Statistica Sinica **21** (2011), 597-624

EARLY DETECTION OF A CHANGE IN POISSON RATE AFTER ACCOUNTING FOR POPULATION SIZE EFFECTS

Yajun Mei, Sung Won Han and Kwok-Leung Tsui

Georgia Institute of Technology

Abstract: Motivated by applications in bio and syndromic surveillance, this article is concerned with the problem of detecting a change in the mean of Poisson distributions after taking into account the effects of population size. The family of generalized likelihood ratio (GLR) schemes is proposed and its asymptotic optimality properties are established under the classical asymptotic setting. However, numerical simulation studies illustrate that the GLR schemes are at times not as efficient as two families of ad-hoc schemes based on either the weighted likelihood ratios or the adaptive threshold method that adjust the effects of population sizes. To explain this, a further asymptotic optimality analysis is developed under a new asymptotic setting that is more suitable to our finite-sample numerical simulations. In addition, we extend our approaches to a general setting with arbitrary probability distributions, as well as to the continuous-time setting involving the multiplicative intensity models for Poisson processes, but further research is needed.

Key words and phrases: Change-point, CUSUM, generalized likelihood ratio, monitoring, Poisson observations, stopping time.

1. Introduction

Early detection of abrupt changes in the properties of stochastic signals and dynamical systems finds application to on-line fault diagnosis in complex technical systems, edge detection in images, monitoring a noisy environment for target tracking or threat assessments, etc. The methodologies for detecting changes are based on results coming from sequential change-point detection, or simply change-point detection theory, which builds on the sequential analysis theory developed by Wald (1947). The classical change-point detection theory is mainly motivated from engineering and manufacturing applications such as quality control. Pioneering foundational work in the field includes Page (1954), Shiryaev (1963), Roberts (1966), Lorden (1971), Pollak (1985), Moustakides (1986), Ritov (1990), Yakir (1994), and Lai (1995). For recent reviews, we refer to Basseville and Nikiforov (1993), Lai (2001), Peskir and Shiryaev (2006), and Poor and Hadjiliadis (2008).

Change-point detection, or more generally sequential methodology, finds challenging new areas of application in the modern information age. In particular, we find that bio and syndromic surveillance provides fruitful new research opportunities for change-point detection and sequential methodology, not only as a source of new "customers" but also as an inspiration for the development of an updated theory. For introductory discussions of bio and syndromic surveillance as well as their statistical challenges, see Fienberg and Shmueli (2005), Woodall (2006), and Tsui et al. (2008).

The specific data motivating this article concerns male thyroid cancer cases (with malignant behavior) in New Mexico during 1973-2005; the data have been studied before in the biosurveillance literature in other contexts, see, for example, Kulldorff (2001) and Sonesson (2007). The data set is available from the Surveillance, Epidemiology, and End Results (SEER) Program at the National Cancer Institute that collects information on cancer incidence, mortality, and survival from the population-based cancer registries in the United States. Figure 1 plots three different curves related to this data set: (1) yearly total number of cancers with malignant behavior; (2) yearly population size (of males) in New Mexico; and (3) yearly (crude) incidence rate per 100,000 (male) population.

From the viewpoint of sequential change-point detection or sequential methodologies, the classical theories and methods are applicable to this data set when one wants to investigate whether the yearly total number of male thyroid cancer cases of this data set increases over time or not, since such a problem can be formulated as detecting a change in the mean of Poisson distributions in the discrete-time setting. From the biosurveillance viewpoint, however, a more interesting goal is to determine whether or not the *risk* for male thyroid cancer increases over time. The term *risk* here essentially means the probability of developing thyroid cancer in a given year, which can be characterized by the incidence rate per 100,000 (male) population; see the plot in the bottom panel of Figure 1. It turns out that for the problem of detecting a change in the risk, the classical change-point detection theory and methods need to be adapted to take into account the effect of population size.

The remainder of this article is as follows. Section 2 states the mathematical formulation of the problem; Section 3 offers the generalized likelihood ratio (GLR) based scheme and establishes its asymptotic optimality properties under the classical asymptotic setting. Section 4 proposes two families of schemes to take into account the effect of population size. The GLR scheme and the two proposed alternative schemes are applied to the male thyroid cancer data in Section 5, and simulation results are reported in Section 6. To gain a deeper insight, and to better reflect finite-sample numerical simulation results, Section 7 presents an asymptotic optimality theory in a new setting and studies the corresponding

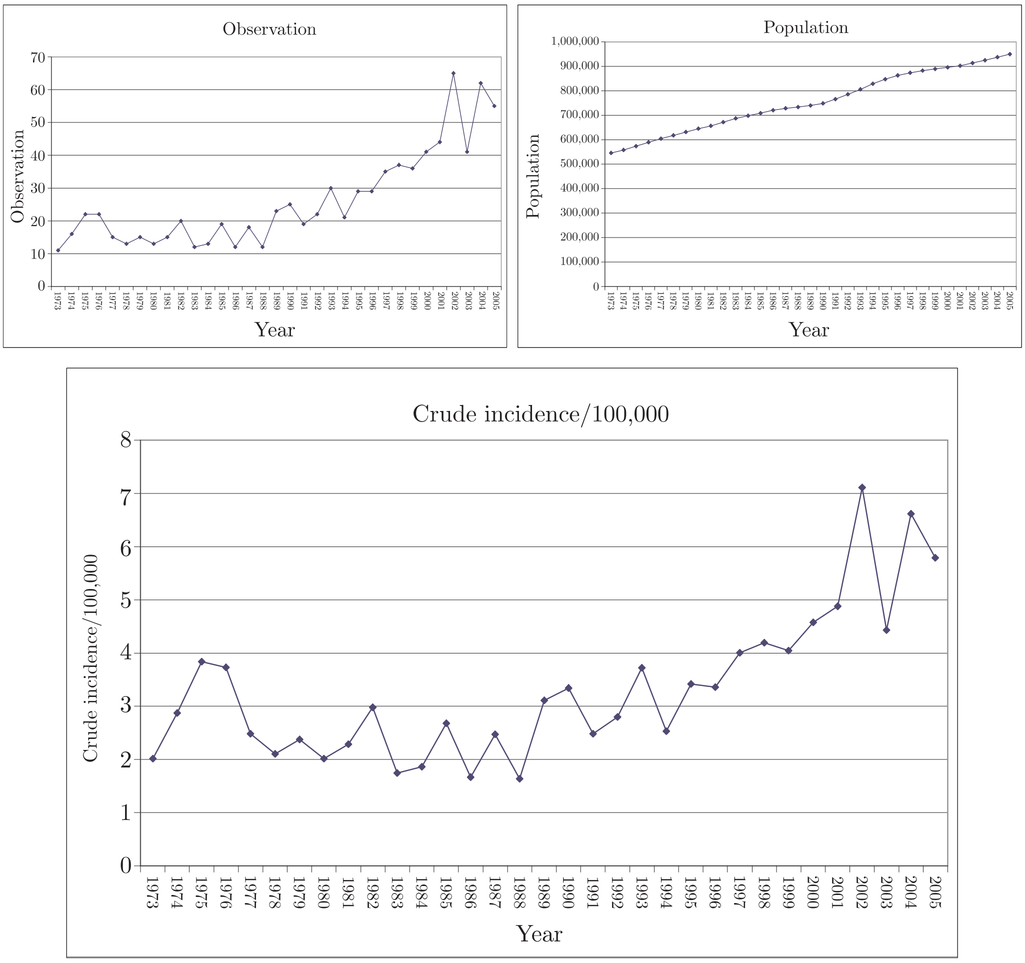

Figure 1. Three time series data of male thyroid cancer in New Mexico during 1973-2005. Top: the left panel plots the total number of male thyroid cancers over years, and the right panel illustrates the trend of the male population. Bottom: the plot is of the crude cancer incidence per 100,000 population over years.

asymptotic properties of the three proposed schemes. Section 8 discusses extensions to general probability functions and the continuous-time setting with the multiplicative intensity model for the Poisson process. Section 9 includes a second thought on optimality theory by incorporating the information of population sizes directly in the performance measures instead of the detection schemes. Section 10 contains some concluding remarks. The proofs of all theorems in Section 7 are included in the Appendix.

2. Mathematical Formulation

In the problem of detecting a change in the risk of male thyroid cancer, it is assumed that one observes two-dimensional random vectors (l_n, Y_n) over time n, where Y_n has a Poisson distribution with mean $\mu_n = l_n \lambda_n$. Here l_n, Y_n, and λ_n can be thought of as the population size (in the units of 100,000 population), the number of disease cases, and the (unobservable true) incidence rate per 100,000 (male) population at the n-th year, respectively. It is useful to think that we model the observation Y_n's by the binomial distribution with parameters l_n and λ_n, which can then be approximated by the Poisson distribution with the same mean. In addition, we also assume that the observations Y_n's are independent conditional on the population sizes l_n's.

Under a very simplified setting, the λ_n's, e.g., the incidence rates per 100,000 (male) population, are assumed to change from one value λ_0 to another value λ_1 at some unknown time ν, and we want to detect such a change as soon as possible if it occurs. Note that we are only interested in detecting a change in the risk λ_n's, and the population sizes l_n's can be either pre-specified constants or observable (possibly dependent) random variables whose distributions are nuisance parameters that are left unspecified.

In the change-point detection problem, a detection scheme is a stopping time T with respect to the observed data sequences $\{(l_n, Y_n)\}_{n\geq 1}$. That is, the decision $\{T = n\}$ only depends on observations in the first n time steps, and $\{T = n\}$ means that we raise an alarm at time n to indicate that a change has occurred somewhere in the first n time steps.

To give the change-point detection problem a rigorous formulation, denote by $\mathbf{P}_\nu$ and E_ν probabilities and expectations when the change in the risk λ_n's occurs at time ν for $\nu = 1, 2, \ldots$, and denote the same by $\mathbf{P}_\infty$ and E_∞ when $\nu = \infty$, i.e., when there are no changes in the λ_n's. Intuitively, we seek a stopping time T that makes the $\mathbf{P}_\nu$-distribution of $(T-\nu)^+$ stochastically small for all $1 \leq \nu < \infty$, subject to the constraint that the $\mathbf{P}_\infty$-distribution of T be stochastically large. Denote by $\mathcal{F}_n$ the sigma algebra generated by all observations up to time n. To incorporate the uncertainty of the change-point ν, in the literature it is standard to consider the following "worst case" detection delay criterion, proposed by Lorden (1971),

$$\overline{\mathrm{E}}_1(T) = \sup_{1\leq\nu<\infty} \text{ess sup } \mathrm{E}_\nu\Big((T-\nu+1)^+ | \mathcal{F}_{\nu-1}\Big),$$

where the essential superum takes the "worst possible observed data before the change" in the sense of providing no (or possibly false) information about the true change. A standard minimax formulation of the change-point detection problem is then to minimize the detection delay $\overline{\mathrm{E}}_1(T)$ under Lorden's criterion subject to a constraint on the average run length (ARL) to false alarm

$$\mathrm{E}_\infty(T) \geq \gamma, \tag{2.1}$$

for some given (large) constant γ.

3. The GLR Scheme and Its Asymptotic Optimality Properties

In the change-point detection problems, a basic tool to construct statistical tests or procedures is the generalized/maximum likelihood ratios (GLR) method. Note that the change-point detection problems can be thought of as testing the null hypothesis $H_0 : \nu = \infty$ (no change) against the composite alternative hypothesis $H_1 : 1 \leq \nu < \infty$ (a change occurs), the logarithm of the corresponding GLR statistic of the first n observations, $\{(l_i, Y_i)\}_{i=1}^n$, is given by

$$W_n = \max_{1 \leq \nu < \infty} \log \frac{d\mathbf{P}_\nu}{d\mathbf{P}_\infty}\Big((l_1, Y_1), \cdots, (l_n, Y_n)\Big).$$

Now given the l_i's, the Y_i's are conditionally independent with a conditional probability density function (pdf) $f_0(Y_i|l_i) = e^{-l_i\lambda_0}(l_i\lambda_0)^{Y_i}/(Y_i!)$ if $i < \nu$, but with a conditional pdf $f_1(Y_i|l_i) = e^{-l_i\lambda_1}(l_i\lambda_1)^{Y_i}/(Y_i!)$ if $i \geq \nu$. Moreover, the distribution of the l_n's is assumed to be the same under $\mathbf{P}_\infty$ or $\mathbf{P}_\nu$, and for the first n observations, $\{(l_i, Y_i)\}_{i=1}^n$, their $\mathbf{P}_\nu$-distribution is the same as their $\mathbf{P}_\infty$-distribution when $\nu > n$, due to the uniqueness of the pre-change distribution. Hence, the logarithm of the GLR statistic can be rewritten as

$$\begin{aligned} W_n &= \max_{1 \leq k \leq n+1} \sum_{i=k}^n \log \frac{f_1(Y_i|l_i)}{f_0(Y_i|l_i)} \\ &= \max_{1 \leq k \leq n+1} \sum_{i=k}^n \Big[Y_i \log \frac{\lambda_1}{\lambda_0} - l_i(\lambda_1 - \lambda_0)\Big], \end{aligned} \tag{3.1}$$

where $\sum_{i=n+1}^n = 0$. Thus, under our setting, the GLR scheme raises an alarm at time

$$T_{GLR}(a) = \text{ first } n \geq 1 \text{ such that } W_n \geq a, \tag{3.2}$$

($= \infty$ if such n does not exist), where the constant a is chosen to satisfy the false alarm constraint in (2.1). For the purpose of online implementation, it is easy to see that W_n in (3.1) enjoys a recursive formula of the classical CUSUM statistics:

$$W_n = \max\Big\{0, W_{n-1} + \Big[Y_n \log \frac{\lambda_1}{\lambda_0} - l_n(\lambda_1 - \lambda_0)\Big]\Big\}.$$

It is interesting to note that no matter whether the population sizes l_n's are (observable) random variables or pre-specified constants, the form of the GLR scheme $T_{GLR}(a)$ is the same. To gain deeper understanding of the effects of population sizes from the theoretical viewpoint, from now on we assume that the population sizes l_n's are pre-specified constants, implying that the observations Y_n's are independent but not necessarily identically distributed Poisson random variables.

The following theorem establishes the asymptotic optimality properties of the GLR scheme $T_{GLR}(a)$ under the classical asymptotic setting:

Theorem 3.1. *Assume that as $n \to \infty$, the population sizes l_i's satisfy*

$$\frac{1}{n} \sum_{i=k+1}^{k+n} l_i \to l^* > 0 \qquad \text{uniformly for all } k \geq 0. \tag{3.3}$$

Then for any stopping time $T(\gamma)$ satisfying the false alarm constraint in (2.1), we have

$$\overline{\mathrm{E}}_1(T(\gamma)) \geq (1 + o(1)) \frac{\log \gamma}{l^* I(\lambda_1, \lambda_0)}, \tag{3.4}$$

as $\gamma \to \infty$, where

$$I(\lambda_1, \lambda_0) = \lambda_1 \log\left(\frac{\lambda_1}{\lambda_0}\right) - (\lambda_1 - \lambda_0). \tag{3.5}$$

Moreover, the family of the GLR schemes $\{T_{GLR}(a)\}$ in (3.2) attains the information bound (3.4) asymptotically.

Proof. The proof follows the lines of Lorden (1971) or Lai (1998), and thus is omitted here (similar arguments are in the proofs of Section 7). Also see Yao (1993) for similar results in the context of linear regression for normal distributions.

Note that the uniform convergence assumption in (3.3) is needed to prove the asymptotic optimality of the GLR schemes $\{T_{GLR}(a)\}$, but a much weaker condition is sufficient to derive the information bound (3.4). See Section 7 below for more discussion of the assumption (3.3).

4. Two Alternative Methods

So far we have "solved" the problem by our favorite GLR methods, but perhaps not solved it in practice. To illustrate that despite its nice asymptotic optimality properties, the GLR scheme may not necessarily be as effective as one expects in application, we propose two ad-hoc methods for comparison.

Intuitively, two features of the GLR scheme seem questionable in the context of non-stationary population sizes: (i) the GLR statistic W_n in (3.1) assigns the same weight to the individual log-likelihood ratio statistic $\log f_1(Y_i|l_i)/f_0(Y_i|l_i)$ regardless of population size l_i's, although the Y_i's with larger population sizes l_i's surely provide more information, and (ii) the GLR scheme $T_{GLR}(a)$ uses the constant threshold value a over time. Accordingly, we propose two alternative detection schemes to take into account the effects of population sizes.

The first scheme is based on the quasi-log-likelihood ratio statistics that normalize each term $\log f_1(Y_i|l_i)/f_0(Y_i|l_i)$ in (3.1) by their (conditional) variances, or equivalently, by the population sizes l_i's (up to a constant). This leads to the detection statistic

$$\begin{aligned}\hat{W}_n &= \max_{1\le k\le n+1} \sum_{i=k}^{n} \frac{1}{l_i} \log \frac{f_1(Y_i|l_i)}{f_0(Y_i|l_i)} \\ &= \max_{1\le k\le n+1} \sum_{i=k}^{n} \Big[\frac{Y_i}{l_i} \log \frac{\lambda_1}{\lambda_0} - (\lambda_1 - \lambda_0)\Big]. \end{aligned} \tag{4.1}$$

Thus, for any given constant b, we can define the weighted likelihood ratio (WLR) scheme

$$T_{WLR}(b) = \text{ first } n \ge 1 \text{ such that } \hat{W}_n \ge b. \tag{4.2}$$

Another motivation for the WLR scheme $T_{WLR}(b)$ in (4.2) is based on Y_n/l_n, a natural estimator of the *risk* or the disease rate per 100,000 population. To see this, if we pretend that Y_n/l_n is Poisson distributed with mean λ_n (this is not true under our setting, but we can still use it to construct detection schemes), then the problem becomes the classical problem of detecting a change in the Poisson mean from λ_0 to λ_1, and the corresponding GLR (or CUSUM) procedure is just the WLR scheme $T_{WLR}(b)$ in (4.2).

The second scheme we propose is to use the GLR-based statistic W_n in (3.1), but with adaptive thresholds to take into account population size effects. Ideally, one would like to use the optimal thresholds or boundaries, say, by some Bayesian or non-Bayesian arguments, but such boundaries seem to be too complicated to derive explicitly. For simplicity, we use the linear boundaries: $l_n c$ (see Section 8.1 below for more explanation). Specifically, the proposed adaptive threshold method (ATM) raises an alarm at time

$$T_{ATM}(c) = \text{ first } n \ge 1 \text{ such that } W_n \ge l_n c, \tag{4.3}$$

for some constant $c > 0$, where W_n is the GLR statistic defined in (3.1).

It is important to point out that when the population sizes l_n's are equal to a constant $l > 0$, then the three detection schemes, $T_{GLR}(a), T_{WLR}(b)$ and

$T_{ATM}(c)$, not only are equivalent (when $a = lb = lc$), but also hold the exact optimality properties of Page's CUSUM procedures proved in Moustakides (1986) and Ritov (1990).

When the population sizes l_n's vary, the conclusions are obscure. On the one hand, under the uniform convergence assumption of Theorem 3.1, the population sizes l_n's converge to a constant value l^*, and thus these three schemes seem to be equivalent and efficient from the asymptotic viewpoint. On the other hand, when the population sizes l_n's vary, these detection schemes are generally not equivalent and thus likely have different finite-sample properties; this is the focus of the remainder of this article.

5. Example Revisited

The purpose of this section is to illustrate that the GLR scheme does not necessarily work as effectively as the ad hoc schemes WLR and ATM for the male thyroid data in New Mexico. Of course, we should acknowledge that our proposed methods are oversimplified for this data set. Nevertheless, as a starting point, we can look at the performances of the three procedures for this data set. Other simulation results will be presented in the next section.

5.1. Model for population growth

In our application or in simulations, we need a model to generate population sizes beyond the observed ones so that we can determine the threshold values of detection schemes that satisfy the false alarm constraint (2.1). In the literature, it is common (e.g., Pinheiro and Bates (2000)) to model the growth curve by the logistic model

$$l_n = \psi(n) + \epsilon_n = \frac{\phi_1}{1 + \exp[-(n - \phi_2)/\phi_3]} + \epsilon_n, \tag{5.1}$$

where $E[\epsilon_n] = 0$ and $Var[\epsilon_n] = \sigma^2$. Here ϕ_1 indicates an asymptotic upper limit of population size, ϕ_2 the middle point of the S-shaped curve, and ϕ_3 the scale adjustment of time periods.

In our specific application, we fit (5.1) to the observed population sizes by a nonlinear least-squares method (we treat year 1972 as time 0, and the population sizes are in the units of 100,000). Using the statistical software R version 2.8.0, the estimated parameters for the logistic model (5.1) are summarized in Table 1.

Figure 2 plots the actual observed population sizes and the estimated growth curve in New Mexico during 1973-2005. From the plot, one sees that the two curves are close to each other, implying that the logistic model is reasonable in our application. One may also wonder that population sizes seem to increase linearly and ask why not just fit a linear model for population sizes? The answer

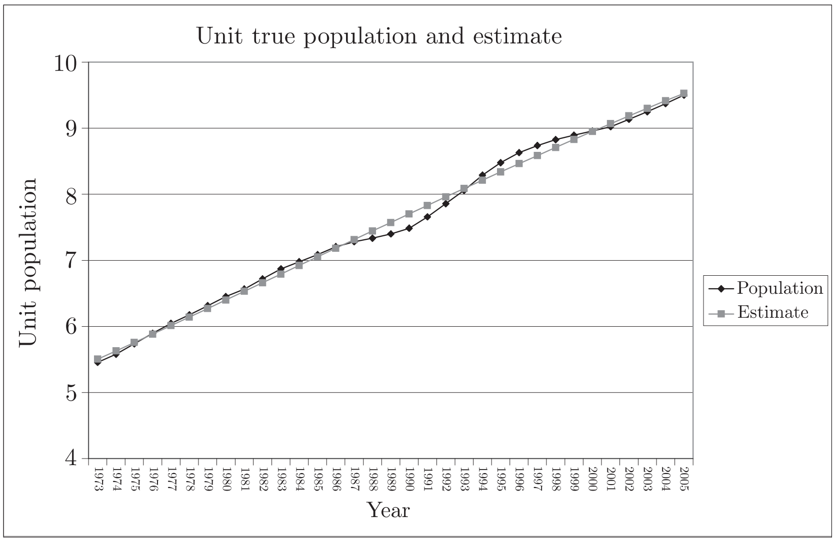

Figure 2. Population and estimate in New Mexico during 1973-2005. The plot shows the actual observed and model-estimated male population sizes in New Mexico during 1973-2005.

Table 1. Estimated parameters for the population growth model in (5.1)

Parameter	ϕ_1	ϕ_2	ϕ_3	σ
Estimate	13.8065 $\pm$ 0.9552	11.8532 $\pm$ 3.7438	26.4037 $\pm$ 2.3127	0.0907

depends on how large the false alarm constraint γ in (2.1) is. When γ is around 30, the linear model may be reasonable, since the observed 32 population sizes indeed seems to be linear. However, if γ is moderately large, say 100, it seems unrealistic to assume that population sizes increase linearly over 100 years, and the logistic model (5.1) may be more suitable. In our simulations, we are more interested in a moderately large value of γ, and thus we use the mean curve of (5.1) with the estimated parameters in Table 1 to generate population sizes.

5.2. Parameters in the change-point problem and detection schemes

To implement our schemes, we need to specify the pre-change rate λ_0 and the post-change rate λ_1. This is not an easy task, especially in the context of bio and syndromic surveillance where the baseline model is not well-defined. To have some estimated values (not necessarily the best ones) for these parameters, one possible approach is to use the time period 1973-1983 as a training period, and then to estimate the pre-change rate λ_0 and the post-change rate λ_1 by the median and the maximum of the crude incidence rate per 100,000 during 1973-

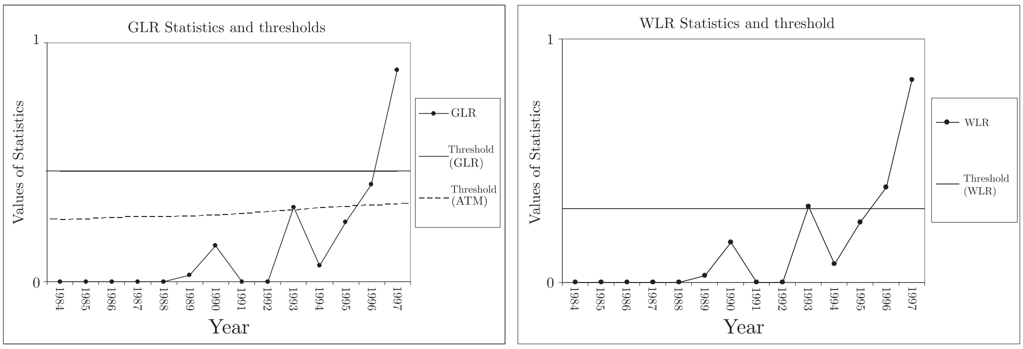

Figure 3. The *left* panel plots the GLR statistics W_n over time n, as well as the alarm boundaries of $T_{GLR}(a)$ (solid line) and $T_{ATM}(c)$ (the dotted line) when $\gamma = 300$. The *right* panel plots the WLR statistic $\hat{W}_n$ as well as the boundary of $T_{WLR}(b)$.

1983, respectively. For our data, we have $\hat{\lambda}_0 = 2.4$ and $\hat{\lambda}_1 = 3.8$. Thus, we chose $\lambda_0 = 2.4$ and $\lambda_1 = 3.8$ when implementing our proposed schemes. These choices are intended only for illustrations.

We also need to specify the false alarm constraint γ in (2.1), since the choice of γ clearly affects the detection thresholds a, b, and c in the proposed schemes. Unfortunately, there are no well-accepted guidelines to choose the constraint γ. Here we tried $\gamma = 100, 200$ or 300. That is, on average we wanted all detection schemes to take at least 100 (or 200 or 300) years before raising the first false alarm when the risk of cancer is $\lambda_0 = 2.4$ per 100,000 population and when there are no changes on the disease risk. Numerical simulation was then conducted to find the detection threshold values to satisfy the false alarm constraint γ in (2.1) with each different value of γ when population sizes were generated from (5.1) with the parameters in Table 1. For instance, when $\gamma = 300$, the corresponding threshold values for $T_{GLR}(a)$ in (3.2), $T_{WLR}(b)$ in (4.2) and $T_{ATM}(c)$ in (4.3) were $a = 3.6870, b = 0.2975$, and $c = 0.2975$, respectively (based on 100,000 replicates). When $\gamma = 100$ or 200, the detection threshold values were smaller than those for $\gamma = 300$, but not by much.

5.3. When to raise alarm?

We then applied the GLR scheme and two alternative schemes, the WLR and ATM, to monitor the cancer risk for the male thyroid cancer data in New Mexico starting from year 1984. Figure 3 plots the GLR statistics W_n in (3.1) and the WLR statistic $\hat{W}_n$ in (4.1) over the time, as well as the detection boundaries of the schemes when the false alarm constraint is $\gamma = 300$. From the plots, the WLR and ATM schemes, $T_{WLR}(b)$ and $T_{ATM}(c)$, trigger an alarm in year 1993, whereas the GLR scheme $T_{GLR}(a)$ raises an alarm in year 1997. In addition, when the false alarm constraint $\gamma = 100$ or 200, we reached the same conclusion. These results seem to suggest that for the male thyroid cancer data in New Mexico, the ad-hoc alternative schemes, $T_{WLR}(b)$ and $T_{ATM}(c)$, are better than the GLR scheme $T_{GLR}(a)$ in the sense of raising an alarm earlier when the false alarm constraint $\gamma = 100, 200$ or 300.

6. More Simulation Study

In this section we perform further simulation studies to compare the GLR scheme with two alternatives. In particular, we want to see whether or not the poor performance of the GLR scheme is a fluke.

In this numerical simulation study, we borrowed from the previous section, but with two modifications. The first one is that we took the post-change risk $\lambda_1 = 2.7$ instead of $\lambda_1 = 3.8$ (we still held the pre-change risk at $\lambda_0 = 2.4$). The reason was to investigate a smaller change, since a larger change may be easily detected by any reasonable method. Indeed, in the previous section the detection statistics W_n or $\hat{W}_n$ were 0 before crossing the detection thresholds for the WLR and ATM detection schemes, indicating that the detection delays of these methods is very small (at most 1) due to a larger change. The second modification was to increase the false alarm constraint to $\gamma =$ 1,000. We hoped that a larger value of γ might lead to a larger detection delay, so that we could better understand the properties of the three detection schemes.

In our simulations, we took the population sizes from the logistic model in (5.1) with different parameter values. The following three cases were studied ($\hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3$ and $\hat{\sigma}$ are the estimates in Table 1):

- Case A (Increasing): $l_n = \hat{\phi}_1/\{1 + \exp[-(n - \hat{\phi}_2)/\hat{\phi}_3]\}$; this is the model we used in our data example.
- Case B (Fast Increasing): $l_n = 2\hat{\phi}_1/\{1 + \exp[-(n - (\hat{\phi}_2 + 26))/\hat{\phi}_3]\}$, where $\hat{\phi}_2 + 26$ is chosen so that population sizes l_n's increase quickly over time to a stationary value, as compared to the data application.
- Case C (Decreasing): $l_n = [\hat{\phi}_1/2.4]/\{1 + \exp[(n - \hat{\phi}_2)/\hat{\phi}_3]\} + 1$; this leads to population sizes l_n that decrease over time. Here the constant 1 ensures that the population size l_n decreases to a nonzero constant.

Note that in the models of population sizes, $2\hat{\phi}_1$ in Case B and $\hat{\phi}_1/2.4$ in Case C are necessary to make sure that the initial population size l_0 is the same as the observed value l_0. The *top left* panel of Figure 4plots the three population size models.

For each population size model, we first determined the detection threshold values through 100,000 replicates, so that the detection schemes satisfy the

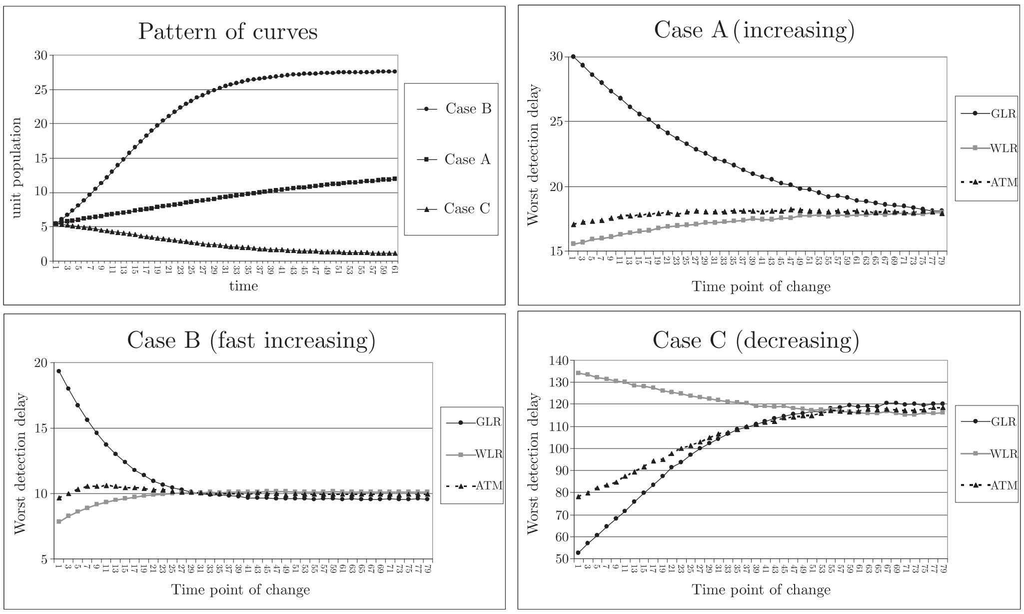

Figure 4. The population sizes are from the smooth model in (5.1). The *top left* panel plots three different population size curves that correspond to the three cases considered. The other three panels illustrate the detection delays of the three detection schemes as a function of change-point ν under different cases of the population size models.

false alarm constraint (2.1) with $\gamma \approx 1{,}000$ (within sampling error) and the pre-change risk $\lambda_0 = 2.4$. Then we simulated the detection delay ess sup$\mathrm{E}_\nu\big((T-\nu+1)^+\big|\mathcal{F}_{\nu-1}\big)$ at different change-point ν with the post-change risk $\lambda_1 = 2.7$. The simulated detection delays were based on 50,000 replicates.

The detection delays of the three schemes are plotted in Figure 4 for the three population size models. From Figure 4, it is interesting to note that if the population sizes are decreasing (Case C), the GLR scheme $T_{GLR}(a)$ seems to be the best in the sense of smallest detection delay at each change-point ν, while the WLR scheme $T_{WLR}(b)$ seems to be the worst. However, if the population sizes are increasing (Cases A and B), the order is reversed.

Under Lorden's worst-case detection delay $\overline{\mathrm{E}}_1(T)$ criterion, similar conclusions still hold. Specifically, the GLR scheme is the best scheme with the smallest worst-case detection delay $\overline{\mathrm{E}}_1(T)$ if the population sizes are decreasing, but it is the worst scheme if the population sizes are increasing; the WLR scheme is the worse scheme if the population sizes are decreasing, but the best if the population sizes are increasing. The adaptive threshold scheme $T_{ATM}(c)$ seems to be *robust* in the sense of small detection delays $\overline{\mathrm{E}}_1(T)$ under Lorden's criterion, no matter whether the population sizes are increasing or decreasing.

It is also interesting to see from Figure 4 that for $T_{GLR}(a)$, the detection delays ess sup$\mathrm{E}_\nu\big((T-\nu+1)^+\big|\mathcal{F}_{\nu-1}\big)$ seem to be decreasing (increasing) as a function of the change-point ν when the populations sizes are increasing (decreasing). However, the detection delays of the WLR scheme $T_{WLR}(b)$ seem to be an increasing (decreasing) function of the change-point ν when the populations sizes are increasing (decreasing). In all simulations, the three schemes had similar detection delays, ess sup$\mathrm{E}_\nu\big((T-\nu+1)^+\big|\mathcal{F}_{\nu-1}\big)$, when the change-point ν was large, but they had very different detection delays when the change-point ν occurred at an earlier stage.

In summary, our simulations suggest that when one has prior information that population sizes are increasing or decreasing, one use the best among these three schemes. When there is uncertainty about the trends of population sizes, one may want to use the adaptive threshold scheme $T_{ATM}(c)$ to take advantage of its robustness properties. In particular, despite its asymptotic optimality properties, the GLR scheme indeed can perform very poorly in finite-sample numerical simulations, especially in the typical scenarios of biosurveillance when the population sizes are increasing.

7. New Asymptotic Analysis

The main purpose of this section is to develop some theory to “explain” our simulation results in the previous section. To find an appropriate setting, we note that the GLR scheme is at times efficient and at times inefficient in our simulations, and thus it is natural reaction to check whether the condition (3.3) required in Theorem 3.1 holds or not. On the one hand, (3.3) holds since in our simulations, the population sizes are from the logistic model and monotonically increase or decease to the stationary value l^*. Hence the GLR scheme is asymptotically optimal in our simulation study, regardless of whether the population sizes are increasing or decreasing. On the other hand, our simulations violate the *spirit* of (3.3). To be more specific, the false alarm constraint γ is only moderately large in our simulation in view of the uniform convergence in (3.3). That is, since the false alarm constraint γ is not too large, the post-change sample size n is generally not too large, and thus the value of $(1/n)\sum_{i=k+1}^{k+n} l_i$ for small k can be very different from the corresponding value for large k.

We now assume that the population sizes l_n's reach the stationary value l^* at some finite time ω. Then Theorem 3.1 deals with ω much much smaller than γ, i.e., fix ω and let the false alarm constraint γ go to ∞. However, simulations had ω comparable to the false alarm constraint γ. For instance, when fitting the logistic model to the data, the population sizes were close to the stationary value, numerically, around time $\omega \approx 120$ while the false alarm constraint $\gamma = 1{,}000$.

In order to reflect finite-sample numerical results, we consider a new asymptotic setting in which the population sizes reach the stationary value l^* at some finite time ω, where $\omega = \omega_\gamma \leq C\gamma$ for some constant $0 < C < 1$, as the false alarm constraint γ goes to ∞. Note then that as $n \to \infty$, $(1/n)\sum_{i=k+1}^{k+n} l_i$ converges to l^* point-wise for each k, but this convergence is no longer uniform over k as required by (3.3). Section 7.1 presents an asymptotic theory under the new setting, and Section 7.2 reports on further simulations to illustrate our theory. The proofs of theorems in this section are postponed to the appendix.

7.1. Asymptotic optimality analysis

We first construct an asymptotic lower bound on the detection delays as the false alarm constraint γ in (2.1) goes to ∞. Then we see whether a specific family of schemes attains the lower bound asymptotically.

Theorem 7.1. *Assume that the population sizes l_n's reach the stationary value $l^* > 0$ at some finite time ω, where $\omega = \omega_\gamma < C\gamma$ for some constant $0 < C < 1$. Then for any stopping time $T(\gamma)$ satisfying the false alarm constraint in* (2.1), (3.4) *holds.*

Theorem 7.2. (i) *For the GLR scheme, we have* $\mathrm{E}_\infty(T_{GLR}(a)) \geq \exp(a)$ *for all* $a > 0$.
(ii) *For the WLR scheme, if* $\inf_{n\geq 1} l_n = l_* > 0$, *then as* $b \to \infty$,

$$\overline{\mathrm{E}}_1(T_{WLR}(b)) \leq \frac{b}{I(\lambda_1, \lambda_0)} + M,$$

where $I(\lambda_1, \lambda_0)$ is defined at (3.5) *and*

$$M = \sqrt{1 + \frac{\lambda_1}{l_*}\Big(\lambda_1 - \frac{\lambda_1 - \lambda_0}{\log\lambda_1 - \log\lambda_0}\Big)^{-2}}.$$

In order to derive the asymptotic optimality properties of the GLR or WLR schemes subject to the false alarm constraint γ in (2.1), we need to derive the detection delays of the GLR scheme $T_{GLR}(a)$ and/or to establish the relationship between the threshold value b in the WLR scheme $T_{WLR}(b)$ and the false alarm constraint γ. These are challenging problems. In addition, for the ATM scheme, it is non-trivial to investigate its false alarm or detection delay properties. We make two simplifications in order to go forward.

First, we focus on two kinds of changes on the disease risk: when the change occurs at time $\nu = 1$ and when the change occurs at $\nu = \omega$. The change-point $\nu = 1$ is used to indicate the detection delays for small values of change-point ν; the change-point $\nu = \omega$ is interesting because the detection delay ess sup $\mathrm{E}_\nu\big((T-$

$\nu + 1)^+ | \mathcal{F}_{\nu-1}\big)$ is the same for all change-point $\nu \geq \omega$, since the observations Y_n are i.i.d. with the same population sizes $l_n = l^{(1)}$ for $n \geq \omega$. This motivates us to consider

$$D(T) = \max\left[\mathrm{E}_{\nu=1}(T),\ \text{ess sup } \mathrm{E}_{\omega}\Big((T - \omega + 1)^+ | \mathcal{F}_{\omega-1}\Big)\right],$$

which provides a lower bound on Lorden's worst-case detection delay $\overline{\mathrm{E}}_1(T)$.

The second simplification is to assume that the initial population size l_n are constant, $l^{(0)}$, for a reasonably long period so that we can use the classical results on Page's CUSUM procedures to derive the detection delay $\mathrm{E}_{\nu=1}(T)$ for the GLR and ATM schemes. Specifically, we now assume that population sizes can be modeled by the step function

$$l_n = \begin{cases} l^{(0)}, & \text{if } n < \omega', \\ \text{some values}, & \text{if } \omega' \leq n < \omega, \\ l^*, & \text{if } n \geq \omega, \end{cases} \tag{7.1}$$

where $\omega' >> \log \gamma$ and $\omega < C\gamma$ for some $0 < C < 1$, and the l_n between time $[\omega', \omega]$ can be some arbitrary positive values bounded between two pre-specified positive constants L_0 and L_1.

There is a subtle but important issue on the step function (7.1) for the population sizes model: we do not want ω' in (7.1) to be too large compared to the false alarm constant γ, since otherwise the problem is essentially the classical change-point detection problem with constant population sizes $l^{(0)}$. Thus, to be meaningful, we need to make sure that, for the proposed three schemes, the false alarms during the initial stage with the constant population $l^{(0)}$ are negligible as compared to those during the latter stage with the constant population l^*. We assume $\omega - \omega' = O(1)$ and $\omega' = o(\gamma^{(1-\eta)l^{(0)}/l^*})$ for some constant $0 < \eta < 1$. These additional assumptions are a little more restrictive than what one may prefer, but they allow us to derive the asymptotic optimality properties of the three proposed schemes.

Theorem 7.3. *Assume that the population sizes* l_n *obey* (7.1) *such that* $\inf_{n \geq 1} l_n = l_* > 0$, *and the observable times* $\omega' = \omega'_\gamma, \omega = \omega_\gamma$ *and the false alarm constraint* γ *in* (2.1) *satisfy* $\log \gamma << \omega' \leq \omega < C\gamma,\ \omega - \omega' = O(1)$ *and* $\omega' = o(\gamma^{(1-\eta)l^{(0)}/l^*})$ *for some* $0 < \eta < 1$ *as* $\gamma \to \infty$. *Then, subject to the false alarm constraint in* (2.1),

$$D[T_{GLR}(a)] = (1 + o(1)) \frac{\log \gamma}{\min\{l^{(0)}, l^*\} I(\lambda_1, \lambda_0)},$$
$$D[T_{WLR}(b)] = (1 + o(1)) \frac{\log \gamma}{l^* I(\lambda_1, \lambda_0)},$$

$$D[T_{ATM}(c)] = (1 + o(1)) \frac{\log \gamma}{l^* I(\lambda_1, \lambda_0)},$$

as $\gamma \to \infty$, *where* $I(\lambda_1, \lambda_0)$ *is defined at* (3.5).

We can now state the asymptotic optimality or sub-optimality of the WLR and GLR schemes under Lorden's worst-case detection delay criterion.

Corollary 7.1. *Under the assumptions of Theorem* 7.3*, the WLR scheme* $T_{WLR}(b)$ *is asymptotically optimal under Lorden's worst-case detection delay criterion* $\overline{\mathrm{E}}_1(T)$*, subject to the false alarm constraint in* (2.1)*. Moreover, when* $l^{(0)} < l^*$*, the GLR scheme* $T_{GLR}(a)$ *is asymptotically* suboptimal *under Lorden's worst-case detection delay criterion.*

Proof. For the WLR scheme $T_{WLR}(b)$, by Theorem 7.3, the threshold $b \sim \log \gamma$ is sufficient to satisfy the false alarm constraint γ in (2.1). Combing this with Theorem 7.2 yields $\overline{\mathrm{E}}_1(T_{WLR}(b)) \le (1 + o(1)) \log \gamma / (l^* I(\lambda_1, \lambda_0))$. Therefore, the WLR scheme attains the lower bound on the detection delays in Theorem 7.1, and is asymptotically optimal under Lorden's worst-case detection delay criterion $\overline{\mathrm{E}}_1(T)$.

Meanwhile, if $l^{(0)} < l^*$, the suboptimality properties of the GLR scheme $T_{GLR}(a)$ follow at once from Theorem 7.3, its comparison with the asymptotic optimal scheme $T_{WLR}(b)$, and the fact that $\overline{\mathrm{E}}_1(T) \ge D(T)$:

$$\lim_{\gamma \to \infty} \frac{\overline{\mathrm{E}}_1(T_{GLR}(a))}{\overline{\mathrm{E}}_1(T_{WLR}(b))} \ge \lim_{\gamma \to \infty} \frac{D(T_{GLR}(a))}{(1 + o(1)) \log \gamma / (l^* I(\lambda_1, \lambda_0))} = \frac{l^*}{\min(l^{(0)}, l^*)} > 1,$$

completing the proof.

Remarks. The above results are consistent with our finite sample simulation results: when the population sizes increase, the WLR scheme is the best scheme and the GLR scheme is the worst. On the other hand, when the population sizes decrease, i.e., when $l^{(0)} > l^*$, the three detection schemes are (first-order) asymptotically equivalent in that $D(T) \sim \log \gamma / [l^* I(\lambda_1, \lambda_0)]$ for any of them. However, our simulations in Section 6 do not support this claim. This is partly because the asymptotic optimality theorem for the WLR scheme requires that all population sizes l_n's are bounded below by $l_* > 0$, but the lower bound l_* in our numerical simulations is too small ($l_* = 1$). We conduct a further simulation below to check the case of $l^{(0)} > l^*$.

Note that $T_{WLR}(b)$ and $T_{ATM}(c)$ are (first-order) "equalizer rules" in the sense that ess sup $\mathrm{E}_\nu \Big((T - \nu + 1)^+ | \mathcal{F}_{\nu - 1} \Big)$ is the same (up to the first-order) for $\nu = 1$ or ω, but the $T_{GLR}(a)$ is not. The property of "equalizer rule" is

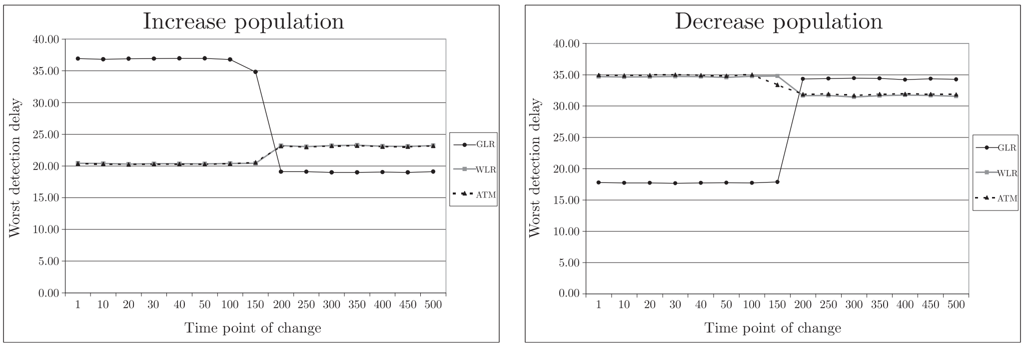

Figure 5. The detection delays of the three proposed detection schemes at different change-points ν when the population sizes are given by the step functions. Left Panel: the step function is increasing. Right Panel: the step function is decreasing.

essential to the establishment of the exact optimality of the CUSUM procedure in the simplest i.i.d. models, and our results suggest that $T_{GLR}(a)$ may lose this property in the finite-sample setting when the population sizes vary.

7.2. Numerical simulations

Here we conduct a numerical study when the population sizes are modeled by (7.1) with $\omega' = \omega = 200$. We assume that the false alarm constraint is $\gamma = 1{,}000$, so that the choice of $\omega' = \omega = 200$ is consistent with $\log \gamma << \omega' \leq \omega < C\gamma$ for some $0 < C < 1$. Two cases are considered: increasing with $l^{(0)} = 6$ and $l^* = 12$, and decreasing with $l^{(0)} = 12$ and $l^* = 6$. As in Section 5, we assume that the pre-change and post-change risks are $\lambda_0 = 2.4$ and $\lambda_1 = 2.7$, respectively.

For the increasing case, our simulations showed that, subject to the false alarm constraint (2.1) with $\gamma \approx 1{,}000$, the threshold values for $T_{GLR}(a)$ in (3.2), $T_{WLR}(b)$ in (4.2) and $T_{ATM}(c)$ in (4.3) were $a = 4.540, b = 0.453$, and $c = 0.452$, respectively. For the decreasing case, the corresponding thresholds were $a = 4.265, b = 0.661$, and $c = 0.665$, respectively.

Figure 5 illustrates the detection delay at different change-points ν for each of the three detection schemes. For the increasing case, note that the detection delay of the GLR scheme $T_{GLR}(a)$ is a decreasing function of ν (decreasing from 36.9 ± 0.1 at $\nu = 0$ to 19.1 ± 0.1 at $\nu = 200$). However, the (simulated) detection delay of either $T_{WLR}(b)$ or $T_{ATM}(c)$ is an increasing function of ν (increasing from 20.4 ± 0.1 at $\nu = 0$ to 23.1 ± 0.1 at $\nu = 200$). Hence, under Lorden's worst-case detection delay criterion, the worst-case detection delays $\overline{\mathrm{E}}_1(T)$ of $T_{GLR}(a), T_{WLR}(b)$ and $T_{ATM}(c)$ were $36.9 \pm 0.1, 23.1 \pm 0.1, 23.1 \pm 0.1$, respectively. In the increasing case, the schemes $T_{WLR}(b)$ and $T_{ATM}(c)$ are better than the scheme $T_{GLR}(a)$ in the sense of smaller detection delays under Lorden's worst-case detection delay criterion; this is consistent with our asymptotic theory for the increasing population sizes with $l^{(0)} < l^*$.

In the decreasing case, we get the reverse pattern: the detection delay is an increasing function of the ν for $T_{GLR}(a)$, but a decreasing function of ν for either $T_{WLR}(b)$ or $T_{ATM}(c)$. Under Lorden's worst-case detection delay criterion, the worst-case detection delays $\overline{E}(T)$ of $T_{GLR}(a), T_{WLR}(b)$ and $T_{ATM}(c)$ were $34.4 \pm 0.1, 35.0 \pm 0.1, 34.7 \pm 0.1$, respectively. Hence, in the decreasing case, despite significantly different (individual) detection delay curves as illustrated in the right panel of Figure 5, the three schemes have similar properties under Lorden's criterion. This is also consistent with the asymptotic theory developed in Section 7.1 for the decreasing case.

Finally, Figure 5 shows that, as compared to the GLR scheme, the detection delay curves for $T = T_{WLR}(b)$ or $T_{ATM}(c)$ look to flatter with respect to ν, no matter whether the population sizes are increasing or decreasing. This supports our claim that the GLR scheme $T_{GLR}(a)$ is not an "equalizer rule" even at the first-order.

8. Extensions

Here we discuss two extensions of our methods and ideas: the first is to allow other distributions than the Poisson, and the second is to the continuous-time setting with non-homogenous Poisson processes.

8.1. General setting

We begin by noting that if one uses some variance-stabilizing transformation, e.g., the square root $Y_n^* = 2\sqrt{Y_n}$ or the Anscombe transformation $Y_n^* = 2\sqrt{Y_n + 3/8}$, that transforms Poisson observations (with large mean) into approximate normal observations with constant variance, then the problem of detecting disease risks becomes the problem of detecting a change in slope in a simple linear regression where the (independent) observations Y_n^* are i.i.d. $N(x_i\beta_0, 1)$ before the change and i.i.d. $N(x_i\beta_1, 1)$ after the change, where $x_i = 2\sqrt{l_i}$, $\beta_0 = \sqrt{\lambda_0}$, and $\beta_1 = \sqrt{\lambda_1}$. Yao (1993) shows that under (3.3), the GLR scheme is asymptotically optimal as the false alarm constraint γ goes to ∞. Still, since the post-change distributions depend on the change-point ν through non-homogeneous population sizes, even for normal distributions, some modifications of the GLR scheme can be made to improve finite sample performances.

In general, suppose that there are two sequences of known densities: $\{f_n\}_{n=1}^{\infty}$ and $\{g_n\}_{n=1}^{\infty}$ with $f_n \neq g_n$. Assume that we observe a sequence of independent random variables $Y_1, Y_2, \ldots$ such that the density of Y_n is f_n if $n < \nu$ and g_n if

$n \geq \nu$, where the change-point ν is unknown. Under this general setting, the GLR and WLR statistics can be written as

$$W_n = \max_{1 \leq k \leq n+1} \sum_{i=k}^{n} \log \frac{g_n(Y_n)}{f_n(Y_n)} \quad \text{and} \quad \hat{W}_n = \max_{1 \leq k \leq n+1} \sum_{i=k}^{n} \frac{1}{I(g_n, f_n)} \log \frac{g_n(Y_n)}{f_n(Y_n)},$$

where $I(g_n, f_n) = \mathrm{E}_{g_n}\big\{\log[g_n(Y)/f_n(Y)]\big\}$ is the Kullback-Leibler (K-L) information number. Hence, the GLR and WLR schemes can be taken as comparing appropriate statistics with some constant thresholds. For the ATM scheme, a simple choice of the threshold boundary uses the K-L information number by choosing $c_n = I(g_n, f_n)c$ for some constant $c > 0$, which can lead to "equalizer rules" under certain conditions.

8.2. Continuous-time model: multiplicative intensity model

We have studied the change-point detection problems in the discrete-time setting and now look at the corresponding continuous-time version involving a Poisson process. This may be useful in the modern information age when bio-surveillance systems can observe events (cancer cases) in (nearly) continuous time. The problem of detecting the intensity of a Poisson process has been studied in the literature, and some recent advances can be found in Baron and Tartakovsky (2006) and Peskir and Shiryaev (2006). We look to extend the Poisson process model to the multiplicative intensity model of Aalen (1978) that has been applied to survival analysis, birth and death processes, and time-continuous Markov chains on finite state spaces.

In the simplest form of the multiplicative intensity model, one observes counting processes $Y = (Y_t)_{t \geq 0}$ and $l = (l_t)_{t \geq 0}$, where $Y = (Y_t)_{t \geq 0}$ is a Poisson process with intensity $\Lambda(t) = \lambda l_t$ for some nonnegative constant λ. In the context of change-point detection, it is assumed that λ changes from λ_0 to λ_1 at some unknown time $\nu > 0$. Based on the continuously observed trajectories of both Y and l, one wants to raise an alarm as soon as possible after the change occurs.

It is not difficult to see that, for this continuous-time model, the log-likelihood process in testing $\mathbf{P}_\infty$ (i.e., no change) against $\mathbf{P}_{\nu=0}$ (i.e., a change occurs at time 0) is

$$Z_t = \log(\frac{\lambda_1}{\lambda_0})Y_t - (\lambda_1 - \lambda_0)\int_0^t l_s ds.$$

Hence, the log-GLR process is given by $W_t^* = Z_t - \inf_{0 \leq s \leq t} Z_s$. Similarly, the continuous-time version of the WLR scheme is based on the process $\hat{W}_t^* = \hat{Z}_t - \inf_{0 \leq s \leq t} \hat{Z}_s$, where the "normalized" log-likelihood ratio process is

$$\hat{Z}_t = \log(\frac{\lambda_1}{\lambda_0})\frac{Y_t}{l_t} - (\lambda_1 - \lambda_0)t.$$

With these detection statistics, it is straightforward to define the corresponding GLR, WLR and ATM detection schemes in the continuous-time version.

In biosurveillance applications, a typical scenario has $\lambda_1 > \lambda_0$ (risk increases) when the population size l_t is a non-decreasing function of time t. From the asymptotic viewpoint, the arguments for the discrete-time model can be easily extended to the continuous-time model in this case, since it is asymptotically optimal to raise alarms only at the times when a new event Y occurs. As in the discrete-time models, the detection schemes $T^*_{WLR}(b)$ and $T^*_{ATM}(c)$ seem to be more efficient than the GLR scheme $T^*_{GLR}(a)$ in the finite sample setting (it is straightforward if l_t is a step function, and it would be interesting to find a rigorous proof of this in general).

It is also be interesting to investigate the situation if $\lambda_1 < \lambda_0$ or if l_t is decreasing. In these scenarios, one may need to raise an alarm even though no new events Y arrive. Hence it is very challenging to develop the optimal detection schemes under either the minimax or Bayesian formulation, and it is beyond the scope of this article.

9. A Second Thought on the Optimality Theory

We have looked at procedures evaluated by two classical performance measures: the ARL to false alarm, and Lorden's worst-case detection delay. It is now well-known that, while these two classical performance measures are meaningful and informative for detecting a change in the i.i.d. model, they may be misleading or inappropriate when the observations are dependent; see, for example, Mei (2008). As one reviewer pointed out, it is natural to question their usefulness in our context where the observations are independent but not identically distributed.

For instance, with non-homogenous (Poisson) observations, the ARL to false alarm can no longer be interpreted as "raising a false alarm once every γ time units" as in the i.i.d. models (though it can still be interpreted as "taking at least γ time units before raising the *first* false alarm"). In addition, one would expect to have a smaller detection delay in a region where more information is available, but Lorden's worst-case detection delay criterion puts the same weights at different individual detection delays. Since the two alternative schemes, WLR and ATM, are designed specifically to outperform the GLR scheme under the classical performance measures (when the population sizes are increasing), it is natural to ask whether the improvement is due to the choice of the classical performance measures? In other words, is it possible that even under the new asymptotic setting in Section 7, the GLR scheme is still (asymptotically) optimal under some alternative performance measures?

The answer is a resounding "Yes." As mentioned in Pollak (2008), instead of minimizing detection delays, one can also minimize the expected number of cases after a change occurs until an alarm is raised. That is, when the change occurs at time ν and a scheme raises an alarm at time $T \geq \nu$, the detection delay criterion $T - \nu = \sum_{i=\nu}^{T} 1$ can be replaced by $\lambda_1(\sum_{i=\nu}^{T} l_i)$. Likewise, the ARL to false alarm criterion $\mathrm{E}_{\infty}(T)$ can be replaced by $\mathrm{E}_{\infty}(\sum_{i=1}^{T} l_i)$. It is interesting to see that in our context, these alternative criteria are just the so-called Kullback-Leibler divergence criteria in Moustakides (2004), which proves that the GLR scheme is (exactly) optimal in detecting changes in the drift of a Brownian motion. In view of Section 8.1, in the problem of detecting a change in the drift of a Brownian motion from $x_t\beta_0$ to $x_t\beta_1$, it follows from Moustakides (2004) that under the asymptotic setting in Section 7, the GLR scheme $T_{GLR}(a)$ is still optimal under the new Kullback-Leibler divergence criteria. Further research is needed for Poisson processes or discrete-time models, and will be investigated elsewhere.

The above discussions lead a tough question in practice: should we incorporate the information of unequal population sizes in the detection schemes themselves, as in the WLR or ATM scheme, or should we incorporate such information directly in the performance measures by considering some new performance measures such as the Kullback-Leibler divergence criteria? We think the answer will depend on the goal or objective of the specific applications. If one is more interested in the speed of detecting a change, say, based on low-frequency observations (e.g., yearly data), it seems to be more appropriate to keep the classical performance measures and use the detection schemes such as the WLR or ATM. On the other hand, with high-frequency observations (e.g., daily data), one might be more interested in minimizing the expected number of post-change cases until an alarm is raised. In such a case, instead of the classical performance measures, it would be better to evaluate the schemes under some alternative performance measures such as the new Kullback-Leibler divergence criteria, for which our favorite GLR schemes are still efficient.

10. Conclusion

In this paper we have studied the problem of detecting a change in the mean of Poisson distributions after taking into account the effect of population sizes. Such a problem has an important application in biosurveillance when one is interested in monitoring whether a disease risk changes or not. Despite its asymptotic optimality properties under the classical asymptotic setting, the GLR scheme can have poor finite sample properties as compared to two alternative adhoc schemes: weighted likelihood ratio (WLR) and adaptive threshold method

(ATM). To understand why the GLR is at times efficient and is at times inefficient, a new asymptotic setting was studied by assuming that the time at which the population sizes reach the stationary value is comparable to the false alarm constraint. Asymptotic properties of the three schemes were established under the new asymptotic setting, and consistent with our finite-sample simulations. Extensions to other probability functions or to the continuous-time setting, as well as new alternative performance measures, were briefly discussed.

There are a lot of open interesting questions for future research. It is important to find efficient robust schemes, or at least to develop a protocol or guideline to derive such schemes, when the population sizes are observable random variables instead of pre-specified constants. Additionally, in practice the assumption of known pre-change and/or post-change risks λ's is too restrictive, and it will be interesting to develop efficient detection schemes when the λ's are partially specified or unknown. Moreover, in biosurveillance, one is generally monitoring multiple locations (states, counties, or cities) to see whether there is any change in any locations, and thus the extension to spatio-temporal detection is essential for global monitoring in applications. We hope this paper will inspire further research on sequential change-point detection in the modern information age.

Acknowledgement

The authors would like to thank Professor George Moustakides for stimulating discussions. Y. Mei's research was supported in part by the AFOSR grant FA9550-08-1-0376 and the NSF Grants CCF-0830472 and DMS-0954704. K. Tsai's research was supported in part by the NSF Grant CMMI 0927592 and GRC/GRF Grant 9041578.

Appendix

Appendix A: Proof of Theorems 7.1 and 7.2

The proof of Theorem 7.1 is similar to that of Theorem 1 in Lai (1998) with a very minor twist. Let m be the largest integer $\leq (\log(\gamma-\omega))^2$, and thus $m < \gamma - \omega$ when γ is sufficiently large. A key observation is that if $\mathrm{E}_\infty(T) \geq \gamma$, then for some $\nu \geq \omega$,

$$\mathbf{P}_\infty(T \geq \nu) > 0 \quad \text{and} \quad \mathbf{P}_\infty(T < \nu + m | T \geq \nu) \leq \frac{m}{\gamma - \omega}.$$

Now the observations Y_n's are i.i.d. Poisson distributed with constant population size l^* for all $n \geq \nu$. Following the arguments in Lai (1998), it is straightforward to show that

$$\mathrm{E}_\nu(T - \nu | T \geq \nu) \geq (1 + o(1)) \frac{\log(\gamma - \omega)}{l^* I(\lambda_1, \lambda_0)}.$$

Theorem 7.1 then follows at once from the fact that $\log(\gamma - \omega) = \log\gamma + O(1)$ when $0 < \omega < C\gamma$ for some constant $0 < C < 1$, since $\log\gamma \ge \log(\gamma-\omega) \ge \log(\gamma - C\gamma) = \log\gamma + \log(1-C)$.

To prove Theorem 7.2, the false alarm properties of the GLR scheme in part (i) is just a special case of Theorem 4 in Lai (1998). Let us now focus on part (ii) for the detection delays of the WLR scheme. Let $\xi_i = Y_i/l_i \log \lambda_1/\lambda_0 - (\lambda_1 - \lambda_0)$, and define $S_n = \sum_{i=1}^n \xi_i$ for $n \ge 1$ and $S_0 = 0$. Consider the open-ended test

$$\tau = \inf\Big\{n \ge 1 : \sum_{i=1}^{n} \xi_i \ge b\Big\}.$$

Following the ideas in Lorden (1971), denote by τ_k the stopping time obtained by applying τ to $Y_k, Y_{k+1}, \ldots$. Then $T_{WLR}(b) = \inf_{k\ge1}\{\tau_k + k - 1\}$, and thus it suffices to show that $[b/I(\lambda_1,\lambda_0)] + M$ is a uniform bound on $\mathrm{E}_{\nu=k}(\tau_k)$ over all $k \ge 1$.

To prove this, let us focus on the open-ended test τ under the probability measure $\mathbf{P}_{\nu=1}$. To simplify notation, in the following proof we simply use E to denote the expectation under $\mathbf{P}_{\nu=1}$. Letting $V = \lambda_1(\log(\lambda_1/\lambda_0))^2$, it is easy to see that $\mathrm{E}\,(\xi_i) = I(\lambda_1,\lambda_0) > 0$ and $Var(\xi_i) = V/l_i$ is uniformly bounded by V/l_*. Hence $\tau < \infty$ a.s. By a modification of Wald's equation (on the independent random variables with the same mean), $\mathrm{E}\,(S_\tau) = I(\lambda_1,\lambda_0)\mathrm{E}\,(\tau)$. One the one hand, by the definition of τ, we have $S_\tau \ge b$ and thus $\mathrm{E}\,(\tau) \ge b/I(\lambda_1,\lambda_0)$. On the other hand, using the fact that $S_{\tau-1} < b$, we have

$$\begin{aligned} I(\lambda_1,\lambda_0)\mathrm{E}\,(\tau) &= \mathrm{E}\,(S_\tau) = \mathrm{E}\,(S_{\tau-1} + \xi_\tau) < b + \mathrm{E}\,|\xi_\tau| \\ &\le b + \sup_{i\ge1} \mathrm{E}\,|\xi_i| \le b + \sup_{i\ge1}\sqrt{\mathrm{E}\,(\xi_i^2)} \\ &= b + \sup_{i\ge1}\sqrt{(\mathrm{E}\,\xi_i)^2 + Var(\xi_i)} = b + \sqrt{(I(\lambda_1,\lambda_0))^2 + \frac{V}{l_*}}, \end{aligned}$$

which implies that $b/I(\lambda_1,\lambda_0) \le \mathrm{E}\,(\tau) \le b/I(\lambda_1,\lambda_0) + M$, where M is the constant given in the theorem. Part (ii) of the theorem then follows at once by applying this relation to the τ_k's for all k, and from the fact that $\inf_{i\ge k} l_i \ge l_*$ holds uniformly over all $k \ge 1$.

Appendix B: Proof of Theorem 7.3

Let us first illustrate the fundamental ideas before presenting a rigorous proof. It is well-known that the detection delays are of the order of $\log\gamma$, subject to the false alarm constraint in (2.1). Since $\log\gamma << \omega'$, for each of the proposed schemes, the detection delay $\mathrm{E}_{\nu=1}(T)$ when the change-point occurs at $\nu = 1$ is mainly determined by the first stage when the population size $l_n = l^{(0)}$. Thus all three schemes become Page's CUSUM procedure with constant population sizes $l^{(0)}$ and the classical results can be applied to determine $\mathrm{E}_{\nu=1}(T)$. Similarly, the detection delay ess sup $\mathrm{E}_{\omega}\big((T-\omega+1)^{+}|\mathcal{F}_{\omega-1}\big)$ is just the detection delay in the classical change-point detection problem with the i.i.d. models in which one detects a change in Poisson mean from $l^{*}\lambda_0$ to $l^{*}\lambda_1$. As the thresholds a, b, and c go to ∞, the detection delays of the three proposed schemes when the change occurs at time $\nu = 1$ or ω can be summarized in the following table:

detection scheme T	$\mathrm{E}_1(T)$	ess sup $\mathrm{E}_{\omega}\Big((T-\omega+1)^{+}|\mathcal{F}_{\omega-1}\Big)$
$T_{GLR}(a)$	$(1+o(1))\frac{a}{l^{(0)}I(\lambda_1,\lambda_0)}$	$(1+o(1))\frac{a}{l^{*}I(\lambda_1,\lambda_0)}$
$T_{WLR}(b)$	$(1+o(1))b/I(\lambda_1,\lambda_0)$	$(1+o(1))b/I(\lambda_1,\lambda_0)$
$T_{ATM}(c)$	$(1+o(1))c/I(\lambda_1,\lambda_0)$	$(1+o(1))c/I(\lambda_1,\lambda_0)$

We also need to establish the relationship between the false alarm constraint γ in (2.1) and the threshold values a, b, and c in the three schemes $T_{GLR}(a), T_{WLR}(b)$ and $T_{ATM}(c)$. Intuitively, since we now assume $\omega << \gamma$, the stage with constant population sizes l^{*} is the most important stage in order to satisfy the false alarm constraint in (2.1), and we have to take at least $\gamma - \omega \sim \gamma$ observations from the stage with constant population sizes l^{*}. Now at the stage with constant population size l^{*}, all three schemes become the classical CUSUM procedure under the i.i.d. model, and thus the classical results imply that $a \sim \log\gamma, b \sim \log\gamma/l^{*}$, and $c \sim \log\gamma/l^{*}$ (the rigorous proof is tedious but straightforward, and is thus omitted). Here $x \sim y$ means that $x = (1+o(1))y$ as y goes to ∞. Combining these with the above results on detection delays yields the desired result in the theorem.

Let us now provide some rigorous arguments. To highlight our main ideas, we rigorously prove $b \sim \log\gamma/l^{*}$ for the WLR scheme $T_{WLR}(b)$, as the methods can be easily extended to all other arguments on the detection delay or false alarm properties of the three schemes. Note that for any stopping time T,

$$\mathrm{E}_{\infty}(T) \le \omega + \mathrm{E}_{\infty}(T-\omega|T\ge\omega)\mathbf{P}_{\infty}(T\ge\omega) \le \omega + \mathrm{E}_{\infty}(T-\omega|T\ge\omega).$$

If the stopping time is $T = \inf\{n : R_n \ge A\}$, with a recursive form for nonnegative detection statistics $R_n = h(R_{n-1}, Y_n)$ and $R_0 = 0$, then we can define a new statistic R_n^{*} with the same Makov structure but with an initial value $R_0^{*} = x$, and consider the corresponding stopping time $T_x^{*} = \inf\{n : R_n^{*} \ge A\}$. In many interesting cases, including the GLR and WLR statistics considered in this paper, we have $T_x^{*} \le T_{x'}^{*}$ when $0 \le x < x'$. Now condition on time ω, let F^{ω} denote the distribution of the statistic $R_{n=\omega}$ conditional on the event $\{T \geq \omega\}$, so that

$$\mathrm{E}_{\infty}(T-\omega|T \geq \omega) = \int_0^A \mathrm{E}_{\infty}(T_x^*)dF^{\omega}(x) \leq \int_0^A \mathrm{E}_{\infty}(T_{x=0}^*)dF^{\omega}(x) = \mathrm{E}_{\infty}(T_{x=0}^*).$$

Now for the WLR scheme T, the corresponding $T_{x=0}^*$ simply becomes Page's CUSUM procedure with constant population sizes l^* and the detection threshold l^*b, since all incoming new post-change observations have the constant population size l^*. Thus, $\mathrm{E}_{\infty}(T_{x=0}^*) = C_1(1+o(1))e^{l^*b}$ for some constant C_1, and this implies that

$$\mathrm{E}_{\infty}(T) \leq \omega + \mathrm{E}_{\infty}(T-\omega|T \geq \omega) \leq \omega + C_1(1+o(1))e^{l^*b}.$$

To satisfy the false alarm constraint γ in (2.1), it is necessary to have

$$b \geq \frac{[\log(\gamma-\omega) - \log(C_1) - \log(1+o(1))]}{l^*} = \frac{(1+o(1))\log\gamma}{l^*}$$

since $\log(\gamma-\omega) = \log\gamma + O(1)$ when $0 < \omega < C\gamma$.

Now it remains to show that $b = (1+o(1))\log\gamma/l^*$ is also sufficient, which requires us to find a lower bound on $\mathrm{E}_{\infty}(T_{WLR}(b))$. Recall that such a lower bound on $\mathrm{E}_{\infty}(T_{GLR}(a))$ is presented in Theorem 7.2 for the GLR scheme, but it involves more mathematics for the WLR or ATM schemes. Note that for any stopping time T,

$$\mathrm{E}_{\infty}(T) \geq \mathrm{E}_{\infty}(T; T \geq \omega) \geq \mathrm{E}_{\infty}(T-\omega; T \geq \omega).$$

Now when $T = T(A) = \inf\{n : R_n \geq A\}$, using the previous notation, for some constant $d \in (0, A)$ we have

$$\begin{aligned}\mathrm{E}_{\infty}(T-\omega|T \geq \omega) &= \int_0^A \mathrm{E}_{\infty}(T_x^*)dF^{\omega}(x) \geq \int_0^d \mathrm{E}_{\infty}(T_x^*)dF^{\omega}(x)\\ &= \mathrm{E}_{\infty}(T_{x=d}^*)\mathbf{P}_{\infty}(R_{\omega} \leq d|T \geq \omega),\end{aligned}$$

and thus

$$\begin{aligned}\mathrm{E}_{\infty}(T) &\geq \mathrm{E}_{\infty}(T-\omega; T \geq \omega) \geq \mathrm{E}_{\infty}(T_{x=d}^*)\mathbf{P}_{\infty}(R_{\omega} \leq d; T \geq \omega)\\ &\geq \mathrm{E}_{\infty}(T_{x=d}^*)\mathbf{P}_{\infty}(\max_{1 \leq i \leq \omega} R_{\omega} \leq d)\\ &= \mathrm{E}_{\infty}(T_{x=d}^*)\mathbf{P}_{\infty}(T(d) \geq \omega),\end{aligned}$$

where the definition of $T(d)$ is similar to that of $T = T(A)$ except with a new detection threshold d.

With a suitable choice of d, the problem now reduces to finding lower bounds on both $\mathrm{E}_{\infty}(T_{x=d}^*)$ and $\mathbf{P}_{\infty}(T(d) \geq \omega)$ for the WLR scheme. To find a lower bound on $\mathrm{E}_{\infty}(T_{x=d}^*)$, note that for the WLR scheme, $T_{x=d}^*$ is just Page's CUSUM procedure with an initial value l^*d and the detection threshold l^*b, and thus it is bounded below by the corresponding Shiryaev-Robert procedure. Specifically, consider the Shiryaev-Robert procedure $N^*_{SR} = \inf\{n : U_n \geq e^{l^*b}\}$ with $U_0 = e^{l^*d}$ and

$$U_{n+1} = (1+U_n)\frac{f_1(Y_n|l_n)}{f_0(Y_n|l_n)} = (1+U_n)\Big[(\frac{\lambda_1}{\lambda_0})^{Y_n} e^{-l_n(\lambda_1-\lambda_0)}\Big].$$

Then, it is clear that $T^*_{x=d} \geq N^*_{SR}$. Observe that $U_n - n$ is a martingale for the Shiryaev-Robert statistics U_n's under $\mathbf{P}_\infty$; by the Optional Stopping Theorem, we have $\mathrm{E}_\infty(U_N - N) = \mathrm{E}_\infty(U_0)$ for the stopping time $N = N^*_{SR}$, see Pollak (1985). Thus,

$$\mathrm{E}_\infty(T^*_{x=d}) \geq \mathrm{E}_\infty(N^*_{SR}) = \mathrm{E}_\infty(U_{N^*_{SR}}) - \mathrm{E}_\infty(U_0) \geq e^{l^*b} - e^{l^*d},$$

since $U_{N^*_{SR}} \geq e^{l^*b}$ by the definition of N^*_{SR} and $U_0 = e^{l^*d}$ is a constant.

Meanwhile, the lower bound on $\mathbf{P}_\infty(T(d) \geq \omega)$ for the WLR scheme involves a few technical details. Over the time interval $n \in [\omega', \omega]$, define a sequence of new detection statistics $R^*_n = R^*_n(x)$ recursively as in the WLR statistics $\hat{W}_n$ except that the "initial" value at time $n = \omega'$ is x. That is, $R^*_n = \max(0, R^*_{n-1} + (1/l_n)\log f_1(Y_n|l_n)/f_0(Y_n|l_n))$ for $n \geq \omega'$, and $R^*_{\omega'} = x$. Since it is assumed that $\omega - \omega' = O(1)$ and the sample sizes l_n's are bounded over the interval $[\omega', \omega]$, i.e., $|\omega - \omega'| \leq D$ for some constant $D > 0$, and each observation only generates finite information, it is evident that $\max_{\omega' \leq i \leq \omega} \hat{R}^*_n(0)$ has a finite distribution when the "initial" value of $R^*_{\omega'} = 0$. Hence, for any $0 < \epsilon < 1$ there exists a constant $K > 0$ such that $\mathbf{P}_\infty(\max_{\omega' \leq i \leq \omega} R^*_n(0) > K) \leq \epsilon$. In addition, for any "initial" values $x \geq 0$, it is easy to verify that $R^*_n(x) \leq x + R^*_n(0)$ for the WLR-type detection statistics. Thus, for $d > K$, whenever $\max_{1 \leq i \leq \omega'} \hat{W}_i \leq d - K$, i.e., when $T(d-K) \geq \omega'$, the probability that $T(d) \leq \omega$ is at most ϵ. Hence, for $d > K$,

$$\begin{aligned}\mathbf{P}_\infty(T(d) \geq \omega) &\geq \mathbf{P}_\infty(T(d) \geq \omega; T(d-K) \geq \omega')\\ &= [1 - \mathbf{P}_\infty(T(d) \leq \omega | T(d-K) > \omega')]\mathbf{P}_\infty(T(d-k) \geq \omega')\\ &= (1-\epsilon)\mathbf{P}_\infty(T(d-k) \geq \omega').\end{aligned}$$

For the WLR scheme, at the stage with constant population sizes $l^{(0)}$, the corresponding $T(d-K)$ is just Page's CUSUM procedure with detection threshold $l^{(0)}(d-K)$, and thus

$$\mathbf{P}_\infty(T \geq \omega') = (1+o(1))\exp\Big(-\omega' e^{-l^{(0)}(d-K)}\Big).$$

This holds since it is well-known that Page's CUSUM procedure with a detection threshold d' is asymptotically exponentially distributed with mean $\exp(d')$ under $\mathbf{P}_\infty$, see, for example, Siegmund (1985).

Therefore, for the WLR scheme $T = T_{WLR}(b)$, as $b \to \infty$, let $d = (1-\eta/2)b$, where η is given in the theorem such that $\omega' = o(\gamma^{(1-\eta)l^{(0)}/l^*})$. Combining the above results yields

$$\mathrm{E}_{\infty}(T_{WLR}(b)) \geq (e^{l^*b} - e^{l^*(1-\eta)b})(1-\epsilon)(1+o(1)) \exp\Big(-\omega' e^{-l^{(0)}(d-K)}\Big).$$

A simple calculation shows that when $\gamma \to \infty$, the choice of $b = (1+o(1))\log\gamma/l^*$ is sufficient to guarantee that $\mathrm{E}_{\infty}(T_{WLR}(b)) \geq (1+o(1))\gamma$. Treating $(1+o(1))\gamma$ as the true false alarm constraint and using the fact that $(1+o(1))\log\gamma/(1+o(1)) = (1+o(1))\log\gamma$, we can see that the choice of $b = (1+o(1))\log\gamma/l^*$ is also sufficient to satisfy the false alarm constraint in (2.1). This completes the proof of the claim that $b \sim (1+o(1))\log\gamma/l^*$ for the WLR scheme.

References

Aalen, O. (1978). Nonparametric inference for a family of counting processes. *Ann. Statist.* **6**, 701-726.

Baron, M. and Tartakovsky, A. G. (2006). Asymptotic optimality of change-point detection schemes in general continuous-time models. *Sequential Anal.* **25**, 257-296.

Basseville, M. and Nikiforov, I. V. (1993). *Detection of Abrupt Changes - Theory and Application.* Prentice-Hall, Englewoood Cliffs, NJ.

Fienberg, S. E. and Shmueli, G. (2005). Statistical issues and challenges associated with rapid detection of bio-terrorist attacks. *Statist. Medicine* **24**, 513-529.

Kulldorff, M. (2001). Prospective time periodic geographical disease surveillance using a scan statistic. *J. Roy. Statist. Soc. Ser. A* **164**, 61-72.

Lai, T. L. (1995). Sequential change-point detection in quality control and dynamical systems (with discussions). *J. Roy. Statist. Soc. Ser. B* **57**, 613-658.

Lai, T. L. (1998). Information bounds and quick detection of parameter changes in stochastic systems. *IEEE Trans. Inform. Theory* **44**, 2917-2929.

Lai, T. L. (2001). Sequential analysis: Some classical problems and new challenges. *Statist. Sinica* **11**, 303-408.

Lorden, G. (1971). Procedures for reacting to a change in distribution. *Ann. Math. Statist.* **42**, 1897-1908.

Mei, Y. (2008). Is average run length to false alarm always an informative criterion? (with discussion). *Sequential Anal.* **27**, 354-419.

Moustakides, G. V. (1986). Optimal stopping times for detecting changes in distribution. *Ann. Statist.* **14**, 1379-1387.

Moustakides, G. V. (2004). Optimality of the CUSUM procedure in continuous time. *Ann. Statist.* **32**, 302-315.

Page, E. S. (1954). Continuous inspection schemes. *Biometrika* **41**, 100-115.

Peskir and Shiryaev(2006). *Optimal Stopping and Free-Boundary Problems.* Birkhäuser, Basel, Switzerland.

Pinheiro, J. C. and Bates, D. M. (2000). *Mixed-Effects Models in S and S-PLUS.* Springer, New York.

Pollak, M. (1985). Optimal detection of a change in distribution. *Ann. Statist.* **13**, 206-227.

Pollak, M. (2008). Discussion on "Is average run length to false alarm always an informative criterion?" by Yajun Mei. *Sequential Anal.* **27**, 389-391.

Poor, H. V. and Hadjiliadis, O. (2008). *Quickest Detection.* Cambridge University Press, Cambridge, UK.

Ritov, Y. (1990). Decision theoretic optimality of the CUSUM procedure. *Ann. Statist.* **18**, 1464-1469.

Roberts, S. W. (1966). A comparison of some control chart procedures. *Technometrics* **8**, 411-430.

Shiryaev, A. N. (1963). On optimum methods in quickest detection problems. *Theory Probab. Appl.* **8**, 22-46.

Siegmund, D. (1985). *Sequential Analysis.* Springer, New York.

Sonesson, C. (2007). A CUSUM framework for detection of space-time disease clusters using scan statistics. *Statist. Medicine* **26**, 4770-4789.

Tsui, K.-L., Chiu, W., Gielich, P., Goldsman, D., Liu, X. and Maschek, T. (2008). A review of healthcare, public health, and syndromic surveillance. *Quality Engineering* **20**, 435-450.

Woodall, W. H. (2006). The use of control charts in health-care and public-health surveillance. *J. Quality Tech.* **38**, 89-104.

Wald, A. (1947). *Sequential Analysis.* John Wiley and Sons, New York.

Yakir, B. (1994). Optimal detection of a change in distribution when the observations form a markov chain with a finite state space. In *Change-Point Problems* **23** (Edited by E. Carlsten, H-G. Muller, and D. Siegmund), 346-358. IMS Lecture Notes Monograph Series.

Yao, Q. (1993). Asymptotically optimal detection of a change in a linear model. *Sequential Anal.* **12**, 201-210.

H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, 765 Ferst Drive NW, Atlanta, GA 30332-0205, U.S.A.

E-mail: ymei@isye.gatech.edu

H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, 765 Ferst Drive NW, Atlanta, GA 30332-0205, U.S.A.

E-mail: shan@isye.gatech.edu

H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, 765 Ferst Drive NW, Atlanta, GA 30332-0205, U.S.A.

E-mail: ktsui@isye.gatech.edu

(Received July 2009; accepted December 2009)