This would only change the notation in our analysis. We would obtain the same values for the true minimal Bayes risk $U(\pi)$ and the Bayes risk $R_{\widetilde{\tau}^*}(\pi)$ incurred by the modeler; see (38) and (39). When $q_0 = \widetilde{q}_0 = 0.5$, this observation gives us a symmetry in the computation of these functions for the $(q_1, \widetilde{q}_1)$ values around the point $(0.5, 0.5)$. In the last two panels of Figure 1, the diagonal symmetry in the tolerance ranges is a consequence of this symmetry.

4.2. **Example II: Gaussian observations with misidentified post-disorder mean and variance.** Suppose now that observations before the change time have a normal distribution with mean $\mu_0$ and variance $\sigma_0^2$. After the change, they are again normally distributed with a different pair of parameters $\mu_1$ and $\sigma_1^2$ respectively. We assume that the modeler incorrectly sets one or both of these post-disorder parameters as $\widetilde{\mu}_1$ and $\widetilde{\sigma}_1^2$ respectively and her other choices are correct.

Under the correct formulation of the model, we have

$$f(x) = \Big(\frac{\sigma_0^2}{\sigma_1^2}\Big)^{\frac{1}{2}}\ e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}+\frac{(x-\mu_0)^2}{2\sigma_0^2}}, \text{ for } x \in \mathbb{R}, \text{ and } Z_k = \Big(\frac{\sigma_0^2}{\sigma_1^2}\Big)^{\frac{k}{2}}\ e^{-\frac{1}{2\sigma_1^2}\sum_{i=1}^k (X_i-\mu_1)^2+\frac{1}{2\sigma_0^2}\sum_{i=1}^k (X_i-\mu_0)^2}. \tag{40}$$

The process $\Phi$ is as in (10) with the explicit form of the process $Z$ above. The value function iterations in (14) can be written as

$$V_n(\phi) = \min\left\{0\ ,\ g(\phi) + (1-p)\int_{-\infty}^{\infty} V_{n-1}\Big(f(x)\Big(\frac{\phi+p}{1-p}\Big)\Big)\frac{1}{\sqrt{2\pi\sigma_0^2}}e^{-\frac{(x-\mu_0)^2}{2\sigma_0^2}}\,dx\right\} \tag{41}$$

and the iterations start with $V_0(\phi) = 0$. Replacing $\mu_1$ and $\sigma_1^2$ in (40-41) with $\widetilde{\mu}_1$ and $\widetilde{\sigma}_1^2$, we obtain the expressions computed by the modeler.

In Figure 2, we give examples where the observations before the change have the standard normal distribution; that is, $\mu_0 = 0$, $\sigma_0^2 = 1$. In the first row, true post-disorder parameters are $\mu_1 = 1.25$, $\sigma_1^2 = 1.75^2$ whereas the modeler incorrectly assumes $\widetilde{\mu}_1 = 1.5$, $\widetilde{\sigma}_1^2 = 2^2$. Similar to the first plot in Figure 1 in the previous subsection, Panel (a) here shows functions $V$ and $\widetilde{V}$ with the corresponding thresholds $\phi^*$ and $\widetilde{\phi}^*$ respectively.

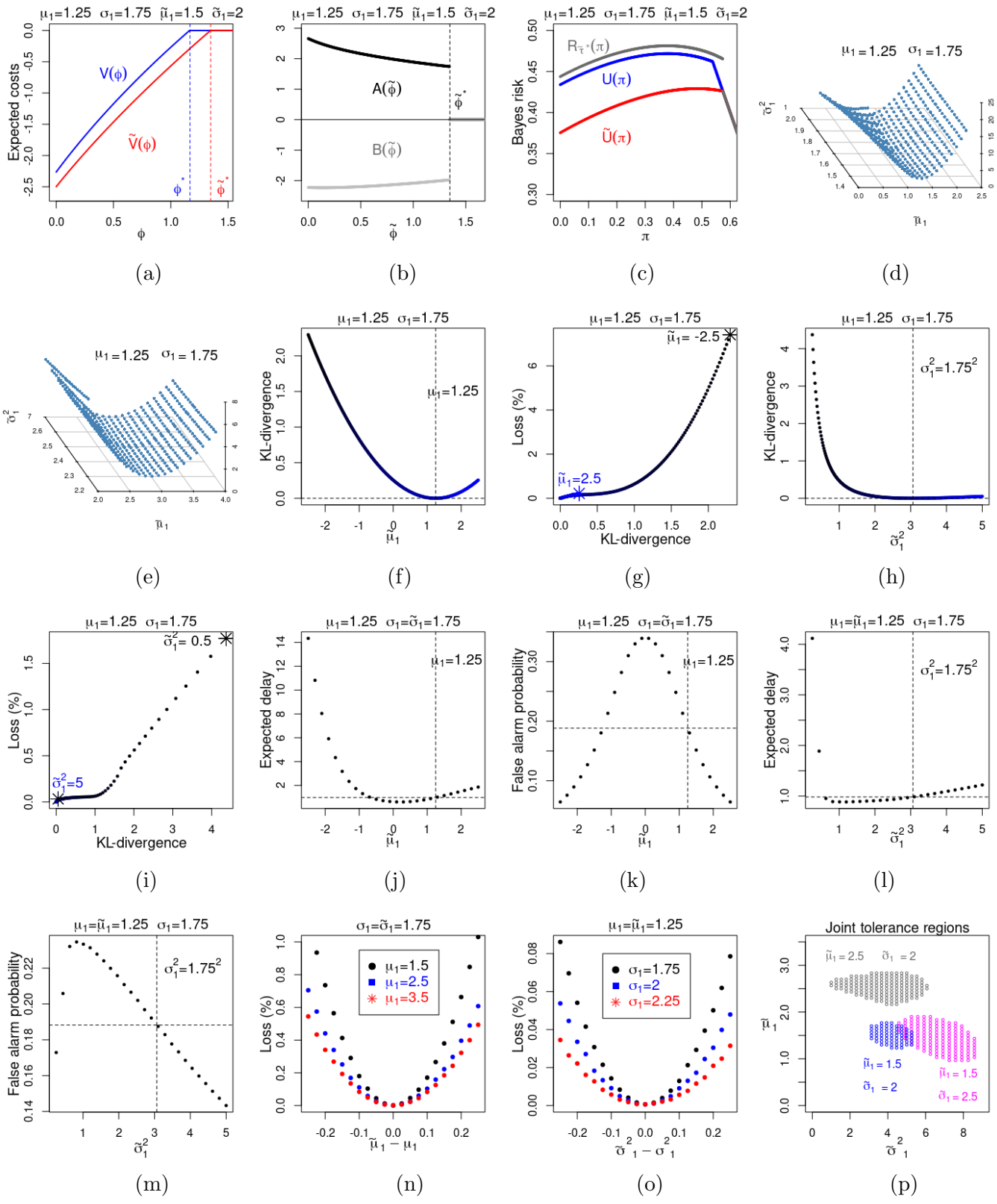


FIGURE 2. Plots for the examples where pre- and post-disorder observations are normally distributed. Prior to the change point, they have the standard normal distribution. As in Section 4.1, we set $p = 0.2$ and $c = 0.25$ in our calculations.

The Bayes risk of the modeler is given by the expressions in (16) and (18-19) with $\pi = \widetilde{\pi}$. The recursive relations (28) in the current setup become

$$
\begin{aligned}
A_n(\widetilde{\phi}) &= 1 + \int_{-\infty}^{\infty} A_{n-1}\Big(\widetilde{f}(x)\big(\frac{\widetilde{\phi}+p}{1-p}\big)\Big)\frac{1}{\sqrt{2\pi\sigma_1^2}}e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}\,dx \\
B_n(\widetilde{\phi}) &= -\frac{p}{c} + p[A_n(\widetilde{\phi})-1] + (1-p)\int_{-\infty}^{\infty} B_{n-1}\Big(\widetilde{f}(x)\big(\frac{\widetilde{\phi}+p}{1-p}\big)\Big)\frac{1}{\sqrt{2\pi\sigma_0^2}}e^{-\frac{(x-\mu_0)^2}{2\sigma_0^2}}\,dx
\end{aligned}
\tag{42}
$$

for $\widetilde{\phi} < \widetilde{\phi}^*$, and clearly, $A_n = B_n = A = B = 0$ for $\widetilde{\phi} \geq \widetilde{\phi}^*$.

Panel (b) in Figure 2 illustrates the functions $A$ and $B$ (again for $\mu_0 = 0$, $\sigma_1^2 = 1^2$, $\mu_1 = 1.25$, $\sigma_1^2 = 1.75^2$, $\widetilde{\mu}_1 = 1.5$, $\widetilde{\sigma}_1^2 = 2^2$). In the next panel, we have the true minimal Bayes risk and the one incurred by the modeler. Compared to Figure 1, the functions $A$, $B$, and $R_{\widetilde{\tau}^*}$ have a more regular structure here. Unlike the binary jump structure of the Bernoulli case, the integrals in (42) have a smoothing effect on the computations.

In Panels (d) and (e), we show 3d scatter plots, for $\pi = 0 = \widetilde{\pi}$, of the percentage loss in (2) as modeler's choices $\widetilde{\mu}_1$, $\widetilde{\sigma}_1^2$ vary jointly around the true values $\mu_1$, $\sigma_1^2$ in two different problem instances. In both panels, we see that the plotted surface has a steeper curvature along the $\widetilde{\mu}_1$-axis. We observe a similar pattern also in other examples in our numerical experiments. Here, we give two plots only. Recall that the mean of the normal distribution is a location parameter and the variance is a non-negative scale parameter. Hence, unit errors in their estimated values have naturally different consequences.

For the first problem instance (with $\mu_1 = 1.25$ and $\sigma_1 = 1.75$), Panels (f-i) illustrates the KL-divergence of a modeler selecting either the post-change mean or variance incorrectly. Plots are similar to those in Panels (h-i) in Figure (1). The KL-divergence for the distribution of the modeler with an incorrect post-change mean is simply $D_{KL}(\mu_1||\widetilde{\mu}_1) = (\widetilde{\mu}_1 - \mu_1)^2/2\sigma_1^2$. In Panels (f-g), we see the KL-divergence on the range $\widetilde{\mu}_1 \in [-2.5, 2.5]$ and the percentage loss in Bayes risk as a function of this divergence. Similarly, the KL-divergence for a modeler with an incorrect post-disorder variance is

$$
D_{KL}(\sigma_1^2 \,||\, \widetilde{\sigma}_1^2) = \frac{1}{2}\Big[\ln\big(\frac{\widetilde{\sigma}_1^2}{\sigma_1^2}\big) + \frac{\sigma_1^2}{\widetilde{\sigma}_1^2} - 1\Big]
$$

Panels (h-i) illutrate this divergence and its effect on the percentage loss in the Bayes risk. Naturally, we see in Panels (g) and (i) that as the KL divergence increases, so does the loss of the modeler with an incorrect parameter.

The next four plots in Panels (j-m) show in the same problem instance (for $\pi = 0$ again) the expected delay and false alarm probability when one of the parameters is misselected. Horizontal dashed lines show the correct values under the true model, and vertical dashed lines point to the correct parameter value in each case. When the mean is overestimated, even though the post-disorder regime makes larger observations more likely, the modeler would expect to observe higher values. However, such observations are relatively less likely in the true post-change environment. Hence, this potentially causes a further delay in her detection decisions. We see this behavior in Panel (j) for values $\widetilde{\mu}_1 > \mu_1$. For negative values of $\widetilde{\mu}_1$, since $\mu_1 = 1.25 > 0$, there is a greater mismatch between her expectations for the observations and what the actual post-disorder regime offers. Hence, we see that expected delay increases as $\widetilde{\mu}_1$ decreases, and we observe higher values on this range.

The symmetry in Panel (k) for the false alarm probabilities is a natural consequence of the zero mean for observations in both the true pre-change regime and the one assumed by the modeler. In this case, reverting the sign of $\widetilde{\mu}_1$ does not change the probability law of $\widetilde{\Phi}$ under $\mathbb{P}_\infty$. Hence, we have the same function $\widetilde{V}$, same threshold $\widetilde{\phi}^*$, and the same iterations in (34) for the false alarm probability. Her expected delay, however, is effected by this change as $\mu_1$ is fixed. As noted in the previous subsection with Bernoulli observations for changing the value of $\widetilde{q}_1$ only (when $q_0 = \widetilde{q}_0 = 0.5$), changing only the value of $\widetilde{\mu}_1$ in this setup yields a different joint law for the pair $(\Phi, \widetilde{\Phi})$. Hence, the expected delay in (5) with $\tau = \widetilde{\tau}^*$ is different, and therefore there is no symmetry in the plot of expected delays.

The shape of the plots in Panels (l) and (m) could also be explained intuitively by using the mismatch again between the likelihoods of the post-change observations in the true and incorrect models. When $\widetilde{\sigma}_1^2$

is large, the incorrect model anticipates more outliers with the new regime. On the other hand, when $\widetilde{\sigma}_1^2$ is very small, observations are expected to be rather concentrated around the post-disorder mean value. In both cases, the modeler potentially stops later, and this yields higher delays and smaller probabilities for early false alarms.

In Panels (n) and (o), we see the effect of the level of a parameter on the percentage loss. In the first one, the x-axis is the error $\widetilde{\mu}_1 - \mu_1$ of the modeler. Clearly, the loss increases as the absolute error increases. The curves for different values of $\mu_1 = 1.5$, 2.5, and 3.5 show that the same amount of (small) error in the estimation yields a lower percentage loss when $\mu_1$ is higher (i.e., further away from its pre-disorder value). Although we use a different formulation, what we observe here is analogous to similar observations in Du et al. [16] which considers a non-Bayesian framework with Gaussian observations and presents a case study for the effect of drift misspecification on the problem of minimizing the *stationary average detection delay* (subject to a lower bound constraint on the *average run length* on false alarms) using the *Shiryaev-Roberts procedure*. In the numerical examples in Tables 2-4 in Du et al. [16], we see that the same error in the post-change drift gives smaller percentage changes in the stationary average detection delay (for fixed average run length) as the true drift increases.

We observe a similar pattern in Panel (o) for the error in estimating the variance. Once again, the percentage loss increases with higher estimation error, and as the post-disorder variance is further away from its pre-disorder value, the same (small) error yields a lower loss.

Finally, in the last plot, we construct 1% joint tolerance regions for three different choices of $(\widetilde{\mu}_1, \widetilde{\sigma}_1^2)$ by the modeler. Consistent with our observations in the paragraph above, the regions enlarge (with different central points) when one of the parameters gets further away from its pre-change value.

4.3. **Example III: Post-disorder Erlang distribution misclassified as exponential.** As a last example, we assume that the observations $X_1, X_2, \ldots$ are at first exponentially distributed with rate $\lambda_0$, and after the change time, they are Erlang distributed with parameters $(m, \lambda_1)$, for some integer $m \geq 1$. The modeler correctly sets up her formulation except that she erroneously selects an exponential distribution for the post-disorder regime with a possibly different rate; that is, she sets $\widehat{m} = 1$ and take $\widehat{\lambda}_1$ as the rate. A possible mistake is to set $\widehat{\lambda}_1 = \lambda_1/m$ so that both the correct post-disorder distribution and her choice have the same mean $1/\lambda_1$. This means that the modeler ignores the change in the distribution and focuses only on the change in the mean.

In the true formulation, we have

$$f(x) = \beta_m x^{m-1} e^{-(\lambda_1 - \lambda_0)x}, \text{ for } x \in \mathbb{R}_+, \text{ and } Z_k = \beta_m^k P_k^{m-1} e^{-(\lambda_1 - \lambda_0)S_k}$$

where

$$\beta_m := \frac{\lambda_1^m}{\lambda_0 (m-1)!}, \quad P_k := \prod_{i=1}^{k} X_i, \quad \text{and} \quad S_k = \sum_{i=1}^{k} X_i \text{ as in (36).}$$

The odds-ratio process is updated as in (14), and the recursive equations (11) to compute the value function have the form

$$V_n(\phi) = \min\Big\{0\ ,\ g(\phi) + (1-p)\int_0^\infty V_{n-1}\Big(f(x)\ (\frac{\phi + p}{1-p})\Big)\lambda_0 e^{-\lambda_0 x}\ dx\Big\}.$$

Replacing $m$ and $\lambda_1$ above with $\widetilde{m} = 1$ and $\widetilde{\lambda}_1$, we obtain the computations from the perspective of the modeler. The first panel in Figure 3 illustrates the functions $V$ and $\widetilde{V}$ on an example where $\lambda_0 = 1$, $m = 3$, $\lambda_1 = 2$, $\widetilde{\lambda}_1 = 2/3$. Vertical lines point to the thresholds $\phi^*$ and $\widetilde{\phi}^*$.

In the current setup, iterations in (28) become

$$\begin{aligned} A_n(\widetilde{\phi}) &= 1 + \int_0^\infty A_{n-1}\Big(\widetilde{f}(x)(\frac{\widetilde{\phi} + p}{1-p})\Big)\ \frac{\lambda_1^m x^{m-1}}{(m-1)!} e^{-\lambda_1 x}\ dx \\ B_n(\widetilde{\phi}) &= -\frac{p}{c} + p[A_n(\widetilde{\phi}) - 1] + (1-p)\int_0^\infty B_{n-1}\Big(\widetilde{f}(x)(\frac{\widetilde{\phi} + p}{1-p})\Big)\ \lambda_0 e^{-\lambda_0 x} dx \end{aligned} \tag{43}$$

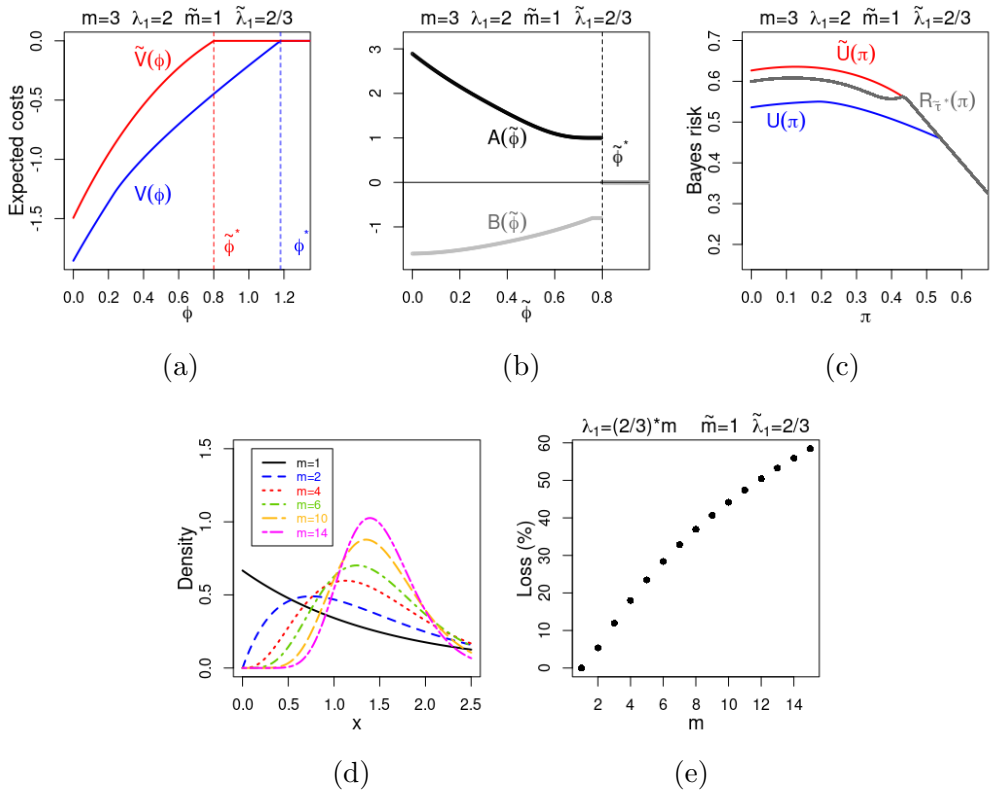


FIGURE 3. Plots for the examples where pre-disorder observations are exponentially distributed with rate $\lambda_0$. Post-disorder observations have the Erlang distribution with parameters $m$ and $\lambda_1$. In each plot, common parameters are $p = 0.2$, $\lambda_0 = 1$, and $c = 0.25$.

for $\widetilde{\phi} < \widetilde{\phi}^*$, and the functions are zero for $\widetilde{\phi} \geq \widetilde{\phi}^*$. With $\widetilde{\phi} = \phi$, we compute the function $W(\phi, \phi) = \phi A(\phi) + B(\phi)$ as in (19), and modeler's Bayes risk is given by $R_{\widetilde{\tau}^*} = (1-\pi) + c(1-\pi)W(\frac{\pi}{1-\pi}, \frac{\pi}{1-\pi})$ as in (30).

Panel (b) of Figure 3 shows the functions $A$ and $B$ (again for the case $\lambda_0 = 1$, $m = 3$, $\lambda_1 = 2$, $\widetilde{\lambda}_1 = 2/3$). As in Section 4.2, in comparison to Figure 1, the functions $A$ and $B$ have a more regular structure due to the integral terms in (43). Panel (c) in the row gives the functions $U$, $\widetilde{U}$, and the actual Bayes risk $B_{\widetilde{\tau}}$.

For a fixed mean level, the consequence of ignoring the correct nature of the post-disorder distribution (and assuming an exponential distribution instead) is illustrated in the second row of the figure. Here, we fix the ratio $\lambda_1/m$ as $2/3$; as $m$ gets higher, we also increase the value of $\lambda_1$ by the same factor. The modeler's choice $\widetilde{\lambda}_1$ remains the same as $2/3$. As shown in Panel (d), as $m$ increases, the Erlang distribution differs significantly from modeler's exponential distribution. We observe in the last panel that the percentage error (shown for $\pi = 0$) indeed increases considerably.

## 5. Concluding remarks

The current paper shows how to compute the actual Bayes risk when a modeler sets up pre- and post-disorder distributions and/or other parameters incorrectly in the classical Bayesian change detection problem studied in Shiryaev [42, 39]. To the best of our knowledge, this is the first paper which considers the computation of a modeler's exact Bayes risk and therefore her error in a Bayesian formulation. The paper shows that this is a challenging task even in such a relatively simple formulation. This formulation has the advantage that the true solution is one dimensional. Therefore, the computation of the modeler's Bayes risk can be carried out in the two dimensional state space of true and wrong odds-ratio processes. Linear dependence of the true odds-ratio process on its initial value allows us to reduce the computations to a pair of one dimensional computations as shown in Section 3.

In implementation, a modeler can use the method of this paper to compare her choices with other alternatives. In particular, she can construct tolerance regions for estimated parameters as illustrated in Section 4. Each of the three examples/settings considered in this section can potentially be explored further, especially numerically, as the subject of a separate paper. Other specific examples can also be considered depending on the application area. Our purpose here is mainly to illustrate our method and highlight some key observations.

In practice, as an alternative to constructing tolerance regions, the modeler could also use a robust change detection formulation in which the pre- and post-change parameters/distributions belong to some uncertainty sets, and the aim is to minimize the worst case Bayes risk. In such an approach, these uncertainty sets could also be misidentified; that is, they may not include true values/distributions. Therefore the misspecification issue could also arise in that approach, and we leave the investigation of this case as a future research subject.

The analysis in this paper can be extended for other change detection problems. As already mentioned in the introduction section, continuous time formulations with Brownian motion and compund Poisson processes will be discussed in subsequent papers. For the classical change detection problem for a Wiener process, Shiryaev [38, 42] can be consulted. The corresponding formulation with simple Poisson observations is given by Gal'chuk and Rozovskii [20] and later solved by Peskir and Shiryaev [32]. Dayanik and Sezer [14] gave the solution for compound Poisson processes; see also Gapeev [21]. Clearly, more complex change detection problems and more generally other sequential diagnosis problems can be studied as further extensions. In diagnosis problems, we are interested in not only detecting a change but also identifying its type; see Lai [24], Dayanik et al. [12], Ma et al. [26], and Dayanik and Kazutoshi [15] for some examples. In such problems, the parameters and distributions are subject to misidentification as well, and it is important to understand the consequences of these errors.

As a final remark, we would like to note that similar misspecification issues may be also be considered in various other stochastic optimization problems in which certain parameters or distributions need to be estimated prior to implementation. In particular, Bayesian formulations involving unobservable processes whose probability laws depend on these parameters/distributions bring new and interesting challenges. Problems employing hidden Markov models form a well-known class of such problems for which a short list of citations would be insufficient to highlight all important contributions. The reader may consider referring to the recent review paper Mor et al. [28] and the books Elliot et al. [18], Cappé et al. [9] for a good starting point to discover a much richer literature. We believe that the current paper provides a new perspective for researchers working on these problems as well.

## Appendix A: Auxiliary results and proofs for Section 3

**A.1. Uniform bounds for the first moments of the exit times of $\widetilde{\Phi}$ under $\mathbb{P}_0$ and $\mathbb{P}_\infty$.** We show here that the exit time $\widetilde{\tau}_\alpha = \inf\{k \in \mathbb{N} : \widetilde{\Phi}_k \geq \alpha\}$, for $\alpha > 0$, has uniformly bounded first moment (over all starting points $\widetilde{\phi} \in \mathbb{R}_+$) under the measures $\mathbb{P}_0$ and $\mathbb{P}_\infty$ and under the assumptions that $\nu_1$ is absolutely continuous with respect to $\nu_0$, $\nu_0$ and $\widetilde{\nu}_0$ are equivalent, and also so are $\nu_1$ and $\widetilde{\nu}_1$. The arguments are similar for both measures. We first give the proof for $\mathbb{P}_0$. Note that $X_i$s are i.i.d. random variables under both measures (with different distributions).

Since i) $\nu_1$ is absolutely continuous with respect to $\nu_0$, ii) $\nu_0$ and $\widetilde{\nu}_0$ are equivalent, and iii) $\nu_1$ and $\widetilde{\nu}_1$ are also equivalent, it follows that $\widetilde{\nu}_1$ is absolutely continuous with respect to $\widetilde{\nu}_0$. Clearly,

$$1 = \int \widetilde{f}(x)\ \widetilde{\nu}_0(dx) = \int \widetilde{f}(x) 1_{\{\widetilde{f}(x)>0\}}\ \widetilde{\nu}_0(dx),$$

and therefore $\{\widetilde{f}(x) \geq 1\}$ is a non-negligible set under the restriction of $\widetilde{\nu}_0$ to the set $\{\widetilde{f}(x) > 0\}$. Hence, $\{\widetilde{f}(x) \geq 1\}$ is a non-negligible subset under $\widetilde{\nu}_1$ as well and also under $\nu_1$ by the equivalence of these two measures. This gives us (see (22))

$$0 < \mathbb{P}_0(D_i) = \gamma_0, \quad \text{if we define} \quad D_i := \Big\{\omega \in \Omega : \widetilde{f}(X_i(\omega)) \geq 1\Big\}, \qquad i \in \mathbb{N}.$$

Applying the recursion in (11) with modeler's choices, we have for any $\ell, k \in \mathbb{N}$

$$\widetilde{\Phi}_{\ell+k} = \frac{\prod_{i=1}^{k} \widetilde{f}(X_{\ell+i})}{(1-\widetilde{p})^k}\, \widetilde{\Phi}_\ell + \sum_{i=1}^{k} \frac{\widetilde{p}}{(1-\widetilde{p})^i} \prod_{j=k-i+1}^{k} \widetilde{f}(X_{\ell+j}) \geq \sum_{i=1}^{k} \frac{\widetilde{p}}{(1-\widetilde{p})^i} \prod_{j=k-i+1}^{k} \widetilde{f}(X_{\ell+j}). \tag{44}$$

The sum $\sum_{i=1}^{\infty} \frac{\widetilde{p}}{(1-\widetilde{p})^i}$ is divergent, and for $k \geq k_\alpha := \lceil -\ln(1+\alpha)/\ln(1-\widetilde{p}) \rceil$, the finite sum $\sum_{i=1}^{k} \frac{\widetilde{p}}{(1-\widetilde{p})^i}$ is greater than or equal to $\alpha$. Taking $\ell = 0$ in (44), this gives

$$\{\widetilde{\tau}_\alpha > k_\alpha\} \subset (\cap_{i=1}^{k_\alpha} D_i)^c \quad \text{and therefore} \quad \mathbb{P}_0^{(\cdot,\widetilde{\phi})}\{\widetilde{\tau}_\alpha > k_\alpha\} \leq 1 - \gamma_0^{k_\alpha} < 1.$$

Since $X_i$s are independent and identically distributed (with distribution $\nu_1$) under $\mathbb{P}_0$, we obtain similarly by taking $\ell = mk_\alpha$ in (44) for an integer $m \geq 1$ that

$$\mathbb{P}_0^{(\cdot,\widetilde{\phi})}\{\widetilde{\tau}_\alpha > (m+1)k_\alpha \mid \widetilde{\tau}_\alpha > mk_\alpha\} \leq 1 - \gamma_0^{k_\alpha},$$

from which the inequality

$$\mathbb{P}_0^{(\cdot,\widetilde{\phi})}\{\widetilde{\tau}_\alpha > mk_\alpha\} \leq (1-\gamma_0^{k_\alpha})^m, \qquad m \geq 1,$$

follows inductively. The inequality is also valid trivially for $m = 0$. Hence, it holds for all $m \in \mathbb{N}$. Using this inequality, we write

$$\mathbb{E}_0^{(\cdot,\widetilde{\phi})}[\widetilde{\tau}_\alpha] = \sum_{k\in\mathbb{N}} \mathbb{P}_0^{(\cdot,\widetilde{\phi})}\{\widetilde{\tau}_\alpha > k\} \leq k_\alpha \sum_{m\in\mathbb{N}} \mathbb{P}_0^{(\cdot,\widetilde{\phi})}\{\widetilde{\tau}_\alpha > mk_\alpha\} \leq k_\alpha \sum_{m\in\mathbb{N}} (1-\gamma_0^{k_\alpha})^m = k_\alpha/\gamma_0^{k_\alpha} < \infty,$$

which shows that the expectation of $\widetilde{\tau}_\alpha$ under $\mathbb{P}_0$ is uniformly bounded over $\widetilde{\phi}$ for any fixed $\alpha > 0$.

The result for $\mathbb{P}_\infty$ can be established similarly by computing the probabilities above under this measure. Because of $\widetilde{\nu}_0$ and $\nu$ are equivalent, $\{\widetilde{f}(x) \geq 1\}$ is also a non-negligible set of $\nu$. This gives $\gamma_\infty = \mathbb{P}_\infty(D_i) > 0$. Hence, replicating the steps above, we obtain $\mathbb{E}_\infty^{(\cdot,\widetilde{\phi})}[\widetilde{\tau}_\alpha] \leq k_\alpha/\gamma_\infty^{k_\alpha} < \infty$.

### A.2. Proofs.

***Proof of Lemma 1***. The non-negativity and monotonicity of the function $A$ is an immediate consequence of the non-negativity of the likelihood process $Z$ and the monotone dependence of the exit time $\widetilde{\tau}^*$ in (15) on the starting point $\widetilde{\phi}$. To establish (21), we recall that the measure $\mathbb{P}_0$ and $\mathbb{P}_\infty$ are related locally as in (24). Then, we have

$$\mathbb{E}_\infty^{(\cdot,\widetilde{\phi})}\Big[\sum_{k=0}^{\widetilde{\tau}^*-1} Z_k\Big] = \sum_{k\in\mathbb{N}} \mathbb{E}_\infty^{(\cdot,\widetilde{\phi})}\Big[1_{\{k<\widetilde{\tau}^*\}} Z_k\Big] = \sum_{k\in\mathbb{N}} \mathbb{E}_0^{(\cdot,\widetilde{\phi})}\Big[1_{\{k<\widetilde{\tau}^*\}}\Big] = \mathbb{E}_0^{(\cdot,\widetilde{\phi})}\Big[\sum_{k\in\mathbb{N}} 1_{\{k<\widetilde{\tau}^*\}}\Big] = \mathbb{E}_0^{(\cdot,\widetilde{\phi})}[\widetilde{\tau}^*]$$

where the first and third equalities are due to monotone convergence theorem. Second equality follows by the usual change of measure arguments since $\{k < \widetilde{\tau}^*\} \in \mathcal{F}_k$. Last equality is self-evident. The uniform bounds in (22) follows by the arguments in Appendix 5 above on the finiteness of the first moments of the exit times.

The lower bound $-\frac{1}{c}$ on the function $B$ is an immediate consequence of its definition in (20). The upper bound $A(0)$ follows easily from (23). To prove (23), it is sufficient to verify that

$$\mathbb{E}_\infty^{(\cdot,\widetilde{\phi})} \sum_{k=0}^{\widetilde{\tau}^*-1} Z_k \sum_{i=1}^{k} \frac{(1-p)^{i-1}}{Z_{i-1}} = \mathbb{E}_\infty^{(\cdot,\widetilde{\phi})} \sum_{k=0}^{\widetilde{\tau}^*-1} (1-p)^k [A(\widetilde{\Phi}_k) - 1]. \tag{45}$$

To that end, we note by the Markov property that

$$\mathbb{E}_\infty^{(\cdot,\widetilde{\phi})}\Big[\sum_{k=i}^{\widetilde{\tau}^*-1} \frac{Z_k}{Z_i}\Big|\mathcal{F}_i\Big] = A(\widetilde{\Phi}_i) \quad \text{and} \quad 1_{\{i<\widetilde{\tau}^*\}} A(\widetilde{\Phi}_i) = 1_{\{i<\widetilde{\tau}^*\}}\big(1 + \mathbb{E}_\infty^{(\cdot,\widetilde{\phi})}\big[f(X_{i+1})A(\widetilde{\Phi}_{i+1})\big|\mathcal{F}_i\big]\big), \quad \mathbb{P}^{(\cdot,\widetilde{\phi})}\text{-a.s.},$$

where the second equality uses the recursion $A(\widetilde{\phi}) = 1_{[0,\widetilde{\phi}^*]}\big(1+\mathbb{E}_\infty^{(\cdot,\widetilde{\phi})}\big[f(X_1)A(\widetilde{\Phi}_1)\big|\mathcal{F}_i\big]\big)$, which can be shown easily using the definition of $A$. Then, we write the left hand side of (45) by repeated application of the monotone convergence theorem as

$$\begin{aligned}
\mathbb{E}_\infty^{(\cdot,\widetilde{\phi})}\sum_{k=0}^{\widetilde{\tau}^*-1} Z_k \sum_{i=1}^{k}\frac{(1-p)^{i-1}}{Z_{i-1}} &= \sum_{k=1}^{\infty}\sum_{i=1}^{\infty}\mathbb{E}_\infty^{(\cdot,\widetilde{\phi})}1_{\{i\le k<\widetilde{\tau}^*\}}(1-p)^{i-1}\frac{Z_k}{Z_{i-1}} \\
&= \sum_{i=1}^{\infty}\mathbb{E}_\infty^{(\cdot,\widetilde{\phi})}1_{\{i<\widetilde{\tau}^*\}}\sum_{k=i}^{\widetilde{\tau}^*-1}(1-p)^{i-1}\frac{Z_k}{Z_{i-1}} = \sum_{i=1}^{\infty}\mathbb{E}_\infty^{(\cdot,\widetilde{\phi})}1_{\{i<\widetilde{\tau}^*\}}(1-p)^{i-1}f(X_i)\ \mathbb{E}_\infty^{(\cdot,\widetilde{\phi})}\Big[\sum_{k=i}^{\widetilde{\tau}^*-1}\frac{Z_k}{Z_i}\big|\mathcal{F}_i\Big] \\
&= \sum_{i=1}^{\infty}\mathbb{E}_\infty^{(\cdot,\widetilde{\phi})}1_{\{i\le\widetilde{\tau}^*\}}(1-p)^{i-1}f(X_i)A(\widetilde{\Phi}_i) = \sum_{i=0}^{\infty}\mathbb{E}_\infty^{(\cdot,\widetilde{\phi})}1_{\{i<\widetilde{\tau}^*\}}(1-p)^{i}f(X_{i+1})A(\widetilde{\Phi}_{i+1}) \\
&= \sum_{i=0}^{\infty}\mathbb{E}_\infty^{(\cdot,\widetilde{\phi})}1_{\{i<\widetilde{\tau}^*\}}(1-p)^{i}\mathbb{E}_\infty^{(\cdot,\widetilde{\phi})}\Big[f(X_{i+1})A(\widetilde{\Phi}_{i+1})\big|\mathcal{F}_i\Big] = \sum_{i=0}^{\infty}\mathbb{E}_\infty^{(\cdot,\widetilde{\phi})}1_{\{i<\widetilde{\tau}^*\}}(1-p)^{i-1}[A(\widetilde{\Phi}_i)-1].
\end{aligned}$$

The last expression is the right hand side in (45), and this completes proof. □

***Proof of Proposition 2.*** The inequalities in (32) follow directly by the characterizations of the functions $A$ and $A_n$ in (21) and (29) respectively. Since $\tau^*$ has a uniformly finite first moment under $\mathbb{P}_0$, it follows that the upper bound converges to zero thanks to dominated convergence theorem as $n\to\infty$.

For the bounds on the convergence of $B_n$s, we note by (23) and (29) that

$$\begin{aligned}
B(\widetilde{\phi}) - B_n(\widetilde{\phi}) = \mathbb{E}_\infty\left[\sum_{k=0}^{\widetilde{\tau}^*\wedge n-1}(1-p)^k p[A(\widetilde{\Phi}_k) - A_{n-k}(\widetilde{\Phi}_k)]\right] \\
+ \mathbb{E}_\infty 1_{\{\tau>n\}}\left[\sum_{k=n}^{\widetilde{\tau}^*-1}(1-p)^k\Big(p[A(\widetilde{\Phi}_k)-1]-\frac{p}{c}\Big)\right]
\end{aligned} \tag{46}$$

The first expectation is non-negative (since $A_n\le A$) and it is bounded above by

$$\sum_{k=0}^{n}(1-p)^k p\|A-A_{n-k}\| = \sum_{k=0}^{\infty}(1-p)^k p 1_{\{k\le n\}}\|A-A_{n-k}\|. \tag{47}$$

Since $1_{\{k\le n\}}\|A-A_{n-k}\|\to 0$ as $n\to\infty$ and $1_{\{k\le n\}}\|A-A_{n-k}\|\le A(0)$, it follows by dominated convergence that the sum in (47) converges to zero as $n\to\infty$ again. For the second expectation in (46), since $A(\widetilde{\Phi}_k)-1\ge 0$ on the event $\{k<\widetilde{\tau}^*\}$, the expectation is bounded from below by $-\frac{1}{c}\sum_{k=n}^{\infty}(1-p)^k p = -\frac{1}{c}(1-p)^n$ and above by $\sum_{k=0}^{n}(1-p)^k pA(0) = (1-p)^n A(0)$. Inequalities in (33) now follow from the bounds on the expectations in (46). □

## Appendix B. The case without the absolute continuity of $\nu_1$ with respect to $\nu_0$

In this case, the measure $\mathbb{P}_\infty$ is not absolutely continuous with respect to the physical measure $\mathbb{P}$ anymore. Therefore, we work with the physical measure. The flow of arguments below is similar to that in Sections 2 and 3. The major difference is that the linearity (in $\pi$) is established by induction.

As studied and solved in Shiryaev [42, 39], the detection problem is an optimal stopping problem under $\mathbb{P}$ for the process $\Pi$, and we have

$$U(\pi) = \inf_{\tau\in\mathbb{F}}\mathbb{E}\Big[\sum_{k=0}^{\tau-1}c\Pi_k + (1-\Pi_\tau)\Big]. \tag{48}$$

The process $\Pi$ can be shown to be a Markov process with the one-step updating rule for $k\ge 0$

$$\Pi_{k+1} = J(X_{k+1},\Pi_k) \quad\text{where}\quad J(x,\pi) := \frac{[\pi+(1-\pi)p]f_1(x)}{(1-\pi)(1-p)f_0(x)+[\pi+(1-\pi)p]f_1(x)}. \tag{49}$$

The densities $f_0$ and $f_1$ are respectively the Radon Nykodym densities of $\nu_0$ and $\nu_1$ with respect to another measure $\eta$ with the property that both $\nu_0$ and $\nu_1$ are absolutely continuous with respect to this measure; we can, for example, take $\eta = \nu_0 + \nu_1$.

In (48), truncating the time horizon with the $n$th observation, we obtain as usual a sequence of functions $(U_n)_{n\in\mathbb{N}}$. These functions form a decreasing collection. They converge to $U$ and they satisfy the iterations

$$U_n(\pi) = \min\left\{1-\pi\ ,\ c\pi + \mathbb{E}[U_{n-1}(\Pi_1)]\right\}, \qquad n \geq 1. \tag{50}$$

There exists a point $\pi^*$ such that $U(\pi) = 1-\pi$ for $\pi \geq \pi^*$ and $U(\pi) < 1-\pi$ for for $\pi < \pi^*$. The first time the process $\Pi$ exceeds this threshold is an optimal stopping time in (48); see Theorem 7 and its proof in Shiryaev [39, Chapter 4.3] for rigorous justifications of these results.

A modeler with some incorrect parameters/distributions computes the iterations above with her choices and obtains an incorrect function $\widetilde{U}$ with a wrong threshold $\widetilde{\pi}^*$. She follows an incorrect conditional process $\widetilde{\Pi}$ updated according to a wrong one-step operator $\widetilde{J}$. The operator $\widetilde{J}$ is obtained by replacing the parameters/densities in the definition of $J$ in (49) with the ones chosen by the modeler.

For a bounded Borel function $h$, one can verify that

$$\mathbb{E}[h(\Pi_{k+1}, \widetilde{\Pi}_{k+1})|\mathcal{F}_k] = \mathcal{S}[h](\Pi_k, \widetilde{\Pi}_k) \qquad \mathbb{P}\text{-almost surely,}$$

with the operator

$$\mathcal{S}[h](\pi,\widetilde{\pi}) := \int_E h\big(J(x,\pi), \widetilde{J}(x,\widetilde{\pi})\big)\big[(\pi+(1-\pi)p)f_1(x) + (1-\pi)(1-p)f_0(x)\big]\eta(dx).$$

This shows that $(\Pi, \widetilde{\Pi})$ is a Markov pair, and we can define the Bayes risk of the modeler as a function of their starting point $(\Pi_0, \widetilde{\Pi}_0) = (\pi, \widetilde{\pi})$ as

$$L(\pi,\widetilde{\pi}) := \mathbb{E}\Big[\sum_{k=0}^{\widetilde{\tau}^*-1} c\Pi_k + (1-\Pi_{\widetilde{\tau}^*})\Big],$$

where $\widetilde{\tau}^*$ (as in Section 3) denotes the stopping of the modeler. Once again, truncating the time horizon, we obtain the sequence

$$L_n(\pi,\widetilde{\pi}) := \mathbb{E}\Big[\sum_{k=0}^{\widetilde{\tau}^*\wedge n-1} c\Pi_k + (1-\Pi_{\widetilde{\tau}^*\wedge n})\Big], \qquad n\in\mathbb{N},$$

converging to $L$ pointwise thanks to monotone and bounded convergence theorems (applied respectively to the separate terms inside the expectation above). By Markov property, we have the iterations

$$L_n(\pi,\widetilde{\pi}) = 1_{\{\widetilde{\pi}<\widetilde{\pi}^*\}}\Big\{c\pi + \mathbb{E}\big[L_{n-1}(\Pi_1,\widetilde{\Pi}_1)\big]\Big\} + 1_{\{\widetilde{\pi}\geq\widetilde{\pi}^*\}}(1-\pi), \qquad n\geq 1, \tag{51}$$

with the first element $L_0(\pi,\widetilde{\pi}) = 1-\pi$.

Clearly, $L_0$ is linear in $\pi$. Next, assume that for an arbitrary $n\geq 1$ we have $L_{n-1}(\pi,\widetilde{\pi}) = \pi G_{n-1}(\widetilde{\pi}) + H_{n-1}(\widetilde{\pi})$ for some functions $G_{n-1}$ and $H_{n-1}$ on $[0,1]$. Then, by using this linear form on the right hand side in (51) (with the distribution of $X_1$ under $\mathbb{P}$), one can show after some careful computations that the linearity in $\pi$ also holds for $L_n$ and we have $L_n(\pi,\widetilde{\pi}) = \pi G_n(\widetilde{\pi}) + H_n(\widetilde{\pi})$ where

$$G_n(\widetilde{\pi}) := 1_{\{\widetilde{\pi}<\widetilde{\pi}^*\}}\Big\{c + (1-p)\mathbb{E}_0[G_{n-1}(\widetilde{\Pi}_1)] + (1-p)\mathbb{E}_\infty[H_{n-1}(\widetilde{\Pi}_1)] - (1-p)\mathbb{E}_0[H_{n-1}(\widetilde{\Pi}_1)]\Big\} - 1_{\{\widetilde{\pi}\geq\widetilde{\pi}^*\}} \tag{52}$$

and

$$\begin{aligned} H_n(\widetilde{\pi}) &:= 1_{\{\widetilde{\pi}<\widetilde{\pi}^*\}}\Big\{p\mathbb{E}_0[G_{n-1}(\widetilde{\Pi}_1)] + p\mathbb{E}_\infty[H_{n-1}(\widetilde{\Pi}_1)] + (1-p)\mathbb{E}_0[H_{n-1}(\widetilde{\Pi}_1)]\Big]\Big\} + 1_{\{\widetilde{\pi}\geq\widetilde{\pi}^*\}} \\ &= 1_{\{\widetilde{\pi}<\widetilde{\pi}^*\}}\left\{\frac{p}{1-p}[G_n(\widetilde{\pi}) - c] + \mathbb{E}_0[H_{n-1}(\widetilde{\Pi}_1)]\right]\right\} + 1_{\{\widetilde{\pi}\geq\widetilde{\pi}^*\}}. \end{aligned} \tag{53}$$

In the recursions above, the expectations $\mathbb{E}_0$ and $\mathbb{E}_\infty$ are respectively according to the measures $\mathbb{P}_0$ and $\mathbb{P}_\infty$ introduced in Sections 2 and 3. Under $\mathbb{P}_0$, $X_1$ has the distribution $\nu_1$, and under $\mathbb{P}_\infty$ this distribution is $\nu_0$.

Pointwise convergence of $(L_n)_{n\in\mathbb{N}}$ and the characterization $L_n(\pi,\widetilde{\pi}) = \pi G_n(\widetilde{\pi}) + H_n(\widetilde{\pi})$, for $n \in \mathbb{N}$, give us the convergence, again pointwise, of the sequences $(G_n)_{n\in\mathbb{N}}$ and $(H_n)_{n\in\mathbb{N}}$; taking $\pi = 0$ establishes the convergence of the former sequence, and then the result for the latter is obvious. Hence, the function $L$ is linear in $\pi$.

The false alarm probability of the true optimal stopping time can be found by the iterations

$$f_n(\pi) = 1_{\{\pi<\pi^*\}}\mathbb{E}[f_{n-1}(\Pi_1)]\Big] + 1_{\{\pi\geq\pi^*\}}(1-\pi)$$

starting with $f_0(\pi) = 1-\pi$. The false alarm probability of the stopping time of the modeler, on the other hand, can be obtained by setting $c = 0$ in the iterations in (51) (and (52-53)). For each stopping rule, once we have the Bayes risk and the false alarm probability, the expected delay is computed by subtracting the false alarm probability from the Bayes risk and dividing the resulting expression by $c$.

## References


[1] T. Banerjee and V. V. Veeravalli. Data-efficient minimax quickest change detection with composite post-change distribution. *IEEE Transactions on Information Theory*, 61(9):5172–5184, 2015.

[2] M. Basseville and I. V. Nikiforov. *Detection of Abrupt Changes: Theory and Application*. Prentice-Hall, 1993.

[3] E. Bayraktar, S. Dayanik, and I. Karatzas. The standard Poisson disorder problem revisited. *Stochastic Processes and their Applications*, 115(9):1437 – 1450, 2005.

[4] E. Bayraktar, S. Dayanik, and I. Karatzas. Adaptive Poisson disorder problem. *The Annals of Applied Probability*, 16(3):1190–1261, 2006.

[5] M. Beibel. A note on sequential detection with exponential penalty for the delay. *The Annals of Statistics*, 28(6):1696–1701, 2000.

[6] C. Blanchet-Scalliet, A. Diop, R. Gibson, D. Talay, and E. Tanré. Technical analysis techniques versus mathematical models: Boundaries of their validity domains. In H. Niederreiter and D. Talay, editors, *Monte Carlo and Quasi-Monte Carlo Methods 2004*, pages 15–30, Berlin, Heidelberg, 2006. Springer Berlin Heidelberg.

[7] C. Blanchet-Scalliet, A. Diop, R. Gibson, D. Talay, and E. Tanré. Technical analysis compared to mathematical models based methods under parameters mis-specification. *Journal of Banking & Finance*, 31(5):1351–1373, 2007.

[8] B. Buonaguidi, A. Mira, H. Bucheli, and V. Vitanis. Bayesian quickest detection of credit card fraud. *Bayesian Analysis*, pages 1–30, 2021.

[9] O. Cappé, E. Moulines, and T. Ryden. *Inference in Hidden Markov Models*. Springer Series in Statistics. Springer-Verlag, Berlin, Heidelberg, 2005.

[10] M. Davis. A note on the Poisson disorder problem. *Banach Center Publications*, 1(1):65–72, 1976.

[11] S. Dayanik and C. Goulding. Sequential detection and identification of a change in the distribution of a Markov-modulated random sequence. *IEEE Transactions on Information Theory*, 55(7):3323–3345, 2009.

[12] S. Dayanik, C. Goulding, and H. V. Poor. Bayesian sequential change diagnosis. *Mathematics of Operations Research*, 33(2):475–496, 2008.

[13] S. Dayanik, H. V. Poor, and S. O. Sezer. Multisource Bayesian sequential change detection. *Annals of Applied Probability*, 18(2):552–590, 2008.

[14] S. Dayanik and S. O. Sezer. Compound Poisson disorder problem. *Mathematics of Operations Research*, 31(4):649–672, 2006.

[15] S. Dayanik and K. Yamazaki. Detection and identification of changes of hidden Markov chains: asymptotic theory. *Statistical Inference for Stochastic Processes*, 2021.

[16] W. Du, A. S. Polunchenko, and G. Sokolov. On robustness of the shiryaev-Roberts change-point detection procedure under parameter misspecification in the post-change distribution. *Communications in Statistics. Simulation and Computation*, 46(3):2185–2206, 2017.

[17] E. Ekström and J. Vaicenavicius. Monotonicity and robustness in Wiener disorder detection. *Sequential Analysis*, 38(1):57–68, 2019.

[18] R. J. Elliott, L. Aggoun, and J. B. Moore. *Hidden Markov Models: Estimation and Control*. Applications of mathematics; vol. 29. Springer-Verlag, New York, 1995.

[19] C. Fuh and A. G. Tartakovsky. Asymptotic Bayesian theory of quickest change detection for hidden Markov models. *IEEE Transactions on Information Theory*, 65(1):511–529, 2019.

[20] L. Gal'chuk and B. Rozovskii. The 'disorder' problem for a Poisson process. *Theory of Probability & Its Applications*, 16(4):712–716, 1971.

[21] P. V. Gapeev. The disorder problem for compound Poisson processes with exponential jumps. *Ann. Appl. Probab.*, 15(1A):487–499, 02 2005.

[22] O. Hadjiliadis. Optimality of the 2-CUSUM drift equalizer rules for detecting two-sided alternatives in the Brownian motion model. *Journal of Applied Probability*, 42(4):1183–1193, 2005.

[23] O. Hadjiliadis and V. Moustakides. 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(1):75–85, 2006.

[24] T. L. Lai. Sequential multiple hypothesis testing and efficient fault detection-isolation in stochastic systems. *IEEE Transactions on Information Theory*, 46(2):595–608, 2000.

[25] G. Lorden. Procedures for reacting to a change in distribution. *The Annals of Mathematical Statistics*, 42(6):1897–1908, 1971.

[26] X. Ma, L. Lai, and S. Cui. Two-stage Bayesian sequential change diagnosis. *IEEE Transactions on Signal Processing*, 69:6131–6147, 2021.

[27] T. L. Molloy and J. J. Ford. Misspecified and asymptotically minimax robust quickest change detection. *IEEE Transactions on Signal Processing*, 65(21):5730–5742, 2017.

[28] B. Mor, S. Garhwal, and A. Kumar. A systematic review of hidden Markov models and their applications. *Archives of Computational Methods in Engineering*, 28(3):1429–1448, 2021.

[29] G. V. Moustakides. Optimal stopping times for detecting changes in distributions. *The Annals of Statistics*, 14(4):1379–1387, 1986.

[30] G. V. Moustakides. Detecting changes in hidden Markov models. In *2019 IEEE International Symposium on Information Theory (ISIT)*, pages 2394–2398, 2019.

[31] A. A. Muravlev and A. N. Shiryaev. Two-sided disorder problem for a Brownian motion in a Bayesian setting. *Proceedings of the Steklov Institute of Mathematics*, 287(1):202–224, 2015.

[32] G. Peskir and A. N. Shiryaev. *Solving the Poisson disorder problem*, pages 295–312. Springer, Berlin, Heidelberg, 2002.

[33] M. Pollak and D. Siegmund. A diffusion process and its applications to detecting a change in the drift of brownian motion. *Biometrika*, 72(2):267–280, 1985.

[34] A. S. Polunchenko and A. G. Tartakovsky. State-of-the-art in sequential change-point detection. *Methodology and Computing in Applied Probability*, 14(3):649–684, 2012.

[35] H. V. Poor. Quickest detection with exponential penalty for delay. *The Annals of Statistics*, 26(6):2179 – 2205, 1998.

[36] H. V. Poor and O. Hadjiliadis. *Quickest Detection*. Cambridge University Press, 2008.

[37] S. O. Sezer. On the Wiener disorder problem. *The Annals of Applied Probability*, 20(4):1537–1566, 2010.

[38] A. Shiryaev. Some exact formulas in the "disorder" problem. *Theory Probab. Appl.*, 10:348–354, 1965.

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

[40] A. Shiryaev. *Stochastic Disorder Problems*. Probability Theory and Stochastic Modelling. Springer, 2019.

[41] A. Shiryaev, M. Zhitlukhin, and W. Ziemba. When to sell Apple and the NASDAQ? Trading bubbles with a stochastic disorder model. *Journal of Portfolio Management*, 40(2):54–63, 2014.

[42] A. N. Shiryaev. On optimum methods in quickest detection problems. *Theory of Probability & Its Applications*, 8(1):22–46, 1963.

[43] A. N. Shiryaev. Quickest detection problems: Fifty years later. *Sequential Analysis*, 29(4):345–385, 2010.

[44] A. N. Shiryaev, M. V. Zhitlukhin, and W. T. Ziemba. Land and stock bubbles, crashes and exit strategies in Japan circa 1990 and in 2013. *Quantitative finance*, 15(9):1449–1469, 2015.

[45] M. S. Srivastava and Y. Wu. Comparison of EWMA, CUSUM and Shiryayev-Roberts procedures for detecting a shift in the mean. *The Annals of Statistics*, 21(2):645–670, 1993.

[46] A. Tartakovsky. *Sequential Change Detection and Hypothesis Testing: General Non-i.i.d. Stochastic Models and Asymptotically Optimal Rules*. Chapman & Hall/CRC Monographs on Statistics and Applied Probability. CRC Press, 2020.

[47] A. Tartakovsky, I. Nikiforov, and M. Basseville. *Sequential Analysis: Hypothesis Testing and Change-point Detection*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. CRC Press, 2014.

[48] A. G. Tartakovsky. Multidecision quickest change-point detection: Previous achievements and open problems. *Sequential Analysis*, 27(2):201–231, 2008.

[49] A. G. Tartakovsky. Asymptotic optimality of mixture rules for detecting changes in general stochastic models. *IEEE Transactions on Information Theory*, 65(3):1413–1429, 2019.

[50] A. G. Tartakovsky and G. V. Moustakides. State-of-the-art in Bayesian changepoint detection. *Sequential Analysis*, 29(2):125–145, 2010.

[51] Q. Xu, Y. Mei, and G. V. Moustakides. Optimum multi-stream sequential change-point detection with sampling control. *IEEE Transactions on Information Theory*, 67(11):7627–7636, 2021.

[52] H. Yang, O. Hadjiliadis, and M. Ludkovski. Quickest detection in the Wiener disorder problem with post-change uncertainty. *Stochastics*, 89(3–4):654–685, 2017.

[53] S. Zacks. Survey of classical and Bayesian approaches to the change-point problem: fixed sample and sequential procedures of testing and estimation. In M. H. Rizvi, J. S. Rustagi, and D. Siegmund, editors, *Recent Advances in Statistics*, pages 245–269. Academic Press, 1983.

[54] M. V. Zhitlukhin and W. T. Ziemba. Exit strategies in bubble-like markets using a changepoint model. *Quantitative Finance Letters*, 4(1):47–52, 2016.

[55] S. Zou, V. V. Veeravalli, J. Li, and D. Towsley. Quickest detection of dynamic events in networks. *IEEE Transactions on Information Theory*, 66(4):2280–2295, 2020.