Monitoring Poisson Count Data with Probability Control Limits when Sample Sizes are Time-Varying

Xiaobei Shen[1], Fugee Tsung[1], Changliang Zou[2], Wei Jiang[3]

[1] *Department of Industrial Engineering and Logistics Management, Hong Kong University of Science and Technology, Hong Kong*

[2] *School of Mathematical Sciences, Nankai University, China*

[3] *Antai College of Economics and Management, Shanghai Jiaotong University, China*

Abstract

This article considers the problem of monitoring Poisson count data with time-varying sample sizes without assuming *a prior* knowledge of sample sizes in advance. Traditional control charts, whose control limits are often determined before the control charts start working, are constructed based on perfect knowledge of sample sizes. In practice, however, there is often little foreknowledge about future sample sizes. An inappropriate assumption of the distribution function may lead to unexpected performance of control charts, e.g., excessive false alarms at early runs of the control charts which in turn hurt an operator's confidence in valid alarms. To overcome this problem, we propose the use of probability control limits, which are determined based on the realization of sample sizes on-line. The conditional probability that the charting statistic exceeds the control limit at present given that there is no alarm before the current time point can be guaranteed to attain a specified false alarm rate. Our simulation studies show that the proposed control chart is able to deliver satisfactory run-length performance for any time-varying sample sizes. The idea presented in this paper can be applied to any effective control charts such as the CUSUM chart.

Keywords: Average run length; EWMA; False alarm rate; Health care; Run length distribution; Statistical process control

1 Introduction

Statistical process control (SPC) charts have been widely used in various applications ranging from industrial quality control, service operations management, to healthcare surveillance (Sonesson and Bock, 2003; Woodall, 2006). In particular, monitoring occurrence of a rare event from a sequence of stochastic processes has received considerable attention recently, e.g., the detection of non-conformities in precise machining and manufacturing, the detection of increase in the rate of people visiting an emergency room, mortality of heart surgery (Poloniecki *et al.*, 1998), and number of cancer patients (Nancy, 2008). In general, the detection aims to issue an out-of-control (OC) signal as early as possible once an adverse event occurs after SPC charts start working. Meanwhile, the false alarm rate needs to be controlled in a desired level so that practitioners would not be bothered by excessive false alarms to investigate their root causes.

To detect changes in the occurrence rate of an adverse event, both the count of events recorded in regular time intervals and the corresponding sample size need to be collected. For example, in manufacturing quality control, a sample of products with size n_t is inspected and the number of non-conformities in the sampled products is monitored for the interest of detecting possible increases of incidence rate of the non-conformities. Usually, it assumes that the count of events or non-conformities follows an (conditionally) independent Poisson distribution given the corresponding sample size. When the sample size is a constant, detecting a change in the rate may be characterized as detecting a change in Poisson mean. Several control charts have been proposed including the Shewhart chart (Montgomery, 1990), the cumulative sum (CUSUM) chart (Lucas, 1985; White and Keats, 1996), and the exponentially weighted moving average (EWMA) chart (Gan, 1990; Frisén and De Maré, 1991; Huwang *et al.* 2009). These control charts have been successfully employed in manufacturing quality control in practice.

In some applications such as healthcare surveillance, sample size refers to population at risk, which however often changes over time. Increasing attentions have thus been paid to the problem of monitoring the occurrence rate of an adverse event with time-varying sample sizes in Phase II analysis. Under the assumption that the sample size can be characterized by a (deterministic) logistic function, Mei *et al.* (2011) proposed three CUSUM-based control charts taking into account the time-varying sample sizes. Shu *et al.* (2011) compared a weighted CUSUM and conventional CUSUM procedures. Dong *et al.* (2008) proposed to monitor the EWMA statistic of incidence rate estimation. Ryan and Woodall (2010) compared CUSUM methods and the EWMA chart by Dong *et al.* (2008) assuming that the sample size follows a Uniform distribution, and suggested a modified EWMA chart by adding a lower reflecting barrier. Zhou *et al.* (2012) proposed a new EWMA method based on weighted likelihood estimation and testing. All of these works were built on the ground that the sample size is assumed to follow a pre-specified random or deterministic model, which is known *a priori* when establishing appropriate control limits before the control charts initiate. As Zhou *et al.* (2012) pointed out, unfortunately, traditional control charts are very sensitive to the correct specification of sample sizes.

In practice, our knowledge about the time-varying sample sizes is rarely available. A practical solution is to estimate the sample size distribution function based on a set of historical observations. When the historical observations are limited, such estimation is inevitably unreliable and model mis-specification and/or estimation errors would lead to unacceptable performance when implementing the control charts (Zhou *et al.*, 2012). To overcome this drawback, this paper proposes the use of probability control limits in an EWMA control chart for monitoring Poisson count data with time-varying sample sizes in Phase II. Notice that, though only the EWMA-type chart is discussed in particular in this paper, the key idea can be similarly applied to any effective traditional control charts such as the CUSUM chart. No matter what the (unknown) time-varying sample sizes are, the proposed EWMA chart *always* shares identical run length distribution with the Geometric distribution and is thus called EWMAG chart. Essentially the EWMAG chart uses dynamic control limits which are determined online and depend only on the current and past sample size observations. It does not need to specify any sample size models before implementation except the desired false alarm rate. The main idea is to maintain the conditional probability (the probability that the charting statistic exceeds the control limit given that there is no alarm before the current time point) to the specified false alarm rate at each time point. To dynamically determine the probability control limit online, a simulation-based procedure and a Markov chain procedure are discussed.

The remainder of this paper is organized as follows. We first discuss the statistical model and some previous work in Section 2. Then the new EWMA control chart with probability control limits is proposed in Section 3, followed by a performance study of the proposed control chart in Section 4. A healthcare surveillance example is visited to demonstrate the application of the EWMAG chart in Section 5. Finally, several remarks draw the paper to its conclusion in Section 6.

2 The Statistical Model

Let X_t be the count of an adverse event during the fixed time period $(t-1, t]$, for simplicity, we will call it the count of event at time t. Suppose X_t independently follows the Poisson distribution with the mean θn_t conditional on n_t, where θ and n_t denote the occurrence rate of the event and sample size at time t, respectively. To detect an abrupt change in the occurrence rate from θ_0 to another unknown value $\theta_1 > \theta_0$ at some unknown time τ, we use the following change point model,

$$X_i \overset{\text{i.d}}{\sim} \begin{cases} \text{Poisson}(\theta_0 n_i | n_i) & \text{for} \quad i = 1, \ldots, \tau - 1 \\ \text{Poisson}(\theta_1 n_i | n_i) & \text{for} \quad i = \tau, \ldots, \end{cases} \tag{1}$$

where the symbol $\overset{\text{i.d}}{\sim}$ means "independently distributed". The objective is to detect the change as soon as possible when it occurs through the sequential counts.

In the change point detection problem, a detection rule is often characterized by a charting statistic $a(\mathbf{n}_t, \mathbf{X}_t)$ and a control limit $h(\mathbf{n}_t)$ determined based on the historical data set $\{n_i, X_i\}_{1 \le i \le t}$, where $\mathbf{n}_t = \{n_i : \ 1 \le i \le t\}$ and $\mathbf{X}_t = \{X_i : \ 1 \le i \le t\}$. The stopping time T is defined as

$$T = \min\{t : a(\mathbf{n}_t, \mathbf{X}_t) > h(\mathbf{n}_t)\}. \tag{2}$$

$T = t$ means that an alarm (OC signal) is issued at time t the first time to declare that a change has occurred somewhere during the time period $[1, t]$. Similarly to the literature of health surveillance, we focus on the detection of increasing occurrence rate, that is the situation $\theta_1 > \theta_0$. Thus only upward shift is studied in this paper, while the detection of downward and two-side shifts can be constructed similarly without much difficulty. Note that the control limit $h(\mathbf{n}_t)$ depends only on $\mathbf{n}_t$, not $\mathbf{X}_t$.

Several control charts have been developed for Poisson count data in previous studies. In the following, we will discuss two EWMA charts proposed by Dong *et al.* (2008) and Ryan and Woodall (2010) and one CUSUM chart suggested by Mei *et al.* (2011) in detail.

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

$$Z_i = (1-\lambda)Z_{i-1} + \lambda \frac{X_i}{n_i}, \tag{3}$$

where $i = 1, 2, \ldots, t$, $Z_0 = \theta_0$, and $\lambda \in (0, 1]$ is a smoothing parameter which determines the weights of past observations. Derived from the EWMA sequence, the first EWMA control chart proposed by Dong *et al.* (2008), termed EWMAe, has a stopping rule,

$$T_{\text{EWMAe}} = \min\{t; Z_t \geq \theta_0 + L\sigma_t, t \geq 1\}, \quad \sigma_t^2 = \lambda^2 \sum_{i=1}^{t} (1-\lambda)^{2t-2i} \frac{\theta_0}{n_i}. \tag{4}$$

The control limit constant L is determined by the nominal value of ARL_0. To avoid the inertial problems, Ryan and Woodall (2010) modified the EWMAe method by adding a lower reflecting barrier at $Z_t = \theta_0$, i.e.,

$$T_{\text{EWMAM}} = \min\{t; \tilde{Z}_t \geq L\sigma_t, t \geq 1\}, \tag{5}$$

where

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

This method is referred as EWMA-modified (EWMAM) control chart. The CUSUM chart proposed by Mei *et al.* (2011) is given by

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

which signals when $W_t > L_C$, where $W_0 = 0$. Its control limit L_C is determined based on the desired ARL_0.

To determine the control limits for these control charts with time-varying sample sizes, the distribution of the sample size n_i is assumed to be known *a priori*. As discussed in the introduction, there is usually very little foreknowledge about the distribution of sample sizes in the future, especially when population-at-risk may be subject to sudden changes due to certain events such as wars. Once the assumption of the distribution deviates significantly from reality, the control limit determined based on the assumption will not be appropriate and result in undesired false alarm rates accordingly, which in turn hurt an operator's confidence in valid alarms. To address this issue, an EWMA control chart with probability control limits is discussed in the next section.

3 An EWMA Chart With Probability Control Limits

3.1 EWMAG chart

We use the EWMA-type control chart statistic (3) as the charting statistic in the following discussion of the probability control limits. The proposed exponentially weighted moving average control chart with probability control limits is called EWMAG chart since its in-control (IC) run length distribution is theoretically identical to the Geometric distribution, i.e., the false alarm rate does not depends on the time of the monitoring, nor the sample sizes being monitored.

The control limit of the EWMAG chart is set so that the conditional probability, i.e., the probability that the charting statistic exceeds the control limit given there is no alarm before the current time point, equals to a specified value of false alarm rate. To be more specific, we want to find the control limits satisfying the following equations,

$$\begin{aligned} &\Pr(Z_1 > h_1(\alpha) \mid n_1) = \alpha, \\ &\Pr(Z_t > h_t(\alpha) \mid Z_i < h_i(\alpha), 1 \leq i < t, n_t) = \alpha \quad \text{for} \quad t > 1, \end{aligned} \tag{6}$$

where α is the pre-specified false alarm rate. This is equivalent to performing a hypothesis test with the type-I error α at each time point t. Therefore the corresponding IC run length distribution is exactly the Geometric distribution (Hawkins and Olwell, 1998). At time t, the probability control limit is determined right after we observe the value of n_t. Consequently, the EWMAG chart does not need the assumption of future sample sizes and does not suffer from wrong model assumptions. This property makes the proposed EWMAG chart significantly different from previous control charts.

It is worth noting that the idea of using time-varying control limits was employed in the literature of self-starting control schemes. As indicated in those studies, the probabilities of false alarms from a chart may increase dramatically after short-runs if a fixed control limit is applied. The approach of using dynamic control limits is originally proposed by Margavio et al. (1995) and Lai (1995) and has been successfully formalized and utilized by Hawkins *et al.* (2003) in the parametric change-point based control charts with unknown IC parameters. See also Zou and Tsung (2010) for a related discussion. However, it should be emphasized here that our procedure distinguishes from those use of dynamic control limits mentioned above in the sense that the control limits in our procedure are determined on-line along with the process observations rather than decided before monitoring. That is, those control limits are data-dependent. This is a unique feature due to the time-varying population size n_t.

Due to the intricacy of the conditional probability (6), it seems impossible to solve $h_t(\alpha)$ analytically. Thus two computational procedures, simulation-based and Markov-chain, are suggested to approximate $h_t(\alpha)$ in this paper. Though only the EWMA-type control chart is discussed here, the two computational procedures can be applied to other charts such as the CUSUM chart with probability control limits.

3.2 Computational procedures for probability control limits

First we introduce the simulation-based procedure for computing the probability control limits. To explain this procedure clearly, let us start with the consideration of time $t = 1$. Under the IC condition, X_1 should follow the Poisson distribution with mean $\theta_0 n_1$, where n_1 is known exactly. Therefore we can obtain the control limit at the first time point by randomly generating $\hat{X}_{1,i}$, where $i = 1, \ldots, M$ and M is a sufficiently large integer, from the distribution Poisson($\theta_0 n_1$) and correspondingly calculating M values of pseudo Z_1 from (3) with $Z_0 = \theta_0$, say $\hat{Z}_{1,1}, \ldots, \hat{Z}_{1,M}$. We then sort those values in ascending order and store them in a vector $\hat{\boldsymbol{Z}}_{1M}$. The control limit $h_1(\alpha)$ can be approximated as the $M' = \lfloor M(1-\alpha) \rfloor$ largest value in $\hat{\boldsymbol{Z}}_{1M}$, where $\lfloor A \rfloor$ denotes the largest integer less than or equal to A. Theoretically, if $M \to \infty$, the equation $\Pr(Z_1 > h_1(\alpha)) = \alpha$ can be exactly satisfied. In this paper M is set to half a million to obtain an appropriate control limit at each time point. After determining the control limit $h_1(\alpha)$, we compare the value of $\hat{Z}_1$, which is calculated based on the observed X_1 and n_1, with $h_1(\alpha)$. An OC signal is issued if $\hat{Z}_1 > h_1(\alpha)$. Otherwise, we continue to the next time point $t = 2$.

According to (6), in order to determine the control limit $h_2(\alpha)$, we should ensure that the value of pseudo Z_1 is less than or equal to h_1. Hence only the ranked values $\hat{Z}_{1,(1)}, \ldots \hat{Z}_{1,(M')}$ should be kept to determine $h_2(\alpha)$. We store the M' ranked pseudo Z_1 into a vector $\hat{\boldsymbol{Z}}'_{1M'}$. Given n_2, a vector $\hat{\boldsymbol{Z}}_{2M}$ with the dimension M can then be obtained by

$$\hat{Z}_{t,i} = (1-\lambda)\hat{Z}_{t-1,j} + \lambda \frac{\hat{X}_{t,i}}{n_2} \tag{7}$$

where $t = 2$, $i = 1, \ldots, M$, $\hat{Z}_{1,j}$ is uniformly selected from $\hat{\boldsymbol{Z}}'_{1M'}$ with $j \in \{1, \ldots, M'\}$, and $\hat{X}_{2,i}$ are randomly generated from Poisson$(\theta_0 n_2)$. Sort the M elements of $\hat{\boldsymbol{Z}}_{2M}$ in ascending order, we can obtain the control limit $h_2(\alpha)$ by setting it at the $(1-\alpha)$-quantile of the M elements. Again we keep the ranked statistics $\hat{Z}_{2,(1)}, \ldots, \hat{Z}_{2,(M')}$ to the next stage $t = 3$. Repeat the above procedure by simulating M samples of Poisson$(\theta_0 n_3)$, ... etc. The simulation-based procedure can be summarized as the following algorithm:

Algorithm 1 **(Simulation-based procedure)**

1. If there is no OC signal at time $t-1$ $(t = 1, 2, \ldots)$, $\hat{X}_{t,i}$ $(i = 1, \ldots, M)$ are generated from the distribution Poisson$(\theta_0 n_t)$ where n_t is known exactly. Accordingly, M values of the pseudo charting statistic Z_t are obtained through (7).

2. Sort them in ascending order and the α upper empirical quantile of those M values is used for approximating the control limit $h_t(\alpha)$.

3. Compare the value of $\hat{Z}_t$, which is calculated based on observed X_t and n_t, with $h_t(\alpha)$ to decide whether to issue an OC signal or to continue to the next time point.

4. If continue, set $M' = [M(1-\alpha)]$ and eliminate the values $\hat{Z}_{t,(M'+1)}, \ldots, \hat{Z}_{t,(M)}$. Go back to step 1.

Next we turn attention to the Markov-chain procedure for computing the probability control limits. The Markov-chain model described here can be considered as an extension of the methods proposed by Brook and Evans (1972) and Borror *et al.* (1998). However, different from the previous methods, we particularly design the Markov-chain procedure for monitoring the occurrence rate with time-varying sample sizes in this paper. Before discussing this procedure in detail, we first introduce a critical idea of this procedure, as well as some concepts including bounds of charting statistics and states, in the following.

At time t, the value of charting statistic Z_t is inside an interval with two-side bounds. Since $Z_t = (1-\lambda)Z_{t-1} + \lambda X_t/n_t$ and $X_t \geq 0$, we can simply set the lower bound $L = 0$ and search for the upper bound U with the constraint $\Pr(Z_t \leq U) = 1 - \xi$, where $\xi > 0$ is a sufficiently small constant *at each time point*. In this study, we set $\xi = 1e^{-16}$. After determining the two-side bounds, we divide the interval (L, U) equally into K partitions (states). Then the i^{th} state is the subinterval (L_i, U_i), where

$$L_i = L + \frac{(i-1)(U-L)}{K} \quad \text{and} \quad U_i = L + \frac{i(U-L)}{K}.$$

The corresponding midpoint, m_i, of the i^{th} state can be determined by $m_i = L + (2i-1)(U-L)/2K$. When Z_t falls into the i^{th} state at the time t, we have $L_i < (1-\lambda)Z_{t-1} + \lambda X_t/n_t \leq U_i$, which is equivalent to

$$\frac{n_t[L_i - (1-\lambda)Z_{t-1}]}{\lambda} < X_t \leq \frac{n_t[U_i - (1-\lambda)Z_{t-1}]}{\lambda}. \tag{8}$$

The probability that Z_t is within the i^{th} state, conditioned on $Z_{t-1} = m_j$ $(i, j \in \{1, \ldots, K\})$, can be obtained by calculating the corresponding probability of X_t since its probability density function is $f(X_t) = e^{-\theta_0 n_t}(\theta_0 n_t)^{X_t}/X_t!$.

To clearly describe the procedure, we again start with the consideration of time $t = 1$. Since Z_0 is specified as θ_0, we consider that Z_0 is in the point θ_0 with probability 1 at time $t = 0$. At time $t = 1$, given Z_0 and n_1, we first determine the upper bound U with the constraint

$$\Pr(Z_1 \leq U) = \Pr\left(X_1 \leq \frac{n_1(U - (1-\lambda)Z_0)}{\lambda}\right) = 1 - \xi,$$

and then divide the interval into K states. Also the probability of Z_1 falls into i^{th} state can be easily calculated by

$$\Pr(Z_1 \in \text{state } i \mid Z_0) = \begin{cases} \Pr\left(\frac{n_1[L_i-(1-\lambda)Z_0]}{\lambda} < X_1 \leq \frac{n_1[U_i-(1-\lambda)Z_0]}{\lambda}\right), \\ \qquad \text{if } \lfloor\frac{n_1[L_i-(1-\lambda)Z_0]}{\lambda}\rfloor < \lfloor\frac{n_1[U_i-(1-\lambda)Z_0]}{\lambda}\rfloor, \\ 0, \qquad \text{otherwise,} \end{cases} \tag{9}$$

where $i = 1, \ldots, K$. We then obtain a vector of K probabilities at time $t = 1$, $\boldsymbol{P}_1 = (p_1, p_2, \ldots, p_K)'$, where the element p_i equals to the value of $\Pr(Z_1 \in \text{state } i \mid Z_0)$. The control limit $h_1(\alpha)$ at time $t = 1$ is determined as the upper bound U_r of the r^{th} state, where

$$r = \arg\min\{j : \sum_{i=1}^{j} p_i \geq 1 - \alpha\}, \; j = 1, 2, \ldots, K. \tag{10}$$

If $\hat{Z}_1 < h_1(\alpha)$ at time $t = 1$, where $\hat{Z}_1$ is calculated based on observed X_1 and n_1, the process is declared as in-control and we can proceed to the next stage $t = 2$. Otherwise, an alarm should be issued.

At time $t = 2$, we first deal with Z_1 since Z_2 is partially dependent on it. The restriction of $Z_1 < h_1(\alpha)$ requires us to keep only the r states of Z_1 and store their normalized probabilities $\tilde{p}_j = p_j / \sum_{j=1}^{r} p_j$ $(j = 1, \ldots, r)$ into a vector $\tilde{\boldsymbol{P}}_1 = (\tilde{p}_1, \ldots, \tilde{p}_r)'$. For any value of Z_1 in the j^{th} state, it is represented with $Z_{1j} = m_j$ where $j = 1, \ldots, r$. Then all possible values of Z_1 considered at time $t = 2$ are the r elements of $\boldsymbol{Z}_1 = (Z_{11}, Z_{12}, \ldots, Z_{1r})'$. Given $Z_1 = Z_{1r}$ and n_2, the upper bound U of Z_2 can then be determined with the constraint

$$\Pr(Z_2 \leq U) = \Pr\left(X_2 \leq \frac{n_2[U - (1-\lambda)Z_{1r}]}{\lambda}\right) = 1 - \xi. \tag{11}$$

Again, we divide the interval $(0, U)$ into K states. When $Z_1 = Z_{1j}$, K conditional probabilities of Z_2 are obtained and stored in a vector $\boldsymbol{P}_{2j} = (p_{1j}, p_{2j}, \ldots, p_{Kj})'$, where $p_{ij} = \Pr(Z_2 \in \text{state } i \mid Z_{1j}, n_2)$, $i = 1, \ldots, K$ and $j \in \{1, \ldots, r\}$. Let p_i be the probability of that Z_2 falls into the i^{th} state given that Z_1 is in-control. The vector of conditional probabilites of Z_2, $\boldsymbol{P}_2 = (p_1, p_2, \ldots, p_K)'$, is calculated as

$$\boldsymbol{P}_2 = (\boldsymbol{P}_{21}, \boldsymbol{P}_{22}, \ldots, \boldsymbol{P}_{2r})_{(K \times r)} \, \tilde{\boldsymbol{P}}_{1(r \times 1)}. \tag{12}$$

With the conditional probabilities of Z_2, the control limit $h_2(\alpha)$ at time $t = 2$ is approximated by the upper bound of the r^{th} state satisfying (10). The procedure is summarized in the following Algorithm 2.

Algorithm 2 **(Markov-chain procedure)**

1. If there is no OC signal at time $t - 1$ $(t = 1, 2, \ldots)$, calculate the vector $\tilde{\boldsymbol{P}}_{t-1}$ storing r normalized probabilities and the vector $\boldsymbol{Z}_{t-1}$ containing the values of r midpoints. Both of the two vectors are with the size $(r \times 1)$. Specially, at time 0, we have $r = 1$, $\tilde{P}_0 = 1$ and $Z_0 = \theta_0$.

2. Always set the lower bound, L, to be 0 and search for the upper bound, U, of Z_t based on the value of $Z_{(t-1)r}$ and divide the interval (L, U) into K states.

3. For each $Z_{(t-1)j}$, $j = 1, \ldots, r$, compute the probability vector $\boldsymbol{P}_{tj}$ with the size $(K \times 1)$. Then the vector of conditional probabilities can be obtained through formula (12).

4. Compare the value of $\hat{Z}_t$ calculated based on observed X_t and n_t with the determined control limit at time t. The control limit, $h_t(\alpha)$, is chosen to be the upper bound of r^{th} state, where $r = \arg\min\{j \mid \sum_{i=1}^{j} p_i \geq 1 - \alpha\}$. If $\hat{Z}_t > h_t(\alpha)$, an OC signal should be issued. Otherwise, go back to step 1 and continue to the next time stage.

3.3 Comparison of the two computational procedures

As discussed before, theoretically the EWMAG chart has identical IC run-length distribution with that of the Geometric random variable, although its design does not utilize any information of n_t. To verify this statement and compare the two computational procedures, we conduct simulation studies under various scenarios of sample size discussed as follows. In healthcare surveillance, Mei *et al.* (2011) suggested to model population growth by the logistic model. Constant sample size (Dong *et al.*, 2008) and uniformly distributed n_t (Ryan and Woodall, 2010) have also been considered in modeling the population size. In particular, the following four scenarios are used in our simulation studies.

(I) Increasing Scenario: $n_t = \frac{c_1}{C(0.5+\exp\{-(t-c_2)/c_3\})}$, where $C = 1$ or 8;

(II) Decreasing Scenario: $n_t = \frac{c_1/2.4}{1+\exp\{(t-c_2)/c_3\}} + C$, where $C = 1$ or 7;

(III) Constant Scenario: $n_t = 4.5$ or $n_t = 10$;

(IV) Uniform Scenario: $n_t \sim U(1, 4)$ or $n_t \sim U(5, 18)$,

In Scenarios (I) and (II) $c_1 = 13.8065$, $c_2 = 11.8532$ and $c_3 = 26.4037$ which are the same as those in Mei *et al.* (2011). Notice that each scenario is set to have different parameters, e.g., different C's and different constant values. For conciseness, we just choose one setting in each scenario for testing in this section.

Under each scenario, we set the desired false alarm rate $\alpha = 0.0027$ and accordingly the desired IC ARL (ARL_0) should approximately be 370. For illustration, the control chart performance is summarized using ARL_0, percentiles of the marginal distribution of the run length, and standard deviation of the run length (SDRL). Besides these quantities, we study the false alarm rate for the first 30 observations, $P(T \le 30 \mid \text{in-control})$, as well. We set $\theta_0 = 1$ and the smoothing parameter $\lambda = 0.1$ and use Monte Carlo simulations of $50,000$ replications to estimate the run length distribution of the EWMAG chart. The Fortran codes for implementing the EWMAG chart is available from the authors upon request.

Tables 1 and 2 summarize the simulation results of the EWMAG charts with the simulation-based and Markov-chain procedures, respectively. M in the simulation-based procedure is set as $500,000$ and K in the Markov-chain procedure is chosen as $3,000$. We use the notations SE, $Q(.10)$, $Q(.90)$ and FAR for the standard error of ARL_0 estimation, 10^{th} percentile, 90^{th} percentile and false alarm rate, respectively. The IC run-length distribution is considered to be satisfactory here if it is close to the Geometric distribution or more generally its variation is less than that of a Geometric distribution. As a reference, 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 39, 852 and 0.078, respectively.

Table 1: IC performance of the EWMAG chart with simulation-based procedure; $\theta_0 = 1$

Distribution	ARL_0	SE	SDRL	Q(.10)	Median	Q(.90)	FAR
(I) [C=8]	372	1.66	371.75	40	258	857	0.0781
(II) [C=1]	371	1.67	373	39	256	854	0.0805
(III) [n_t=4.5]	370	1.65	369	39	258	849	0.0805
(IV) [U(1,4)]	369	1.66	370	37	256	846	0.0826
Geometric	370	-	370	39	256	852	0.078

Table 2: IC performance of the EWMAG chart with Markov-chain procedure; $\theta_0 = 1$

Distribution	ARL_0	SE	SDRL	Q(.10)	Median	Q(.90)	FAR
(I) [C=8]	370	1.65	368.81	39	258	843	0.0785
(II) [C=1]	372	1.66	370.19	40	259	858	0.0765
(III) [n_t=4.5]	371	1.67	373	39	257	852	0.0796
(IV) [U(1,4)]	372	1.67	373	40	257	852	0.0769
Geometric	370	-	370	39	256	852	0.078

Under Scenarios (I)-(IV), the values of ARL_0 obtained from the EWMAG charts based on the two procedures are apparently close to the desired value of 370 (the slight difference is due to simulation errors). SDRL, $Q(.10)$, $Q(.90)$, Median and FAR are all approximately equivalent to the theoretical values. That is, the EWMAG chart has identical IC run-length distribution with that of the Geometric distribution and the two proposed calculation procedures have the similar design and performance. Therefore we will use only the simulation-based procedure in the following studies when evaluating the EWMAG chart because it is faster in implementation.

4 Performance Comparison

In this section, the performance of the EWMAG chart is compared with that of the EWMAe, EWMAM, and CUSUM charts respectively under the four scenarios of time-varying sample sizes discussed previously. We set the false alarm rate $\alpha = 0.0027$ and choose the smoothing parameter of EWMA-type control charts as $\lambda = 0.1$, and $\theta_0 = 1$ of the CUSUM chart as in Mei *et al.* (2011). In order to make a comprehensive comparison, we consider the performance under both the IC and OC situations in the following.

4.1 IC performance

Assume that population sizes can be known exactly. Figure 1 presents the IC run lengths of the four control charts (EWMAe, EWMAM, CUSUM and EWMAG) under the four scenarios considered in Section 3.3. As expected, the run length distribution curve of the EWMAG chart merges with that of the Geometric distribution, which verifies again that the IC run length distribution of the proposed EWMAG chart is exactly the Geometric distribution. In contrast, the curves of the other three control charts deviate significantly from the Geometric distribution curve in different degrees under various scenarios. In particular, the EWMAe chart often has higher early false alarms than the Geometric distribution, especially under Scenarios (I), (III), and (IV), while the CUSUM chart has considerably lower false alarms than the Geometric distribution under Scenario (I).

Figure 1: Comparison of IC run length distribution among the EWMAe, EWMAM, CUSUM and EWMAG charts

Table 3: The effect of mis-specified population sizes on IC ARLs of the EWMAe, EWMAM and CUSUM charts; $\theta_0 = 1$

Assumed Dist.	Real Distribution							
	(I)[C=8]	(I)[C=1]	(II)[C=1]	(II)[C=7]	(III)[n_t=4.5]	(III)[n_t=10]	(IV)[U(1,4)]	(IV)[U(15,18)]
				EWMAe				
(I)[C=8]	／	461	307	407	384	424	343	437
(II)[C=1]	450	566	／	502	471	521	411	553
(III)[n_t=4.5]	352	436	295	390	／	408	327	426
(IV)[U(1,4)]	393	492	325	437	416	461	／	486
				EWMAM				
(I)[C=8]	／	486	294	418	392	441	331	479
(II)[C=1]	467	648	／	549	503	580	423	615
(III)[n_t=4.5]	347	459	286	398	／	420	319	435
(IV)[U(1,4)]	408	545	329	472	437	495	／	526
				CUSUM				
(I)[C=8]	／	1032	503	424	402	390	394	607
(II)[C=1]	282	986	／	256	292	374	300	481
(III)[n_t=4.5]	314	1036	499	423	／	394	390	589
(IV) [U(1,4)]	297	1019	466	419	291	393	／	561

The IC ARLs with mis-specified models are reported in Table 3. Obviously, the observed ARL_0's would deviate from the nominal one (370) to certain degrees when the distributional model of population sizes does not match the reality. Even with appropriate models, mis-specified parameters in the distribution function also result in poor IC performance. As pointed out before, accurate information about future population sizes can rarely be obtained in many applications. Therefore, control charts constructed on the basis that distribution functions of varying population sizes are exactly known will result in an unacceptable run-length distributions as shown in this table. This clearly indicates the advantage of our EWMAG chart in practice.

4.2 OC performance

To investigate the OC performance, only the EWMAe chart is used for the comparison since it has identical form as the EWMAG chart except the control limits. We assume that the population sizes are known exactly here since it is unfair to compare different procedures in terms of OC ARL when their IC run-length distributions differ significantly. In general, OC run length distributions depend on the OC conditions (that is, the rate of event occurrence changes from θ_0 to θ_1 at time τ) and the occurrence time τ. Therefore, in the following, OC performance is studied under different values of θ_1 and different occurrence time τ successively.

Assuming $\tau = 21$, Table 4 presents the OC ARLs of the EWMAG and EWMAe charts with different values of θ_1 under Scenarios (I)-(IV). It is easy to see that the OC performance of the two control charts are comparable regardless of the population scenarios. This demonstrates that the proposed EWMAG chart is able to ensure desired IC run length performance without degradation of its change detection ability.

Setting $\theta_1 = 1.2$, Figure 2 shows the OC ARLs of the two charts under Scenarios (I)-(IV) when the occurrence time τ ranges from 1 to 100. It shows that both charts are sensitive to the occurrence time of the change, especially when population monotonely increases or decreases. In general, the EWMAG chart always has smaller OC ARL values than the EWMAe chart, except when τ is very small under Scenario (I).

It is worth to indicate that the main objective of the proposed dynamic procedure is to

Table 4: OC ARL Comparison of the EWMAG chart and the EWMAe chart under Scenarios (I)-(IV) with $\tau = 21$

θ_1	(I)[C=8]		(II)[C=1]		(III)[$n_t = 4.5$]		(IV)[U(1,4)]	
	EWMAe	EWMAG	EWMAe	EWMAG	EWMAe	EWMAG	EWMAe	EWMAG
1.025	249	241	284	285	233	224	266	259
1.050	171	168	220	217	149	146	194	188
1.075	126	124	172	169	104	99.9	146	141
1.100	95.3	94.9	137	133	73.8	72.2	111	107
1.200	44.5	43.8	55.2	52.5	27.7	27.8	48.0	45.9
1.300	27.2	27.1	26.3	25.0	15.2	15.1	26.7	25.9
1.400	19.2	19.1	15.1	14.5	10.0	9.88	17.4	17.1
1.500	14.5	14.5	10.3	9.89	7.31	7.25	12.6	12.6
1.750	8.77	8.81	5.50	5.32	4.19	4.15	7.11	7.18
2.000	6.18	6.18	3.64	3.49	2.83	2.80	4.81	4.90
2.500	3.69	3.72	2.03	1.96	1.59	1.59	2.77	2.94
3.000	2.56	2.58	1.33	1.27	1.02	1.02	1.87	2.06
4.000	1.48	1.48	0.66	0.63	0.47	0.47	1.01	1.23

make the IC run length distribution of a control chart attaining the theoretical Geometric distribution rather than to improve the detection ability of the chart. Therefore we suggest to use the proposed EWMAG chart in practice due its desired IC run length performance and competitive OC performance.

5 An Example in Healthcare Surveillance

In this section, an example of female thyroid cancer in healthcare surveillance is used to demonstrate the application of the proposed EWMAG chart. According to the report from the National Cancer Institute, there are about $37,000$ new cases of thyroid cancer each year in the U.S. and females are most likely to have thyroid cancer at a ratio of three to one. Thyroid cancer may occur in any age group, although it is most common after age 30, and its aggressiveness increases significantly in older patients.

Figure 2: OC comparison of the EWMAe and EWMAG charts with different τ

The data, provided by the New York State Cancer Registry through its official website[1], include the number of female thyroid cancer cases and the incidence rate each year in the New York State. Based on the provided data, the corresponding population size each year can be simply derived. In Figure 3, (a) and (b) show the time series plots of the counts (in the units of 100 cases) and the incidence rates per 100 million population of female thyroid cancer in the New York State respectively. It can be observed that the incidence rate remains quite stable before year 1982 and exhibits a slight increase after 1983. The increase tendency becomes more significant starting from 1990. In Figure 3 (c), the population size of female increases from 8.95 million in 1976 to 9.88 million in 1995 significantly.

Based on the pattern of incidence rate discussed above, the period from 1976 to 1982 is chosen as the Phase-I reference sample and accordingly the nominal incidence rate is

[1]http://www.health.ny.gov/statistics/cancer/registry/table2/tb2thyroidnys.htm

(a) Population

(b) Case

(c) Incidence Rate

(d) Estimated population

Figure 3: Female thyroid cancer incidence data and the expected population size estimated by the fitted logistic model. (a) Female population (b) number of thyroid cancer cases (c) incidence rate and (d) Estimated population sizes based on the fitted logistic model.

estimated as $\theta_0 = 0.45$. A calibration sample of this size may be smaller than ideal to determine fully the in-control parameter but it suffices to illustrate the use of the method in a real-world setting. Our target is to monitor the incidence rate of female thyroid cancer from 1983 to 1995 and compare the performance of the EWMAG chart and that of the EWMAe chart in this example.

Before the monitoring, we set $\alpha = 0.0027$ or equivalently $\mathrm{ARL}_0 = 370$. We fit a logistic model to the observed population sizes in Phase I by a nonlinear least square method (year 1975 is treated as time 0 and the population sizes are in the units of $1,000,000$) and obtain $n_t = 5.856/[0.5 + \exp\{-(t + 86.295)/45.645\}]$ (Scenario I). Figure 3 (d) shows the real population sizes and the expected population sizes well estimated by the fitted logistic regression model with random variations. It indicates that in general the natural character of the population growth can be well described by this logistic model. To show the adverse impacts of inappropriately estimated population sizes, we further assume that the population sizes are constant with $n_t = 9.0$ (Scenario III) or uniformly distributed with $n_t \sim U(7.0, 10.0)$ (Scenario IV) for comparison.

Under the three different scenarios of population sizes discussed above, the control limits of the EWMAe chart are determined as $\theta_0 + L\sigma_t$ with $L_{(I)} = 2.533$, $L_{(III)} = 2.547$, and $L_{(IV)} = 2.553$ respectively. Figure 4 plots the charting statistics (the solid curves connecting the dots) and the corresponding control limits (the dashed curves) of the EWMAG chart and the EWMAe charts.

(a) EWMAG chart

(b) EWMAe chart under Scenario I

(c) EWMAe chart under Scenario III

(d) EWMAe chart under Scenario IV

Figure 4: The EWMAG chart and the EWMAe charts for monitoring the female thyroid cancer incidence data set

A significant increase can be observed in incidence rate from 1990. Therefore an alarm should be issued as soon as possible after 1990. From the plots, it can be seen that the EWMAG chart exceeds its control limit in year 1994 and it remains above the control limit afterwards. The EWMAe chart in Scenario I also triggers a signal in the same year. It is rational to have similar detection results from the two control charts since the natural population sizes are appropriately modeled in Scenario I. On the other hand, the EWMAe chart issues a delayed OC signal in year 1995 in both Scenarios III and IV, which is caused by the inappropriate distributions of population sizes. The example justifies the usefulness of the EWMAG chart in reality since its detection capability does not depend on the estimation of population sizes, which is difficult to be determined in advance.

6 Conclusion

As indicated in Section 4, there is a significant shortage of traditional control charts for monitoring Poisson count data with time-varying sample sizes. That is, their need for knowing the distribution or model of sample sizes before monitoring since we rarely have such foreknowledge in real applications and an inappropriate assumption or estimation may lead to poor run-length performance of the traditional control charts. To this end, we suggest to use probability control limits, which are determined based on the observed real sample sizes on-line, for traditional control charts so that the conditional probability that the charting statistic exceeds the control limit at present given that there is no alarm before the current time point can be guaranteed to attain a specified false alarm rate. In this paper, a EWMA-type control chart with the probability control limits, termed EWMAG chart, is discussed in detail. However, we should emphasize that the presented idea can be readily applied to any effective control charts such as the CUSUM chart. The proposed EWMAG chart is able to deliver robust and satisfactory IC and OC run-length performance under various situations, which has been proved by the simulation studies in this paper.

For the future research, there are some valuable directions discussed as follows. First of all, recall that in this paper we apply the proposed EWMAG chart to monitor the occurrence rate of adverse events assuming that the count of events follows a Poisson distribution when given the sample size on-line. Clearly, the EWMAG chart can be also extended to more general cases in which the observations follow a conditional distribution given some related parameters/covariates whose information can be obtained and updated on-line as well. For example, when monitoring and predicting shopping quantity in retail data mining, frequency of purchase and other demographics play an important role to determine the baseline purchasing frequency and quantity (Rossi *et al.*, 1996).

In addition, it is well recognized that the performance of the EWMA-type chart depends on the smoothing parameter λ, which is simply set to a constant value in this paper. One of our ongoing work is to sequentially determine optimal values of λ in the EWMAG chart. Moreover, the occurrence rates, which correspond to different sample sizes, are regarded as equally informative in the current study. That the EWMAG pays the same attention to a ratio R_i based on a small at-risk n_i as to one based on a large n_i, even though the later is more informative in some cases. A control chart with sample-size-varying smoothing parameters would be more reasonable.

Finally, 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 EWMAG chart which can simultaneously update parameter estimates and check for OC conditions (Zantek and Nestler, 2009).

Acknowledgment

The authors would like to thank the Editor, Associate Editor and two anonymous referees for their many helpful comments that have resulted in significant improvements in the article.

References

Borror, C.M., Champ, C.W. and Rigdon, S.E. (1998). "Poisson EWMA Control Charts". *Journal of Quality Technology* 30, pp. 352-361.

Brook, D. and Evans, D. A. (1972). "An Approach to the Probability Distribution of CUSUM Run Length". *Biometrika* 59, pp. 539-549.

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, pp. 843-853.

Frisén, M. and De Maré, J. (1991). "Optimal Surveillance". *Biometrika* 78, pp. 271-280.

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

Hawkins, D. M. and Olwell, D. H. (1998). *Cumulative Sum Charts and Charting for Quality Improvement*, 1st edition. New York, NY: Springer-Verlag.

Hawkins, D. M., Qiu, P., and Kang, C. W. (2003). "The Changepoint Model for Statistical Process Control". *Journal of Quality Technology* 35, pp. 355-366.

Huwang, L., Wang, Y. H., Yeh, A. B., and Chen, Z. S. (2009). "On the Exponentially Weighted Moving Variance". *Naval Research Logistics* 56, pp. 659-668.

Lai, T. L. (1995). "Sequential Changepoint Detection in Quality Control and Dynamical Systems". *Journal of the Royal Statistical Society, Series B* 57, pp. 613-658.

Lucas, J. M. (1985). "Counted Data CUSUMs". *Technometrics* 27, pp. 129-144.

Margavio, T. M., Conerly, M. D., Woodall, W. H., and Drake, L. G. (1995). "Alarm Rates for Quality Control Charts". *Statistics & Probability Letters* 24, pp. 219-224.

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, pp. 597-624.

Montgomery, D. C. (1990). *Introduction to Statistical Quality Control*, 2nd edition. New York, NY: Wiley.

Nancy, K. (2008). "Hormone Therapy and the Rise and Perhaps Fall of US Breast Cancer Incidence Rates: Critical Reflections". *International Journal of Epidemiology* 37, pp.627-637.

Poloniecki, J.; Valencia, O.; and Littlejohns, P. (1998). "Cumulative Risk Adjusted Mortality Chart for Detecting Changes in Death Rate: Observational Study of Heart Surgery". *British Medical Journal* 316, pp. 1697-1700.

Rossi, P. E., McCulloch, R., and Allenby, G. (1996). "The Value of Purchase History Data in Target Marketing". *Marketing Science* 15, pp. 321-340.

Ryan, A. G. and Woodall, W. H. (2010). "Control Charts for Poisson Count Data with Varing Sample Sizes". *Journal of Quality Technology* 42, pp. 260-274.

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, pp. 725-741.

Sonesson, C. and Bock, D. (2003). "A Review and Discussion of Prospective Statistical Surveillance in Public Health". *Journal of the Royal Statistical Society, Ser. A* 166, pp. 5-21.

White, C. H. and Keats, J. B. (1996). "ARLs and Higher-Order Run-Length Moments for the Poisson CUSUM". *Journal of Quality Technology* 28, pp. 363-369.

Woodall, W. H. (2006). "The Use of Control Charts in Health-Care and Public-Health Surveillance". *Journal of Quality Technology* 38, pp. 89-134.

Zantek, P. F. and Nestler, S. T. (2009). "Performance and Properties of Q-statistic Monitoring Schemes". *Naval Research Logistics* 56, pp. 279-292.

Zou, C. and Tsung, F. (2010). "Likelihood ratio based distribution-free EWMA schemes". *Journal of Quality Technology* 42, pp. 174-196.

Zhou, Q., Zou, C., Wang, Z. and Jiang, W. (2012). "Likelihood-based EWMA charts for monitoring Poisson count data with time-varying sample sizes". *Journal of the American Statistical Association* to appear.