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

Qin Zhou\textsuperscript{1,3}, Changliang Zou\textsuperscript{1}, Zhaojun Wang\textsuperscript{1} and Wei Jiang\textsuperscript{2}

\textsuperscript{1}LPMC and Department of Statistics, School of Mathematical Sciences, Nankai University, Tianjin, China
\textsuperscript{2}Antai College of Economics and Management, Shanghai Jiaotong University, Shanghai, China
\textsuperscript{3}Department of Mathematics, Xuzhou Normal University, Xuzhou, China

Abstract

Many applications involve monitoring incidence rates of the Poisson distribution when 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

\textsuperscript{*}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. This research was supported by the NNSF of China Grants 11001138, 11071128, 11131002, 11101306, 71172131 and the RFDP of China Grant 20110031110002. Zhou's work is partially supported by PAPD of Jiangsu Higher Education Institutions. Zou also thanks for the support of the National Center for Theoretical Sciences, Math Division. The first two authors contributed equally to this work. Wang is the corresponding author: zjwang@nankai.edu.cn
sizes. Under the framework of weighted likelihood ratio test, this paper 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.

**Keywords:** Average run length; Healthcare; Run length distribution; EWMA; Short-run processes; Poisson count data; 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 (c.f., 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 whereas controlling the rate of false alarms (c.f., 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 (i.i.d.) 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 (1986) and Montgomery (2009) for discussions. The cumulative sum (CUSUM) chart, which has received considerable attention for detecting small changes, can be derived based on likelihood ratio test
principle, see Lucas (1985), Lai (1995), and White and Keats (1996). Gan (1990) considers 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) propose a likelihood ratio method and show 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 et al. (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 dynamic changes in the mean number of events, Mei et al. (2011) further develop some variations of the CUSUM method and propose three CUSUM-based charts. Shu et al. (2011) compare weighted CUSUM and conventional CUSUM procedures in the presence of monotone changes in population size. In addition to the CUSUM techniques, Dong et al. (2008) consider the EWMA methods to address the issue of non-constant population size. Ryan and Woodall (2010) compare 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 et al. (2008) and Ryan and Woodall (2010) is to divide the observed counts by the corresponding sample sizes to account for the variability of the sample sizes.

In this paper, 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 et al. (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 et al. (2011) and the two modifications of the EWMA chart in Dong et al. (2008), although 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 paper is organized as follows. We first describe the mathematical formulation of the problem and existing works 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 paper 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 materials.

2 The Statistical Model and Existing Works

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 $\theta$ denote the size of the population at time $t$ and the incidence rate of a rare event, respectively. Although other distribution 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 $\theta$ changes from $\theta_0$ to another unknown value $\theta_1$ at some unknown time $\tau$, that is, the observations collected come from the following change-point model

\[
X_t \, \text{indep} \left\{ \begin{array}{ll}
\text{Poisson}(n_t \theta_0), & \text{for } t = 1, \ldots, \tau, \\
\text{Poisson}(n_t \theta_1), & \text{for } t = \tau + 1, \ldots,
\end{array} \right.
\]

where "\text{indep}" 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 $a(n_t, X_t)$ and an alarm limit
\( g(t) \), where \( n_t = \{n_i; i \leq t\} \) and \( X_t = \{X_i; i \leq t\} \). The time of an alarm, \( T \), is defined as

\[
T = \min\{t; a(n_t, X_t) \geq g(t)\}.
\]

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 of incidence rate, i.e., \( \theta_1 > \theta_0 \), but the lower-sided and two-sided charts can be constructed without difficulty.

The EWMA-type control chart statistic proposed by Dong et al. (2008) is

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

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 et al. (2008) develop three different stopping rules, EWMAe, EWMAa1 and EWMAa2 control charts, as follows,

\[
T_{\text{EWMAe}} = \min\{t; Z_t \geq \theta_0 + L\sigma_t, t \geq 1\}, \quad \sigma_t^2 = \frac{\lambda}{n_t} \sum_{i=1}^{t} (1 - \lambda)^{2t-2i}\theta_0; \]

\[
T_{\text{EWMAa1}} = \min\{t; Z_t \geq \theta_0 + L\sigma^*_t, t \geq 1\}, \quad \sigma_t^2 = \frac{\theta_0}{n_0} \frac{\lambda}{n_0} \frac{2 - \lambda}{1 - (1 - \lambda)^{2t}}; \]

\[
T_{\text{EWMAa2}} = \min\{t; Z_t \geq \theta_0 + L\sigma^*_t, t \geq 1\}, \quad \sigma_t^2 = \frac{\theta_0}{n_0} \frac{\lambda}{n_0} \frac{2 - \lambda}{2};
\]

where the control limit coefficient \( L \) 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 confusions, we use the generic notation \( L \) to represent the control limit coefficient for different control chart. 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 \( Z_t = \theta_0 \),

\[
T_{\text{EWMAAM}} = \min\{t; Z_t \geq L\sigma_t, t \geq 1\},
\]

where

\[
\tilde{Z}_t = \max\{\theta_0, (1 - \lambda)\tilde{Z}_{t-1} + \frac{\lambda X_t}{n_t}\}, \quad \tilde{Z}_0 = \theta_0.
\]
We will refer it as EWMA-modified (EWMAM) method henceforth. Ryan and Woodall (2010) argue that the EWMAM performs better than EWMAe, but in their comparison the weighting parameters of the two EWMA charts are chosen differently. As shown in Section 4, given the same values of $\lambda$, 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 et al. (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 et al. (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

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

We notice that the CUSUM, WLR and ATM methods are all equivalent when the sample size is constant. It is also worth to point out that design of the above three charts requires the specification of not only the pre-change rate $\theta_0$ but also the post-change rate $\theta_1$. Of course, when $\theta_1$ 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, i.e., 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 those 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, e.g., \((X_t - n_t\theta_0)/\sqrt{n_t\theta_0}\). 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, 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 a LRT for a simple hypothesis. However, for EWMA-type charts, it seems difficult to have connection with a 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 \(\theta_0\) 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 \(\theta\), 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 et al. (2010) for profile monitoring with arbitrary design points. In that paper, the authors propose an exponentially weighted least-squared function to on-line update the regression function and construct monitoring statistics.

Given the value of \(\lambda\), the maximum weighted likelihood estimate (MWLE) of \(\theta\) 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×logarithm of weighted LRT (WLRT) statistic

\[
R_{t,\lambda} = 2 \left[ Y_t(\hat{\theta}_t; \lambda) - Y_t(\theta_0; \lambda) \right]
\]

\[
= 2 \sum_{j=0}^{t} \omega_{j,\lambda} \left[ l_j(\hat{\theta}_t) - l_j(\theta_0) \right]
\]

\[
= 2 \left[ Y_{c,t} \log \frac{Y_{c,t}}{\theta_0 Y_{p,t}} - Y_{c,t} + Y_{p,t} \theta_0 \right],
\]

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 re-expressed as the following equivalent formulations

\[
Y_{c,t} = \lambda X_j + (1 - \lambda) Y_{c,t-1},
\]

\[
Y_{p,t} = \lambda n_j + (1 - \lambda) Y_{p,t-1},
\]

where the initial values are \(Y_{c,0} = \theta_0 n_1\) and \(Y_{p,0} = n_1\), respectively, based on the pseudo sample \((X_0, n_0)\) defined before. Hence, the WEWMA control chart can still be conducted in
a recursive fashion as the traditional EWMA charts do. Under some conditions imposed on \( n_t \) (c.f., Mei et al. 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 \( \lambda t \to \infty \),

\[
\frac{\sum_{i=1}^{t} w_i n_i}{\sum_{i=1}^{t} w_i^2 n_i} R_{t,\lambda} \xrightarrow{d} \chi^2_1.
\]

When \( \lambda \) 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 in 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 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 reduces to the Poisson EWMA chart (equivalently speaking) investigated by Borror et al. (1998). A straightforward proof can be found in the 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 above, 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 above is applicable in most SPC monitoring problems. It can be used as a standard tool to derive the EWMA chart under certain complex circumstance 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 et al. 2011)

\[
\hat{\theta}_t = \hat{\theta}_t I(\hat{\theta}_t > \theta_0) + \theta_0 I(\hat{\theta}_t \leq \theta_0)
\]
since the function $Y_t(\theta; \lambda)$ is monotonically decreasing on the right side of $\theta_0$ when $\hat{\theta}_t \leq \theta_0$. Accordingly, by substituting $\hat{\theta}_t$ into the WLRT, the WEWMA monitoring statistic becomes

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

Finally, our proposed one-sided WEWMA control chart is

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

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 incident rate without any difficulty.

In general, for EWMA-type control charts, a small value of $\lambda$ leads to optimal detection of small shifts (c.f., e.g., 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 to point out that all the control charts for discrete data share a common shortcoming, i.e., there may not be an exact control limit to achieve certain values of IC ARL. As mentioned by some authors (c.f., Borror et al. 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 $\lambda$ is chosen. In our experience, the WEWMA chart’s IC ARL can always be attained quite exactly if $\lambda \leq 0.2$.

Before ending this section, we present WEWMA’s asymptotic bounds of ARL. As pointed out in Mei et al. (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 et al. (2011). Denote

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

and $g^{-1}(\cdot)$ as its inverse function. Denote by $\text{ARL}(T(h))$ 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 $T_{\text{WEWMA}}$ we have

(i). When the process is in control, then

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

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

(ii). When the process is out of control, we have

$$\frac{1}{E_1}\left\{\eta^2 - 4d_1 \eta \sqrt{2\log(\eta)}\right\} \leq \text{ARL}(T_{\text{WEWMA}}(h)) \leq \frac{1}{E_2}\left\{\eta^2 + 4d_1 \eta \sqrt{\log(\eta)}\right\},$$

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

This theorem can be considered as an application of Theorem 2 in Han and Tsung (2006) which presents a unified framework for the asymptotic analysis of any stopping time satisfying certain conditions. Thus, the proof of this theorem amounts to rewriting $T_{\text{WEWMA}}$ 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 is available from the authors upon request.

In our simulation study below, we investigate the performance of different control charts under various scenarios of time-varying sample sizes. For health surveillance, Mei et al. (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 : $n_t = \frac{c_1}{1 + \exp[-(t-c_2)/c_3]}$;
(II) Fast Increasing Scenario: \( n_t = \frac{2c_1}{1+\exp[-(t-(c_2+26))/c_3]}; \)

(III) 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 et al. (2011), Scenario (I) is the estimated curve from a real-data set discussed in the next section, Scenario (II) corresponds to the case that the population size increases quickly, and Scenario (III) is the case that the population size decreases rapidly to the stationary value. Dong et al. (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:

(IV) Constant Scenario: \( n_t = 10 \) for all \( t; \)

(V) Uniform Scenario: \( n_t \sim U(10, 15). \)

Note that the case (V) 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,

(VI) 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 point \( t \) (cf., Jones et al. 2001, Mei 2008). As an alternative, the control chart performance will be summarized using ARL, percentiles of the marginal distribution of the run length, standard deviation of the run length (SDRL). The control limits are set so that \( ARL_0 \approx 300, \) which is consistent with Mei et al. (2011). We also study the false-alarm rate for the first 30 observations, \( Pr_{IC}(T \leq 30) \) for each chart. We use notation CL, 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 \( \text{ARL}_0 \) 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 (I)-(IV), i.e., the sample sizes are increasing, fast increasing, decreasing and fixed, in Tables 1-4, respectively. The results for Scenario (V) and (VI) are similar to (IV) 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 \( \theta_1 \) is chosen as 2 in all the CUSUM-type charts, consistent with Mei et al. (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 a quite reasonable approximation to the IC run length distributions of the EWMAe, CUSUM, EWMAM and WEWMA charts. This 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 a similar evidence to that of Tables 1-2, i.e., 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 pre-specified 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 \(\lambda\) and IC ARL to check whether the above observations still hold in other settings. The simulation results show that these charts have quite satisfactory performance in other cases as well.

### Table 2: IC ARL comparison under Scenario (II)

<table>
<thead>
<tr>
<th></th>
<th>(L)</th>
<th>ARL(_0)</th>
<th>SE</th>
<th>SDRL</th>
<th>(Q(.10))</th>
<th>Median</th>
<th>(Q(.90))</th>
<th>FAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>EWMAe</td>
<td>2.371</td>
<td>299</td>
<td>2.16</td>
<td>306</td>
<td>23</td>
<td>205</td>
<td>694</td>
<td>0.1125</td>
</tr>
<tr>
<td>EWMAa1</td>
<td>1.212</td>
<td>299</td>
<td>3.81</td>
<td>540</td>
<td>1</td>
<td>32</td>
<td>978</td>
<td>0.4987</td>
</tr>
<tr>
<td>EWMAa2</td>
<td>1.183</td>
<td>298</td>
<td>3.47</td>
<td>491</td>
<td>5</td>
<td>51</td>
<td>932</td>
<td>0.4161</td>
</tr>
<tr>
<td>CUSUM</td>
<td>2.802</td>
<td>300</td>
<td>2.71</td>
<td>383</td>
<td>11</td>
<td>148</td>
<td>812</td>
<td>0.2552</td>
</tr>
<tr>
<td>WLR</td>
<td>0.155</td>
<td>332</td>
<td>6.11</td>
<td>863</td>
<td>2</td>
<td>12</td>
<td>1228</td>
<td>0.7175</td>
</tr>
<tr>
<td>ATM</td>
<td>0.155</td>
<td>332</td>
<td>6.11</td>
<td>864</td>
<td>2</td>
<td>12</td>
<td>1225</td>
<td>0.7212</td>
</tr>
<tr>
<td>EWMAM</td>
<td>2.609</td>
<td>300</td>
<td>2.13</td>
<td>302</td>
<td>27</td>
<td>207</td>
<td>698</td>
<td>0.0942</td>
</tr>
<tr>
<td>WEWMA</td>
<td>2.757</td>
<td>300</td>
<td>2.23</td>
<td>316</td>
<td>29</td>
<td>199</td>
<td>712</td>
<td>0.1025</td>
</tr>
</tbody>
</table>

### Table 3: IC ARL comparison under Scenario (III)

<table>
<thead>
<tr>
<th></th>
<th>(L)</th>
<th>ARL(_0)</th>
<th>SE</th>
<th>SDRL</th>
<th>(Q(.10))</th>
<th>Median</th>
<th>(Q(.90))</th>
<th>FAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>EWMAe</td>
<td>2.550</td>
<td>299</td>
<td>2.13</td>
<td>302</td>
<td>38</td>
<td>204</td>
<td>697</td>
<td>0.0797</td>
</tr>
<tr>
<td>EWMAa1</td>
<td>2.342</td>
<td>300</td>
<td>1.52</td>
<td>214</td>
<td>105</td>
<td>238</td>
<td>580</td>
<td>0.0002</td>
</tr>
<tr>
<td>EWMAa2</td>
<td>2.341</td>
<td>299</td>
<td>1.49</td>
<td>211</td>
<td>106</td>
<td>238</td>
<td>574</td>
<td>0.0003</td>
</tr>
<tr>
<td>CUSUM</td>
<td>3.705</td>
<td>302</td>
<td>2.33</td>
<td>329</td>
<td>24</td>
<td>191</td>
<td>726</td>
<td>0.1210</td>
</tr>
<tr>
<td>WLR</td>
<td>3.257</td>
<td>302</td>
<td>1.47</td>
<td>208</td>
<td>115</td>
<td>240</td>
<td>570</td>
<td>0.0000</td>
</tr>
<tr>
<td>ATM</td>
<td>3.260</td>
<td>299</td>
<td>1.47</td>
<td>208</td>
<td>113</td>
<td>237</td>
<td>571</td>
<td>0.0000</td>
</tr>
<tr>
<td>EWMAM</td>
<td>2.809</td>
<td>300</td>
<td>2.07</td>
<td>293</td>
<td>36</td>
<td>212</td>
<td>677</td>
<td>0.0761</td>
</tr>
<tr>
<td>WEWMA</td>
<td>2.660</td>
<td>300</td>
<td>1.95</td>
<td>275</td>
<td>40</td>
<td>220</td>
<td>659</td>
<td>0.0812</td>
</tr>
</tbody>
</table>
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, i.e., Scenario (IV). It is easy to see that the WEWMA method performs quite stable under all cases of population model. 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 (II). The ARL$_0$ values of the EWMAe chart is also far away from 300 under Scenarios (III). 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.

Table 5: The sensitivity comparison of the control limit.

<table>
<thead>
<tr>
<th></th>
<th>Scenario (I)</th>
<th>Scenario (II)</th>
<th>Scenario (III)</th>
<th>Scenario (V)</th>
<th>Scenario (VI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EWMAe</td>
<td>2.401</td>
<td>306(314)</td>
<td>320 (332)</td>
<td>228(231)</td>
<td>296 (301)</td>
</tr>
<tr>
<td>CUSUM</td>
<td>3.863</td>
<td>372(289)</td>
<td>999(1129)</td>
<td>355(386)</td>
<td>375 (371)</td>
</tr>
<tr>
<td>EWMAM</td>
<td>2.640</td>
<td>312(316)</td>
<td>324(330)</td>
<td>217(213)</td>
<td>298 (302)</td>
</tr>
<tr>
<td>WEWMA</td>
<td>2.688</td>
<td>293(300)</td>
<td>283 (293)</td>
<td>307(287)</td>
<td>300 (297)</td>
</tr>
</tbody>
</table>

NOTE: Standard errors are in parentheses.
4.2 OC performance comparison

In this subsection, 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 $\text{ARL}_0 = 300$ for the illustration purpose. Results with other commonly used $\text{ARL}_0$ values, such as 500 or 800, are provided in the supplemental file. As shown in the last subsection, the EWMAa1, EWMAa2, WLR and ATM charts have unacceptable IC run length distribution. 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 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 words, the 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 of the aforementioned control charts, we consider $\gamma_t \equiv \Pr_{\text{OC}}(T \leq t) - \Pr_{\text{IC}}(T \leq t)$, i.e., the “pure” probability that a stopping time $T$ detects an OC condition before time point $t$ beyond randomness. We compare the above control charts using the values of $\gamma_t$ for $t \leq 100$ 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 $\gamma_t$ 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 (I)-(III) are shown in Figures 1-3, respectively. In each figure, the first plot depicts the cumulative distribution function (CDF) of IC run-length distributions, i.e., $\Pr_{\text{IC}}(T \leq t)$ and the other three plots show the $\gamma_t$ 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 $\lambda$ for all the EWMA-type chart is fixed as 0.1 and the tuning parameter $\theta_1$ 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 \( \gamma_t \) curve increases much faster after a change occurs. Note that under Scenarios (III), the \( \gamma_t \) curve of the WEWMA tends to be lower than those of the other three modifications when \( t \) becomes large. This is due to that when \( \theta \) is large (e.g., \( \theta = 1.5 \)), the values of \( \Pr_{OC}(T \leq t) \) increase to one very quickly and accordingly a chart with larger false alarm rate will have a smaller value of \( \gamma_t \) 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 ARL\(_0\) = 800 in the supplemental file).

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

any series in which a signal occurs before the $(\tau+1)$-th observation is discarded (c.f., 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

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

where $N$ is the total number of shifts considered, $\text{ARL}_{\delta_l}$ is the OC ARL of the given control chart when detecting a parameter shift of magnitude $\delta_l$, and $\text{MARL}_{\delta_l}$ is the smallest among all OC ARL values of the charts considered when detecting the shift $\delta_l$. So $(\text{ARL}_{\delta_l} - \text{MARL}_{\delta_l})/\text{MARL}_{\delta_l}$ could be considered as a relative efficiency measure of the given control chart, compared to the best chart, when detecting the shift $\delta_l$, 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. Here the RMI values are evaluated at
Figure 3: Performance comparison between the WEWMA, EWMAa1, EWMAa2 and WLR under Scenario (III): (a) four in-control CDF curves along with Geometric distribution (with expectation 300); (b)-(d) Curves of $\gamma_t \equiv \Pr_{OC}(T \leq t) - \Pr_{IC}(T \leq t)$ when $\theta = 1.05, 1.2$ and 1.5, respectively.

For a relatively fair comparison, we choose appropriate values of $\theta_1$ for the CUSUM chart under different scenarios. To be more specific, we set $\theta_1$ equal to the shift level at which the WEWMA with $\lambda = 0.1$ is approximately the best of detection across all the values of $\lambda$. For example, in Scenario (I), we found by simulations that $\theta = 1.4$ is the shift which the WEWMA with $\lambda = 0.1$ is roughly the optimal in the sense that with other values of $\lambda$ the WEWMA cannot be (or significantly) better than that with 0.1. By doing this, $\theta_1$ is chosen as 1.4, 1.3, 1.4, 1.2, 1.2, and 1.3 for Scenarios (I)-(VI), 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 better performance compared with all other control charts for shifts up to $\theta = 1.4$. The EWMAM chart outperforms the WEWMA chart when $\theta$ 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.

Table 6: OC ARL comparison under Scenarios (I) and (II)

<table>
<thead>
<tr>
<th>$\theta$</th>
<th>Scenario (I)</th>
<th>Scenario (II)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>EWMAe</td>
<td>CUSUM</td>
</tr>
<tr>
<td>$\theta_1 = 1.4$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.025</td>
<td>144 (137)</td>
<td>178 (175)</td>
</tr>
<tr>
<td>1.050</td>
<td>81.8 (72.0)</td>
<td>112 (107)</td>
</tr>
<tr>
<td>1.100</td>
<td>37.2 (28.3)</td>
<td>53.6 (47.4)</td>
</tr>
<tr>
<td>1.200</td>
<td>15.7 (9.85)</td>
<td>18.8 (14.5)</td>
</tr>
<tr>
<td>1.300</td>
<td>9.65 (5.37)</td>
<td>10.1 (6.86)</td>
</tr>
<tr>
<td>1.400</td>
<td>7.01 (3.56)</td>
<td>6.70 (3.98)</td>
</tr>
<tr>
<td>1.500</td>
<td>5.48 (2.62)</td>
<td>4.96 (2.63)</td>
</tr>
<tr>
<td>1.700</td>
<td>3.87 (1.71)</td>
<td>3.32 (1.52)</td>
</tr>
<tr>
<td>2.000</td>
<td>2.77 (1.13)</td>
<td>2.29 (0.93)</td>
</tr>
<tr>
<td>3.000</td>
<td>1.56 (0.57)</td>
<td>1.23 (0.43)</td>
</tr>
<tr>
<td>4.000</td>
<td>1.18 (0.38)</td>
<td>1.02 (0.12)</td>
</tr>
<tr>
<td>RMI</td>
<td>0.119</td>
<td>0.158</td>
</tr>
</tbody>
</table>

NOTE: Standard deviations are in parentheses.

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 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 (VI) 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 a quite 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 et al. (2010), is
## Table 7: OC ARL comparison under Scenarios (III) and (IV)

<table>
<thead>
<tr>
<th>$\theta$</th>
<th>EWMAe</th>
<th>CUSUM</th>
<th>EWMAM</th>
<th>WEWMA</th>
<th>EWMAe</th>
<th>CUSUM</th>
<th>EWMAM</th>
<th>WEWMA</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_1 = 1.4$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.025</td>
<td>238 (238)</td>
<td>235 (267)</td>
<td>240 (238)</td>
<td>230 (220)</td>
<td>155 (153)</td>
<td>160 (158)</td>
<td>167 (166)</td>
<td>152 (150)</td>
</tr>
<tr>
<td>1.050</td>
<td>185 (185)</td>
<td>179 (206)</td>
<td>197 (198)</td>
<td>182 (179)</td>
<td>88.0 (82.5)</td>
<td>93.4 (89.1)</td>
<td>97.9 (94.9)</td>
<td>87.3 (81.9)</td>
</tr>
<tr>
<td>1.100</td>
<td>114 (120)</td>
<td>102 (125)</td>
<td>129 (136)</td>
<td>114 (116)</td>
<td>38.3 (32.6)</td>
<td>40.6 (35.9)</td>
<td>43.5 (39.3)</td>
<td>38.0 (32.6)</td>
</tr>
<tr>
<td>1.200</td>
<td>47.4 (51.9)</td>
<td>39.3 (48.6)</td>
<td>55.6 (62.9)</td>
<td>48.3 (51.6)</td>
<td>14.8 (9.97)</td>
<td>14.8 (10.3)</td>
<td>15.3 (11.3)</td>
<td>14.8 (10.1)</td>
</tr>
<tr>
<td>1.300</td>
<td>24.2 (24.2)</td>
<td>19.3 (20.5)</td>
<td>27.3 (30.8)</td>
<td>24.3 (23.8)</td>
<td>8.80 (5.05)</td>
<td>8.66 (4.92)</td>
<td>8.52 (5.28)</td>
<td>8.76 (5.01)</td>
</tr>
<tr>
<td>1.400</td>
<td>14.9 (12.6)</td>
<td>11.9 (10.9)</td>
<td>15.6 (14.8)</td>
<td>14.8 (11.9)</td>
<td>6.26 (3.24)</td>
<td>6.05 (3.04)</td>
<td>5.87 (3.15)</td>
<td>6.22 (3.23)</td>
</tr>
<tr>
<td>1.500</td>
<td>10.6 (7.44)</td>
<td>8.38 (6.17)</td>
<td>10.7 (8.38)</td>
<td>10.7 (7.43)</td>
<td>4.88 (2.35)</td>
<td>4.73 (2.18)</td>
<td>4.48 (2.22)</td>
<td>4.84 (2.34)</td>
</tr>
<tr>
<td>1.700</td>
<td>6.77 (3.91)</td>
<td>5.43 (3.36)</td>
<td>6.46 (4.00)</td>
<td>6.94 (3.99)</td>
<td>3.44 (1.51)</td>
<td>3.32 (1.35)</td>
<td>3.10 (1.35)</td>
<td>3.42 (1.49)</td>
</tr>
<tr>
<td>2.000</td>
<td>4.46 (2.22)</td>
<td>3.57 (1.87)</td>
<td>4.17 (2.18)</td>
<td>4.55 (2.29)</td>
<td>2.46 (0.98)</td>
<td>2.36 (0.85)</td>
<td>2.21 (0.85)</td>
<td>2.45 (0.98)</td>
</tr>
<tr>
<td>3.000</td>
<td>2.25 (0.93)</td>
<td>1.82 (0.74)</td>
<td>2.06 (0.83)</td>
<td>2.32 (0.95)</td>
<td>1.41 (0.52)</td>
<td>1.33 (0.47)</td>
<td>1.25 (0.44)</td>
<td>1.41 (0.52)</td>
</tr>
<tr>
<td>4.000</td>
<td>1.63 (0.62)</td>
<td>1.31 (0.49)</td>
<td>1.46 (0.55)</td>
<td>1.67 (0.65)</td>
<td>1.08 (0.28)</td>
<td>1.03 (0.16)</td>
<td>1.02 (0.13)</td>
<td>1.08 (0.27)</td>
</tr>
<tr>
<td>RMI</td>
<td>0.194</td>
<td>0.002</td>
<td>0.219</td>
<td>0.204</td>
<td>0.058</td>
<td>0.046</td>
<td>0.036</td>
<td>0.052</td>
</tr>
</tbody>
</table>

NOTE: Standard deviations are in parentheses.

Developed under the framework of LRT and change-point detection. It is efficient in certain sense due to 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 health-care 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, $\theta_0$, $\lambda$ and $\tau$, to check whether the above conclusions would change in other cases. Some representative simulation results are reported in the supplemental file 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.
Table 8: OC ARL comparison under Scenarios (V) and (IV)

<table>
<thead>
<tr>
<th>θ</th>
<th>EWMAe</th>
<th>CUSUM</th>
<th>EWMA</th>
<th>WEWMA</th>
<th>EWMAe</th>
<th>CUSUM</th>
<th>EWMA</th>
<th>WEWMA</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.025</td>
<td>153 (150)</td>
<td>159 (158)</td>
<td>168 (167)</td>
<td>151 (148)</td>
<td>192 (189)</td>
<td>179 (179)</td>
<td>199 (198)</td>
<td>164 (160)</td>
</tr>
<tr>
<td>1.050</td>
<td>87.4 (81.4)</td>
<td>90.7 (90.7)</td>
<td>99.7 (96.1)</td>
<td>86.9 (82.1)</td>
<td>124 (121)</td>
<td>115 (113)</td>
<td>135 (131)</td>
<td>100 (96.2)</td>
</tr>
<tr>
<td>1.100</td>
<td>38.7 (33.9)</td>
<td>40.5 (35.9)</td>
<td>43.3 (39.0)</td>
<td>37.9 (32.5)</td>
<td>60.9 (54.3)</td>
<td>54.8 (51.4)</td>
<td>67.1 (63.5)</td>
<td>46.4 (40.4)</td>
</tr>
<tr>
<td>1.200</td>
<td>15.0 (10.2)</td>
<td>14.9 (10.5)</td>
<td>15.3 (11.4)</td>
<td>14.7 (9.98)</td>
<td>23.7 (18.1)</td>
<td>19.7 (15.8)</td>
<td>25.4 (20.6)</td>
<td>18.4 (13.1)</td>
</tr>
<tr>
<td>1.300</td>
<td>8.81 (5.10)</td>
<td>8.67 (4.90)</td>
<td>8.56 (5.28)</td>
<td>8.80 (5.04)</td>
<td>14.0 (8.87)</td>
<td>10.9 (7.27)</td>
<td>14.0 (9.34)</td>
<td>11.0 (6.56)</td>
</tr>
<tr>
<td>1.400</td>
<td>6.28 (3.26)</td>
<td>6.08 (3.03)</td>
<td>5.90 (3.18)</td>
<td>6.25 (3.23)</td>
<td>9.78 (5.16)</td>
<td>7.44 (4.22)</td>
<td>9.73 (5.48)</td>
<td>7.75 (4.07)</td>
</tr>
<tr>
<td>1.500</td>
<td>4.88 (2.35)</td>
<td>4.74 (2.19)</td>
<td>4.47 (2.21)</td>
<td>4.86 (2.36)</td>
<td>7.80 (3.79)</td>
<td>5.68 (2.88)</td>
<td>7.55 (3.86)</td>
<td>6.11 (2.99)</td>
</tr>
<tr>
<td>1.700</td>
<td>3.45 (1.51)</td>
<td>3.31 (1.36)</td>
<td>3.12 (1.35)</td>
<td>3.44 (1.50)</td>
<td>5.53 (2.60)</td>
<td>3.92 (1.77)</td>
<td>5.33 (2.56)</td>
<td>4.33 (1.94)</td>
</tr>
<tr>
<td>2.000</td>
<td>2.46 (0.98)</td>
<td>2.37 (0.86)</td>
<td>2.21 (0.85)</td>
<td>2.44 (0.97)</td>
<td>3.92 (1.83)</td>
<td>2.76 (1.16)</td>
<td>3.70 (1.74)</td>
<td>3.09 (1.31)</td>
</tr>
<tr>
<td>3.000</td>
<td>1.41 (0.51)</td>
<td>1.32 (0.47)</td>
<td>1.25 (0.44)</td>
<td>1.41 (0.51)</td>
<td>2.01 (0.96)</td>
<td>1.28 (0.64)</td>
<td>1.79 (0.89)</td>
<td>1.62 (0.89)</td>
</tr>
<tr>
<td>4.000</td>
<td>1.08 (0.27)</td>
<td>1.03 (0.16)</td>
<td>1.02 (0.14)</td>
<td>1.08 (0.27)</td>
<td>1.33 (0.59)</td>
<td>1.01 (0.12)</td>
<td>1.12 (0.38)</td>
<td>1.10 (0.39)</td>
</tr>
<tr>
<td>RMI</td>
<td>0.060</td>
<td>0.047</td>
<td>0.040</td>
<td>0.052</td>
<td>0.336</td>
<td>0.045</td>
<td>0.320</td>
<td>0.064</td>
</tr>
</tbody>
</table>

NOTE: Standard deviations are in parentheses.

5 Analytical Bounds for ARL

Dong et al. (2008) gave the analytical bounds of ARL₀ and ARL₁ 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 et al. (2008). To give a broad picture of the two methods, we consider two commonly used values of λ = 0.1 and 0.2, and calculate ARL₁ with ARL₀ = 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 et al. (2008). The ARL bounds of the WEWMA chart are derived from Theorem 1. Following Dong et al. (2008), we study the situation in which θ₀ = 1, θ₁ = 2, and nₜ = 10 (in units of thousand) for all t ≥ 1, and τ = 20.

Table 9 presents ARL₀ and its lower bound, ARL₁ and its upper bound for different fixed values of ARL₀. 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 ARL₁ should be always smaller than the corresponding ARL₀. Hence, the analytic bounds for ARL₁ which are larger than the ARL₀s are useless from practical viewpoints. In Table 9, the entries with the symbol “−” represent that the values are larger than the corresponding ARL₀.
It is clearly seen that all the upper bounds of EWMAe are larger than the corresponding ARL\(_0\). This is consistent with the results shown in Dong et al. (2008). In fact, in some cases, the upper bounds of ARL\(_1\) provided by Dong et al. (2008) are not finite (see some detailed discussions on Theorem 2 in Dong et al. 2008). In comparison, the analytic bounds give by Theorem 1 are 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, those 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.

<table>
<thead>
<tr>
<th>(\lambda = 0.1)</th>
<th>(\lambda = 0.2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARL(_0) lower bound</td>
<td>ARL(_1) lower bound</td>
</tr>
<tr>
<td>EWMAe</td>
<td>WEWMA</td>
</tr>
<tr>
<td>100</td>
<td>30</td>
</tr>
<tr>
<td>300</td>
<td>122</td>
</tr>
<tr>
<td>500</td>
<td>240</td>
</tr>
<tr>
<td>800</td>
<td>448</td>
</tr>
<tr>
<td>1000</td>
<td>590</td>
</tr>
<tr>
<td>2000</td>
<td>1450</td>
</tr>
<tr>
<td>3000</td>
<td>2466</td>
</tr>
<tr>
<td>4000</td>
<td>3689</td>
</tr>
</tbody>
</table>

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 et al. (2011) and Shu et al. (2011).

Known risk factors for thyroid cancer include the exposure to ionizing radiation during childhood, radiation treatment, 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 Figure 4(b)-(c), respectively. It is clear that the incidence rate remains relatively stable before 1994 and exhibits an increasing tendency beginning in 1994. Readers may refer to Mei et al. (2011) and Shu et al. (2011) and the references therein for details.

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

Because the incidence rate is relatively stable during the period from 1973 to 1983, Mei et al. (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 et al. (2011), we also use the IC incidence rate of $\theta_0 = 2.4$. The estimated sample size function is just the function given in Scenario (I) in Section 4. We set $\lambda = 0.1$ for the WEWMA chart and the simulation leads to a control limit 2.713 to attain $\text{ARL}_0 = 300$.

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 et al. (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 et al. (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 post-change size and only consider a common value of $\lambda$. 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.

![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](image)

7 Concluding Remarks

In this paper, 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 similar computational effort 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 paper 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 rate. 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 on-line determined 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 \(-2\)logarithm of WLRT becomes

\[
R_{t, \lambda} = 2 \left[ Y_{c,t} \log\frac{Y_{c,t}}{n\theta_0} - \frac{1}{\sum_{j=0}^{t} \omega_{j, \lambda}} - Y_{c,t} + n\theta_0 \sum_{j=0}^{t} \omega_{j, \lambda} \right].
\]

When \( t \) is large \( \sum_{j=0}^{t} \omega_{j, \lambda} \approx 1 \). Thus, \( R_{t, \lambda} \) can be re-written 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 \( R_{t, \lambda} \) 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, using the test above at each time point leads to the EWMA control chart studied by Borror et al. (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 i.i.d. 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 \( n_i \in (n_{\min}, n_{\max}) \),

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

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,

\[
R_{t,\lambda} = 2Y_{p,t} \theta_0 [\hat{\alpha}_t \log(\hat{\alpha}_t) - (\hat{\alpha}_t - 1)]
\]

\[
= 2Y_{p,t} \theta_0 \{\hat{\alpha}_t[(\hat{\alpha}_t - 1) - 2^{-1}(\hat{\alpha}_t - 1)^2] - (\hat{\alpha}_t - 1)\} + o_p(Y_{p,t} \text{var}(\hat{\alpha}_t))
\]

\[
= Y_{p,t} \theta_0 (1 - \hat{\alpha}_t)^2 (2 - \hat{\alpha}_t) + o_p(Y_{p,t} \text{var}(\hat{\alpha}_t))
\]

\[
= \left( \frac{1 - \hat{\alpha}_t}{[\text{var}(\hat{\alpha}_t)]^{1/2}} \right)^2 (2 - \hat{\alpha}_t) Y_{p,t} \theta_0 \cdot \text{var}(\hat{\alpha}_t) + o_p(Y_{p,t} \text{var}(\hat{\alpha}_t)).
\]

By noting \( 2 - \hat{\alpha}_t \xrightarrow{p} 1 \) and using Slutsky’s Theorem, we have

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

Appendix C: The proof of Theorem 1

When \( t \) is larger than \( M \), the observations are i.i.d. 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

\[
\hat{R}_{t,\lambda} = R_{t,\lambda} I(\hat{\theta}_t > \theta_0) > h
\]

27
is essentially equivalent to the test
\[
\sqrt{\frac{2 - \lambda Y_{c,t} - n^*\theta_0}{\lambda}} > \sqrt{\frac{2 - \lambda g^{-1}(\frac{h}{2}) - n^*\theta_0}{\lambda}}.
\]
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, \( h_{i,j}(\xi) = E(e^{\xi X_j}) = e^{n_j \theta_0 (e^\xi - 1)}, (1 \leq j \leq t) \) are analytic and \(|\log h_{i,j}(\xi)| = n_j \theta_0 (eH - 1) \equiv a_j \) for \(|\xi| < H \), and that
\[
\limsup \left\{ \frac{1}{t} \sum_{j=1}^{t} a_{j}^{3/2} \right\} = \left[ n^* \theta_0 (eH - 1) \right]^{3/2} < +\infty,
\]
\[
\liminf \left\{ \frac{1}{t} \sum_{j=1}^{t} \text{var}(X_j) \right\} = n^* \theta_0 > 0.
\]
Let \( S_t(\xi) = t^{-1} \sum_{j=1}^{t} \log h_{i,j}(\xi) \), there exist two moment-generating functions, \( h_j(\xi) = E(e^{\xi Z_j}), i = 1, 2, \) and a positive number \( \xi^* = e^{-1} \) such that \( \log h_1(\xi) \leq S_t(\xi) \leq \log h_2(\xi) \) for \( t \geq 1, h_1(e^{-1}) = 1 \), where \( Z_1 \sim \text{Poisson}(n_{\min} \theta_0) \) and \( Z_2 \sim \text{Poisson}(n_{\max} \theta_0) \), and \( \lim S^*_t(e^{-1}) = n^* \theta_0 e < \infty \) and \( \lim S^*_t(e^{-1}) / t = 0 \). So far, we have validated the conditions (I)-(III) and (V) in Han and Tsung (2006). Following their arguments, it is straightforward to show that the relation (i) and (ii) hold. □

References:


