An MEWMA control chart based on multivariate Poisson distribution data

Yizhuo Xiao¹, Xuemin Zi²

¹School of Science, Tianjin University of Technology and Education, Tianjin, Tianjin, 300222, China
²School of Science, Tianjin University of Technology and Education, Tianjin, Tianjin, 300222, China

*E-mail: rkchp@zhobk.cn

Abstract. Count data appears widely in finance, biomedical science, psychology, actuarial insurance, road safety and many other areas of life, and it is often multivariate. In this paper, we consider using the multivariate exponentially weighted moving average (MEWMA) control chart method to monitor count data under the assumption of multivariate Poisson distribution, which is derived by He [1] from another form of multivariate gamma distribution. We illustrate the proposed scheme by numerical simulation. In this paper, we provide two tables of average run length (ARL) values of the MEWMA control chart with different parameter shifts. The performance of the proposed control chart is evaluated by the ARL, given a parameter shift occurred, which is able to deliver satisfactory control efficiency.

1. Introduction

In recent years, due to control charts could be used widely in the process of every productions, it have created a certain practical application values in the real world. Control charts are convenient to monitor and diagnose the production process or product quality, which enable the managers to find and solve the problems as soon as possible, and make the production enterprise obtain considerable economic benefits. Therefore, the research on process control charts has received more attention.

In general, dealing with several quality characteristics of a product is a common work in the manufacturing process. Multivariate statistical process control (MSPC) is a method which aims to monitor multiple process variables in many practical applications. In order to deal with multiple quality characteristics at the same time, the focus on control charts has gradually shifted from univariate to multivariate.

MEWMA control chart is a classical method of multivariate control chart, which is considered under the different specified distributions. MEWMA control chart can be used to improve the detection of small and medium shifts in multivariate statistical process control. In a general situation, the quality characteristics of counting processes are considered to be binomial distribution or Poisson distribution. Poisson distribution is a kind of distribution commonly used to describe count data information. For example, in the field of insurance, the claim frequency data of different types of insurance policies which is obviously following the multivariate Poisson distribution should be monitoring seriously in order to meet the actuaries' demand for pricing problems of different policies. Liang et al. [2] analyzed the number of forest fires in different regions by using the multivariate generalized Poisson model. Patel [3] first proposed a control chart of multivariate Poisson distribution.
By extending Hotelling control procedure, Patel [3] proposed a monitor scheme which was suitable for both multivariate binomial data and Poisson data. Patel also did some studies on time-dependent and time-independent samples. However, Patel's scheme was too complex to be applied in practice. The multivariate Poisson model adopted by Chiu and Kuo [4] was given under the extension of Holgate's bivariate model.

Zhang [5] introduced another form of multivariate gamma distribution and its properties. Based on his research, He [1] proposed the definition of discrete distribution of multivariate Poisson distribution, and discussed a series of properties of multivariate Poisson distribution in detail. He [1] defined the new multivariate Poisson distribution by using the density function, and then he discussed some properties of multivariate Poisson distribution from the study of its characteristic functions. At present, there are many researchers who combine various of multivariate Poisson models with common control charts to construct a new control chart for monitoring multivariate discrete count data, such as MP-CUSUM and MP-EWMA. The application of MEWMA control chart method to data with multivariate Poisson distribution has been discussed in some literatures. Laungrungrong et al. [6] proposed a method to evaluate and implement the MEWMA scheme with multivariate Poisson framework. They used a special form of model in which the covariances between any two variables were equal. They also showed that the MEWMA control chart based on the multivariate Poisson distribution was superior to the MEWMA control chart based on the normal theory in terms of the in-control (IC) ARL.

In this paper, we will construct an MEWMA control chart for the data coming from the multivariate Poisson distribution which was proposed by He [1]. The covariance matrix of multivariate Poisson random variables proposed by He [1] is used to assist in constructing statistics. We use the new control chart to monitor the mean value for given Poisson distribution and detect the shifts as soon as possible.

This paper is organized as follows. In Section 2, we give a brief review of the MEWMA control chart and describe the new form of multivariate Poisson distribution proposed by He [1]. We present a numerical example using the proposed control chart to detect different mean shifts in Section 3. Some conclusions about the content of this paper are given in Section 4.

2. Methodology

2.1. MEWMA Control Chart

Lowry et al. [7] expanded the univariate EWMA proposed by Roberts [8] to MEWMA. The MEWMA vector is defined as

$$Z_t = RX_t + (1 - R)Z_{t-1}, \ t = 1, 2, \ldots$$

where $Z_0$ is the p-dimensional zero vector, $X_t$ is the $t^{th}$ observation vector, $I$ is a q-dimensional identity matrix, $R = \text{diag}(r_1, r_2, \ldots, r_q)$, $0 < r_j \leq 1$ is a smooth parameter and $j = 1, 2, \ldots, q$. Usually we take the same smooth parameter for all the components so that $r_1 = r_2 = \cdots = r_q = r$. Choices of $r$ were presented by Lucas and Saccucci [9]. Then

$$Z_t = rX_t + (1 - r)Z_{t-1}.$$  

The form of the asymptotic covariance matrix is given as follows:

$$\Sigma_{Z_t} = \left\{ \frac{r}{2-r} \right\} \Sigma.$$

The MEWMA chart gives an out-of-control (OC) signal as soon as

$$T_t^2 = Z_t^2 \Sigma_{Z_t}^{-1} Z_t > H,$$

where $H > 0$ is a control limit that satisfies a particular given IC ARL. Lowry et al. [7] showed that the values of non-centrality parameters are critical to the ARL performance of MEWMA control chart. Here the non-centrality parameter is defined as

$$\delta = \left[ (\mu - \mu_0) \Sigma^{-1}(\mu - \mu_0) \right]^{1/2},$$

where $\mu_0$ and $\mu$ are the mean vectors before and after shift respectively. $\Sigma$ is the covariance matrix of multivariate random variables. On the other hand, this non-central parameter formula is also a Mahalanobis distance formula.
2.2. The Multivariate Poisson Distribution

Suppose that $X = (X_1, X_2, \ldots, X_q)'$ is a q-variate Poisson random variable, and the multivariate Poisson probability distribution function has the form (He [1]):

$$f_X(x) = \frac{\theta_1^{x_1} \theta_2^{x_2} \cdots x_{q}^{x_q}}{x_1!(x_2-x_1)! \cdots (x_q-x_{q-1})!} e^{-(\theta_1+\theta_2+\cdots+\theta_q)}$$

where $\theta_k > 0, k = 1, 2, \cdots, q$. $x_1 \leq x_2 \leq \cdots \leq x_q, x_k = 0, 1, \cdots, k = 1, 2, \cdots, q$. This distribution is denoted as $X \sim P_{q}(\theta_1, \theta_2, \cdots, \theta_q)$.

Suppose that $X = (X_1, X_2, \ldots, X_q)' \sim P_{q}(\theta_1, \theta_2, \cdots, \theta_q)$. Then the characteristic function of $X$ is defined as:

$$\varphi_X(w) = \exp\left\{-(-\theta_1 + \theta_2 + \cdots + \theta_q) + \left[\theta_1 e^{i(w_1+w_2+\cdots+w_q)} + \theta_2 e^{i(w_2+\cdots+w_q)} + \cdots + \theta_q e^{i w_q}\right]\right\},$$

where $w \in \mathbb{R}^q$. $i$ is an imaginary unit. Let $w_1 = \cdots = w_{k-1} = w_{k+1} = \cdots = w_q = 0$, and we can derive the characteristic function of $X_k$,

$$\varphi_{X_k}(w_k) = \varphi_X(0, \cdots, 0, w_k, 0, \cdots, 0) = \exp\left\{-(-\theta_1 + \theta_2 + \cdots + \theta_q) + (\theta_1 + \theta_2 + \cdots + \theta_q)e^{iw_k}\right\}.$$ 

Each $X_k$ follows a Poisson distribution with mean $(\theta_1 + \cdots + \theta_k)$. After a series of derivations, we can get the covariance matrix. The covariance matrix of $X_1, X_2, \ldots, X_q$ has diagonal elements, $\text{Var}\left(X_i\right)$, and off-diagonal elements, $\text{Cov}\left(X_j, X_k\right), j \neq k$. Elements of the covariance matrix are

$$\text{Cov}\left(X_j, X_k\right) = \text{Var}\left(X_j\right) = \theta_1 + \cdots + \theta_j, 1 \leq j \leq q,$$

$$\text{Cov}\left(X_j, X_k\right) = \theta_1 + \cdots + \theta_j, 1 \leq j < k \leq q.$$ 

Thus the covariance matrix of $X = (X_1, X_2, \ldots, X_q)'$ can be written as

$$\text{Cov}(X) = \begin{pmatrix}
\theta_1 & \theta_1 + \theta_2 & \cdots & \theta_1 + \theta_2 + \cdots + \theta_q \\
\theta_2 & \theta_1 + \theta_2 & \cdots & \theta_1 + \theta_2 + \cdots + \theta_q \\
\vdots & \vdots & \ddots & \vdots \\
\theta_1 & \theta_1 + \theta_2 & \cdots & \theta_1 + \theta_2 + \cdots + \theta_q
\end{pmatrix}.$$ 

3. Performance

The Control charts are mainly used to help the managers monitor and judge whether the production process is IC or OC. Traditionally, the average number of sample groups taken from the control chart from the beginning of detection to the warning of a problem in the process, namely ARL, which is often used to evaluate the effectiveness of a control chart. Provided the nominated IC ARL (ARL0), the OC ARL is smaller that means the control chart is more effective.

In this section, we will provide two tables of ARL values of the MEWMA control chart with different parameters shift. The performance of the proposed control chart is evaluated by the ARL, given a parameter shift occurred. Here we consider the case of a four-variate Poisson variable($q = 4$), the parameters of the Poisson distribution $\theta_1, \theta_2, \theta_3, \theta_4$ are all 3. And a smoothing parameter $r = 0.1$ is used. The shift size is calculated according to the non-centrality parameter formula mentioned earlier. Each MEWMA ARL value in our simulation is obtained by using 30000 simulations.

As for the determination of control limit $H$, the method we choose is Monte Carlo simulation, and the idea of the binary segmentation approach is applied in it. The specific steps are as follows:

1. Set the initial parameter values $\theta_1, \theta_2, \theta_3, \theta_4$ to 3.
2. The target value IC ARL0 is fixed at 200, and set the lower bounds HL and upper bounds HU of the initial control range as 0 and 100 respectively.
3. Calculate the length of ARL when HM is equal to $(H_U + H_L)/2$.
4. If the calculated length of ARL is equal to the preset ARL0, or the difference between ARL and ARL0 is less than 1, then HM is the required control limit $H$, and the whole process ends.
5. If the resulting length of ARL is greater than the preset length of ARL0, then $H_U = H_M$. If not, set $H_L = H_M$, and go back to step 3.
In step 3, the ARL of the control chart can be obtained by the Monte Carlo simulation or the Markov chain method. In our study, we choose the Monte Carlo simulation method. When the fixed IC ARL0 is 200, the control limit $H$ is 12.84 by simulation calculation. And we obtain the actual IC ARL0 value 199.293 as well.

For the OC case, as the four parameters $\theta_1, \theta_2, \theta_3, \theta_4$ are all related to the mean value, we consider the case that the number of parameters with shift is 1, 2, 3, 4 respectively. There are 15 different cases in total. We use the OC mean vector $\mu$ after shift to subtract the IC mean vector $\mu_0$, and then calculate the corresponding $\delta$ value according to the Mahalanobis distance formula. In our study, the covariance matrix of the four-variate Poisson variable refers to IC state. Then we generated a set of data by setting the values of control limits $H$ and $\delta$. And changing the mean value of the corresponding position of random variable $X$. Similarly, we calculated the OC ARL value after 30,000 simulations. In addition, we discuss two different cases of parameter shift with one unit and parameter shift with two units respectively. We list these simulation results in Table 1 and Table 2.

Table 1. ARL values for MEWMA control chart with one unit parameter shift ($q = 4$).

| No. | Number of parameter shift | Parameter shift $\theta_1, \theta_2, \theta_3, \theta_4$ | Shift size $\delta$ | ARL       |
|-----|----------------------------|-----------------------------------------------|-------------------|-----------|
| 1   | 0                          | 3, 3, 3, 3                                  | 0                 | 199.293   |
| 2   | 1                          | 4, 3, 3, 3                                  | 0.577             | 56.763    |
| 3   | 1                          | 3, 4, 3, 3                                  | 0.577             | 56.962    |
| 4   | 1                          | 3, 3, 4, 3                                  | 0.577             | 57.308    |
| 5   | 1                          | 3, 3, 3, 4                                  | 0.577             | 57.289    |
| 6   | 2                          | 4, 4, 3, 3                                  | 0.816             | 20.491    |
| 7   | 2                          | 4, 3, 4, 3                                  | 0.816             | 20.541    |
| 8   | 2                          | 4, 3, 4, 4                                  | 0.816             | 20.611    |
| 9   | 2                          | 3, 4, 4, 3                                  | 0.816             | 20.634    |
| 10  | 2                          | 3, 4, 3, 4                                  | 0.816             | 20.626    |
| 11  | 2                          | 3, 3, 4, 4                                  | 0.816             | 20.529    |
| 12  | 3                          | 4, 4, 4, 3                                  | 1                 | 11.539    |
| 13  | 3                          | 4, 4, 3, 4                                  | 1                 | 11.501    |
| 14  | 3                          | 4, 3, 4, 4                                  | 1                 | 11.501    |
| 15  | 3                          | 3, 4, 4, 4                                  | 1                 | 11.578    |
| 16  | 4                          | 4, 4, 4, 4                                  | 1.155             | 7.952     |

Table 2. ARL values for MEWMA control chart with two units parameter shift ($q = 4$).

| No. | Number of parameter shift | Parameter shift $\theta_1, \theta_2, \theta_3, \theta_4$ | Shift size $\delta$ | ARL       |
|-----|----------------------------|-----------------------------------------------|-------------------|-----------|
| 1   | 0                          | 3, 3, 3, 3                                  | 0                 | 199.293   |
| No. | Unit Change | Case Mean | IC ARL | OC ARL |
|-----|-------------|-----------|--------|--------|
| 1   | 1           | (3, 3, 3, 3) | 1.155  | 8.403  |
| 2   | 1           | (3, 3, 3, 3) | 1.155  | 8.438  |
| 3   | 1           | (3, 3, 3, 3) | 1.155  | 8.459  |
| 4   | 1           | (3, 3, 3, 3) | 1.155  | 8.445  |
| 5   | 2           | (5, 5, 3, 3) | 1.633  | 3.850  |
| 6   | 2           | (5, 5, 3, 3) | 1.633  | 3.844  |
| 7   | 2           | (5, 5, 3, 3) | 1.633  | 3.839  |
| 8   | 2           | (5, 5, 3, 3) | 1.633  | 3.835  |
| 9   | 2           | (3, 5, 5, 3) | 1.633  | 3.841  |
| 10  | 2           | (3, 5, 5, 3) | 1.633  | 3.834  |
| 11  | 2           | (3, 5, 5, 3) | 1.633  | 3.838  |
| 12  | 3           | (5, 5, 5, 3) | 2       | 2.579  |
| 13  | 3           | (5, 5, 5, 3) | 2       | 2.578  |
| 14  | 3           | (5, 3, 5, 5) | 2       | 2.589  |
| 15  | 3           | (3, 5, 5, 5) | 2       | 2.589  |
| 16  | 4           | (5, 5, 5, 5) | 2.309  | 2.027  |

Figure 1. ARL curve for the MEWMA control chart with one unit parameter shift.

Figure 2. ARL curve for the MEWMA control chart with two unit parameter shift.

The rows of No.1 in Table 1 and Table 2 are under the IC state, while the rest of No.2 to No.16 are under the OC state. Table 1 displays the cases of parameter shift with one unit. To make it easier to understand, here are two examples. Firstly, when a parameter shift with one unit, such as the value of $\theta_3$ is changed from 3 to 4, then the IC mean vector $\mu_0' = (3, 6, 9, 12)$, the mean vector $\mu' = (4, 7, 10, 13)$ after a shift has occurred, so that mean vector shift $(\mu - \mu_0)' = (1, 1, 1, 1)$. In this case, $\delta = 0.577$, the OC ARL is 56.763. In addition, we consider the case where two parameters shift with one unit, such as the values of $\theta_3$ and $\theta_4$ are both changed from 3 to 4, then the IC mean vector $\mu_0' = (3, 6, 9, 12)$, the mean vector $\mu' = (3, 6, 10, 14)$ after a shift has occurred, so that mean vector shift $(\mu - \mu_0)' = (0, 0, 1, 2)$. Under this situation, we obtained the shift size $\delta$ is 0.816, the OC ARL is 20.529. Notice that ARL values for MEWMA control chart under two units parameter shift are given in Table 2. When three parameters shift with two unit, such as the values of $\theta_3$, $\theta_2$ and $\theta_3$ are changed...
from 3 to 5, then the IC mean vector $\mu_0^\prime = (3, 6, 9, 12)$, the mean vector $\mu^\prime = (5, 10, 15, 18)$ after a shift has occurred, so that mean vector shift $(\mu - \mu_0)^\prime = (2, 4, 6, 6)$. In this case, $\delta = 2$, the OC ARL is 2.579. For the case where two unit shift occurs for all four parameters, $\mu_0^\prime = (3, 6, 9, 12)$, and $\mu^\prime = (5, 10, 15, 20)$, then $(\mu - \mu_0)^\prime = (2, 4, 6, 8)$, the shift size $\delta$ and the OC ARL are calculated to be 2.309 and 2.027 respectively.

By observing Table 1 and Table 2, we can see that the OC ARL of the control chart we proposed is far less than the IC ARL value, which indicates that this control chart can be used to detect the shift of parameters. When the number of parameter shift is the same, their shift size $\delta$ value is the same, and their OC ARL value is very close. With the increase of shift size $\delta$ value, the value of OC ARL becomes smaller and smaller, and the monitoring speed becomes faster and more effective.

According to the shift size and ARL values in Table 1, we draw ARL curve for the MEWMA control chart in Figure 1, in which we calculate the average value of several ARL with the same shift size. Similarly, Figure 2 is drawn based on the data in Table 2. From Figure 1 and Figure 2, we can intuitively see that the larger the shift value, the smaller the OC ARL value for the most cases.

4. Conclusions
In this paper, we study the control chart method of count data follow Poisson distribution. We combine a Poisson model with MEWMA control chart to construct a new control chart, which extends the research on monitoring multivariate discrete counting data to a certain extent. We also illustrate the performance of the proposed control chart by presenting two ARL tables.

Here, we only give the results when the dimension of the variable is 4 and the four parameters $\theta_1, \theta_2, \theta_3, \theta_4$ related to the mean are all 3. We can make further research on the results of other different cases. In addition, we can also compare the performance of our proposed control charts with other control charts in the later work.

Acknowledgments
This paper was supported by the NNSF of China, grants No. 11771332 and 11771220, Youth Program of Tianjin Natural Science Foundation, Grant No. 18JCQNJC70100.

References
[1] He, D.J. (2006). Multivariate Poisson Distribution and its Properties. Journal of Anhui Normal University (Natural Science), 29: 122-126.
[2] Liang, X.L., Li, E.Q. and Tian, M.Z. (2017). The parametric estimation and diagnostics of the multivariate generalized Poisson distribution. Systems Science and Mathematics, 37: 1319-1334.
[3] Patel, H.I. (1973). Quality control methods for multivariate binomial and poisson distributions. Technometrics, 15: 103-112.
[4] Chiu, J.E. and Kuo, T.I. (2007). Attribute control chart for multivariate Poisson distribution. Communications in Statistics – Theory and Methods, 37: 146-158.
[5] Zhang, G.Y. (2000). Another form of multivariate $\Gamma$ distribution. Appl. Math. J. Chinese Univ. Ser. A, 15: 65-71.
[6] Laungrungrong, B., Borror, C.M. and Montgomery, D.C. (2011). EWMA control charts for multivariate Poisson-distributed data. International Journal of Quality Engineering and Technology, 2:185-211.
[7] Lowry, C.A., Woodall, W.H., Champ, C.W. and Rigdon, S.E. (1992). A multivariate exponential weighted moving average control chart. Technometrics, 34: 46-53.
[8] Roberts S.W. (1959). Control Chart Tests Based on Geometric Moving Averages. Technometrics, 1: 239-250.
[9] Lucas, J.M. and Saccucci, M.S. (1990). Exponentially Weighted Moving Average Control Schemes: Properties and Enhancements. Technometrics, 32: 1-12.