Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/JASA

Likelihood-Based EWMA Charts for Monitoring Poisson Count Data With Time-Varying Sample Sizes

Qin ZHOU, Changliang ZOU, Zhaojun WANG, and Wei JIANG

Many applications involve monitoring incidence rates of the Poisson distribution when the sample size varies over time. Recently, a couple of cumulative sum and exponentially weighted moving average (EWMA) control charts have been proposed to tackle this problem by taking the varying sample size into consideration. However, we argue that some of these charts, which perform quite well in terms of average run length (ARL), may not be appealing in practice because they have rather unsatisfactory run length distributions. With some charts, the specified in-control (IC) ARL is attained with elevated probabilities of very short and very long runs, as compared with a geometric distribution. This is reflected in a larger run length standard deviation than that of a geometric distribution and an elevated probability of false alarms with short runs, which, in turn, hurt an operator's confidence in valid alarms. Furthermore, with many charts, the IC ARL exhibits considerable variations with different patterns of sample sizes. Under the framework of weighted likelihood ratio test, this article suggests a new EWMA control chart which automatically integrates the varying sample sizes with the EWMA scheme. It is fast to compute, easy to construct, and quite efficient in detecting changes of Poisson rates. Two important features of the proposed method are that the IC run length distribution is similar to that of a geometric distribution and the IC ARL is robust to various patterns of sample size variation. Our simulation results show that the proposed chart is generally more effective and robust compared with existing EWMA charts. A health surveillance example based on mortality data from New Mexico is used to illustrate the implementation of the proposed method. This article has online supplementary materials.

KEY WORDS: Average run length; EWMA; Healthcare; Poisson count data; Run length distribution; Short-run processes; Statistical process control.

1. INTRODUCTION

Control charts are effective tools in statistical process control (SPC) for monitoring the stability of a process over time. Currently, most competitive manufacturing companies are implementing SPC methods in various applications. Statistical approaches to continual surveillance of a rare event of interest are greatly needed in industrial, clinical, and epidemiological environments (cf. Sonesson and Bock 2003). Among them, the problem of detecting a change in the rate of occurrence of an event through sequential observations in a stochastic process is very important. Examples include detection of an increased birth rate of infants with congenital malformations and increased rate of incidence of diseases, nonconformities, or adverse drug reactions. The objective is to detect the change occurring at an unknown time point as early as possible after it has occurred while controlling the rate of false alarms (cf. Woodall 2006).

Considerable research has been developed on detecting changes in the number of events recorded in regular time intervals. A simple method is to model the count of events recorded in regular time intervals by independent and identically distributed (iid) Poisson random variables; therefore, detecting a change in the rate of occurrence of the event may be characterized as detecting a change in the mean of the Poisson process. The Shewhart chart is commonly used for monitoring the Poisson mean. See Duncun (1974) and Montgomery (2009) for discussions. The cumulative sum (CUSUM) chart, which has received considerable attention for detecting small changes, can be derived based on the likelihood ratio test principle, see Lucas (1985), Lai (1995), and White and Keats (1996). Gan (1990) considered the exponentially weighted moving average (EWMA) charts, which have superiority over the Shewhart-type chart in terms of average run length (ARL). Frisén and De Maré (1991) proposed a likelihood ratio method and showed that it is preferable to Poisson Shewhart and Poisson CUSUM charts in the sense of minimizing expected delay. However, this optimality property requires the assumption that the expected number of events is constant over time. See also Frisén and Sonesson (2006) for some analogous discussions. This assumption weakens the potential advantage of the CUSUM method in other applications, such as health surveillance. In many situations, the size of the population at-risk is not constant but varies over time; consequently, the expected number of incidents is no longer a constant but changes over time according to the population size as well as the incidence rate of the event.

Recently, there has been an increasing attention devoted to surveillance of incidence rate with time-varying population sizes. For example, Rossi, Lampugnani, and Marchi (1999) propose an approximate CUSUM procedure. The basic idea is to first standardize the count data by using a normal approximation of the Poisson process, and then to employ the classical CUSUM procedure to monitor the transformed data. To accommodate the

Qin Zhou is Assistant Professor, School of Mathematical Sciences, Jiangsu Normal University, Xuzhou, China (E-mail: *zhouqin8808@163.com*). Changliang Zou is Assistant Professor, School of Mathematical Sciences, Nankai University, Tianjin, China (E-mail: *chlzou@yahoo.com.cn*). Zhaojun Wang is Professor, LPMC and School of Mathematical Sciences, Nankai University, Tianjin, China (E-mail: *zjwang@nankai.edu.cn*). Wei Jiang is Professor, Antai College of Economics and Management, Shanghai Jiaotong University, Shanghai, China (E-mail: *jiangwei08@gmail.com*). The authors thank the editor, associate editor, and three anonymous referees for their many helpful comments that have resulted in significant improvements in the article. We also thank William Woodall for the useful discussions. This research was supported by the NNSF of China grants 11001138, 11071128, 11131002, 11101306, and 71172131 and the RFDP of China grant 20110031110002. Zhou's work is partially supported by PAPD of Jiangsu Higher Education Institutions. Zou thanks the support of the National Center for Theoretical Sciences, Math Division. This work was completed when Zhou was a doctoral student at Nankai University. The first two authors contributed equally to this work. Wang is the corresponding author.

Journal of the American Statistical Association
September 2012, Vol. 107, No. 499, Theory and Methods
DOI: 10.1080/01621459.2012.682811

dynamic changes in the mean number of events, Mei, Han, and Tsui (2011) further develop some variations of the CUSUM method and propose three CUSUM-based charts. Shu, Jiang, and Tsui (2011) compared weighted CUSUM and conventional CUSUM procedures in the presence of monotone changes in population size. In addition to the CUSUM techniques, Dong, Hedayat, and Sinha (2008) considered the EWMA methods to address the issue of nonconstant population size. Ryan and Woodall (2010) compared the performance of EWMA methods with some CUSUM methods under the assumption of random sample sizes and suggest a modified EWMA chart by adding a lower reflecting barrier. The central idea in Dong, Hedayat, and Sinha (2008) and Ryan and Woodall (2010) was to divide the observed counts by the corresponding sample sizes to account for the variability of the sample sizes.

In this article, motivated by the finding that the classical EWMA control chart can be derived under the framework of weighted likelihood, we suggest a new EWMA control chart which naturally integrates time-varying sample sizes with the EWMA scheme. The weighted likelihood method discounts historical evidence about change points and thus grants the EWMA chart superiority in detecting recent parameter changes. Simulation results show that the proposed method is generally more robust in detecting the change of Poisson rate with varying sample sizes over time than the existing EWMA control charts discussed in Dong, Hedayat, and Sinha (2008) and Ryan and Woodall (2010). Moreover, we argue that for some charts such as the two extensions of the CUSUM chart proposed in Mei, Han, and Tsui (2011) and the two modifications of the EWMA chart in Dong, Hedayat, and Sinha (2008), though these perform quite well in terms of ARL, they may not be appealing in practice because of their rather unsatisfactory run length distributions. The probabilities of false alarms of these charts may increase dramatically after short-runs, which also result in excessive variations of run length.

The remainder of the article is organized as follows. We first describe the mathematical formulation of the problem and existing work in Section 2. We then introduce our proposed method followed by its asymptotic bounds of ARL in Section 3. The performance comparison for detecting changes in Poisson rate with time-varying sample sizes is presented in Section 4. Run length distributions are discussed through Monte Carlo simulations. The analytical bounds of ARL are compared with the simulation results in Section 5. Section 6 contains a health surveillance example to illustrate the application of our proposed chart. Several remarks draw the article to its conclusion in Section 7. Some technical details are provided in the Appendices. Some other simulation results are provided in another appendix, which is available online as supplementary material.

2. THE STATISTICAL MODEL AND EXISTING WORK

Let $X_1, X_2, \ldots$ be a sequence of event counts observed during fixed time periods. Assume that the X_t's are independent Poisson observations with mean $\mu_t = n_t\theta$, where n_t and θ denote the size of the population at time t and the incidence rate of a rare event, respectively. Although other distributional assumptions could be made, the Poisson assumption is widely used (Chen 1987). In the context of detecting a change in the incidence rate, it is assumed that θ changes from θ_0 to another unknown value θ_1 at some unknown time τ, that is, the observations collected come from the following change-point model

$$X_t \overset{\text{indep}}{\sim} \begin{cases} \text{Poisson}(n_t\theta_0), & \text{for} \quad t = 1, \ldots, \tau, \\ \text{Poisson}(n_t\theta_1), & \text{for} \quad t = \tau + 1, \ldots, \end{cases} \tag{1}$$

where "$\overset{\text{indep}}{\sim}$" denotes "independently distributed." The objective is to detect the change as early as possible once it occurs through sequential observations.

In the change-point detection problem, a detection scheme is a stopping time T and the control limit with respect to the observed data sequences $(n_t, X_t)_{t\geq 1}$. We use an alarm system consisting of two parts at stage t: a monitoring statistic $a(\mathbf{n}_t, \mathbf{X}_t)$ and an alarm limit $g(t)$, where $\mathbf{n}_t = \{n_i; i \leq t\}$ and $\mathbf{X}_t = \{X_i; i \leq t\}$. The time of an alarm, T, is defined as

$$T = \min\{t; a(\mathbf{n}_t, \mathbf{X}_t) \geq g(t)\}.$$

That is, the decision $T = t$ only depends on the first t observations, and $T = t$ means that the first alarm is triggered at time t to indicate that a change has occurred somewhere in the first t observations. Consistent with the literature, we focus on using an upper-sided chart to detect increases in the incidence rate, that is, $\theta_1 > \theta_0$, but the lower-sided and two-sided charts can be constructed without difficulty.

The EWMA-type control chart statistic proposed by Dong, Hedayat, and Sinha (2008) is

$$Z_t = (1 - \lambda)Z_{t-1} + \lambda\frac{X_t}{n_t}, \ t = 1, 2, \ldots,$$

where $Z_0 = \theta_0, \lambda \in (0, 1]$ is the smoothing parameter which determines the weights assigned to past observations. Based on this EWMA sequence, Dong, Hedayat, and Sinha (2008) developed three different stopping rules, EWMAe, EWMAa1, and EWMAa2 control charts, as follows:

$$\begin{aligned} T_{\text{EWMAe}} &= \min\{t; Z_t \geq \theta_0 + L\sigma_t, t \geq 1\}, \\ \sigma_t^2 &= \lambda^2\sum_{i=1}^{t}(1 - \lambda)^{2t-2i}\frac{\theta_0}{n_i}, \\ T_{\text{EWMAa1}} &= \min\{t; Z_t \geq \theta_0 + L\sigma_t^*, t \geq 1\}, \\ \sigma_t^{*2} &= \frac{\theta_0}{n_0}\frac{\lambda}{2 - \lambda}[1 - (1 - \lambda)^{2t}], \\ T_{\text{EWMAa2}} &= \min\{t; Z_t \geq \theta_0 + L\sigma^*, t \geq 1\}, \\ \sigma^{*2} &= \frac{\theta_0}{n_0}\frac{\lambda}{2 - \lambda}, \end{aligned}$$

where the control limit coefficients L are determined given the nominal value of in-control (IC) ARL (denoted as ARL_0) and the value n_0 is the minimum sample size among all the values of $n_i, i = 1, \ldots, t$. Without confusion, we use the generic notation L to represent the control limit coefficient for different control charts. Note that the EWMAe and EWMAa1 methods are equivalent when the sample size is constant and the EWMAa2 chart is just a variant of EWMAa1 by using the asymptotic variance and has been shown essentially equivalent to EWMAa1 in terms of steady-state ARLs (Ryan and Woodall 2010).

To avoid the inertial problems, Ryan and Woodall (2010) modified the EWMAe method by adding a lower reflecting barrier at $Z_t = \theta_0$

$$T_{\text{EWMAM}} = \min\{t; \tilde{Z}_t \geq L\sigma_t, t \geq 1\},$$

where

$$\tilde{Z}_t = \max\left\{\theta_0, (1-\lambda)\tilde{Z}_{t-1} + \lambda\frac{X_t}{n_t}\right\}, \tilde{Z}_0 = \theta_0.$$

Henceforth, we will refer to it as the EWMA-modified (EWMAM) method. Ryan and Woodall (2010) argued that the EWMAM performs better than EWMAe, but in their comparison the weighting parameters of the two EWMA charts were chosen differently. As shown in Section 4, given the same values of λ, the EWMAM chart does not seem to have significant advantages over EWMAe, especially for detecting small changes of incidence rate.

The CUSUM chart proposed by Mei, Han, and Tsui (2011) is defined by

$$W_t = \max\left\{0, W_{t-1} + \left[X_t \log\frac{\theta_1}{\theta_0} - n_t(\theta_1 - \theta_0)\right]\right\}, \quad W_0 = 0$$

and the corresponding stopping time is

$$T_{\text{CUSUM}} = \min\{t; W_t \geq L, t \geq 1\}.$$

Mei, Han, and Tsui (2011) suggest two modifications to further enhance the performance of the CUSUM chart when n_t varies dramatically, the weighted-likelihood ratio (WLR) and the adaptive threshold method (ATM) whose stopping times are

$$T_{\text{WLR}} = \min\{t; \tilde{W}_t \geq L, t \geq 1\},$$
$$T_{\text{ATM}} = \min\{t; W_t \geq n_t L, t \geq 1\},$$

where

$$\tilde{W}_t = \max\left\{0, \tilde{W}_{t-1} + \left[\frac{X_t}{n_t}\log\frac{\theta_1}{\theta_0} - (\theta_1 - \theta_0)\right]\right\}, \quad \tilde{W}_0 = 0.$$

We notice that the CUSUM, WLR, and ATM methods are all equivalent when the sample size is constant. It is also worth pointing out that the design of the earlier three charts requires the specification of not only the prechange rate θ_0 but also the postchange rate θ_1. Of course, when θ_1 is unknown (in most applications), we can simply assign a reasonable value as in the traditional CUSUM practice (Hawkins and Olwell 1998).

As shown in Section 4, the two modifications of EWMAe charts, EWMAa1 and EWMAa2, and two modifications of CUSUM charts, WLR and ATM, share a similar drawback, that is, they have rather unsatisfactory run length distributions. When the sample sizes vary, the probability of false alarms after short runs may be dramatically increased, which inflates the run length standard deviation and hurts an operator's confidence in valid alarms. This undesirable characteristic has been observed for the traditional control charts with estimated parameters (c.f. Jensen et al. 2006). Too frequent and excessive early false alarms render these charts useless and thus, unacceptable in practice.

Although the EWMAe and EWMAM charts are quite sensitive to the parameter change, one may wonder how to construct a proper EWMA scheme by taking the varying sample sizes into account. Note that in these two charts the observed counts are divided by the corresponding sample sizes. Intuitively speaking, this procedure is to make a sequence of random variables whose expectations are the same over time, analogous to the traditional EWMA charts for normal observations. Furthermore, one may also want to obtain a centered and standardized sequence, for example, $(X_t - n_t\theta_0)/\sqrt{n_t\theta_0}$. Is there any rule we can follow in constructing EWMA-type charts? In the next section, we will answer this question and propose a new EWMA chart for monitoring Poisson count data with time-varying sample sizes.

3. WEIGHTED-LIKELIHOOD-BASED EWMA METHOD

In the statistical context, the maximum likelihood principle is one of the most popular methods in both estimation theory and hypothesis testing. The likelihood ratio test (LRT) is asymptotically optimal (under mild conditions) and is also found to be more efficient than other competitors in finite-sample cases. In quality control or sequential analysis, the CUSUM chart is directly derived from an LRT for a simple hypothesis. However, for EWMA-type charts, it seems difficult to have connection with an LRT. In what follows, we will demonstrate that an appropriate EWMA control chart can be derived under the framework of weighted likelihood ratio test, which naturally incorporates the varying sample sizes into the EWMA scheme for monitoring Poisson count data.

Recall the change-point model (1). The value of θ_0 is usually known, and the monitoring task is to test $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$ at each time point. By ignoring two constant terms with respect to θ, we can express the log-likelihood of the observation X_j as

$$l_j(\theta) = X_j \log\theta - n_j\theta.$$

At any time point t, consider the following exponentially weighted log-likelihood over samples 1 to t

$$Y_t(\theta;\lambda) = \sum_{j=0}^{t} \omega_{j,\lambda} l_j(\theta),$$

where $\lambda \in (0, 1]$ is a smoothing parameter, and $\omega_{j,\lambda} = \lambda(1-\lambda)^{t-j}$ is a sequence of constants to ensure that all the weights sum up to 1 as $t \to \infty$. For $j = 0$, (X_0, n_0) can be viewed as a pseudo "sample" and is chosen as $(n_1\theta_0, n_1)$ here. It does not play any important role in detecting the change but makes the definition of our chart proposed below operate like the traditional EWMA scheme. Obviously, $Y_t(\theta;\lambda)$ makes full use of all available samples up to the current time point t, and different samples are weighted as in an EWMA chart (i.e., the more recent samples receive more weight, and the weight decays exponentially over time). An analogous idea has been used by Qiu, Zou, and Wang (2010) for profile monitoring with arbitrary design points. In that article, the authors propose an exponentially weighted least-squared function to online update the regression function and construct monitoring statistics.

Given the value of λ, the maximum weighted likelihood estimate (MWLE) of θ at the time point t is defined as the solution to the following maximization problem

$$\hat{\theta}_t = \arg\max_{\theta} Y_t(\theta;\lambda).$$

By some simple algebra, we get

$$\hat{\theta}_t = \frac{\sum_{j=0}^{t}\omega_{j,\lambda}X_j}{\sum_{j=0}^{t}\omega_{j,\lambda}n_j} = \frac{Y_{c,t}}{Y_{p,t}},$$

and as a consequence we obtain the following $-2\times$logarithm of weighted LRT (WLRT) statistic

$$\begin{aligned} R_{t,\lambda} &= 2\left[Y_t(\hat{\theta}_t;\lambda) - Y_t(\theta_0;\lambda)\right] \\ &= 2\sum_{j=0}^{t}\omega_{j,\lambda}\left[l_j(\hat{\theta}_t) - l_j(\theta_0)\right] \\ &= 2\left[Y_{c,t}\log\frac{Y_{c,t}}{\theta_0 Y_{p,t}} - Y_{c,t} + Y_{p,t}\theta_0\right], \end{aligned}$$

where $Y_{c,t}$ and $Y_{p,t}$ are the exponentially weighted average of counts and populations, respectively. The WLRT statistic $R_{t,\lambda}$ can thus, be used as the monitoring statistic and the corresponding control chart triggers a signal if $R_{t,\lambda}$ exceeds some specified control limit. Hereafter, this chart is referred to as the WLRT-based EWMA (WEWMA) control chart. Note that $Y_{c,t}$ and $Y_{p,t}$ can be reexpressed as the following equivalent formulations

$$\begin{aligned} Y_{c,t} &= \lambda X_j + (1-\lambda)Y_{c,t-1}, \\ Y_{p,t} &= \lambda n_j + (1-\lambda)Y_{p,t-1}, \end{aligned}$$

where the initial values are $Y_{c,0} = \theta_0 n_1$ and $Y_{p,0} = n_1$, respectively, based on the pseudosample (X_0, n_0) defined earlier. Hence, the WEWMA control chart can still be conducted in a recursive fashion as the traditional EWMA charts do. Under some conditions imposed on n_t (cf. Mei, Han, and Tsui 2011), we can obtain the following proposition, whose proof is shown in Appendix B.

Proposition 1. Suppose there exist two constants $0 < n_{\min} < n_{\max} < \infty$ so that $n_t \in (n_{\min}, n_{\max})$ for all t. As $\lambda \to 0$ and $\lambda t \to \infty$,

$$\frac{\sum_{i=1}^{t} w_i n_i}{\sum_{i=1}^{t} w_i^2 n_i} R_{t,\lambda} \xrightarrow{d} \chi_1^2.$$

When λ is small, we can expect $\sum_{i=1}^{t} w_i n_i / \sum_{i=1}^{t} w_i^2 n_i$ will not change much over time. This result reveals the fact that the marginal distribution of the monitoring statistic $R_{t,\lambda}$ is almost the same from an asymptotic viewpoint, which allows us to use a fixed control limit for the WEWMA chart given the nominal IC ARL. Our simulation results shown in the next section concur with this asymptotic analysis that the IC run length distributions of WEWMA are not very sensitive to the control limit for different sample size patterns.

Note that when $n_j = n$ for all j, WEWMA chart reduces to the Poisson EWMA chart (equivalently speaking) investigated by Borror, Champ, and Rigdon (1998). A straightforward proof can be found in Appendix A. It is also worth pointing out that by taking $l_j(\theta)$ as the likelihood function of the normal distribution and using a similar procedure described earlier, we can show that the WLRT-based scheme will lead to the classical EWMA chart for normal observations (Lucas and Saccucci 1990). Hence, we emphasize here that the weighted-likelihood framework introduced earlier is applicable to most SPC monitoring problems. It can be used as a standard tool to derive the EWMA chart under certain complex circumstances in which it may not be appropriate to directly derive weighted averages of the observations, such as the case of Poisson count data with time-varying sample sizes.

As pointed out before, in practice we are often only interested in detecting an increase in the incident rate and thus, a one-sided chart is desirable. At a first glance, our proposed WEWMA chart is an omnibus one and the one-sided counterpart is not available at hand. In fact, the derivation of one-sided EWMA chart is quite straightforward and just amounts to considering the hypothesis problem: $H_0 : \theta = \theta_0$ versus $H_1 : \theta > \theta_0$. The MWLE in this situation is modified by (Shu, Jiang, and Wu 2012)

$$\tilde{\theta}_t = \hat{\theta}_t I(\hat{\theta}_t > \theta_0) + \theta_0 I(\hat{\theta}_t \le \theta_0)$$

since the function $Y_t(\theta;\lambda)$ is monotonically decreasing on the right side of θ_0 when $\hat{\theta}_t \le \theta_0$. Accordingly, by substituting $\tilde{\theta}_t$ into the WLRT, the WEWMA monitoring statistic becomes

$$\tilde{R}_{t,\lambda} = R_{t,\lambda} I(\hat{\theta}_t > \theta_0).$$

Finally, our proposed one-sided WEWMA control chart is

$$T_{\text{WEWMA}} = \min\left\{t; \tilde{R}_{t,\lambda} > L\frac{\lambda}{2-\lambda}, t \ge 1\right\},$$

where $L > 0$ is a control limit chosen to achieve a specific value of IC ARL. The constant $\frac{\lambda}{2-\lambda}$ in the above stopping time is just to make the control limit coefficient L not too close to zero so that we may search it more conveniently. It is an asymptotic representation of $\sum_{i=1}^{t} w_i n_i / \sum_{i=1}^{t} w_i^2 n_i$ when n_i is a constant. We can extend this chart to detecting decreases in the incidence rate without any difficulty.

In general, for EWMA-type control charts, a small value of λ leads to optimal detection of small shifts (cf. for example, Lucas and Saccucci 1990). This statement is still valid for the WEWMA chart. Based on our simulation results, we suggest choosing $\lambda \in [0.05, 0.2]$, which is a reasonable rule-of-thumb in practice. The computational effort of WEWMA is quite trivial and basically similar to that of the EWMAe and EWMAM methods. The control limit coefficient L can be found easily through simulations with the help of bisection searching algorithms. It is also worth pointing out that all the control charts for discrete data share a common shortcoming, that is, there may not be an exact control limit to achieve certain values of IC ARL. As mentioned by some authors (cf. Borror, Champ, and Rigdon 1998), the CUSUM-type charts suffer from this issue greatly but it can be much alleviated when using EWMA if a relatively small value of λ is chosen. In our experience, the WEWMA chart's IC ARL can always be attained quite closely if $\lambda \le 0.2$.

Before ending this section, we present WEWMA's asymptotic bounds of ARL. As pointed out in Mei, Han, and Tsui (2011), it is usually difficult to derive theoretical bounds for control charts without any assumption on the time-varying sample size n_t. Hence, we impose some conditions on the sample sizes, which follow the settings of Theorems 7.1–7.3 in Mei, Han, and Tsui (2011). Let

$$g(x) = x\log\frac{x}{n^*\theta_0} - x + n^*\theta_0,$$

and let $g^{-1}(\cdot)$ be its inverse function. Denote by ARL$(T(h))$ the ARL of the stopping time T with the control limit h. We assume

that the control limit h is large in the following theorem for asymptotic analysis.

Theorem 1. Assume that the population sizes n_t's reach the stationary value n^* at some finite time M and there exist two constants $0 < n_{\min} < n_{\max} < \infty$ so that $n_t \in (n_{\min}, n_{\max})$ for all t. Then, for the stopping time T_{WEWMA}, we have

1. When the process is in control, then

$$\frac{\sqrt{2\pi}e^{\eta^2/2}}{\eta K}(1+o(1)) \le \text{ARL}(T_{\text{WEWMA}}(h)) \le \sqrt{2\pi}\eta^7 e^{\eta^2/2}\log(\eta)(1+o(1)),$$

where $\eta = \sqrt{2-\lambda}(g^{-1}(\frac{h}{2}) - n^*\theta_0)/\sqrt{\lambda n^*\theta_0}$, $K = \int_0^\infty x\psi^2(x)dx$, $\psi(x) = 2x^{-2}\exp\{-2\sum_{n=1}^\infty \frac{\Phi(-x\sqrt{n}/2)}{n}\}$, and $\Phi(\cdot)$ is the distribution function of $N(0,1)$.

2. When the process is out of control, we have

$$\frac{1}{E_1^2}\left\{\eta^2 - 4d_1\eta\sqrt{2\log(\eta)}\right\} \le \text{ARL}(T_{\text{WEWMA}}(h)) \le \frac{1}{E_1^2}\left\{\eta^2 + 4d_1\eta\sqrt{\log(\eta)}\right\},$$

where $E_1 = \sqrt{n^*}(\theta_1 - \theta_0)/\sqrt{\theta_0}$ and $d_1^2 = \frac{\theta_1}{\theta_0}$.

This theorem can be considered as an application of Theorem 2 in Han and Tsung (2006) which presented a unified framework for the asymptotic analysis of any stopping time satisfying certain conditions. Thus, the proof of this theorem amounts to rewriting T_{WEWMA} into some appropriate forms within that unified framework and verifying the conditions in Han and Tsung (2006). Details of the proof are given in Appendix C.

4. PERFORMANCE COMPARISON

We present some simulation results in this section to compare the performance of the proposed WEWMA chart and some other procedures in the literature. All results in this section are obtained from 20,000 replications. The Fortran codes for implementing the proposed procedure are available from the authors upon request.

In our simulation study later, we investigate the performance of different control charts under various scenarios of time-varying sample sizes. For health surveillance, Mei, Han, and Tsui (2011) suggest to model population growth by the logistic model which is adopted here. In particular, they consider the following three models:

1. Increasing Scenario: $n_t = \frac{c_1}{1+\exp[-(t-c_2)/c_3]}$,
2. Fast Increasing Scenario: $n_t = \frac{2c_1}{1+\exp[-(t-(c_2+26))/c_3]}$,
3. Decreasing Scenario: $n_t = \frac{c_1/2.4}{1+\exp[(t-c_2)/c_3]} + 1$,

where $c_1 = 13.8065$, $c_2 = 11.8532$, and $c_3 = 26.4037$. According to Mei, Han, and Tsui (2011), Scenario (1) is the estimated curve from a realdataset discussed in the next section, Scenario (2) corresponds to the case that the population size increases quickly, and Scenario (3) is the case that the population size decreases rapidly to the stationary value. Dong, Hedayat, and Sinha (2008) and Ryan and Woodall (2010), respectively, consider a case with constant sample size and one with uniformly distributed n_t in their simulations, which are also used here:

4. Constant Scenario: $n_t = 10$ for all t,
5. Uniform Scenario: $n_t \sim U(10, 15)$.

Note that the case (5) involves stationary sample sizes although inhomogeneous. Finally, to appreciate the effectiveness of our WEWMA chart for other "stationary" sample sizes but with time-varying patterns, we consider the following sine function which varies cyclically over time

6. Sine Scenario: $n_t = 10|\sin(t)| + 1$.

We fix $\theta_0 = 1$ which is consistent with the setting in the literature.

4.1 IC Performance Comparison

First, we study IC run length distribution of the WEWMA chart. As recognized in the literature, it is often insufficient to summarize run length behavior by ARL, especially when the marginal distribution of the charting statistic is not the same for all time points t (cf. Jones, Champ, and Rigdon 2001; Mei 2008). As an alternative, the control chart performance will be summarized using ARL, percentiles of the marginal distribution of the run length, and standard deviation of the run length (SDRL). The control limits are set so that $\text{ARL}_0 \approx 300$, which is consistent with Mei, Han, and Tsui (2011). We also study the false alarm rate for the first 30 observations, $\Pr_{\text{IC}}(T \le 30)$ for each chart. We use the notation L, SE, $Q(.10)$, $Q(.90)$, and FAR to denote the control limit coefficient, standard error of ARL estimation, 10th percentile, 90th percentile, and false alarm rate, respectively. Here, the IC run length distribution is considered to be satisfactory if it is close to the geometric distribution (Hawkins and Olwell 1998) or more generally its variation is less than that of a geometric distribution. Note that when the run length distribution is geometric, the SDRL should be approximately equal to ARL_0 and $Q(.10)$, $Q(.90)$, and FAR are about 31, 690 and 0.0953, respectively.

We summarize the results of the control charts discussed in Section 2 under Scenarios (1)–(4), that is, the sample sizes are increasing, fast increasing, decreasing, and fixed, in Tables 1–4, respectively. The results for Scenario (5) and (6) are similar to that for Scenario (4) and thus, omitted here to save space. Note that in Table 4 in which the sample size is a constant, we only present the results of the EWMAe, CUSUM, EWMAM, and WEWMA methods because the EWMAe, EWMAa1, and EWMAa2 methods are equivalent, and the CUSUM, WLR, and ATM methods are equivalent as well. For convenience, here we use the same value $\lambda = 0.1$ for all the EWMA-type control charts and θ_1 is chosen as 2 in all the CUSUM-type charts, consistent with the work by Mei, Han, and Tsui (2011).

From Tables 1 and 2, we can find that the SDRLs and FARs of EWMAa1, EWMAa2, WLR, and ATM are much larger than the desired values. Excessive false alarms at early runs will make the detection results unreliable; consequently, these charts are not acceptable in terms of run length distributions. It can also be clearly seen that the geometric distribution is quite a reasonable approximation to the IC run length distributions of the EWMAe, CUSUM, EWMAM, and WEWMA charts. This

Table 1. IC ARL comparison under Scenario (1)

	L	ARL_0	SE	SDRL	$Q(0.10)$	Median	$Q(0.90)$	FAR
EWMAe	2.391	300	2.14	302	25	206	699	0.1152
EWMAa1	1.618	299	2.90	410	3	126	833	0.3242
EWMAa2	1.587	301	2.69	380	10	155	802	0.2496
CUSUM	3.578	297	2.32	328	23	179	722	0.1313
WLR	0.306	302	3.62	512	4	50	959	0.5433
ATM	0.306	299	3.64	515	4	52	941	0.5323
EWMAM	2.632	300	2.16	306	27	205	700	0.1119
WEWMA	2.721	299	2.16	306	31	202	696	0.0984

confirms that these charts work well under the IC condition and the ARL is a suitable summary of their IC run behavior.

Similarly, in Table 3, when n_t is decreasing, the run length distributions of EWMAa1, EWMAa2, WLR, and ATM are far away from the geometric distribution. In this case, their SDRLs are much smaller than the nominal one 300 which seems to be a benefit. Actually, this benefit comes from the fact that the probabilities of false alarms after short runs are significantly small, making the chart fail to trigger a quick detection of shifts (see the results in Figure 3). Tables 3–4 provide similar evidence to that of Tables 1–2, that is, the EWMAe, CUSUM, EWMAM, and WEWMA charts offer satisfactory in-control run length performance. It is important to point out that it is rather difficult to find the corresponding control limit for the CUSUM chart for the prespecified value of ARL_0 in Table 4. Our simulation result shows that the ARL_0 can only attain around 230 or 377. This is consistent with the previous discussions due to the discreteness of the Poisson distribution. We conducted some other simulations with various combinations of λ and IC ARL to check whether the earlier observations still hold in other settings. The simulation results show that these charts have quite a satisfactory performance in other cases as well.

Generally, the control limit coefficient L not only depends on the control charts, but also the underlying population models (n_t). However, when we set the control limit for a control chart, the actual population model is rarely known in advance. It is important to note that, comparing with other alternative control charts, the WEWMA chart has a relatively consistent control limit coefficient for different underlying population models. To verify this observation, we performed a sensitivity analysis of the control limit against different sample size settings. Table 5 shows the IC ARL values of EWMAe, CUSUM, EWMAM, and WEWMA charts for different underlying population models when the control limit of each chart is set assuming the constant sample size, that is, Scenario (4). It is easy to see that the WEWMA method performs quite stably under all cases of population models. This is due to the nice property of the WLRT statistic discussed in Proposition 1. In contrast, the ARL_0's of other charts have fairly large deviations from 300. For example, the CUSUM chart is very sensitive to the population model. Its ARL_0 value could be as large as 1000 under scenario (2). The ARL_0 values of the EWMAe chart is also far away from 300 under Scenarios (3). That is, if the actual population doesn't follow the assumed model, these charts may have very different IC ARL values than the postulated one. This turns to be a competitive advantage of the WEWMA chart since we don't need to worry too much about the accuracy of the underlying population models in practice.

4.2 OC Performance Comparison

In this section, we consider the out-of-control (OC) ARL comparison. Because a similar conclusion holds for other cases, throughout this section, we only present the results when $ARL_0 = 300$ for the illustration purpose. Results with other commonly used ARL_0 values, such as 500 or 800, are provided in the supplemental file (available online). As shown in the last section, the EWMAa1, EWMAa2, WLR, and ATM charts have unacceptable IC run length distributions. These modifications of the EWMA and CUSUM charts achieve the specified IC ARL with elevated probabilities of very short and very long runs, as compared with a geometric distribution. For instance, for the fast increasing sample-size scenario, the false alarm rates for the first 30 observations of both ATM and WLR charts are as large as 0.7. Under the OC model, the probabilities of very long runs would decrease and consequently they would have quite small ARLs compared to the WEWMA, CUSUM, and EWMAe charts. However, this "advantage" is mainly due to very large short-run false alarms due to randomness. In other words, the

Table 2. IC ARL comparison under Scenario (2)

	L	ARL_0	SE	SDRL	$Q(0.10)$	Median	$Q(0.90)$	FAR
EWMAe	2.371	299	2.16	306	23	205	694	0.1125
EWMAa1	1.212	299	3.81	540	1	32	978	0.4987
EWMAa2	1.183	298	3.47	491	5	51	932	0.4161
CUSUM	2.802	300	2.71	383	11	148	812	0.2552
WLR	0.155	332	6.11	863	2	12	1228	0.7175
ATM	0.155	332	6.11	864	2	12	1225	0.7212
EWMAM	2.609	300	2.13	302	27	207	698	0.0942
WEWMA	2.757	300	2.23	316	29	199	712	0.1025

Table 3. IC ARL comparison under Scenario (3)

	L	ARL_0	SE	SDRL	$Q(0.10)$	Median	$Q(0.90)$	FAR
EWMAe	2.550	299	2.13	302	38	204	697	0.0797
EWMAa1	2.342	300	1.52	214	105	238	580	0.0002
EWMAa2	2.341	299	1.49	211	106	238	574	0.0003
CUSUM	3.705	302	2.33	329	24	191	726	0.1210
WLR	3.257	302	1.47	208	115	240	570	0.0000
ATM	3.260	299	1.47	208	113	237	571	0.0000
EWMAM	2.809	300	2.07	293	36	212	677	0.0761
WEWMA	2.660	300	1.95	275	40	220	659	0.0812

ARL (or expectation of detection delay) is not a good index for the comparison between the WEWMA chart and the four modifications.

To demonstrate the difference between the aforementioned control charts, we consider $\gamma_t \equiv \Pr_{\text{OC}}(T \le t) - \Pr_{\text{IC}}(T \le t)$, that is, the "pure" probability that a stopping time T detects an OC condition before time point t beyond randomness. We compare the aforementioned control charts using the values of γ_t for $t \le 100$ which correspond to early detection. All the control charts are designed to achieve the nominal IC ARL. Apparently, a control chart with a larger value of γ_t is considered better. This quantity reflects the "true" detection capability of a chart and thus, would be a reasonable index for OC comparison given that the run-length distributions of some charts are far away from geometric. Some representative results under Scenarios (1)–(3) are shown in Figures 1–3, respectively. In each figure, the first plot depicts the cumulative distribution function (CDF) of IC run-length distributions, that is, $\Pr_{\text{IC}}(T \le t)$ and the other three plots show the γ_t values for $\theta = 1.05$, 1.2, and 1.5, respectively. Note that the CUSUM, EWMAe, and EWMAM charts are not included in the three figures because their curves are similar to those of the WEWMA chart. Meanwhile the curves of the ATM and WLR charts are not distinguishable in all the plots and thus, only the results of the WLR chart are provided for illustration. The smoothing parameter λ for all the EWMA-type charts is fixed as 0.1 and the tuning parameter θ_1 for the WLR chart is chosen as 1.3 for a relatively fair comparison (see more discussions regarding this choice later). We can see that the IC run-length distribution of the WEWMA chart is quite similar to the geometric distribution in all of the three scenarios, while that of the EWMAa1, EWMAa2, and WLR deviates significantly from the geometric. The WEWMA outperforms the other three charts in the sense that its γ_t curve increases much faster after a change occurs. Note that under Scenarios (3), the γ_t curve of the WEWMA tends to be lower than those of the other three modifications when t becomes large. This is because when θ is large (e.g., $\theta = 1.5$), the values of $\Pr_{\text{OC}}(T \le t)$ increase to 1 very quickly and accordingly a chart with larger false alarm rate will have a smaller value of γ_t for large values of t (in this case, the false alarm rates of those modifications in short-runs are rather small compared to the geometric). For other values of ARL_0, similar patterns can be observed (see the results for $\text{ARL}_0 = 800$ in the supplemental file, available online).

Next, we compare the WEWMA with the EWMAe, CUSUM, and EWMAM charts in terms of OC ARL. Since the zero-state and steady-state ARL (SSARL) comparison results are similar, only the SSARLs are provided. To evaluate the SSARL behavior of each chart, any series in which a signal occurs before the $(\tau + 1)$th observation is discarded (cf. Hawkins and Olwell 1998). Here, we consider $\tau = 20$ for illustration. In order to assess the overall performance of these charts, besides OC ARLs, we also compute their relative mean index (RMI) values. The RMI index of a control chart, suggested by Han and Tsung (2006), is defined as

$$\text{RMI} = \frac{1}{N}\sum_{l=1}^{N}\frac{\text{ARL}_{\delta_l} - \text{MARL}_{\delta_l}}{\text{MARL}_{\delta_l}}, \tag{2}$$

where N is the total number of shifts considered, ARL_{δ_l} is the OC ARL of the given control chart when detecting a parameter shift of magnitude $\boldsymbol{\delta}_l$, and MARL_{δ_l} is the smallest among all OC ARL values of the charts considered when detecting the shift $\boldsymbol{\delta}_l$. So, $(\text{ARL}_{\boldsymbol{\delta}_l} - \text{MARL}_{\boldsymbol{\delta}_l})/\text{MARL}_{\delta_l}$ could be considered as a relative efficiency measure of the given control chart, compared to the best chart, when detecting the shift δ_l, and RMI is the average of all such relative efficiency values. Based on this index, a control chart with a smaller RMI value is considered better in its overall performance. To save space, we only list the OC values of $\theta = 1.025$, 1.05, 1.1, 1.2, 1.4, 2, 3 in the following three tables, but the RMI values are evaluated at OC values of $\theta = 1.025$, 1.05, 1.1, 1.2, 1.3, 1.4, 1.5, 1.7, 2, 3, 4.

For a relatively fair comparison, we choose appropriate values of θ_1 for the CUSUM chart under different scenarios. To be more specific, we set θ_1 equal to the shift level at which the WEWMA with $\lambda = 0.1$ is approximately the best of detection across all

Table 4. IC ARL comparison under Scenario (4)

	L	ARL_0	SE	SDRL	$Q(0.10)$	Median	$Q(0.90)$	FAR
EWMAe	2.401	300	2.18	308	24	204	705	0.1227
CUSUM	3.863	377	2.64	374	40	263	857	0.0748
EWMAM	2.640	299	2.15	304	29	205	695	0.1018
WEWMA	2.688	300	2.09	296	36	208	684	0.0822

Table 5. The sensitivity comparison of the control limit

	L	Scenario (1)	Scenario (2)	Scenario (3)	Scenario (5)	Scenario (6)
EWMAe	2.401	306 (314)	320 (332)	228 (231)	296 (301)	281 (291)
CUSUM	3.863	372 (289)	999 (1129)	355 (386)	375 (371)	308 (306)
EWMAM	2.640	312 (316)	324 (330)	217 (213)	298 (302)	269 (275)
WEWMA	2.688	293 (300)	283 (293)	307 (287)	300 (297)	304 (299)

NOTE: Standard deviations are in parentheses.

the values of λ. For example, in Scenario (1), we found by simulations that $\theta = 1.4$ is the shift in which the WEWMA with $\lambda = 0.1$ is roughly optimal in the sense that with other values of λ the WEWMA cannot be (or significantly) better than that with 0.1. By doing this, θ_1 is chosen as 1.4, 1.3, 1.4, 1.2, 1.2, and 1.3 for Scenarios (1)–(6), respectively. Table 6 presents the SSARL values for various shifts in the Poisson rate, where n_t is in the cases of increasing and fast increasing. The WEWMA chart has a better performance compared with all other control charts for shifts up to $\theta = 1.4$. The EWMAM chart outperforms the WEWMA chart when θ is larger than 1.5 and the WEWMA chart performs generally better than the EWMAe chart. The CUSUM chart has slightly larger OC ARLs than the WEWMA but also provides satisfactory detection ability in all the cases. In terms of the RMI index, WEWMA performs the best overall in these two scenarios.

Next, we turn to the comparison under decreasing, constant, and random population scenarios. The simulation results are summarized in Tables 7–8. We can see that the three EWMA charts provide similar and comparable detection ability in these cases. In general, the EWMAe and WEWMA charts are more sensitive to small shifts whereas the EWMAM chart is more powerful in detecting large shifts. The WEWMA chart offers quite satisfactory performance and the overall performance difference between it and the other two EWMA charts is minor in terms of RMI values. With $\theta_1 = 1.4$, the CUSUM chart

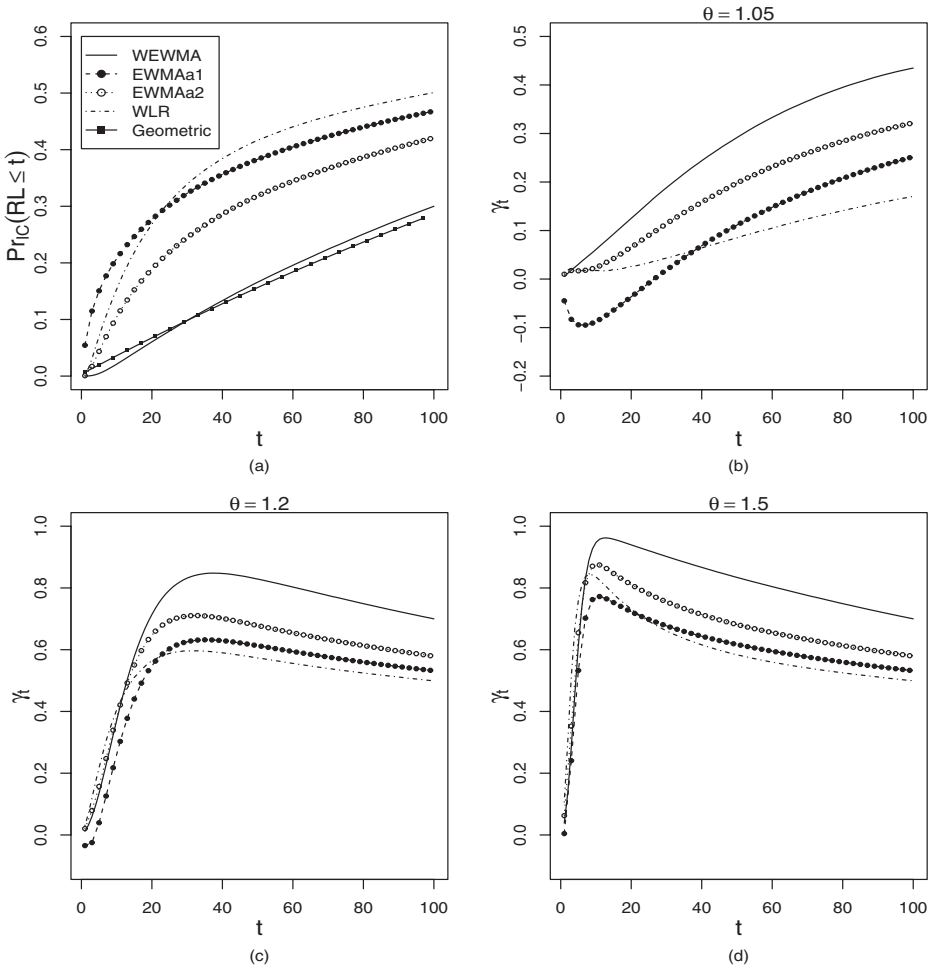

Figure 1. Performance comparison between WEWMA, EWMAa1, EWMAa2, and WLR under Scenario (1): (a) four in-control CDF curves along with geometric distribution (with expectation 300); (b)–(d) Curves of $\gamma_t \equiv \Pr_{OC}(T \le t) - \Pr_{IC}(T \le t)$ when $\theta = 1.05$, 1.2, and 1.5, respectively. The legend in the first plot is applicable for all the others.

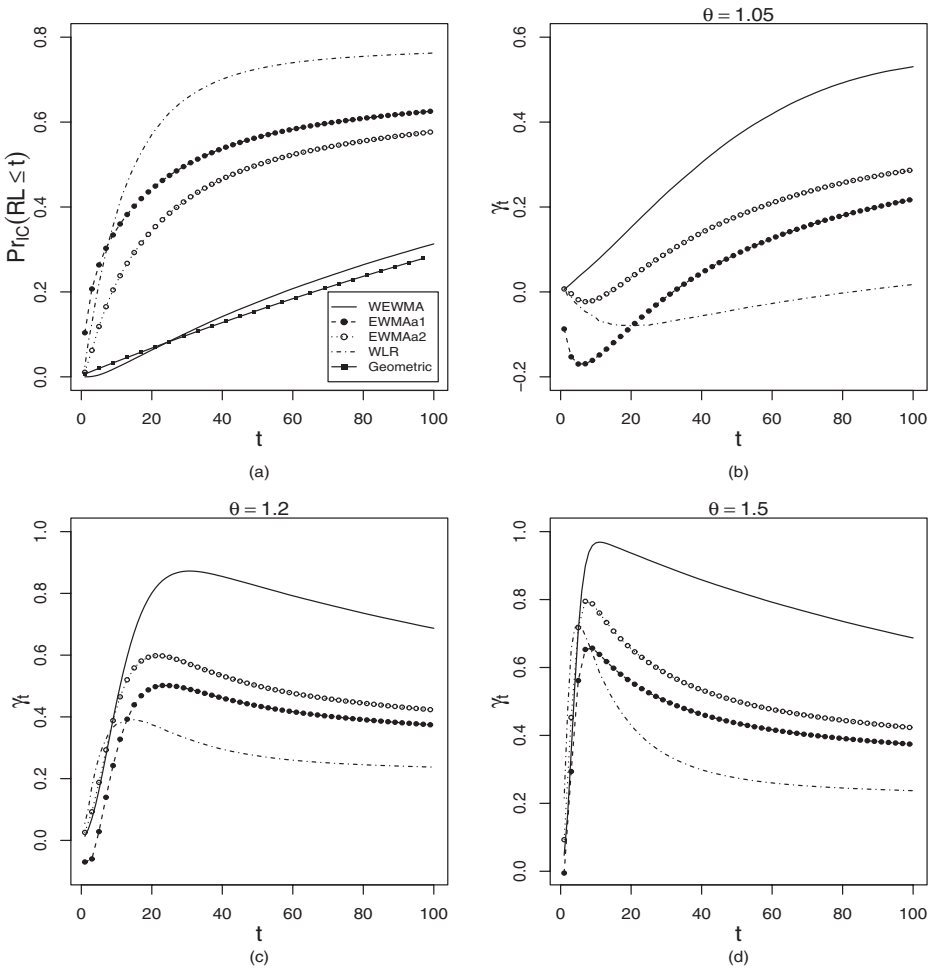

Figure 2. Performance comparison between WEWMA, EWMAa1, EWMAa2, and WLR under Scenario (2): (a) four in-control CDF curves along with the geometric distribution (with expectation 300); (b)–(d) Curves of $\gamma_t \equiv \Pr_{OC}(T \le t) - \Pr_{IC}(T \le t)$ when $\theta = 1.05$, 1.2, and 1.5, respectively.

performs almost uniformly better than the other three charts under the decreasing population scenario. The four control charts have similar detection abilities under the constant and random population scenarios.

Finally, we compare these four charts under Scenarios (6) in which the sample sizes vary according to a sine function. The results are tabulated in the last four columns of Table 8. In this situation, the proposed WEWMA chart outperforms the other two EWMA charts by quite a significant margin. The CUSUM chart works reasonably well in detecting various magnitudes of shifts in this scenario and has similar detection ability to the WEWMA chart. This can be expected because the CUSUM chart, proposed by Mei, Han, and Tsui (2010), was developed under the framework of LRT and change-point detection. It is efficient in a certain sense due to the full utilization of the information from the process. We should emphasize that the real measurement for which the population (sample) size changes as a sine function may rarely be seen in practical healthcare or surveillance applications, but this example reflects the robustness of the WEWMA chart and confirms our arguments that the WLRT-based scheme may be more appropriate than other alternatives in practice for dealing with time-varying sample sizes.

We conducted some other simulations with various IC ARL, θ_0, λ, and τ, to check whether the aforementioned conclusions would change in other cases. Some representative simulation results are reported in the supplemental file (available online) to show that the WEWMA chart works well for other cases as well in terms of the OC ARL. The comparison conclusion still generally holds. To summarize, by considering its efficiency, robustness, ease of construction, and fast computation, the WEWMA chart should be a reasonable alternative for monitoring Poisson count data with time-varying sample sizes.

5. ANALYTICAL BOUNDS FOR ARL

Dong, Hedayat, and Sinha (2008) gave the analytical bounds of ARL_0 and ARL_1 for the EWMAe chart. Here, we present some simulation results to illustrate the performance of our analytical bounds and compare them with those given by Dong, Hedayat, and Sinha (2008). To give a broad picture of the two methods, we consider two commonly used values $\lambda = 0.1$ and 0.2, and calculate ARL_1 with $\text{ARL}_0 = 100$, 500, 800, 1000, 2000, 3000, or 4000. The approximate ARL values of the EWMAe chart are computed from the analytical bounds discussed in Sections 2 and 3 of Dong, Hedayat, and Sinha (2008). The

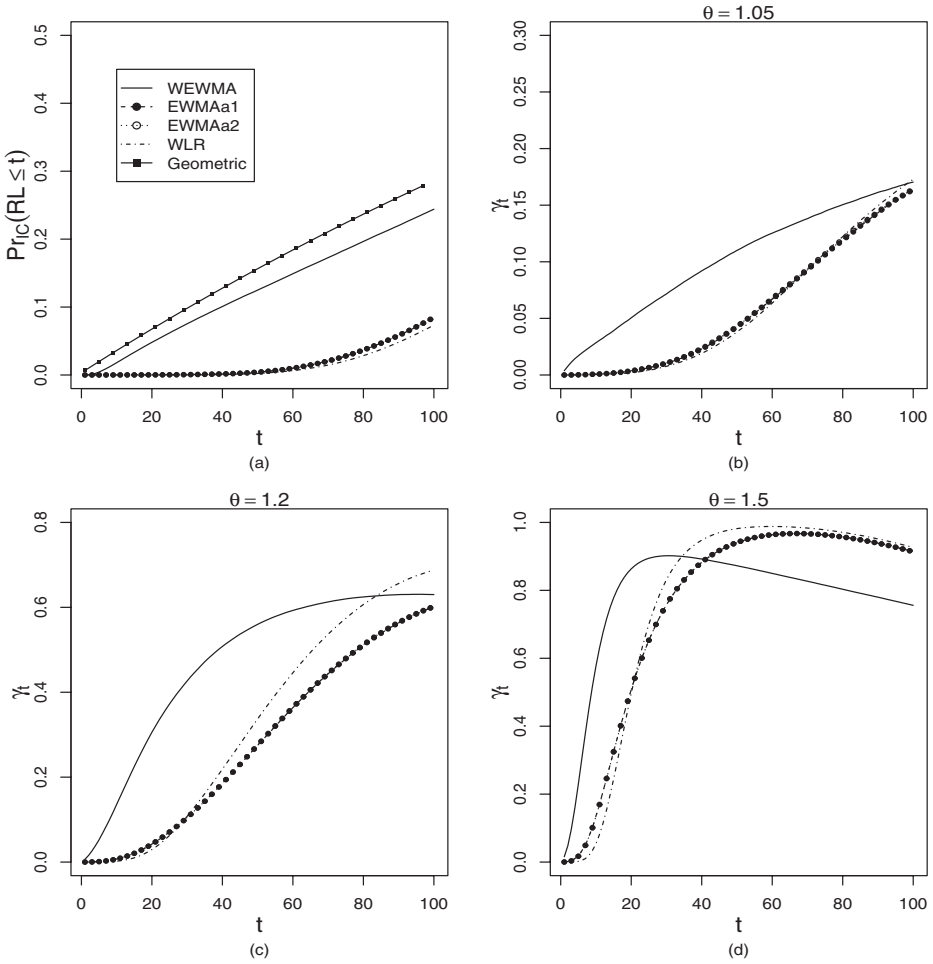

Figure 3. Performance comparison between the WEWMA, EWMAa1, EWMAa2, and WLR under Scenario (3): (a) four in-control CDF curves along with geometric distribution (with expectation 300); (b)–(d) Curves of $\gamma_t \equiv \Pr_{OC}(T \le t) - \Pr_{IC}(T \le t)$ when $\theta = 1.05$, 1.2, and 1.5, respectively.

ARL bounds of the WEWMA chart are derived from Theorem 1. Following Dong, Hedayat, and Sinha (2008), we study the situation in which $\theta_0 = 1$, $\theta_1 = 2$, and $n_t = 10$ (in units of thousand) for all $t \ge 1$, and $\tau = 20$.

Table 9 presents ARL_0 and its lower bound and ARL_1 and its upper bound for different fixed values of ARL_0. From the asymptotic analysis and empirical results shown above, we know that the EWMAe and WEWMA charts are ARL-unbiased. That is, the value of ARL_1 should be always smaller than the corresponding ARL_0. Hence, the analytic bounds for ARL_1 which are larger than the ARL_0s are useless from a practical viewpoint. In Table 9, the entries with the symbol "-" represent that the values are larger than the corresponding ARL_0. It is clearly seen that all the upper bounds of EWMAe are larger than the

Table 6. OC ARL comparison under Scenarios (1) and (2)

	Scenario (1)				Scenario (2)			
θ	EWMAe	CUSUM $\theta_1 = 1.4$	EWMAM	WEWMA	EWMAe	CUSUM $\theta_1 = 1.3$	EWMAM	WEWMA
1.025	144 (137)	178 (175)	158 (153)	138 (137)	118 (104)	154 (145)	129 (118)	109 (104)
1.050	81.8 (72.0)	112 (107)	90.5 (82.4)	76.6 (70.8)	62.2 (47.7)	89.9 (78.9)	69.1 (56.0)	57.9 (47.9)
1.100	37.2 (28.3)	53.6 (47.4)	40.8 (32.7)	34.6 (27.3)	29.6 (19.4)	39.7 (30.6)	31.5 (21.7)	27.5 (18.8)
1.200	15.7 (9.85)	18.8 (14.5)	16.0 (10.9)	14.8 (9.50)	13.6 (7.71)	15.1 (9.86)	13.4 (8.06)	12.4 (7.43)
1.400	7.01 (3.56)	6.70 (3.98)	6.56 (3.52)	6.50 (3.45)	6.26 (3.02)	5.96 (3.05)	5.84 (2.89)	5.66 (2.82)
2.000	2.77 (1.13)	2.29 (0.93)	2.47 (0.97)	2.56 (1.08)	2.58 (1.01)	2.25 (0.81)	2.29 (0.87)	2.33 (0.93)
3.000	1.56 (0.57)	1.23 (0.43)	1.38 (0.50)	1.48 (0.55)	1.48 (0.54)	1.26 (0.44)	1.29 (0.46)	1.36 (0.50)
RMI	0.119	0.158	0.087	0.049	0.107	0.163	0.067	0.016

NOTE: Standard deviations are in parentheses.

Table 7. OC ARL comparison under Scenarios (3) and (4)

	Scenario (3)				Scenario (4)			
θ	EWMAe	CUSUM $\theta_1 = 1.4$	EWMAM	WEWMA	EWMAe	CUSUM $\theta_1 = 1.2$	EWMAM	WEWMA
1.025	238 (238)	235 (267)	240 (238)	230 (220)	155 (153)	160 (158)	167 (166)	152 (150)
1.050	185 (185)	179 (206)	197 (198)	182 (179)	88.0 (82.5)	93.4 (89.1)	97.9 (94.9)	87.3 (81.9)
1.100	114 (120)	102 (125)	129 (136)	114 (116)	38.3 (32.6)	40.6 (35.9)	43.5 (39.3)	38.0 (32.6)
1.200	47.4 (51.9)	39.3 (48.6)	55.6 (62.9)	48.3 (51.6)	14.8 (9.97)	14.8 (10.3)	15.3 (11.3)	14.8 (10.1)
1.400	14.9 (12.6)	11.9 (10.9)	15.6 (14.8)	14.8 (11.9)	6.26 (3.24)	6.05 (3.04)	5.87 (3.15)	6.22 (3.23)
2.000	4.46 (2.22)	3.57 (1.87)	4.17 (2.18)	4.55 (2.29)	2.46 (0.98)	2.36 (0.85)	2.21 (0.85)	2.45 (0.98)
3.000	2.25 (0.93)	1.82 (0.74)	2.06 (0.83)	2.32 (0.95)	1.41 (0.52)	1.33 (0.47)	1.25 (0.44)	1.41 (0.52)
RMI	0.194	0.002	0.219	0.204	0.058	0.046	0.036	0.052

NOTE: Standard deviations are in parentheses.

corresponding ARL_0. This is consistent with the results shown in Dong, Hedayat, and Sinha (2008). In fact, in some cases, the upper bounds of ARL_1 provided by Dong, Hedayat, and Sinha (2008) are not finite (see some detailed discussions on Theorem 2 in Dong, Hedayat, and Sinha 2008). In comparison, the analytic bounds given by Theorem 1 work fairly well for both ARL_0 and ARL_1. When ARL_0 is large, both the lower bound for ARL_0 and the upper bound for ARL_1 are quite close to the actual values, especially when $\lambda = 0.1$. Hence, these bounds are useful for approximating the ARL behavior of the WEWMA chart and the lower bound for ARL_0 also serves as a good starting point for finding the control limits.

6. A HEALTH SURVEILLANCE EXAMPLE

In this section, we demonstrate the proposed methodology by applying it to the male thyroid cancer incidence dataset. The dataset was collected by the New Mexico Tumor Registry for the Surveillance, Epidemiology, and End Results (SEER) program at the National Cancer Institute in New Mexico from 1973 to 2005. This example has been studied by Mei, Han, and Tsui (2011) and Shu, Jiang, and Tsui (2011).

Known risk factors for thyroid cancer include the exposure to ionizing radiation during childhood, radiation treatment, and radioactivity from nuclear explosions or other sources. The observed variables in the dataset were the number of male thyroid cancer cases together with the age-specific population size for each year and each county. The population size was estimated based on the decennial US census. The data were geographically aggregated into 32 counties. The total male population increased from 546,000 in 1973 to 946,000 in 2005, as shown in Figure 4(a). The thyroid cancer incidence rate is low. In New Mexico, a total of 863 cases was reported during the period 1973–2005, which is rare as compared to the population size. The time series plots of the counts and the (estimated) incidence rate (per 100,000) of male thyroid cancer in New Mexico are shown in Figures 4(b)–4(c), respectively. It is clear that the incidence rate remained relatively stable before 1994 and exhibits an increasing tendency beginning in 1994. Readers may refer to Mei, Han, and Tsui (2011) and Shu, Jiang, and Tsui (2011) and the references therein for details.

Because the incidence rate is relatively stable during the period from 1973 to 1983, Mei, Han, and Tsui (2011) used this period of data to estimate the IC incidence rate as 2.4 per 100,000. The remaining data from 1984 to 2005 were treated as Phase II data assuming that they were available sequentially afterwards. Following Mei, Han, and Tsui (2011), we also used the IC incidence rate of $\theta_0 = 2.4$. The estimated sample size function is just the function given in Scenario (1) in Section 4. We set $\lambda = 0.1$ for the WEWMA chart and the simulation leads to a control limit 2.713 to attain $ARL_0 = 300$.

Figure 5 presents the resulting WEWMA chart (solid curve connecting the dots) along with its control limit (the solid horizontal line). The corresponding CUSUM chart used by Mei,

Table 8. OC ARL comparison under Scenarios (5) and (4)

	Scenario (5)				Scenario (6)			
θ	EWMAe	CUSUM $\theta_1 = 1.2$	EWMAM	WEWMA	EWMAe	CUSUM $\theta_1 = 1.3$	EWMAM	WEWMA
1.025	153 (150)	159 (158)	168 (167)	151 (148)	192 (189)	179 (179)	199 (198)	164 (160)
1.050	87.4 (81.4)	93.7 (90.7)	99.7 (96.1)	86.9 (82.1)	124 (121)	115 (113)	135 (131)	100 (96.2)
1.100	38.7 (33.0)	40.5 (35.9)	43.3 (39.0)	37.9 (32.5)	60.9 (54.3)	54.8 (51.4)	67.1 (63.5)	46.4 (40.4)
1.200	15.0 (10.2)	14.9 (10.5)	15.3 (11.4)	14.7 (9.98)	23.7 (18.1)	19.7 (15.8)	25.4 (20.6)	18.4 (13.1)
1.400	6.28 (3.26)	6.08 (3.03)	5.90 (3.18)	6.25 (3.23)	9.78 (5.16)	7.44 (4.22)	9.73 (5.48)	7.75 (4.07)
2.000	2.46 (0.98)	2.37 (0.86)	2.21 (0.85)	2.44 (0.97)	3.92 (1.83)	2.76 (1.16)	3.70 (1.74)	3.09 (1.31)
3.000	1.41 (0.51)	1.32 (0.47)	1.25 (0.44)	1.41 (0.51)	2.01 (0.96)	1.28 (0.64)	1.79 (0.89)	1.62 (0.89)
RMI	0.060	0.047	0.040	0.052	0.336	0.045	0.320	0.064

NOTE: Standard deviations are in parentheses.

Table 9. ARL_0 lower bound and ARL_1 upper bound comparison

	$\lambda = 0.1$						$\lambda = 0.2$					
	ARL_0 lower bound		ARL_1		ARL_1 upper bound		ARL_0 lower bound		ARL_1		ARL_1 upper bound	
ARL_0	EWMAe	WEWMA	EWMAe	WEWMA	EWMAe	WEWMA	EWMAe	WEWMA	EWMAe	WEWMA	EWMAe	WEWMA
100	30	75	2.08	2.04	–	1.11	61	63	1.81	1.79	–	1.47
300	122	225	2.46	2.45	–	1.86	251	185	2.10	2.10	–	2.18
500	240	385	2.64	2.63	–	2.20	478	310	2.25	2.25	–	2.48
800	448	637	2.79	2.80	–	2.50	910	503	2.37	2.37	–	2.76
1000	590	794	2.89	2.89	–	2.63	1200	638	2.44	2.44	–	2.89
2000	1450	1685	3.11	3.12	–	3.04	3056	1321	2.63	2.63	–	3.28
3000	2466	2643	3.27	3.26	–	3.28	5075	2061	2.77	2.75	–	3.50
4000	3689	3667	3.41	3.42	–	3.44	7420	2899	2.85	2.86	–	3.67

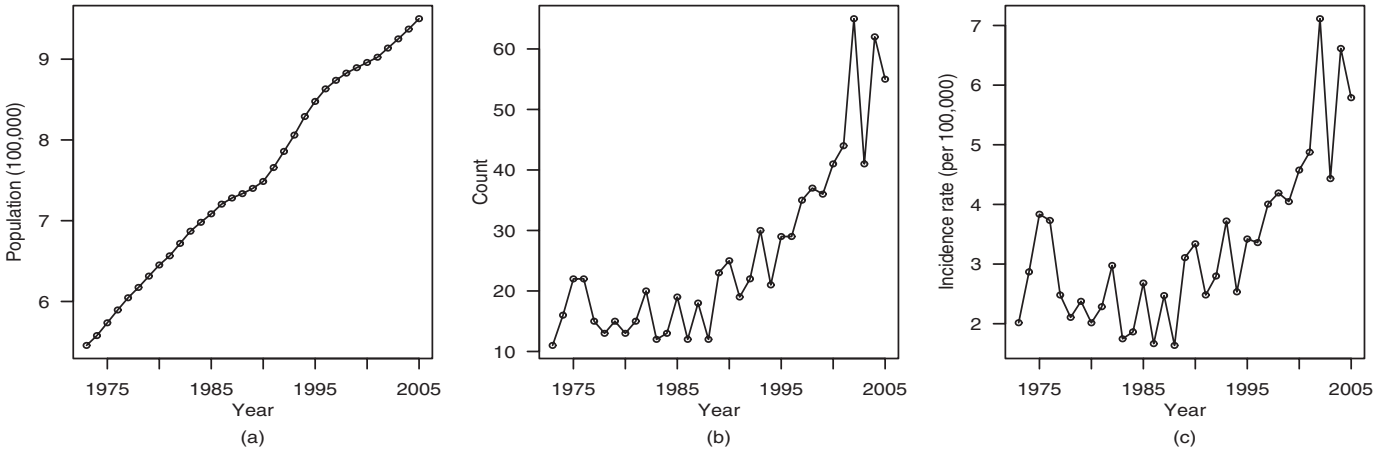

Figure 4. Male thyroid cancer incidence data: (a) Male population, (b) thyroid cancer counts, and (c) incidence rate per 100,000 persons.

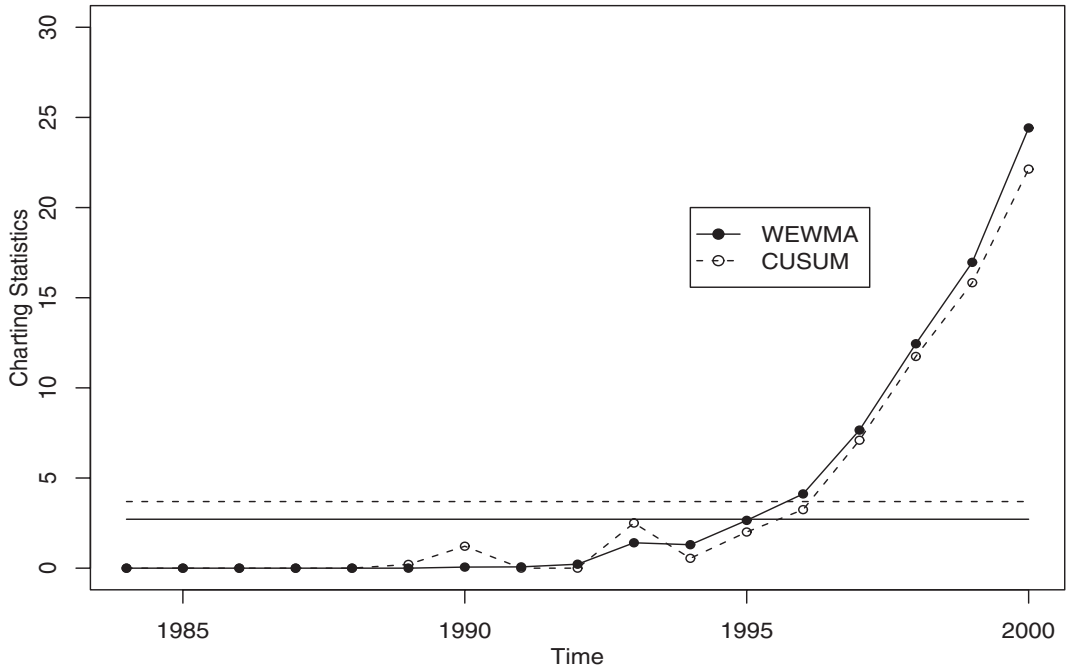

Figure 5. The WEWMA and CUSUM control charts for monitoring the male thyroid cancer incidence dataset. The solid and dashed horizontal lines indicate their control limits, respectively

Han, and Tsui (2011) with $\theta_1 = 3.8$ (dashed curve connecting circles) is also presented in the figure, along with its control limits of 3.694 by dashed line. Note that $\theta_1 = 3.8$ is recommended by Mei, Han, and Tsui (2011) and is estimated based on the "future" Phase II observations. From the plot, it can be seen that the WEWMA chart exceeds its control limit at year 1996 and it remains above the control limit all along. This excursion suggests that a marked step change has occurred. In comparison, the CUSUM chart gives a signal at year 1997. Note that in the design of our WEWMA chart, we haven't tuned its parameter to the postchange size and only consider a common value of λ. Nevertheless, the WEWMA chart provides a similar ability as the CUSUM chart in detecting the increasing risk of male thyroid cancer, which justifies its usefulness in real applications.

7. CONCLUDING REMARKS

In this article, we propose a new EWMA scheme, WEWMA, for monitoring Poisson count data with time-varying sample sizes. This chart is derived based on the weighted likelihood ratio test and naturally integrates the varying sample sizes with the EWMA scheme. With updating formulations, the proposed scheme is fast to compute with a computational effort similar to other EWMA charts. Compared with existing methods, it is not only more robust in IC and OC performance, but also generally more sensitive to the small and moderate parameter changes. In many cases, the improvement is quite remarkable. Especially when the sample sizes vary significantly over time, it significantly outperforms other competitors.

This article focuses on Phase II monitoring only and presumes that all historical observations used for estimating the IC parameters follow independent Poisson distributions with identical incidence rates. In many practical applications, there is no such assurance. Hence, it requires more research to extend our method to Phase I analysis, in which detection of outliers or change-points in a historical dataset and estimation of the baseline incidence rate would be of great interest. Moreover, it is known that the performance of all control charts is affected by the amount of data in the reference dataset. Thus, the determination of required Phase I sample sizes to ensure reasonable performance of the control charts with estimated parameters is needed. Furthermore, future research needs to be directed to develop a self-starting version of the WEWMA chart which can simultaneously update parameter estimates and check for OC conditions (e.g., Quesenberry 1995). Finally, in light of the importance of robust IC ARL performance with different patterns of variation in the sample sizes, a possible topic for future research would be a chart that is optimal in this property. Since the information content varies with the sample size, the control limit would logically also vary and should be determined online in terms of some criterion given the observations n_t.

APPENDIX A: THE EQUIVALENCE BETWEEN THE WEWMA AND THE POISSON EWMA CHART WHEN THE SAMPLE SIZE IS FIXED

When $n_j = n$ for all j, the –2logarithm of WLRT becomes

$$R_{t,\lambda} = 2\left[Y_{c,t}\log\frac{Y_{c,t}}{n\theta_0\sum_{j=0}^{t}\omega_{j,\lambda}} - Y_{c,t} + n\theta_0\sum_{j=0}^{t}\omega_{j,\lambda}\right].$$

When t is large, $\sum_{j=0}^{t}\omega_{j\lambda} \approx 1$. Thus, $R_{t,\lambda}$ can be rewritten as

$$R_{t,\lambda} \approx 2\left[Y_{c,t}\log\frac{Y_{c,t}}{n\theta_0} - Y_{c,t} + n\theta_0\right].$$

By taking derivatives of $R_{t,\lambda}$ with respect to $Y_{c,t}$, we can easily see that $R_{t\lambda}$ is monotonically increasing (decreasing) on the right (left) side of $n\theta_0$. Thus, the test based on $R_{t,\lambda}$ is essentially equivalent to the test

$$|Y_{c,t} - n\theta_0| > C,$$

where C is some given critical value. Obviously, by noting that $Y_{c,t}$ admits the classical EWMA updating formulas, and using the earlier test at each time point leads to the EWMA control chart studied by Borror, Champ, and Rigdon (1998).

APPENDIX B: THE PROOF OF PROPOSITION 1

Denote $\hat{\alpha}_t = \hat{\theta}_t/\theta_0$. Under H_0, we have

$$E(\hat{\alpha}_t) = 1, \quad \text{and} \quad \text{var}(\hat{\alpha}_t) = \frac{\sum_{i=1}^{t} w_i^2 n_i}{\theta_0 Y_{p,t}^2}.$$

Note that $\hat{\alpha} - 1$ can be expressed as a linear combination of iid variables, say

$$\hat{\alpha} - 1 = \frac{1}{\theta_0 Y_{p,t}}\sum_{i=1}^{t} w_i\sqrt{n_i\theta_0}\,\frac{X_i - n_i\theta_0}{\sqrt{n_i\theta_0}}.$$

When $n_i \in (n_{\min}, n_{\max})$

$$\max_{1\le i\le t}\frac{(w_i\sqrt{n_i\theta_0})^2}{\sum_{i=1}^{t}(w_i\sqrt{n_i\theta_0})^2} \le \frac{n_{\max}}{n_{\min}\sum_{i=1}^{t} w_i^2} \to 0,$$

as $\lambda t \to \infty$ and $\lambda \to 0$. Thus, by the Hajek–Sidak's Theorem, we have

$$[\text{var}(\hat{\alpha}_t)]^{-1/2}(\hat{\alpha}_t - 1) \xrightarrow{d} N(0, 1).$$

Hence, by the second-order Taylor expansion

$$\begin{aligned} R_{t,\lambda} &= 2Y_{p,t}\theta_0[\hat{\alpha}_t\log(\hat{\alpha}_t) - (\hat{\alpha}_t - 1)] \\ &= 2Y_{p,t}\theta_0\left\{\hat{\alpha}_t[(\hat{\alpha}_t - 1) - 2^{-1}(\hat{\alpha}_t - 1)^2] - (\hat{\alpha}_t - 1)\right\} \\ &\quad + o_p(Y_{p,t}\text{var}(\hat{\alpha}_t)) \\ &= Y_{p,t}\theta_0(1 - \hat{\alpha}_t)^2(2 - \hat{\alpha}_t) + o_p(Y_{p,t}\text{var}(\hat{\alpha}_t)) \\ &= \left(\frac{1 - \hat{\alpha}_t}{[\text{var}(\hat{\alpha}_t)]^{1/2}}\right)^2 (2 - \hat{\alpha}_t)Y_{p,t}\theta_0 \cdot \text{var}(\hat{\alpha}_t) + o_p(Y_{p,t}\text{var}(\hat{\alpha}_t)). \end{aligned}$$

By noting $2 - \hat{\alpha}_t \to^p 1$ and using Slutsky's Theorem, we have

$$\frac{\sum_{i=1}^{t} w_i n_i}{\sum_{i=1}^{t} w_i^2 n_i} R_{t,\lambda} \xrightarrow{d} \chi_1^2.$$

APPENDIX C: THE PROOF OF THEOREM 1

When t is larger than M, the observations are iid from Poisson($n^*\theta_0$). Moreover, $Y_{p,t} = \sum_{j=0}^{t}\omega_{j,\lambda}n_j \to n^*$ as $t \to \infty$. Similar to Proposition 1, given a sufficiently large control limit $h > 0$, the test

$$\tilde{R}_{t,\lambda} = R_{t,\lambda}I(\widehat{\theta}_t > \theta_0) > h$$

is essentially equivalent to the test

$$\sqrt{\frac{2-\lambda}{\lambda}}\frac{Y_{c,t} - n^*\theta_0}{\sqrt{n^*\theta_0}} > \sqrt{\frac{2-\lambda}{\lambda}}\frac{g^{-1}(\frac{h}{2}) - n^*\theta_0}{\sqrt{n^*\theta_0}}.$$

Let $A_t = \sqrt{2-\lambda}(Y_{c,t} - n^*\theta_0)/\sqrt{\lambda n^*\theta_0}$. When the process is in control, $E(A_t) = 0$, $\text{var}(A_t) \approx 1$. When the process is out of control, $E(A_t) = \sqrt{(2-\lambda)n^*}(\theta_1 - \theta_0)/\sqrt{\lambda\theta_0}$, $\text{var}(A_t) \approx \frac{\theta_1}{\theta_0}$.

We note that $X_j, (j = 1, \ldots, n)$ are mutually independent random variables and there exist positive constants, $H, a_1, a_2, \ldots$ such that the moment-generating functions, $h_{t,j}(\xi) = E(e^{\xi X_j}) = e^{n_j\theta_0(e^{\xi}-1)}, (1 \le j \le t)$ are analytic and $|\log h_{t,j}(\xi)| = n_j\theta_0(e^H - 1) \triangleq a_j$ for $|\xi| < H$, and that

$$\limsup\left\{\frac{1}{t}\sum_{j=1}^{t} a_j^{3/2}\right\} = [n^*\theta_0(e^H - 1)]^{3/2} < +\infty,$$

$$\liminf\left\{\frac{1}{t}\sum_{j=1}^{t} \mathrm{var}(X_j)\right\} = n^*\theta_0 > 0.$$

So far, we have validated the conditions (I), (II), and (V) in Han and Tsung (2006). Following their arguments, it is straightforward to show that the relation (i) and (ii) hold. □

[Received March 2011. Revised November 2011.]

REFERENCES

Borror, C. M., Champ, C. W., and Rigdon, S. E. (1998), "Poisson EWMA Control Charts," *Journal of Quality Technology*, 30, 352–361. [1052,1061]

Chen, R. (1987), "The Relative Efficiency of the Sets and the Cusum Techniques in Monitoring the Occurrence of a Rare Event," *Statistics in Medicine*, 6, 517–525. [1050]

Dong, Y., Hedayat, A. S., and Sinha, B. K. (2008), "Surveillance Strategies for Detecting Changepoint in Incidence Rate Based on Exponentially Weighted Moving Average Methods," *Journal of the American Statistical Association*, 103, 843–853. [1050,1053,1057,1059]

Duncan, A. J. (1986), *Quality Control and Industrial Statistics* (4th ed.), Homewood, IL: Irwin. [1049]

Frisén, M., and De Maré, J. (1991), "Optimal Surveillance," *Biometrika*, 78, 271–280. [1049]

Frisén, M., and Sonesson, C. (2006), "Optimal Surveillance Based on Exponentially Weighted Moving Averages," *Sequential Analysis*, 25, 379–403. [1049]

Gan, F. F. (1990), "Monitoring Poisson Observations Using Modified Exponentially Weighted Moving Average Control Charts," *Communications in Statistics—Simulation and Computation*, 19, 103–124. [1049]

Han, D., and Tsung, F. (2006), "A Reference-Free Cuscore Chart for Dynamic Mean Change Detection and a Unified Framework for Charting Performance Comparison," *Journal of the American Statistical Association*, 101, 368–386. [1053,1055,1062]

Hawkins, D. M., and Olwell, D. H. (1998), *Cumulative Sum Charts and Charting for Quality Improvement*, New York: Springer-Verlag. [1051,1053,1055]

Jensen, W. A., Jones-Farmer, L. A., Champ, C. W., and Woodall, W. H. (2006), "Effects of Parameter Estimation on Control Chart Properties: A Literature Review," *Journal of Quality Technology*, 38, 349–364. [1051]

Jones, L. A., Champ, C. W., and Rigdon, S. E. (2001), "The Performance of Exponentially Weighted Moving Average Charts With Estimated Parameters," *Technometrics*, 43, 156–167. [1053]

Lai, T. L. (1995), "Sequential Changepoint Detection in Quality Control and Dynamical System," *Journal of the Royal Statistical Society,* Series B, 57, 613–658. [1049]

Lucas, J. M. (1985), "Counted Data CUSUMs," *Technometrics*, 27, 129–144. [1049]

Lucas, J. M., and Saccucci, M. S. (1990), "Exponentially Weighted Moving Average Control Schemes Properties and Enhancements," *Technometrics*, 32, 1–29. [1052]

Mei, Y. (2008), "Is Average Run Length to False Alarm Always an Informative Criterion?" (With Discussions), *Sequential Analysis*, 27, 354–419. [1053]

Mei, Y., Han, S. W., and Tsui, K.-L. (2011), "Early Detecting of a Change in Poisson Rate After Accounting for Population Size Effects," *Statistica Sinica*, 21, 597–624. [1050,1051,1052,1053,1057,1059]

Montgomery, D. C. (2009), *Introduction to Statistical Quality Control* (7th ed.), New York: John Wiley & Sons. [1049]

Qiu, P., Zou, C., and Wang, Z. (2010), "Nonparametric Profile Monitoring by Mixed Effects Modeling" (With Discussions), *Technometrics*, 52, 265–277. [1051]

Quesenberry, C. P. (1995), "On Properties of *Q* Charts for Variables," *Journal of Quality Technology*, 27, 184–203. [1061]

Rossi, G., Lampugnani, L., and Marchi, M. (1999), "An Approximate CUSUM Procedure for Surveillance of Health Events," *Statistics in Medicine*, 18, 2111–2122. [1049]

Ryan, A. G., and Woodall, W. H. (2010), "Control Charts for Poisson Count Data With Varing Sample Sizes," *Journal of Quality Technology*, 42, 260–274. [1050,1051,1053]

Shu, L., Jiang, W., and Tsui, K.-L. (2011), "A Comparison of Weighted CUSUM Procedures That Account for Monotone Changes in Population Size," *Statistics in Medicine*, 30, 725–741. [1050,1059]

Shu, L., Jiang, W., and Wu, Z. (2012) "Exponentially Weighted Moving Average Control Charts for Monitoring Increases in Poisson Rate," *IIE Transactions*, 44, 711–723. [1052]

Sonesson, C., and Bock, D. (2003), "A Review and Discussion of Prospective Statistical Surveillance in Public Health," *Journal of the Royal Statistical Society,* Series A, 166, 5–21. [1049]

White, C. H., and Keats, J. B. (1996), "ARLs and Higher-order Run-length Moments for the Poisson CUSUM," *Journal of Quality Technology*, 28, 363–369. [1049]

Woodall, W. H. (2006), "The Use of Control Charts in Health-Care and Public-Health Surveillance," *Journal of Quality Technology*, 38, 89–104. [1049]