Multivariate Binomial/Multinomial Control Chart

Jian Li[1], Fugee Tsung[1], and Changliang Zou[2*]

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

[2] *LPMC and Department of Statistics, School of Mathematical Sciences,*
Nankai University, Tianjin, China

Abstract

This paper considers statistical process control for multivariate categorical processes. In particular, we focus on multivariate binomial and multivariate multinomial processes. More and more real applications involve categorical quality characteristics, which cannot be measured on a continuous scale. These characteristic factors usually correlate with each other, indicating a must for multivariate charting techniques. However, there is a scarcity of research on monitoring multivariate categorical data, and most existing methods lack robustness for some deficiencies. We employ log-linear models for characterizing the relationship among categorical factors, which are adapted into a framework of multivariate binomial and multivariate multinomial distributions. A Phase II control chart is proposed, which is robust to detect various shifts efficiently, especially those in interaction effects representing the dependence among factors. Numerical simulations and a real-data example demonstrate the effectiveness of the chart.

Keywords: Contingency table, EWMA, log-linear model, statistical process control

[*] Corresponding author. Email: chlzou@yahoo.com.cn

1 Introduction

Statistical process control (SPC) has been demonstrated to be an important tool for monitoring process or product quality in manufacturing and service industries. In many applications, quality characteristics that are measured by attribute levels have become increasingly common. This is because having their continuous values is usually expensive or even infeasible. In addition, the attribute levels are often sufficient to represent their values, and such rough measurements do not cost much. For instance, on a production line each item is inspected and classified as conforming or nonconforming to its predefined quality specification. Similarly, a service level can be assessed as excellent, acceptable, or unacceptable. Since the characteristic variables or factors involved usually have two or more attribute levels, they are categorical rather than continuous.

The well known charts for monitoring categorical processes are the p-chart and the np-chart for binomial distributed variables as well as the c-chart and the u-chart for Poisson distributed processes. See Woodall (1997) for a detailed review. Recent results include Huang *et al.* (2012) for monitoring a binomial process using a moving window. These control charts provide a seemingly satisfactory framework for identifying the existence of assignable causes in categorical processes. However, they highlight out-of-control (OC) signals by taking only one quality characteristic into account. In situations where the overall quality of a product or service needs to be evaluated by simultaneously checking multiple correlated variables, multivariate control charts must be considered.

There are at least two reasons for introducing a multivariate control procedure. First, monitoring several variables independently requires a multi-chart to be handled in parallel, where each separate chart has a statistic to be updated and plotted from sample to sample (Woodall and Ncube, 1985). The control limits of the separate charts composing the multi-chart must be chosen, so that they each have a specified individual in-control (IC) average run length (ARL) and jointly achieve a specified overall IC ARL of the multi-chart. Here, the ARL is the average number of samples needed for a control chart to signal. However, the marginal distributions of variables are not necessarily identical. This may make determining the control limits by simulation quite complicated and infeasible even for a general multivariate categorical distribution. Second, multivariate control charts appropriately describe and exploit the correlations among multiple variables and therefore provide general tools for monitoring multivariate processes.

A large amount of literature has been dedicated to monitoring multivariate continuous variables. Refer to Lowry and Montgomery (1995) and Bersimis *et al.* (2007) for an overview. However, this is not the case for multivariate categorical processes, and to the best of our knowledge, available work on their SPC is fairly rare. Patel (1973) proposed a Hotelling's T^2 type χ^2-chart for multivariate binomial or multivariate Poisson populations, by the assumption that given a large sample size, the joint distribution of correlated binomial or Poisson variables can be approximated by a multivariate normal distribution. Lu *et al.* (1998) designed a Shewhart-type *mnp*-chart for monitoring multivariate attribute processes, which uses the weighted sum of the number of nonconforming units of each quality characteristic. In a very similar way to the *mnp*-chart, an *mp*-chart was proposed by Chiu and Kuo (2008) to monitor multivariate Poisson count data. It is noted that all the above charts can merely treat factors with two attribute levels. In Marcucci (1985), a generalized p-chart was developed for multinomial processes, which extends the p-chart from two attribute levels to three or more by adopting the Pearson chi-square statistic. However, it applies to only one factor. Similarly, there is also the multinomial cumulative sum (CUSUM) chart proposed by Ryan et al. (2011), which still applies to one factor based on the likelihood ratio statistic. Recent developed charting techniques for multinomial data also include Chen *et al.* (2011), Li *et al.* (2012), Weiss (2012), and Yashchin (2012). Topalidou and Psarakis (2009) provided a review of surveillance methods for multivariate attribute data, but except the aforementioned ones, most of them are not in the framework of SPC for multivariate categorical processes.

As discussed above, there is no appropriate method for the case of multiple factors, in which at least one has more than two attribute levels. In addition, most of the existing methods suffer from a drawback, in that they focus on only the attribute levels of each factor and neglect the cross-classifications among multiple factors. This makes these methods insensitive to some types of shifts such as changes in two-factor or higher-order interaction effects. There seems to be a severe lack of general methodologies for monitoring multivariate categorical processes.

In this paper, we study monitoring multivariate categorical processes in the systematic framework of multivariate binomial and multivariate multinomial processes, which include the cases of several factors all with two attribute levels and at least one with more than two levels. Here, the categorical data can be summarized in a multi-way contingency table and formulated in terms of log-linear models, which characterize the association patterns among categorical factors. Simply put, a log-linear model relates the logarithms of the expected cell counts in the contingency table to a linear model that is similar to an analysis of variance (ANOVA) model. A detailed discussion of log-linear models and their applications was given by Bishop *et al.* (2007). The introduction of log-linear models to SPC also appeared in Qiu (2008), which designed a distribution-free monitoring scheme for multivariate continuous processes by dichotomizing numerical data and applying log-linear models to the resultant binary data for estimating the IC categorical distribution.

Analogous to multi-way ANOVA, the cross-classification cell counts are determined by main effects and interaction effects. The main effect of one factor reflects mainly its marginal distribution, and the interaction effects of multiple factors represent the dependence among them, which play the same role as the correlation coefficients in multivariate normal distributions. For a log-linear model that has only main effects, no dependence exists among the involved factors, and they are independent of each other. Furthermore, log-linear models can be equivalently rewritten as a regression form, resulting in a one-to-one correspondence between factor effects and coefficient subvectors. Therefore, shifts to OC states, which arise in factor effects, equivalently result in deviations of coefficient subvectors. This provides a practical explanation of shifts in such processes. Based on log-linear models, a Phase II control chart is proposed, which utilizes the log-likelihood function of log-linear models and the exponentially weighted moving average (EWMA) control scheme. It incorporates properly the exponential weights used at different time points in the EWMA scheme into the log-likelihood function, leading to an exponentially weighted log-likelihood ratio testing statistic. This chart is fast to compute and convenient to use, and it applies to the unified framework of multivariate binomial and multivariate multinomial processes. Numerical results further confirm its effectiveness under various conditions, as well as its superiority over existing charts in the literature in detecting shifts in interaction effects that reflect dependence among factors. Therefore, it can be implemented reliably within a diverse range of scenarios.

The remainder of this paper is organized as follows. First, the log-linear modeling for multivariate binomial and multivariate multinomial processes is introduced in Section 2. Our proposed control chart is described in detail in Section 3. Its numerical performance is investigated in Section 4. In Section 5, we demonstrate the method using a manufacturing example, followed by several concluding remarks in Section 6. Some proofs are given in the appendix.

2 Multivariate Binomial/Multinomial Modeling

Statistical process control generally consists of two phases. In Phase I, a set of process data is gathered and analyzed. Any unusual patterns such as outliers or change-points in the data will lead to adjustments and fine tuning of the process. Once all such assignable causes are accounted for, we are left with a clean set of data, gathered under stable operating conditions. This dataset, referred to as the IC dataset hereafter, is then used for estimating the IC parameters of processes. In Phase II SPC, the estimated IC process parameters are used, and the major goal is to detect any changes after an unknown time point. The performance of a Phase II procedure is often measured by the ARL. In this section, we introduce multivariate binomial and multivariate multinomial processes and then their modeling based on log-linear models, which is the basis for proposing the control chart.

2.1 Multivariate Binomial/Multinomial Processes

In the context of multivariate categorical processes, suppose that there are p categorical variables or factors $\mathcal{C} = \{C_1, C_2, \ldots, C_p\}$, with each classification factor C_i taking a number, say h_i, of possible attribute levels. Consider all the cross-classifications among all the level combinations of these factors, which form a p-way $h_1 \times h_2 \times \ldots \times h_p$ cross-classified contingency table with $h = \prod_{i=1}^{p} h_i$ cells. Therefore, each cell corresponds to one level combination of the p factors. For instance, in taking measurements on wooden boxes sampled from a production line, the classification factors may be the quality characteristics of the length, the width, and the height, each assessed as conforming or nonconforming. Here we have $p = 3$ and $h_1 = h_2 = h_3 = 2$. Without loss of generality, for each factor, -1 and 1 are used to represent the two levels "conforming" and "nonconforming", respectively. Therefore, $2^3 = 8$ cells are arranged into a $2 \times 2 \times 2$ cube, and each cell count in this cube represents the count with a certain level combination. For example, $(-1,\ 1,\ -1)$ represents a box with conforming length and height and nonconforming width.

Now we turn to a general p-way contingency table of size $h_1 \times h_2 \times \ldots \times h_p$ (Johnson *et al.*, 1997). Let the probability of obtaining the combination of factor levels $a_1, a_2, \ldots, a_p$ be $p_{a_1 a_2 \ldots a_p}$ ($a_i = 1, \ldots, h_i$ and $i = 1, \ldots, p$). Furthermore, denote the count of observations with the level combination $a_1 a_2 \ldots a_p$ among a total sample of size N by $n_{a_1 a_2 \ldots a_p}$. The marginal counts for the ith factor, which are denoted by $n_{(i)1}, n_{(i)2}, \ldots, n_{(i)h_i}$ and calculated as

$$n_{(i)v} = \sum_{a_1} \sum_{a_2} \cdots \sum_{a_{i-1}} \sum_{a_{i+1}} \cdots \sum_{a_{p-1}} \sum_{a_p} n_{a_1 a_2 \ldots a_{i-1} v a_{i+1} \ldots a_{p-1} a_p}, \quad v = 1, \ldots, h_i,$$

will follow the multinomial distribution $\mathrm{MN}(N; p_{(i)1}, p_{(i)2}, \ldots, p_{(i)h_i})$, where

$$p_{(i)v} = \sum_{a_1} \sum_{a_2} \cdots \sum_{a_{i-1}} \sum_{a_{i+1}} \cdots \sum_{a_{p-1}} \sum_{a_p} p_{a_1 a_2 \ldots a_{i-1} v a_{i+1} \ldots a_{p-1} a_p}, \quad v = 1, \ldots, h_i.$$

Thus, consider the joint distribution of the p sets of variables

$$n_{(i)1}, n_{(i)2}, \ldots, n_{(i)h_i}, \quad i = 1, \ldots, p,$$

each being a multinomial distribution. This joint distribution is the multivariate multinomial distribution (Johnson *et al.*, 1997), $\mathrm{MMN}(N; \boldsymbol{\pi})$, where $\boldsymbol{\pi}$ is the h-variate vector of true cell probabilities. For instance, the p_{ijk} ($i = 1, \ldots, h_1$; $j = 1, \ldots, h_2$; $k = 1, \ldots, h_3$) in Equation (1) compose the cell probability vector $\boldsymbol{\pi}$ of size $h \times 1$ and $h = h_1 \times h_2 \times h_3$. When each factor has two levels, it reduces naturally to the multivariate binomial distribution. Therefore, based on the framework of multivariate binomial and multivariate multinomial distributions, multiple categorical variables can be studied.

2.2 Log-Linear Modeling

To characterize the relationship between each cell count and the factor levels determining it, the log-linear model can be employed to model the cell counts in the contingency table. Before illustrating this, we first turn to multi-way ANOVA, where responses to all factor level combinations are also in a multi-way table. The responses in ANOVA are assumed to have normal distributions and dependent on main factor effects and factor interaction effects. For a simple three-way ANOVA model with its three factors taking h_1, h_2, and h_3 levels, denote the expected response with the first factor at its ith level, the second factor at its jth level, and the third factor at its kth level as y_{ijk} ($i = 1, \ldots, h_1$; $j = 1, \ldots, h_2$; $k = 1, \ldots, h_3$). The three-way ANOVA model is

$$y_{ijk} = u^{(0)} + u_i^{(1)} + u_j^{(2)} + u_k^{(3)} + u_{i,j}^{(1,2)} + u_{i,k}^{(1,3)} + u_{j,k}^{(2,3)} + u_{i,j,k}^{(1,2,3)},$$

where $u^{(0)}$ is the overall mean, $u^{(1)}, u^{(2)}, u^{(3)}$ are the main effects, $u^{(1,2)}, u^{(1,3)}, u^{(2,3)}$ are the two-factor interaction effects, and $u^{(1,2,3)}$ is the three-factor interaction effect. Identifiability requires constraints such as

$$\sum_i u_i^{(1)} = \sum_i u_{i,j}^{(1,2)} = \sum_i u_{i,k}^{(1,3)} = \sum_i u_{i,j,k}^{(1,2,3)} = 0$$

for the first factor along its index i. Similar equations exist for the second and third factors along their indexes j and k, respectively.

We are now ready to consider log-linear models. The total number N of observations is usually fixed, for example, as a convention during the monitoring process in Phase II SPC. So the cell counts collected in a sample are reasonably assumed to follow a multinomial distribution. For simplicity, we take a three-way contingency table for illustration. For an $h_1 \times h_2 \times h_3$ table, denote the observed count by n_{ijk} in Cell(i, j, k) ($i = 1, \ldots, h_1$; $j = 1, \ldots, h_2$; $k = 1, \ldots, h_3$), the expected count by m_{ijk}. We assume here that the cell counts follow a multinomial distribution and have the probability mass function (PMF)

$$f(\{n_{ijk}\}) = \frac{N!}{\prod_{i,j,k} n_{ijk}!} \prod_{i,j,k} p_{ijk}^{n_{ijk}}, \tag{1}$$

where $p_{ijk} = m_{ijk}/N$ is the probability of a random observation being in Cell(i, j, k). Thus, there is a constraint $\sum_{i,j,k} p_{ijk} = 1$.

On the other hand, we know that the ANOVA model is a linear model with normally distributed responses. It has a canonical link function of unity from the view of generalized linear models (GLM). For Poisson distributed responses, the corresponding canonical link function in GLM is the logarithm (McCullagh and Nelder, 1989), which naturally leads to log-linear models. In addition, there is the fact that a series of independent Poisson distributed variables result in a multinomial distribution given the sum of these variables. Therefore, we can employ log-linear models to model the cell counts in a multi-way contingency table. For the above three-way contingency table, the log-linear model for describing the relationship between the expected cell count m_{ijk} and the factor levels indexed with i, j, k is (Bishop *et al.*, 2007)

$$\ln m_{ijk} = u^{(0)} + u_i^{(1)} + u_j^{(2)} + u_k^{(3)} + u_{i,j}^{(1,2)} + u_{i,k}^{(1,3)} + u_{j,k}^{(2,3)} + u_{i,j,k}^{(1,2,3)}. \tag{2}$$

For a log-linear model without any interaction effects, the factors involved are independent of each other. Therefore, dependence among factors is reflected by their interaction effects.

The log-linear model (2) and its identifiability constraints are somewhat wordy and inconvenient. However, it can be rewritten equivalently as a regression form, which is illustrated by a 2 × 3 contingency table. The identifiability constraints allow setting

$$\begin{array}{lll} u^{(0)} = \beta_0, & u_1^{(1)} = \beta_1, & u_2^{(1)} = -\beta_1, \\ u_1^{(2)} = \beta_2, & u_2^{(2)} = \beta_3, & u_3^{(2)} = -\beta_2 - \beta_3, \\ u_{1,1}^{(1,2)} = \beta_4, & u_{1,2}^{(1,2)} = \beta_5, & u_{1,3}^{(1,2)} = -\beta_4 - \beta_5, \\ u_{2,1}^{(1,2)} = -\beta_4, & u_{2,2}^{(1,2)} = -\beta_5, & u_{2,3}^{(1,2)} = \beta_4 + \beta_5. \end{array}$$

Therefore, the cell count expectation m_{ij} ($i = 1, 2$; $j = 1, 2, 3$) will be $\ln m_{ij} = \beta_0 + \sum_{k=1}^{5} \beta_k x_k$, and x_k ($k = 1, 2, \ldots, 5$) takes values on 1, 0, or -1 where appropriate. Obviously, β_1 measures the main effect $u^{(1)}$ of the first factor, $[\beta_2, \beta_3]^T$ measures the main effect $u^{(2)}$ of the second factor, and $[\beta_4, \beta_5]^T$ measures the interaction effect $u^{(1,2)}$ of the two factors. In a word, the factor effects can be represented explicitly by their corresponding coefficient subvectors, and this can be extended to a general case.

By imposing the identifiability constraints, the log-linear model (2) for a p-way contingency table can be expressed as the following regression form

$$\ln \mathbf{m} = \mathbf{1}\beta_0 + \sum_{i=1}^{2^p-1} \boldsymbol{X}_i \boldsymbol{\beta}_i, \tag{3}$$

where $\mathbf{m}$ is the expectation vector of size $h \times 1$, $\mathbf{1}$ is a column vector with appropriate dimensions and 1 as all its elements, $\boldsymbol{X}_i$ is the design submatrix corresponding to the ith main or interaction effect and of size $h \times q_i$ with elements 1, or 0, or -1 where appropriate, and $\boldsymbol{\beta}_i$ is the coefficient subvector of size $q_i \times 1$. In addition, a function such as the logarithm ln and the factorial ! applied to a vector means the vector of the function applied to each of the elements of this vector. Obviously, β_0 is a scalar representing the intercept. By denoting the design matrix by $\boldsymbol{X} = (\mathbf{1}, \boldsymbol{X}_1, \ldots, \boldsymbol{X}_{2^p-1})$ and the coefficient vector by $\boldsymbol{\beta} = (\beta_0, \boldsymbol{\beta}_1^T, \ldots, \boldsymbol{\beta}_{2^p-1}^T)^T$, model (3) can be rewritten as $\ln \mathbf{m} = \boldsymbol{X}\boldsymbol{\beta}$. The design matrix $\boldsymbol{X}$ can guarantee the identifiability conditions, and its derivation can be found in the additional File 1 of Dahinden *et al.* (2007). For convenience, we provide in Appendix A a list of some important notations that appear above for describing log-linear models.

The design submatrices and their corresponding coefficient subvectors are usually arranged in the sequence of the overall mean, the main effects, the two-factor interaction effects, and so on. Take three factors C_1, C_2, and C_3 with 2, 3, and 3 levels, respectively, for illustration. We follow the order of 1, C_1, C_2, C_3, C_1C_2, C_1C_3, C_2C_3, $C_1C_2C_3$. Hence, the coefficient vector is

$$\begin{array}{rclllllll} \boldsymbol{\beta} & = & [& \beta_0 & \beta_1 & \beta_{2_1} & \beta_{2_2} & \beta_{3_1} & \beta_{3_2} \\ & & & \beta_{1,2_1} & \beta_{1,2_2} & \beta_{1,3_1} & \beta_{1,3_2} & \beta_{2_1,3_1} & \beta_{2_1,3_2} \\ & & & \beta_{2_2,3_1} & \beta_{2_2,3_2} & \beta_{1,2_1,3_1} & \beta_{1,2_1,3_2} & \beta_{1,2_2,3_1} & \beta_{1,2_2,3_2} &]^T. \end{array}$$

We will use $\boldsymbol{\beta}$ arranged in similar sequences as above in later numerical simulations. Note that for the multinomial sampling with a fixed sample size N, given all the other entries in the coefficient vector $\boldsymbol{\beta}$, the first entry β_0 can be determined by N, which is guaranteed by the constraint that the cell probabilities in the contingency table sum up to one. Therefore, attention may be paid to only the coefficient subvectors $\boldsymbol{\beta}_i$ ($i = 1, \ldots, 2^p - 1$). According to the above sequence, it is clear that, for example, $\boldsymbol{\beta}_3 = [\beta_{3_1}, \beta_{3_2}]^T$ measures the main effect of the factor C_3, and $\boldsymbol{\beta}_6 = [\beta_{2_1,3_1}, \beta_{2_1,3_2}, \beta_{2_2,3_1}, \beta_{2_2,3_2}]^T$ measures the two-factor interaction effect of C_2 and C_3. In summary, there is a one-to-one correspondence among the ith main or interaction effect, the design submatrix $\boldsymbol{X}_i$, and the coefficient subvector $\boldsymbol{\beta}_i$ ($i = 1, \ldots, 2^p - 1$). Therefore, the cell count expectation vector is essentially determined by the magnitudes of these coefficient subvectors.

A log-linear model for a p-way contingency table is saturated if it involves all the effects of all orders from 1 up to p. In other words, from the main effects to the p-factor interaction effect, they are all included in the log-linear model. See model (2) for a saturated log-linear model example. Usually, a reduced model can be obtained via variable selection. It means that some effects in the saturated model can be dropped while others are retained. Therefore, some of the coefficient subvectors $\boldsymbol{\beta}_i$ together with their corresponding design submatrices $\boldsymbol{X}_i$ in model (3) may be dropped.

For the reduced models, the hierarchy principle should be followed. It means that all the lower-order effects have to be included if a higher-order interaction effect containing them appears in the model. For instance, in a three-way contingency table, if the effect $u^{(2,3)}$ is in the model, $u^{(2)}$ and $u^{(3)}$ should also be adopted. The hierarchical models can be denoted in a simple way, such as [13][23] representing the model without the effects $u^{(1,2)}$ or $u^{(1,2,3)}$ as well as [123] representing the saturated model (2). In the log-linear model, the design matrix $\boldsymbol{X}$ may actually be constructed in other ways. If the hierarchical log-linear model is reparametrized using a different design matrix, all the zero terms in the coefficient vector still remain zero. However, this cannot apply to non-hierarchical ones. In other words, hierarchy is preserved after reparametrization, and all the zero coefficients can be interpreted in terms of conditional independence (Dahinden *et al.*, 2007). This is the main advantage of hierarchical models over non-hierarchical ones.

3 Multivariate Binomial/Multinomial Monitoring

3.1 Online Detection Problem

This paper focuses on Phase II monitoring only and presumes that the coefficient vector $\boldsymbol{\beta}$ in the log-linear model (3) has already been estimated from an IC dataset by variable selection and parameter estimation. Refer to Christensen (1997) to see the stepwise procedures for variable selection and the Newton-Raphson iterative algorithm for parameter estimation. In addition, it should be noted that all of the historical observations used for estimating the IC model are i.i.d.. However, in practical applications there is no such assurance. So it requires much future research to extend our method to Phase I analysis, in which detecting unusual patterns in a historical dataset would be of interest.

Based on the IC coefficient vector in the log-linear model, we present in this section a Phase II monitoring scheme, which is able to detect various shifts among the multiple categorical variables in the systematic framework of multivariate binomial and multivariate multinomial distributions. For simplicity, let $F(\boldsymbol{X};\boldsymbol{\beta})$ represent the pre-specified log-linear model

$$\ln \mathbf{m} = \boldsymbol{X}\boldsymbol{\beta} \quad \text{and} \quad \mathbf{1}^T\mathbf{m} = N, \tag{4}$$

where $\boldsymbol{X}$ is an $h \times s$ matrix with rank s, and N is the sample size in Phase II. It is usually reasonable to assume that the jth online multivariate sampling observation vector, $\mathbf{n}_j$ of size $h \times 1$, is collected over time from the following change-point model

$$\mathbf{n}_j \overset{\text{i.i.d.}}{\sim} \begin{cases} F(\boldsymbol{X};\boldsymbol{\beta}^{(0)}), & \text{for} \quad j = 1, \ldots, \tau, \\ F(\boldsymbol{X};\boldsymbol{\beta}^{(1)}), & \text{for} \quad j = \tau + 1, \ldots, \end{cases}$$

where τ is the unknown change-point, and $\boldsymbol{\beta}^{(0)} \neq \boldsymbol{\beta}^{(1)}$ are the known IC and unknown OC process coefficient vectors, respectively. Thus, the monitoring problem is closely related to the goodness-of-fit test in the context of multinomial analysis (Bishop *et al.*, 2007). Moreover, the coefficient vector $\boldsymbol{\beta}$ summarizes the relationship between the response $\mathbf{m}$ and the explanatory variable $\boldsymbol{X}$, which is essentially a profile. Therefore, model (4) has certain links with the parametric profile monitoring, in which checking the stability of a linear or nonlinear regression model over time is of interest (Zou *et al.*, 2007).

There is a one-to-one correspondence between factor effects and coefficient subvectors. Therefore, shifts of marginal distributions of factors or dependence among multiple factors to OC states, which appear in the form of deviations of their main effects or interaction effects, respectively, are reflected in the changes of the corresponding coefficient subvectors. As a result, the monitoring task is to test if $\boldsymbol{\beta} = \boldsymbol{\beta}^{(0)}$. Based on the likelihood ratio test (LRT) (Christensen, 1997), a naive method that comes to mind for online detection is to use the current sampling observation vector to construct a Shewhart-type chart. However, this would be very inefficient for moderate and small changes, since it completely ignores the past samples. As an alternative, we may consider an EWMA scheme. A natural idea is to first obtain the estimate of the coefficient vector $\boldsymbol{\beta}$ for each sample, and then apply the multivariate EWMA chart (Lowry *et al.*, 1992) and some modifications of the LRT to these estimates of $\boldsymbol{\beta}$ at different time points. However, this naive approach may still be inefficient, since for each sample the coefficient vector $\boldsymbol{\beta}$ is estimated based on only N random observations.

3.2 Log-Linear Multivariate Binomial/Multinomial Chart

Based on the above analysis, we propose an EWMA-type control chart by using the idea of weighted likelihood. The log-likelihood of the observation vector $\mathbf{n}_j$ in the jth sample of size N in Phase II can be written from the PMF of the multinomial distribution and expressed as

$$\begin{aligned} l_j(\boldsymbol{\beta}) = & \sum_{a_1,a_2,\ldots,a_p} n_{a_1a_2\ldots a_p,j} \ln m_{a_1a_2\ldots a_p} - \sum_{a_1,a_2,\ldots,a_p} n_{a_1a_2\ldots a_p,j} \ln N \\ & + \ln(N!) - \sum_{a_1,a_2,\ldots,a_p} \ln(n_{a_1a_2\ldots a_p,j}!) \\ = & \mathbf{n}_j^T \ln \mathbf{m} - \mathbf{n}_j^T \mathbf{1} \ln N + \ln(N!) - \mathbf{1}^T \ln(\mathbf{n}_j!) \\ = & \mathbf{n}_j^T \boldsymbol{X\beta} - N \ln N + \ln(N!) - \mathbf{1}^T \ln(\mathbf{n}_j!). \end{aligned}$$

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

$$w_k(\boldsymbol{\beta};\lambda) = a_{0,k,\lambda}^{-1} \sum_{j=1}^{k} (1-\lambda)^{k-j} l_j(\boldsymbol{\beta}),$$

where $\lambda \in (0,1]$ is a smoothing parameter, and $a_{t_0,t_1,\lambda} = \sum_{j=t_0+1}^{t_1}(1-\lambda)^{t_1-j}$ is a sequence of constants to ensure that all the weights sum up to 1. Obviously, $w_k(\boldsymbol{\beta};\lambda)$ makes full use of all available samples up to the current time point k, and different samples are weighted as in an EWMA chart (i.e., the more recent sample has more weight, and the weight changes exponentially over time). Then given λ, the maximum weighted likelihood estimate of $\boldsymbol{\beta}$ at the time point k, $\widehat{\boldsymbol{\beta}}_k$, is defined as the solution to the following maximization problem,

$$\widehat{\boldsymbol{\beta}}_k = \arg\max_{\boldsymbol{\beta}} w_k(\boldsymbol{\beta};\lambda), \ \text{ subject to } \mathbf{1}^T(\exp(\boldsymbol{X\beta})) = N. \tag{5}$$

Therefore, the weighted likelihoods under the alternative and null hypotheses, evaluated at $\widehat{\boldsymbol{\beta}}_k$ and $\boldsymbol{\beta}^{(0)}$, respectively, are given by

$$w_{1,k} = a_{0,k,\lambda}^{-1} \sum_{j=1}^{k} (1-\lambda)^{k-j} l_j(\widehat{\boldsymbol{\beta}}_k),$$
$$w_{0,k} = a_{0,k,\lambda}^{-1} \sum_{j=1}^{k} (1-\lambda)^{k-j} l_j(\boldsymbol{\beta}^{(0)}).$$

Consequently, the -2LRT statistic will be

$$\begin{aligned} R_k &= 2(w_{1,k} - w_{0,k}) \\ &= 2(\mathbf{z}_k^T(\boldsymbol{X}\widehat{\boldsymbol{\beta}}_k - \boldsymbol{X}\boldsymbol{\beta}^{(0)}), \end{aligned} \tag{6}$$

where

$$\mathbf{z}_k = a_{0,k,\lambda}^{-1} \sum_{j=1}^{k} (1-\lambda)^{k-j} \mathbf{n}_j. \tag{7}$$

Clearly, $\mathbf{z}_k$ is the exponentially weighted average of the observation vectors $\mathbf{n}_j$ $(j = 1, \ldots, k)$ over time. Before proceeding, we need to calculate the weighted likelihood $w_{1,k}$ in which $\mathbf{z}_k^T\boldsymbol{X}\widehat{\boldsymbol{\beta}}_k$ is involved. Although computing power has greatly improved, and it is computationally trivial to perform log-linear model estimation for individual sampling observation vectors, for online process monitoring which generally handles a large amount of samples, fast implementation is important, and some computational issues deserve our careful examination. At first glance, to obtain $\widehat{\boldsymbol{\beta}}_k$ by solving the maximization problem (5) directly requires a considerable amount of computing time especially when k is large. However, notice that the core of the exponentially weighted log-likelihood $w_k(\boldsymbol{\beta}; \lambda)$ over samples 1 to k, which contains $\boldsymbol{\beta}$, is $a_{0,k,\lambda}^{-1} \sum_{j=1}^{k} (1-\lambda)^{k-j} \mathbf{n}_j \boldsymbol{X}\boldsymbol{\beta}$, and that the log-likelihood of $\mathbf{z}_k$ is $\mathbf{z}_k\boldsymbol{X}\boldsymbol{\beta}$ by ignoring some constants. According to Equation (7), these two parts are equal, which concludes the following proposition.

Proposition 1 *The $\widehat{\boldsymbol{\beta}}_k$ is the maximum likelihood estimation (MLE) of the log-linear model (4) with a pseudo-observation vector $\mathbf{z}_k$ in Equation (7).*

We call $\mathbf{z}_k$ the pseudo-observation vector, because its components are not integers so that it cannot be observed in practice. Proposition 1 provides an easy way to evaluate $w_{1,k}$.

Furthermore, since $\mathbf{1}^T\mathbf{z}_k = N$ and the constraint in Equation (5) is identical to that in the log-linear model (4), the maximization in Equation (5) is exactly the same as solving the MLE for model (4) with $\mathbf{z}_k$. Therefore, we can rewrite R_k in Equation (6) as

$$R_k = 2\mathbf{z}_k^T(\ln \widehat{\mathbf{y}}_k - \ln \mathbf{m}^{(0)}), \tag{8}$$

where $\widehat{\mathbf{y}}_k$ is the MLE of the cell count expectation vector over $\mathbf{z}_k$, and $\mathbf{m}^{(0)}$ is the cell count expectation vector in the IC state. Note that the expectation vector $\widehat{\mathbf{y}}_k$ in Equation (8) can be obtained by performing the iterative proportional fitting (IPF) algorithm (Bishop *et al.*, 2007), which is efficient in calculating the MLE of cell count expectations under any hierarchical log-linear models and is included in major statistical softwares such as the subroutine "PRPFT" in Fortran with the IMSL library.

By using Equation (8), a large value of R_k rejects the null hypothesis, and hence our proposed chart triggers an OC signal if

$$R_k > L, \text{ for } k \geq 1,$$

where $L > 0$ is a control limit chosen to achieve a specific IC ARL, denoted by ARL_0. Hereafter, this chart is referred to as the log-linear multivariate binomial/multinomial (LMBM) control chart. The control limit L can be searched by simulation based on the IC model. For a given λ, model $F(\boldsymbol{X}; \boldsymbol{\beta})$, and a desired ARL_0, the computation involved in finding L is not difficult, partly due to the fact that the IPF algorithm used in the MLE computation is efficient. For the searching procedure, some numerical searching algorithms, such as the bisection search, can be applied. For instance, when ARL_0=370, it requires about 30 minutes to complete the bisection searching procedure based on 10,000 simulations when $h = 32$, $s = 21$, and $N = 1,000$ using a Pentium 3.0GHz CPU. The Fortran codes for implementing the proposed procedure are available from the authors upon request.

We now discuss the diagnostic issue of multivariate binomial and multivariate multinomial processes. According to the one-to-one correspondence between factor effects and coefficient subvectors in a log-linear model, shifts in the marginal distribution of one factor lead to deviations of the coefficient subvector corresponding to its main effect, and shifts in the dependence among multiple factors result in deviations of the coefficient subvector reflecting their interaction effect. Possible diagnostic procedures should focus on separating the shifted coefficient subvector once an OC signal is triggered. The least absolute shrinkage and selection operator (LASSO) could be employed to identify this. In particular, Zou *et al.* (2011) provided a general framework for diagnosis, and the log-linear model can be formulated to adapt this framework.

Finally, some asymptotic properties of the charting statistic R_k are given, which can be used theoretically to study the property of the charting statistic, justify the performance of the LMBM chart to a certain degree, and therefore shed some light on the practical design of this chart as well. Theorem 1 below gives the asymptotic behavior of R_k under both the IC and OC models, and its proof is provided in Appendix B. Note that this is an asymptotic result, so it cannot be used in simulations to determine the control limit nor in real examples. Denote $b_{t_0,t_1,\lambda} = \sum_{i=t_0+1}^{t_1}(1-\lambda)^{2(t_1-i)}$ and $c_{t_0,t_1,\lambda} = a^2_{t_0,t_1,\lambda}/b_{t_0,t_1,\lambda}$. Suppose that $\boldsymbol{\pi}^{(0)}$ is the true IC cell probability vector in the process, and when the process is OC, the true cell probability vector is $\boldsymbol{\pi}^{(1)} = \boldsymbol{\pi}^{(0)} + \boldsymbol{\mu}(c_{0,k,\lambda}N)^{-1/2}$.

Theorem 1 *As $N \to \infty$, or $\lambda \to 0$ and $k \to \infty$, we have*

(i) When the process is IC, $c_{0,k,\lambda}R_k \xrightarrow{\mathcal{L}} \chi^2_{s-1}$.

(ii) When the process is OC, $c_{0,k,\lambda}R_k \xrightarrow{\mathcal{L}} \chi^2_{s-1}(\boldsymbol{\mu}^T\mathbf{D}^{-1}_{\boldsymbol{\pi}^{(0)}}\boldsymbol{\mu})$, where $\boldsymbol{\mu}^T\mathbf{D}^{-1}_{\boldsymbol{\pi}^{(0)}}\boldsymbol{\mu}$ is the noncentrality parameter, and $\mathbf{D}_{\boldsymbol{\pi}^{(0)}}$ is an $h \times h$ diagonal matrix with $\boldsymbol{\pi}^{(0)}$ on its diagonal.

Remark 1. If some cells in the contingency table have very small probabilities or the sample size is not large enough, there will be no observations in these cells, and they have zero counts and are noninformative. This is the sparsity phenomenon, which may lead to the nonexistence of MLEs (Fienberg and Rinaldo, 2007). For each observation vector $\mathbf{n}_j$, it is possible that some of its entries have zero counts. However, the locations of these zero cells may differ from sample to sample. By combining the collected samples in an exponentially weighted way, the probability that the pseudo-observation vector $\mathbf{z}_k$ still has exact zero cells is low, especially when k increases as the process proceeds. Therefore, the sparsity is mitigated to a large extent or even eliminated by the EWMA scheme, and the nonexistence of MLEs is avoided.

Remark 2. The IPF method proportionally adjusts the cell counts of a sample to fit a set of margins (e.g., the factor association or interaction structure). Once given the hierarchical structure of log-linear models, this successive proportional adjustment can obtain directly the MLE of the cell count expectation vector $\mathbf{m}$ by skipping the estimation of the coefficient vector $\boldsymbol{\beta}$. In addition, it has the following properties (Biship *et al.*, 2007): (1) converging to the required unique set of MLE; (2) a stopping rule that ensures accuracy to any desired degree in the elementary cell estimates; (3) depending only on the sufficient configurations and no special provision for sporadic cells with no observations (this relates to the sparsity phenomenon above); (4) any set of starting values; (5) yielding the exact estimates in one cycle if direct estimates exist. The IPF algorithm possesses these advantages over the earlier techniques proposed, such as the Newton-Raphson which fails to have properties (3) and (5) above. Therefore, we adopt the IPF for online estimation in Phase II.

4 Simulation Studies

In this section, the performance of the proposed LMBM chart is investigated through some numerical simulations. The simulations are made in scenarios of both multivariate binomial and multivariate multinomial processes. In both cases, the LMBM chart is compared with its counterparts in various cases of shifts, and the advantages of the LMBM chart over others are verified. Throughout the simulation, for fair comparison, the IC ARL is fixed at 370 for each control chart, and all ARL values reported are averages of 10,000 replicated simulations. If the process is OC, a smaller OC ARL of a chart means that it gives rise to an OC signal faster, and that this chart therefore performs better.

4.1 Monitoring a Multivariate Binomial Process

For monitoring multivariate binomial processes, a natural competitive method is Patel's (1973) Hotelling's T^2 type χ^2-chart. However, since the χ^2-chart is the Shewhart-type, and the LMBM chart is the EWMA-type, the comparison may not be fair to the Shewhart-type as we could expect its deficiency in detecting small and moderate shifts due to the fact that it completely ignores the past information. Hence, the EWMA version of the χ^2-chart, which is a typical competitor of the LMBM chart for monitoring the multivariate binomial processes, is developed accordingly for fair comparison.

In this context, only factors with two levels, hence a p-way contingency table with 2^p cells for p factors, are considered. Without loss of generality, we denote the two levels by 1 and 0. For the ith factor, suppose that in the IC state $P(1) = p_{(i)}^{(0)}$ and $P(0) = 1 - p_{(i)}^{(0)}$ $(i = 1, \dots, p)$. Given the sample size N in Phase II, if the process is IC, the Level 1 count $n_{(i)}$ of the ith factor is subject to a binomial distribution $\mathrm{BN}\big(N; p_{(i)}^{(0)}\big)$. Therefore, the joint distribution of these p binomial distributions is a multivariate binomial distribution. By assuming that the IC cell probability combination $p_{a_1 a_2 \dots a_p}^{(0)}$ $(a_i = 0$ or 1, $i = 1, \dots, p)$ is known or has been estimated from an IC dataset, we have

$$p_{(i)}^{(0)} = \sum_{a_1}\sum_{a_2} \cdots \sum_{a_{i-1}}\sum_{a_{i+1}} \cdots \sum_{a_{p-1}}\sum_{a_p} p_{a_1 a_2 \dots a_{i-1} 1 a_{i+1} \dots a_{p-1} a_p}^{(0)}.$$

Also by assuming $i < j$ without loss of generality, let

$$p_{(ij)}^{(0)} = \sum_{a_1}\sum_{a_2} \cdots \sum_{a_{i-1}}\sum_{a_{i+1}} \cdots \sum_{a_{j-1}}\sum_{a_{j+1}} \cdots \sum_{a_{p-1}}\sum_{a_p} p_{a_1 a_2 \dots a_{i-1} 1 a_{i+1} \dots a_{j-1} 1 a_{j+1} \dots a_{p-1} a_p}^{(0)}.$$

Actually, $p_{(ij)}^{(0)}$ is the IC probability of both the ith and the jth factors taking Level 1. For the kth sample in Phase II, denote each cell count by $n_{a_1 a_2 \dots a_p, k}$. The Level 1 count of the ith factor should be

$$n_{(i)k} = \sum_{a_1}\sum_{a_2} \cdots \sum_{a_{i-1}}\sum_{a_{i+1}} \cdots \sum_{a_{p-1}}\sum_{a_p} n_{a_1 a_2 \dots a_{i-1} 1 a_{i+1} \dots a_{p-1} a_p, k}.$$

Let $\mathbf{n}_{\mathrm{MB},k} = \big[n_{(1)k}, \dots, n_{(p)k}\big]^T$ and $\mathbf{p}_{\mathrm{MB}}^{(0)} = \big[p_{(1)}^{(0)}, \dots, p_{(p)}^{(0)}\big]^T$. By employing the EWMA framework, we consider the statistic

$$G_{\mathrm{MB},k} = \frac{1}{N}\big(\mathbf{z}_{\mathrm{MB},k} - N\mathbf{p}_{\mathrm{MB}}^{(0)}\big)^T \boldsymbol{\Sigma}_{\mathrm{MB}}^{-1} \big(\mathbf{z}_{\mathrm{MB},k} - N\mathbf{p}_{\mathrm{MB}}^{(0)}\big),$$

where

$$\mathbf{z}_{\mathrm{MB},k} = a_{0,k,\lambda}^{-1} \sum_{j=1}^{k} (1-\lambda)^{k-j} \mathbf{n}_{\mathrm{MB},j},$$

$$\boldsymbol{\Sigma}_{\mathrm{MB}\,uv} = \begin{cases} p_{(u)}^{(0)}\big(1 - p_{(u)}^{(0)}\big) & \text{if } u = v \\ p_{(uv)}^{(0)} - p_{(u)}^{(0)} p_{(v)}^{(0)} & \text{if } u \neq v \end{cases}.$$

Note that when $\lambda = 1$, this charting statistic reduces to the one employed by Patel (1973). We refer to this chart as the multivariate binomial EWMA (MBE) control chart, and compare it with the LMBM chart in monitoring multivariate binomial processes.

Suppose that during a production process, five quality characteristics each labeled as conforming or nonconforming are being monitored, hence we have a five-way contingency table with 2^5 cells. After model estimation from an IC dataset, we obtain the IC model hierarchy structure [14][123][135][234][235][345] and the IC log-linear model with the coefficient vector

$$\begin{array}{rcrrrrrrrrl} \boldsymbol{\beta}^{(0)} & = & [& \beta_0 & 0.72 & 0.93 & 0.49 & 0.25 & 0.47 & -0.57 & 0.22 \\ & & & 0.11 & -0.14 & 0.15 & -0.16 & 0.41 & 0.16 & -0.19 & 0.33 \\ & & & 0.39 & 0 & 0 & 0 & 0.21 & 0 & 0.45 & 0.33 \\ & & & 0 & 0.27 & 0 & 0 & 0 & 0 & 0 & 0 \ ]^T. \end{array}$$

Here, β_0 is the intercept accommodating the Phase II sample size N because of the constraint that all cell counts sum up to N as expressed in Equation (4). Note that the zeros here mean that the corresponding effect terms are excluded.

It is believed that the possible shifts to OC states arise in the main effect of one factor reflecting its marginal distribution or in interaction effects of multiple factors representing their dependence. According to the one-to-one correspondence between factor effects and coefficient subvectors, shifts to OC states reflect the changes of the corresponding coefficient subvectors in the IC log-linear model. For simplicity, we consider the case where only one coefficient of the log-linear model changes by adding a magnitude δ and study the OC ARL performance. Note that the models before and after the change have the same hierarchy structure, and that the change only occurs on the coefficient magnitude of one retained term. The OC ARLs of the LMBM and MBE charts for various shift magnitudes are presented in Table 1 with the smoothing parameter $\lambda = 0.1$ and the Phase II sample size $N = 1,000$. To save space, not all coefficients are listed. According to Table 1, for the main factor effects, e.g., β_1 and β_4, the MBE chart outperforms the LMBM chart almost uniformly. The MBE chart exactly and completely summarizes the one-way marginal sums for each factor, which are mostly determined by the main effects. Therefore, it is more sensitive than the LMBM chart to the change of each main effect, which reflects the shifts of the one-way marginal sums directly. For the two-factor interaction effects such as $\beta_{1,2}$, $\beta_{2,3}$, $\beta_{2,5}$, and $\beta_{3,4}$ as well as the

Table 1: OC ARL comparison between the LMBM and MBE charts when only one coefficient changes, $\lambda = 0.1$ and $N = 1,000$

δ	LMBM		MBE		LMBM		MBE		LMBM		MBE	
	β_1				β_4				$\beta_{1,2}$			
0.01	232	(2.30)	160	(1.52)	246	(2.40)	172	(1.65)	210	(2.02)	156	(1.51)
0.02	87.8	(0.77)	49.0	(0.39)	100	(0.92)	56.7	(0.47)	71.0	(0.61)	45.4	(0.36)
0.05	14.8	(0.07)	10.2	(0.05)	16.6	(0.08)	11.3	(0.05)	12.5	(0.06)	9.61	(0.04)
0.20	2.99	(0.01)	2.38	(0.01)	3.24	(0.01)	2.56	(0.01)	2.67	(0.01)	2.27	(0.01)
-0.01	224	(2.19)	146	(1.40)	240	(2.37)	156	(1.50)	205	(1.96)	139	(1.34)
-0.02	81.7	(0.72)	44.1	(0.36)	94.8	(0.87)	50.2	(0.42)	67.7	(0.59)	41.6	(0.33)
-0.05	13.5	(0.07)	9.21	(0.04)	15.1	(0.08)	10.2	(0.05)	11.7	(0.05)	8.87	(0.04)
-0.20	2.41	(0.01)	1.95	(0.01)	2.54	(0.01)	2.04	(0.01)	2.26	(0.01)	1.94	(0.01)
	$\beta_{2,3}$				$\beta_{2,5}$				$\beta_{3,4}$			
0.01	259	(2.55)	248	(2.41)	289	(2.88)	322	(3.28)	262	(2.62)	305	(3.01)
0.02	110	(1.01)	106	(0.97)	148	(1.42)	199	(1.93)	114	(1.04)	156	(1.50)
0.05	18.3	(0.10)	19.2	(0.11)	25.9	(0.16)	44.4	(0.35)	19.0	(0.10)	30.3	(0.21)
0.20	3.42	(0.01)	3.54	(0.01)	4.21	(0.01)	5.86	(0.02)	3.49	(0.01)	4.61	(0.02)
-0.01	248	(2.47)	221	(2.20)	279	(2.79)	287	(2.84)	249	(2.47)	259	(2.56)
-0.02	101	(0.91)	89.3	(0.81)	139	(1.29)	170	(1.63)	106	(0.99)	131	(1.24)
-0.05	16.3	(0.09)	16.4	(0.10)	22.4	(0.14)	34.9	(0.26)	16.9	(0.09)	24.5	(0.17)
-0.20	2.63	(0.01)	2.66	(0.01)	3.07	(0.01)	4.01	(0.01)	2.67	(0.01)	3.37	(0.01)
	$\beta_{1,3,5}$				$\beta_{2,3,4}$				$\beta_{3,4,5}$			
0.01	232	(2.33)	252	(2.43)	273	(2.70)	339	(3.38)	264	(2.62)	343	(3.38)
0.02	84.7	(0.74)	109	(1.01)	132	(1.23)	239	(2.33)	118	(1.10)	233	(2.32)
0.05	14.4	(0.07)	20.0	(0.12)	21.7	(0.13)	64.5	(0.56)	19.1	(0.10)	59.7	(0.50)
0.20	2.95	(0.01)	3.56	(0.01)	3.80	(0.01)	7.41	(0.03)	3.55	(0.01)	6.91	(0.03)
-0.01	227	(2.24)	227	(2.21)	267	(2.65)	314	(3.14)	251	(2.46)	302	(2.97)
-0.02	79.6	(0.70)	94.7	(0.87)	120	(1.11)	204	(1.99)	109	(1.03)	194	(1.88)
-0.05	13.2	(0.06)	17.4	(0.10)	19.3	(0.11)	51.3	(0.42)	17.1	(0.09)	46.5	(0.37)
-0.20	2.41	(0.01)	2.82	(0.01)	2.85	(0.01)	5.03	(0.02)	2.70	(0.01)	4.81	(0.02)

NOTE: Standard errors are in parentheses.

three-factor interaction effects including $\beta_{1,3,5}$, $\beta_{2,3,4}$, and $\beta_{3,4,5}$, as the effect order increases, the LMBM chart shows more and more significant superiority over the MBE chart. The change occurring in the high-order interaction effect leads to little shifts of the one-way marginal sums, which are difficult for the MBE chart to detect. However, the LMBM chart is still capable of capturing the potential change of a high-order interaction effect powerfully via the log-likelihood ratio.

The OC ARLs of the LMBM and MBE charts for the same coefficient shift magnitudes as Table 1 under some other choices of the λ and N combination are omitted here for saving space, and they are available from the authors upon request. Generally, they exhibit the same patterns as in Table 1 for various changes of the coefficients, and similar conclusions can be drawn. With a fixed λ, for the same changes, both charts become more powerful when the sample size N increases. Moreover, for a fixed sample size N and the same coefficient, the LMBM chart with a smaller λ detects smaller shifts faster, whereas it has a better performance when detecting larger shifts with a larger λ. This is consistent with the properties of the conventional EWMA chart (Lucas and Saccucci, 1990), and it is further confirmed by Figure 1-(a) and -(b) which show the OC ARL curves (in log-scale) of the LMBM chart with the λ values of 0.05, 0.1, 0.2, and 0.5 when there are shifts in $\beta_{2,5}$ and $\beta_{2,3,4}$, respectively. The above property can in fact guide the selection of λ, and our empirical results show that a reasonable suggestion of λ may be between 0.05 and 0.2.

We assume in the above that the model hierarchy structure in both the IC and OC states are identical. Recall that the IC model hierarchy structure is [14][123][135][234][235][345]. This means the IC log-linear model does not contain the terms $\beta_{1,2,4}$, $\beta_{1,2,5}$, $\beta_{1,3,4}$, $\beta_{1,4,5}$, $\beta_{2,4,5}$, $\beta_{1,2,3,4}$, $\beta_{1,2,3,5}$, $\beta_{1,2,4,5}$, $\beta_{1,3,4,5}$, $\beta_{2,3,4,5}$, $\beta_{1,2,3,4,5}$, which leads to the absence of the three-factor interaction effects $C_1C_2C_4$, $C_1C_2C_5$, $C_1C_3C_4$, $C_1C_4C_5$, and $C_2C_4C_5$, as well as all the four-factor interaction effects and the five-factor interaction effect. However, sometimes the model structure itself changes when the process is OC. Compared to the IC model, the OC model may be either reduced or extended. In either case, the hierarchy principle should not be violated. After model extension by only one term, one nonexistent effect term emerges, such as $\beta_{1,2,4}$, $\beta_{1,3,4}$, and $\beta_{1,4,5}$ with some magnitude δ, and the original IC model does not encompass the current model any longer. Hence, the LMBM chart may not be necessarily

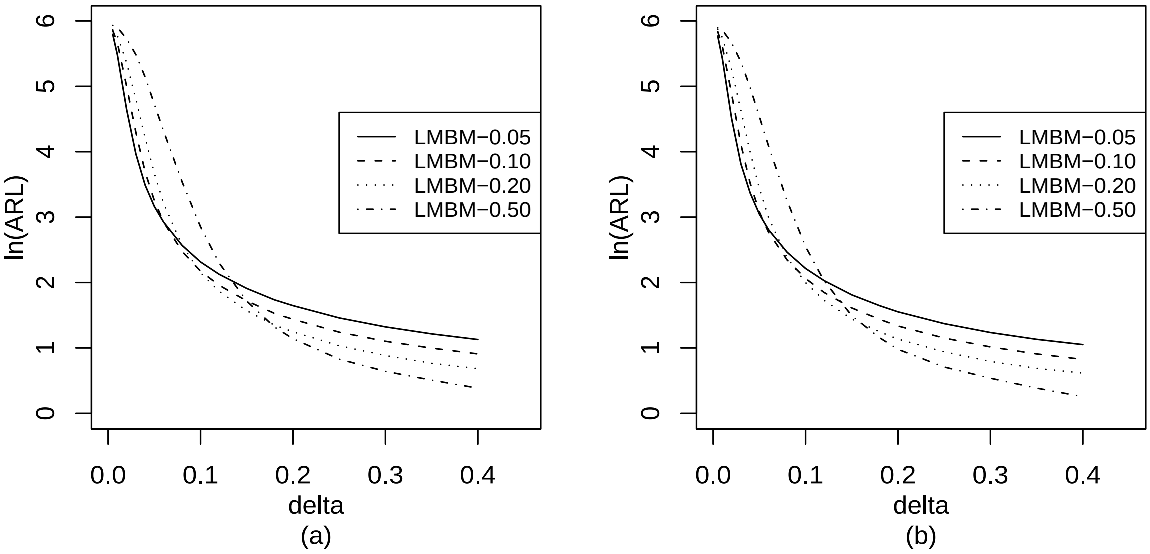

Figure 1: OC ARL curves of the LMBM chart with various values of λ when there are shifts in: (a) $\beta_{2,5}$; (b) $\beta_{2,3,4}$

superior over the MBE chart. The simulation results are illustrated in the upper half of Table 2. On the other hand, after model reduction by only one term, one effect becomes zero or disappears, and the original IC model can still include the reduced model. Therefore, the LMBM chart may still outperform the MBE chart. Some cases of removing merely one existent effect, for instance, $\beta_{1,3,5}$, $\beta_{2,3,4}$, and $\beta_{3,4,5}$, are listed in the lower half of Table 2.

The above simulation results are based on the assumption that the IC parameters are known or have been estimated from a sufficiently large reference dataset, which is not always practical. Here we investigate the effects of the Phase I reference sample size on the IC ARL, which violates this assumption. In this setting, the IC parameters are obtained from an IC reference dataset of size N_0. In particular, in each replication, first an IC sample of size N_0 is generated, and then the IC parameters are computed based on this sample. Finally, an independent sequence of online multivariate observation vectors are generated, and therefore we obtain the run lengths. With the same setting as Table 1, Table 3 lists the IC ARLs and standard deviations of run lengths (SDRL, in parentheses) when the IC parameters are computed from IC historical datasets of various size N_0, and the nominal IC ARL is 370. Here we see that when the reference sample size is relatively small, the actual IC ARL is far away from its nominal level 370. In addition, as the Phase I sample size N_0 increases, this

Table 2: OC ARL comparison between the LMBM and MBE charts for model misspecification, $\lambda = 0.1$, $N = 1,000$

δ	LMBM		MBE		LMBM		MBE		LMBM		MBE	
	$\beta_{1,2,4}$				$\beta_{1,3,4}$				$\beta_{1,4,5}$			
0.01	231	(2.31)	275	(2.74)	233	(2.25)	207	(2.02)	249	(2.49)	267	(2.65)
0.02	87.6	(0.77)	132	(1.27)	89.4	(0.79)	71.8	(0.64)	107	(0.98)	124	(1.17)
0.05	14.8	(0.07)	24.6	(0.16)	15.2	(0.07)	13.4	(0.04)	17.7	(0.09)	22.9	(0.15)
0.20	2.96	(0.01)	3.99	(0.01)	2.97	(0.01)	2.78	(0.01)	3.30	(0.01)	3.84	(0.01)
-0.01	224	(2.17)	239	(2.29)	232	(2.29)	179	(1.74)	248	(2.45)	240	(2.36)
-0.02	83.9	(0.75)	113	(1.06)	86.2	(0.77)	63.2	(0.54)	102	(0.92)	109	(1.01)
-0.05	13.7	(0.07)	21.3	(0.14)	14.1	(0.07)	12.2	(0.06)	16.1	(0.08)	19.8	(0.12)
-0.20	2.47	(0.01)	3.18	(0.01)	2.52	(0.01)	2.32	(0.01)	2.69	(0.01)	3.05	(0.01)
	$\beta_{1,3,5}$				$\beta_{2,3,4}$				$\beta_{3,4,5}$			
	2.26	(0.01)	2.64	(0.01)	1.20	(0.00)	1.94	(0.01)	1.99	(0.01)	3.30	(0.01)

NOTE: Standard errors are in parentheses.

bias decreases.

Table 3: IC ARLs of models estimated from a Phase I sample, $N = 1,000$

N_0	$\lambda = 0.1$		$\lambda = 0.2$	
2,000	910	(845)	916	(862)
3,000	729	(707)	750	(733)
5,000	573	(574)	572	(579)
8,000	485	(484)	492	(494)
10,000	458	(464)	452	(457)
20,000	404	(388)	400	(396)
50,000	375	(360)	376	(372)
100,000	367	(351)	372	(371)

4.2 Monitoring a Multivariate Multinomial Process

Apart from all the factors with two levels, factors at least one with three or more levels may exist in real production or services. A simple example is the attitude of a customer towards a service, which may take on the values of excellent, acceptable, or unacceptable. If four service indexes are taken into account, there will be a four-way contingency table with 3^4 cells for the four factors. In such complex cases, it is challenging to compare the proposed approach with alternative methods. This is because to the best of our knowledge, there is currently no appropriate monitoring approach incorporating the cross-classifications among multiple factors, among which at least one has more than two attribute levels. A naive method that comes into mind for comparison is to monitor the p groups of marginal sums of a p-way contingency table by introducing p individual charts separately.

If only the group of marginal sums for the ith factor ($i = 1, \ldots, p$) is considered, we face the same situation as Marcucci (1985), which treated the monitoring of univariate multinomial processes with the Pearson chi-square statistic. Actually, provided that the IC cell probability combination $p^{(0)}_{a_1 a_2 \ldots a_p}$ ($a_i = 1, \ldots, h_i$ and $i = 1, \ldots, p$) has been established, in the kth sample of size N in Phase II, the group of marginal sums for the ith factor

$$n_{(i,v)k} = \sum_{a_1} \sum_{a_2} \cdots \sum_{a_{i-1}} \sum_{a_{i+1}} \cdots \sum_{a_{p-1}} \sum_{a_p} n_{a_1 a_2 \ldots a_{i-1} v a_{i+1} \ldots a_{p-1} a_p, k}, \quad v = 1, \ldots, h_i,$$

where $n_{a_1 a_2 \ldots a_p, k}$ is the cell count in the jth sample, follow a multinomial distribution

$$\mathrm{MN}\big(N; p^{(0)}_{(i,1)}, p^{(0)}_{(i,2)}, \ldots, p^{(0)}_{(i,h_i)}\big),$$

where

$$p^{(0)}_{(i,v)} = \sum_{a_1} \sum_{a_2} \cdots \sum_{a_{i-1}} \sum_{a_{i+1}} \cdots \sum_{a_{p-1}} \sum_{a_p} p^{(0)}_{a_1 a_2 \ldots a_{i-1} v a_{i+1} \ldots a_{p-1} a_p}, \quad v = 1, \ldots, h_i.$$

The EWMA version of the Pearson chi-square statistic can be defined as

$$G_{\mathrm{MM},(i)k} = \frac{1}{N} \big(\mathbf{z}_{\mathrm{MM},(i)k} - N \mathbf{p}^{(0)}_{\mathrm{MM},(i)}\big)^T \mathbf{\Sigma}^{-1}_{\mathrm{MM},(i)} \big(\mathbf{z}_{\mathrm{MM},(i)k} - N \mathbf{p}^{(0)}_{\mathrm{MM},(i)}\big),$$

where

$$\begin{aligned}
\mathbf{z}_{\mathrm{MM},(i)k} &= a^{-1}_{0,k,\lambda} \sum_{j=1}^{k} (1-\lambda)^{k-j} \mathbf{n}_{\mathrm{MM},(i)j}, \\
\mathbf{n}_{\mathrm{MM},(i)j} &= \big[n_{(i,1)j}, n_{(i,2)j}, \ldots, n_{(i,h_i-1)j}\big]^T, \\
\mathbf{p}^{(0)}_{\mathrm{MM},(i)} &= \big[p^{(0)}_{(i,1)}, p^{(0)}_{(i,2)}, \ldots, p^{(0)}_{(i,h_i-1)}\big]^T, \\
\mathbf{\Sigma}_{\mathrm{MM},(i)\,uv} &= \begin{cases} p^{(0)}_{(i,u)}\big(1 - p^{(0)}_{(i,u)}\big) & \text{if } u = v \\ -p^{(0)}_{(i,u)} p^{(0)}_{(i,v)} & \text{if } u \neq v \end{cases}.
\end{aligned}$$

Consequently, for each of the p factors, we construct an individual chart to monitor its marginal sums, and hence p charting statistics $G_{\mathrm{MM},(i)k}$ $(i = 1, \ldots, p)$ are obtained. By adopting these p separate charts for the p factors, we establish a multi-chart, namely the multivariate multinomial EWMA (MME) chart, in the sense that it signals whenever at least one of these p charts constituting the MME chart does. We compare the LMBM and MME charts in monitoring multivariate multinomial processes.

Assume that a service flow has four quality characteristics under surveillance, with the first two evaluated as satisfactory or dissatisfactory and the last two assessed as excellent, acceptable, or unacceptable. This is a case of factors with mixed levels, and it forms a four-way contingency table of size 2 × 2 × 3 × 3 with 36 cells. After model estimation from an IC dataset, we obtain the IC model hierarchy structure [12][134][234] and the IC log-linear model with the coefficient vector

$$\begin{array}{lll rrrrrrrrr l} \boldsymbol{\beta}^{(0)} & = & [& \beta_0 & 0.73 & 0.72 & 0.70 & 0.12 & 0.72 & 0.11 & 0.17 & 0.12 & \\ & & & -0.15 & 0.19 & -0.14 & 0.23 & 0.07 & 0.16 & -0.14 & 0.23 & -0.30 & \\ & & & -0.17 & 0.14 & 0 & 0 & 0 & 0 & 0.19 & -0.15 & 0.11 & \\ & & & 0.22 & 0.24 & 0.24 & -0.08 & -0.16 & 0 & 0 & 0 & 0 &]^T, \end{array}$$

where β_0 is the intercept accommodating the Phase II sample size N, and the zeros represent the removed effect terms.

The control limits of the MME chart are chosen by simulation so that each individual chart has an identical IC ARL, jointly yielding the overall $\mathrm{ARL}_0 = 370$. Similar to the comparison between the LMBM and MBE charts, we also present the results where only one coefficient changes by adding a shift magnitude of δ and study the OC ARL performance. Some simulation results are listed in Table 4 in the case of $\lambda = 0.1$ and $N = 1,000$. The results for the other coefficients are available from the authors upon request. From Table 4, we see that the MME chart shows better performance than the LMBM chart when the main effects, e.g., β_2 and β_{4_2}, change as we would expect. This is easy to understand, since each of the four individual charts constituting the MME chart collects adequately the one-way marginal sums, which result directly from the main effects. Therefore, the MME chart stands out with a higher sensitivity to the main effects than the LMBM chart. The superiority of the LMBM chart over the MME chart becomes remarkable when the two-factor interaction effects such as $\beta_{1,3_1}$, $\beta_{1,4_2}$, $\beta_{2,3_2}$, $\beta_{2,4_1}$, $\beta_{3_2,4_1}$, and $\beta_{3_2,4_2}$ are focused on, and especially in the

Table 4: OC ARL comparison between LMBM and MME charts when only one coefficient changes, $\lambda = 0.1$, $N = 1,000$

δ	LMBM		MME		LMBM		MME		LMBM		MME	
	β_2				β_{4_2}				$\beta_{1,3_1}$			
0.02	172	(1.66)	79.7	(0.68)	200	(1.99)	134	(1.25)	132	(1.24)	111	(1.04)
0.05	30.4	(0.20)	14.7	(0.08)	41.6	(0.31)	22.4	(0.14)	21.8	(0.13)	20.4	(0.12)
0.20	4.57	(0.01)	3.02	(0.01)	4.73	(0.01)	3.36	(0.01)	3.66	(0.01)	3.56	(0.01)
-0.02	153	(1.48)	68.4	(0.61)	205	(1.97)	141	(1.34)	125	(1.16)	93.6	(0.85)
-0.05	25.7	(0.17)	12.5	(0.07)	42.8	(0.32)	23.1	(0.14)	19.8	(0.11)	17.8	(0.11)
-0.20	3.26	(0.01)	2.26	(0.01)	4.85	(0.01)	3.46	(0.01)	3.00	(0.01)	2.97	(0.01)
	$\beta_{1,4_2}$				$\beta_{2,3_2}$				$\beta_{2,4_1}$			
0.02	201	(1.96)	243	(2.39)	198	(1.90)	230	(2.26)	134	(1.26)	114	(1.09)
0.05	41.0	(0.30)	61.6	(0.52)	39.5	(0.29)	52.8	(0.43)	22.2	(0.13)	20.7	(0.13)
0.20	4.68	(0.01)	5.75	(0.02)	4.57	(0.01)	5.19	(0.02)	3.71	(0.01)	3.59	(0.01)
-0.02	201	(2.00)	242	(2.40)	197	(1.92)	230	(2.25)	123	(1.17)	95.1	(0.87)
-0.05	41.9	(0.31)	64.2	(0.55)	39.6	(0.29)	55.0	(0.45)	20.0	(0.12)	18.1	(0.11)
-0.20	4.72	(0.01)	6.31	(0.02)	4.66	(0.01)	5.96	(0.02)	3.00	(0.01)	2.97	(0.01)
	$\beta_{3_2,4_1}$				$\beta_{3_2,4_2}$				$\beta_{1,3_1,4_2}$			
0.02	234	(2.30)	291	(2.84)	276	(2.74)	343	(3.46)	241	(2.37)	342	(3.42)
0.05	55.4	(0.45)	103	(0.95)	90.8	(0.82)	228	(2.22)	62.7	(0.51)	222	(2.14)
0.20	5.47	(0.02)	7.88	(0.03)	7.32	(0.03)	20.8	(0.13)	5.90	(0.02)	17.5	(0.10)
-0.02	234	(2.38)	278	(2.78)	277	(2.70)	339	(3.41)	243	(2.36)	335	(3.34)
-0.05	56.6	(0.47)	106	(0.99)	92.4	(0.84)	238	(2.34)	63.8	(0.54)	217	(2.12)
-0.20	5.64	(0.02)	9.23	(0.04)	7.58	(0.03)	27.6	(0.18)	5.96	(0.02)	22.0	(0.14)
	$\beta_{1,3_2,4_2}$				$\beta_{2,3_1,4_1}$				$\beta_{2,3_2,4_2}$			
0.02	271	(2.67)	359	(3.58)	143	(1.36)	156	(1.51)	274	(2.69)	362	(3.69)
0.05	90.2	(0.80)	286	(2.81)	23.8	(0.14)	30.6	(0.21)	92.5	(0.83)	300	(3.04)
0.20	7.34	(0.03)	39.6	(0.30)	3.77	(0.01)	4.39	(0.01)	7.36	(0.03)	44.5	(0.35)
-0.02	271	(2.70)	359	(3.62)	136	(1.30)	131	(1.24)	273	(2.74)	367	(3.66)
-0.05	91.4	(0.81)	314	(3.17)	21.9	(0.13)	26.8	(0.19)	91.8	(0.82)	296	(2.92)
-0.20	7.60	(0.03)	102	(0.94)	3.25	(0.01)	3.98	(0.01)	7.48	(0.03)	59.7	(0.51)

NOTE: Standard errors are in parentheses.

cases of changes in the three-factor interaction effects, including $\beta_{1,3_1,4_2}$, $\beta_{1,3_2,4_2}$, $\beta_{2,3_1,4_1}$, and $\beta_{2,3_2,4_2}$, where the MME chart is outperformed by the LMBM chart by quite a large margin. Like the comparison between the LMBM chart and the MBE chart, simulations under some other parameter settings also exhibit the same trends as in Table 4. The effects of the Phase II sample size N and the EWMA smoothing parameter λ on the LMBM chart are similar to those in the multivariate binomial processes.

Combined with the comparison between the LMBM and MBE charts, it is affirmed again that the superiority of the LMBM chart lies in shifts in the interaction effects of multiple factors that represent the dependence among them. In addition, it should be noted that when there is only one factor, the LMBM chart reduces to the EWMA version of the log-likelihood ratio statistic, and the MME chart simplifies to the EWMA version of the Pearson chi-square statistic. Their performance should be more or less the same, since they enjoy the same asymptotic distribution under the null hypothesis as a central χ^2 with an appropriate degree of freedom. Hence, the LMBM chart can work well within the unified framework of the univariate/multivariate binomial/multinomial processes.

5 A Real Application

In this section, the proposed methodology is implemented in an aluminium electrolytic capacitor (AEC) manufacturing process to demonstrate its real utilization. This may be regarded as a typical example to apply the LMBM chart in practice. The process comprises a series of stages, from cutting, winding, drying, impregnating, and assembling, to sleeving, washing, aging, and packaging. The production line manufactures AECs from various raw materials, including anode aluminum foil, cathode aluminum foil, guiding pin, electrolyte sheet, plastic cover, aluminum shell, and plastic tube. Right after each stage, the quality of the stagewise AEC products, namely capacitor elements in terms of appearance condition and functional performance, will be inspected by sampling in order to meet the specifications.

We consider the quality in the aging stage, where the AEC quality monitoring is concentrated on three most important quality characteristics: capacity (CAP), dissipation factor (DF), and leakage current (LC). Each characteristic can be evaluated as conforming or nonconforming to its specification by an electronic device at a very high speed. Obtaining their precise numerical values is possible, but this will cost too much. To this end, for a specific sample size in Phase II, this can be regarded as a multivariate binomial process, which has three factors including CAP, DF, and LC all with two levels. The cross-classification counts with all factor level combinations are stored in a three-way contingency table with $2^3 = 8$ cells. Based on the above information, the log-linear model

$$\ln m_{ijk} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{1,2} x_1 x_2 + \beta_{1,3} x_1 x_3 + \beta_{2,3} x_2 x_3 + \beta_{1,2,3} x_1 x_2 x_3,$$

where $i = 1, 2; j = 1, 2; k = 1, 2$, all the expected counts m_{ijk} sum up to the sample size N, and x_1, x_2, x_3 take 1 or -1 representing the two levels of three factors CAP, DF, and LC, respectively, can adequately characterize the relationship between the cell counts and these level combinations. Therefore, the proposed LMBM chart can be adopted to monitor the three quality characteristics simultaneously in the aging stage.

It is known that each workbench in the aging stage manufactures at least 6,000 AEC elements every day, and that the three quality characteristics CAP, DF, and LC for each AEC element are then inspected automatically as conforming or nonconforming by some electronic devices. We perform model estimation from a historical IC dataset [2,1,19,12,1,75,732,39447] with about 40,000 observations, which contains the 8 cell counts and could have been completed by one workbench within only a few days. Consequently, the IC model has a hierarchy structure [CAP DF][CAP LC]. Furthermore, the estimated coefficient vector is given by $\boldsymbol{\beta}^{(0)}$=[β_0, 1.91, 2.11, 0.88, 1.02, 1.11, 0.00, 0.00]^T. In addition, the AECs are usually inspected in a batch of 500, hence we have a Phase II sample size $N = 500$. Based on this, the IC cell count expectation vector $\mathbf{m}^{(\mathbf{0})}$ is

$$\mathbf{m}^{(0)} = [2.2996, 1.4235, 23.762, 14.710, 1.7174, 92.601, 907.96, 48956]^T \times 10^{-2}.$$

In Phase II, the EWMA smoothing parameter λ is chosen to be 0.1. We obtain the control limit of the LMBM chart, which is 0.83, by simulation, such that the IC ARL is 370. Now the LMBM chart is ready to be constructed to monitor the process. After obtaining new observations, we calculate the charting statistics R_k for each sample, then plot them in the control chart, and compare them with the control limit. The original observation vectors

$\mathbf{n}_k$ for each sample are available from the authors upon request. We take the 9th point for example to illustrate the calculation of charting statistics. Based on $\mathbf{m}^{(0)}$ and the original observation vectors $\mathbf{n}_1, \ldots, \mathbf{n}_9$, for example, $\mathbf{n}_9 = [0, 0, 0, 0, 0, 6, 10, 484]^T$, we first calculate the pseudo-observation vector

$$\mathbf{z}_9 = [0.89090, 0.55151, 22.598, 26.403, 0.66537, 133.15, 873.51, 48942]^T \times 10^{-2}$$

using Equation (7). Then we use the IPF algorithm based on the hierarchy structure [CAP DF][CAP LC] to get the MLE of the cell count expectation vector over $\mathbf{z}_9$, which is

$$\widehat{\mathbf{y}}_9 = [0.67165, 0.77075, 22.817, 26.184, 2.3419, 131.47, 871.83, 48944]^T \times 10^{-2}.$$

Finally, by Equation (8), its charting statistic R_9 is calculated as 0.25332. Figure 2 shows the resulting LMBM chart (solid curve connecting the dots), along with its control limit (solid horizontal line). The LMBM chart signals at the 28th observation and remains above the control limit in the remainder of the sequence.

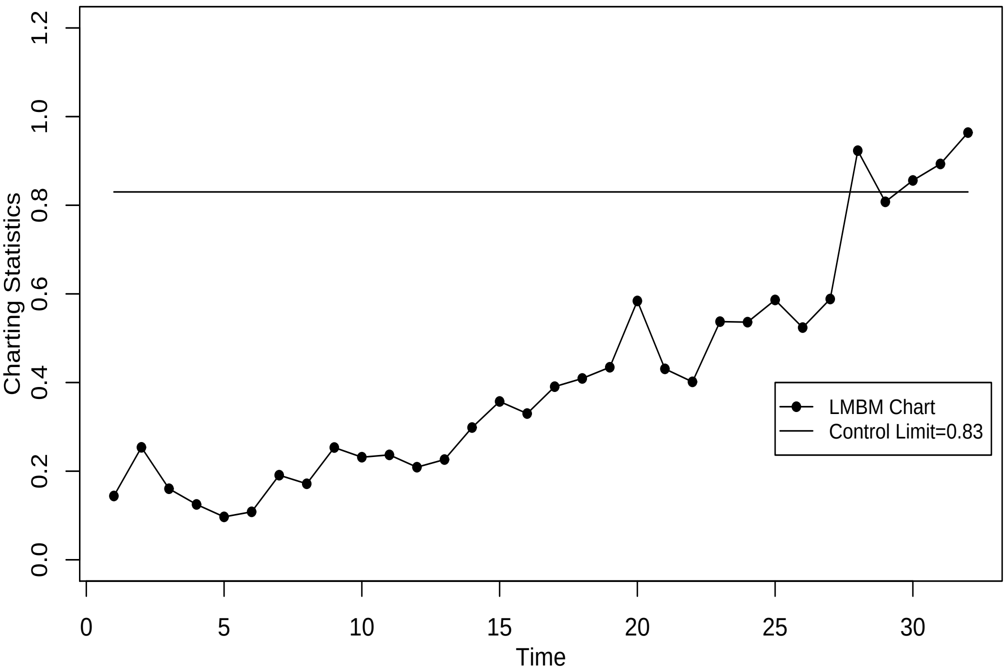

Figure 2: The LMBM control chart for monitoring the AEC process. The solid horizontal line indicates its control limit.

By comparison, we also adopt the MBE chart to monitor the AEC process with $\text{ARL}_0 = 370$, $N = 1,000$, and $\lambda = 0.1$, which is shown in Figure 3. The MBE chart also triggers an OC signal at the 28th sample. However, it is difficult to say which chart performs better based on only this single run. The MBE chart enjoys a charting statistic of a simpler form, but it cannot provide the practical interpretation of shifts according to the one-to-one correspondence between factor effects and coefficient subvectors in a log-linear model. As indicated earlier, this correspondence may exploit insights into multivariate binomial/multinomial processes and assist in further diagnosis.

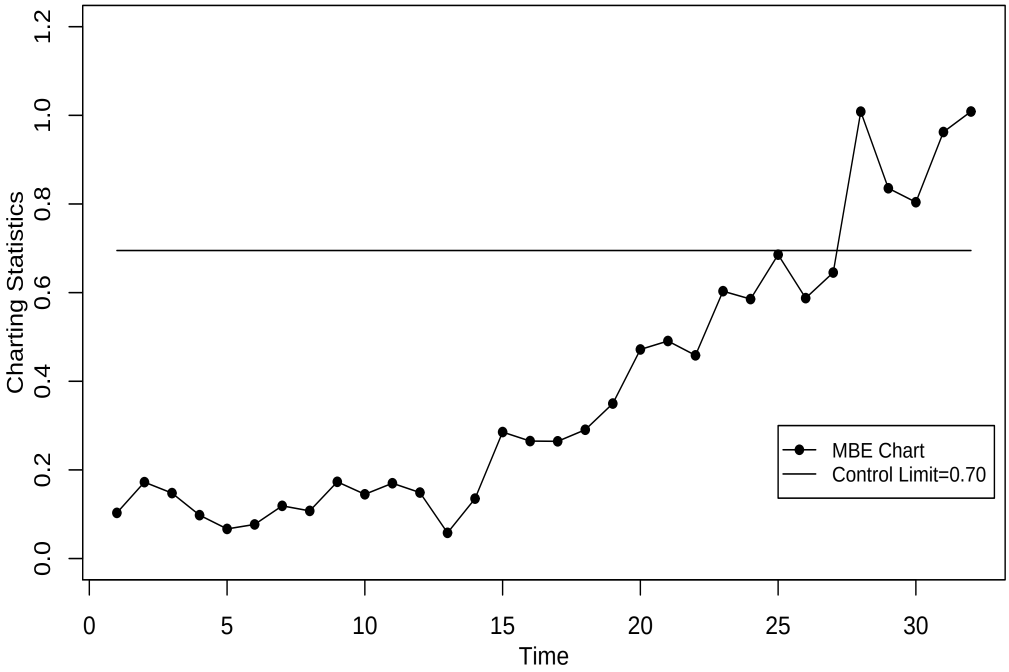

Figure 3: The MBE control chart for monitoring the AEC process. The solid horizontal line indicates its control limit.

6 Conclusion

This paper proposes a Phase II control chart, namely the log-linear multivariate binomial/multinomial control chart, which can be actualized as a general SPC tool for the monitoring of multivariate/univariate binomial/multinomial distributed data. The LMBM chart adopts the EWMA control scheme in terms of the exponentially weighted pseudo-observation vector in Phase II, which exploits the information of the past and current sampling cells adequately and distinguishingly as well as mitigates the potential tendency of sparsity in the multi-way contingency table. In comparison with the existing approaches, numerical simulations demonstrate that at certain expenses of sensitivity to the changes in main effects, the LMBM chart is much more robust than traditional ones, which take only the marginal cell probabilities into account. Instead, the LMBM chart provides much higher detection ability to possible shifts in interaction effects of multiple factors which represent their dependence. In addition, the AEC application shows that the LMBM chart can be implemented effortlessly into real manufacturing and even service industries.

As pointed out previously, we assume in this work that the model hierarchy structures under the IC and OC conditions are the same. However, this is not always the case. Our ongoing research focuses on monitoring simultaneously the possible changes of both the model hierarchy structure and the coefficient vector, such that the potential shifts can be detected and diagnosed more accurately and efficiently. Moreover, the log-linear model may explode when the numbers of factors or their levels are large. As mentioned in Section 2, a reduced model via variable selection is desired accordingly. In such cases, our proposed procedure is still applicable. As shown in Table 3, the performance of the proposed chart is affected by the amount of data in the reference dataset, especially for a large number of factors or their levels. Very large Phase I samples must be collected for the LMBM chart to perform as well as those with known parameters. So determining the Phase I sample size required to remove the effects of estimated parameters is critical and warrants future research.

In addition, the considered log-linear model is related to the so-called context-tree model, which describes dependent categorical data with finite attribute levels in terms of context dependency (Ben-Gal *et al.*, 2003; Brice and Jiang, 2009). Such dependency means that the statistical distribution of a new sample is conditional on a set of the most recent samples that precedes it in a data stream (Brice *et al.*, 2011). We know that the multivariate binomial/multinomial data in this paper are assumed to be temporally independent, but the context dependency should also apply to a multi-way contingency table in that at a time point a cell count may depend spatially on the counts in its neighborhood cells. This may be another point of view that characterizes a multi-way contingency table and deserves the development of a control chart based on it.

The sparsity phenomenon mentioned previously is somewhat similar to that in high-quality processes (e.g., Nelson, 1994; McCool and Joyner-Motley, 1998), where the proportion of nonconforming products is extremely small, say, in the level of parts-per-million. Consequently, the p-chart will be of little use, since the estimated nonconforming fraction is mostly zero. Therefore, it is important to use a dataset large enough to achieve a reasonably accurate estimate of parameters. We refer to Yang *et al.* (2002) for a discussion on the effect of the dataset size in Phase I on estimating the control limits of a geometric chart (Kaminsky *et al.*, 1992), which is developed particularly for monitoring high-quality processes. Based on log-linear models, the LMBM chart should also be conveniently modified to monitor multivariate categorical high-quality processes (Niaki and Abbasi, 2007), which is our future research.

Acknowledgement

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

Appendix

A. Some Notations in Log-Linear Models

Table A.1 below lists important notations that are useful for describing log-linear models.

B. Proof of Theorem 1

Throughout the appendix, we use the following additional notations.

$$\widehat{\mathbf{p}} = \mathbf{z}_k/N, \quad \widehat{\boldsymbol{\pi}} = \exp\{\boldsymbol{X}\widehat{\boldsymbol{\beta}}_k\}/N, \quad \mathbf{A}_\pi = \boldsymbol{X}(\boldsymbol{X}^T\mathbf{D}_{\boldsymbol{\pi}}\boldsymbol{X})^{-1}\boldsymbol{X}^T$$

A diagonal matrix with the elements of the vector $\mathbf{x}$ on its diagonal will be written as $\mathbf{D}_\mathbf{x}$. Moreover, for notational convenience, we will use the subscript $\boldsymbol{\pi}_i$ instead of $\boldsymbol{\pi}^{(i)}$ for $i = 0, 1$ in this section which should not cause any confusion.

Table A.1: Important notations in log-linear models

Notation	Meaning
N	the sample size
p	the number of factors in a general contingency table
h_i	the number of levels of the ith factor
h	the total number of cells in a p-way contingency table
a_i	the index of the level of the ith factor, $a_i = 1, \ldots, h_i$
$n_{a_1 a_2 \ldots a_p}$	the count of observations in the cell$(a_1, a_2, \ldots, a_p)$
$m_{a_1 a_2 \ldots a_p}$	the expectation of $n_{a_1 a_2 \ldots a_p}$
$p_{a_1 a_2 \ldots a_p}$	the probability of an observation in the cell$(a_1, a_2, \ldots, a_p)$
$\mathbf{n}$	the vector of cell counts in a p-way contingency table, of size $h \times 1$
$\mathbf{m}$	the expectation of $\mathbf{n}$, of size $h \times 1$
$\mathbf{p}$	the vector of cell probabilities in a p-way contingency table, of size $h \times 1$
$\boldsymbol{\beta}_i$	the coefficient subvector corresponding to the ith main or interaction effect
$\boldsymbol{X}_i$	the design submatrix corresponding to $\boldsymbol{\beta}_i$
$\boldsymbol{\beta}$	the coefficient vector
$\boldsymbol{X}$	the design matrix

To prove Theorem 1 in Section 2, we need the following lemmas.

Lemma 1 *Suppose $\boldsymbol{\pi}_0$ is the vector of true cell probabilities in the process. Under the conditions in Theorem 1, we have*

$$\sqrt{c_{0,k,\lambda} N}(\widehat{\mathbf{p}} - \boldsymbol{\pi}_0) \xrightarrow{\mathcal{L}} N_h(\mathbf{0}, \mathbf{D}_{\boldsymbol{\pi}_0} - \boldsymbol{\pi}_0 \boldsymbol{\pi}_0^T).$$

Proof. When $N \to \infty$, this lemma can be directly obtained by using Theorem 14.8-2 in Bishop *et al.* (2007) and the independence of $\mathbf{n}_i$ for $i = 1, \ldots, k$; When N is fixed, it is sufficient to prove that, as $\lambda \to 0$ and $k \to \infty$, for any h-dimensional vector $\boldsymbol{\theta}$,

$$\sqrt{c_{0,k,\lambda} N}\boldsymbol{\theta}^T(\widehat{\mathbf{p}} - \boldsymbol{\pi}_0) \xrightarrow{\mathcal{L}} N_1(0, \boldsymbol{\theta}^T(\mathbf{D}_{\boldsymbol{\pi}_0} - \boldsymbol{\pi}_0 \boldsymbol{\pi}_0^T)\boldsymbol{\theta}).$$

Rewrite $\boldsymbol{\theta}^T(\widehat{\mathbf{p}} - \boldsymbol{\pi}_0) = a_{0,k,\lambda}^{-1} \sum_{i=1}^k (1-\lambda)^{k-i} \boldsymbol{\theta}^T(\mathbf{n}/N - \boldsymbol{\pi}_0)$. Obviously,

$$\max_{1 \le i \le k} \frac{\lambda^{2(k-i)}}{\sum_{i=1}^k (1-\lambda)^{2(k-i)}} \to 0 \quad \text{as} \quad n \to \infty,$$

and $\boldsymbol{\theta}^T(\mathbf{n}_i/N - \boldsymbol{\pi}_0)$ for $i = 1, \ldots, k$ are i.i.d. random variables with mean 0 and variance $N^{-1}\boldsymbol{\theta}^T(\mathbf{D}_{\boldsymbol{\pi}_0} - \boldsymbol{\pi}_0\boldsymbol{\pi}_0^T)\boldsymbol{\theta}$. By using the Hájek-Sidak central limit theorem (Serfling, 1980), we know $\boldsymbol{\theta}^T(\widehat{\mathbf{p}} - \boldsymbol{\pi}_0)\xrightarrow{\mathcal{L}} N_1(0, (c_{0,k,\lambda}N)^{-1}\boldsymbol{\theta}^T(\mathbf{D}_{\boldsymbol{\pi}_0} - \boldsymbol{\pi}_0\boldsymbol{\pi}_0^T)\boldsymbol{\theta})$ from which the lemma follows.

Lemma 2 *Suppose $\boldsymbol{\pi}_0$ is the vector of true cell probabilities in the process. Under the conditions in Theorem 1, we have*

$$\widehat{\boldsymbol{\pi}} - \boldsymbol{\pi}_0 = \mathbf{D}_{\boldsymbol{\pi}_0}\mathbf{A}_{\boldsymbol{\pi}_0}(\widehat{\mathbf{p}} - \boldsymbol{\pi}_0) + o_p((c_{0,k,\lambda}N)^{-\frac{1}{2}}).$$

Proof. This lemma can be proved in a similar fashion to Theorem 12.3.3-(3) in Christensen (1997). By the Lagrange multiplier method we can verify that the maximum weighted likelihood estimator $\widehat{\boldsymbol{\beta}}_k$ is a function of $\mathbf{z}_k$ which is defined implicitly as the solution to

$$[\mathbf{z}_k - \exp\{\boldsymbol{X}\boldsymbol{\beta}\}]^T\boldsymbol{X} = \mathbf{0}.$$

Thus, we can write a function $\mu(\mathbf{n}) = \boldsymbol{X}\boldsymbol{\beta}(\mathbf{n})$. Next, the first-order Taylor's expansion of $\mu(\mathbf{z}_k/N)$ at $\mathbf{m}^{(0)}/N$ gives

$$\mu(\mathbf{z}_k/N) - \mu(\mathbf{m}^{(0)}/N) - \mathrm{d}\mu(\boldsymbol{\pi}_0)(\widehat{\mathbf{p}} - \boldsymbol{\pi}_0) = o(||\widehat{\mathbf{p}} - \boldsymbol{\pi}_0||). \tag{A.1}$$

By Lemmas 12.5.2 and 12.5.3 of Christensen (1997), we can show that

$$\mu(\mathbf{z}_k/N) - \mu(\mathbf{m}^{(0)}/N) = \boldsymbol{X}\widehat{\boldsymbol{\beta}}_k - \boldsymbol{X}\boldsymbol{\beta}^{(0)}. \tag{A.2}$$

On the other hand, using similar arguments in the proof of Lemma 12.3.2 of Christensen (1997), we can have

$$\mathrm{d}\mu(\boldsymbol{\pi}_0) = \boldsymbol{X}[\boldsymbol{X}^T\mathbf{D}_{\boldsymbol{\pi}_0}\boldsymbol{X}]^{-1}\boldsymbol{X}^T. \tag{A.3}$$

Combining Equations (A.1)-(A.3) and using Lemma 1 lead to

$$(\boldsymbol{X}\widehat{\boldsymbol{\beta}}_k - \boldsymbol{X}\boldsymbol{\beta}^{(0)}) = \mathbf{A}(\widehat{\mathbf{p}} - \boldsymbol{\pi}_0) + o((c_{0,k,\lambda}N)^{-\frac{1}{2}}). \tag{A.4}$$

Then, by Taylor's expansions, we have

$$\begin{aligned}\widehat{\boldsymbol{\pi}} - \boldsymbol{\pi}_0 &= N^{-1}(\exp\{\boldsymbol{X}\widehat{\boldsymbol{\beta}}_k\} - \exp\{\boldsymbol{X}\boldsymbol{\beta}^{(0)}\}) \\ &= N^{-1}D(\exp\{\boldsymbol{X}\boldsymbol{\beta}^{(0)}\})(\boldsymbol{X}\widehat{\boldsymbol{\beta}}_k - \boldsymbol{X}\boldsymbol{\beta}^{(0)}) + o((c_{0,k,\lambda}N)^{-\frac{1}{2}}) \\ &= \mathbf{D}_{\boldsymbol{\pi}_0}\mathbf{A}_{\boldsymbol{\pi}_0}(\widehat{\mathbf{p}} - \boldsymbol{\pi}_0) + o((c_{0,k,\lambda}N)^{-\frac{1}{2}}).\end{aligned}$$

Proof of Theorem 1: (i) Rewrite

$$\begin{aligned} R_k &= 2\mathbf{z}_k^T(\ln \widehat{\mathbf{y}}_k - \ln \mathbf{m}^{(0)}) \\ &= 2N(\widehat{\mathbf{p}} - \boldsymbol{\pi}_0)^T(\ln \widehat{\boldsymbol{\pi}} - \ln \boldsymbol{\pi}_0) + 2N\boldsymbol{\pi}_0^T(\ln \widehat{\boldsymbol{\pi}} - \ln \boldsymbol{\pi}_0). \end{aligned}$$

By the second-order Taylor expansions, we know

$$\begin{aligned} c_{0,k,\lambda}R_k =& c_{0,k,\lambda}N[2(\widehat{\mathbf{p}} - \boldsymbol{\pi}_0)^T\mathbf{D}_{\boldsymbol{\pi}_0}^{-1}(\widehat{\boldsymbol{\pi}} - \boldsymbol{\pi}_0) + o(||\widehat{\mathbf{p}} - \boldsymbol{\pi}_0|| \cdot ||\widehat{\boldsymbol{\pi}} - \boldsymbol{\pi}_0||) \\ &+ 2\mathbf{1}^T(\widehat{\boldsymbol{\pi}} - \boldsymbol{\pi}_0) - (\widehat{\boldsymbol{\pi}} - \boldsymbol{\pi}_0)^T\mathbf{D}_{\boldsymbol{\pi}_0}^{-1}(\widehat{\boldsymbol{\pi}} - \boldsymbol{\pi}_0) + o(||\widehat{\boldsymbol{\pi}} - \boldsymbol{\pi}_0|| \cdot ||\widehat{\boldsymbol{\pi}} - \boldsymbol{\pi}_0||)]. \end{aligned}$$

Using Lemmas 1 and 2 yields

$$\begin{aligned} c_{0,k,\lambda}R_k &= 2c_{0,k,\lambda}N(\widehat{\mathbf{p}} - \boldsymbol{\pi}_0)^T\mathbf{A}_{\boldsymbol{\pi}_0}(\widehat{\mathbf{p}} - \boldsymbol{\pi}_0) - c_{0,k,\lambda}N(\widehat{\mathbf{p}} - \boldsymbol{\pi}_0)^T\mathbf{A}_{\boldsymbol{\pi}_0}\mathbf{D}_{\boldsymbol{\pi}_0}\mathbf{A}_{\boldsymbol{\pi}_0}(\widehat{\mathbf{p}} - \boldsymbol{\pi}_0) + o_p(1) \\ &= c_{0,k,\lambda}N(\widehat{\mathbf{p}} - \boldsymbol{\pi}_0)^T\mathbf{A}_{\boldsymbol{\pi}_0}(\widehat{\mathbf{p}} - \boldsymbol{\pi}_0) + o_p(1), \qquad (A.5) \end{aligned}$$

where we use the fact $\mathbf{A}_{\boldsymbol{\pi}_0}\mathbf{D}_{\boldsymbol{\pi}_0}\mathbf{A}_{\boldsymbol{\pi}_0} = \mathbf{A}_{\boldsymbol{\pi}_0}$. By Lemma 1, $\sqrt{c_{0,k,\lambda}N}(\widehat{\mathbf{p}} - \boldsymbol{\pi}_0)$ is asymptotically distributed as $N_h(\mathbf{0}, \mathbf{D}_{\boldsymbol{\pi}_0} - \boldsymbol{\pi}_0\boldsymbol{\pi}_0^T)$. As a consequence, it follows from the Cochran's theorem that $2c_{0,k,\lambda}R_k$ has a limiting χ^2 distribution (Anderson, 2003) if

$$(\mathbf{D}_{\boldsymbol{\pi}_0} - \boldsymbol{\pi}_0\boldsymbol{\pi}_0^T)\mathbf{A}_{\boldsymbol{\pi}_0}(\mathbf{D}_{\boldsymbol{\pi}_0} - \boldsymbol{\pi}_0\boldsymbol{\pi}_0^T)\mathbf{A}_{\boldsymbol{\pi}_0}(\mathbf{D}_{\boldsymbol{\pi}_0} - \boldsymbol{\pi}_0\boldsymbol{\pi}_0^T) = (\mathbf{D}_{\boldsymbol{\pi}_0} - \boldsymbol{\pi}_0\boldsymbol{\pi}_0^T)\mathbf{A}_{\boldsymbol{\pi}_0}(\mathbf{D}_{\boldsymbol{\pi}_0} - \boldsymbol{\pi}_0\boldsymbol{\pi}_0^T). \quad (A.6)$$

By noting that

$$(\mathbf{I} - \mathbf{D}_{\boldsymbol{\pi}_0}^{1/2}\mathbf{A}_{\boldsymbol{\pi}_0}\mathbf{D}_{\boldsymbol{\pi}_0}^{1/2})\mathbf{D}_{\boldsymbol{\pi}_0}^{1/2}\mathbf{X} = \mathbf{0}, \qquad (A.7)$$

Equation (A.6) can be verified with some straightforward algebra.

Finally, Theorem 1-(i) follows immediately from

$$\begin{aligned} \mathrm{tr}\{(\mathbf{D}_{\boldsymbol{\pi}_0} - \boldsymbol{\pi}_0\boldsymbol{\pi}_0^T)\mathbf{A}\} &= \mathrm{tr}\{\mathbf{D}_{\boldsymbol{\pi}_0}\mathbf{A}_{\boldsymbol{\pi}}\} - \mathrm{tr}\{\mathbf{D}_{\boldsymbol{\pi}_0}\mathbf{1}\mathbf{1}^T\mathbf{D}_{\boldsymbol{\pi}_0}\mathbf{A}_{\boldsymbol{\pi}_0}\} \\ &= s - \mathrm{tr}\{\mathbf{1}\mathbf{1}^T\mathbf{D}_{\boldsymbol{\pi}_0}\} = s - 1, \end{aligned}$$

where we use $\mathbf{1}^T\boldsymbol{\pi}_0 = 1$ and Equation (A.7) once again.

(ii) When $\boldsymbol{\pi}_1 = \boldsymbol{\pi}_0 + \boldsymbol{\mu}(c_{0,k,\lambda}N)^{-1/2}$, by Lemma 1, $\sqrt{c_{0,k,\lambda}N}(\widehat{\mathbf{p}} - \boldsymbol{\pi}_0)$ is asymptotically distributed as $N_h(\boldsymbol{\mu}, \mathbf{D}_{\boldsymbol{\pi}_1} - \boldsymbol{\pi}_1\boldsymbol{\pi}_1^T)$. Similar to Equation (A.5), we have

$$c_{0,k,\lambda}R_k = c_{0,k,\lambda}N(\widehat{\mathbf{p}} - \boldsymbol{\pi}_0)^T\mathbf{A}_{\boldsymbol{\pi}_1}(\widehat{\mathbf{p}} - \boldsymbol{\pi}_0) + o_p(1).$$

From the first part of the proof, we know $c_{0,k,\lambda}R_k$ has a limiting noncentral χ^2 distribution with $s-1$ degrees of freedom and the noncentrality parameter

$$\psi^2 = \boldsymbol{\mu}^T \mathbf{D}_{\boldsymbol{\pi}_1}^{-1} \boldsymbol{\mu} = \boldsymbol{\mu}^T \mathbf{D}_{\boldsymbol{\pi}_0}^{-1} \boldsymbol{\mu} + o_p(1).$$

References:

Anderson, T.W. (2003) *An Introduction to Multivariate Statistical Analysis*, third edn, Wiley, New York.

Ben-Gal, I., Morag, G. and Shmilovici, A. (2003) Context-based statistical process control: a monitoring procedure for state-dependent processes. *Technometrics*, **45**, 293–311.

Bersimis, S., Psarakis, S. and Panaretos, J. (2007) Multivariate statistical process control charts: an overview. *Quality and Reliability Engineering International*, **23**, 517–543.

Bishop, Y.M.M., Fienberg, S.E. and Holland, P.W. (2007) *Discrete Multivariate Analysis*, Springer.

Brice, P. and Jiang, W. (2009) A context tree method for multistage fault detection and isolation with applications to commercial video broadcasting systems. *IIE Transactions*, **41**, 776–789.

Brice, P., Jiang, W. and Wan, G.H. (2011) A cluster-based context-tree model for multivariate data streams with applications to anomaly detection. *INFORMS Journal on Computing*, forthcoming.

Chen, L., Chang, F.M. and Chen, Y. (2011) The application of multinomial control charts for inspection error. *International Journal of Industrial Engineering*, **18**, 244–253.

Chiu, J. and Kuo, T. (2008) Attribute control chart for multivariate poisson distribution. *Communications in Statistics: Theory and Methods*, **37**, 146–158.

Christensen, R. (1997) *Log-Linear Models and Logistic Regression*, second edn, Springer.

Dahinden, C., Parmigiani, G., Emerick, M.C. and Bühlmann, P. (2007) Penalized likelihood for sparse contingency tables with an application to full-length cDNA libraries. *BMC Bioinformatics*, **8**, 476.

Fienberg, S.E. and Rinaldo, A. (2007) Three centuries of categorical data analysis: log-linear models and maximum likelihood estimation. *Journal of Statistical Planning and Inference*, **137**, 3430–3445.

Huang, W., Reynolds, M.R., Jr. and Wang, S. (2012) A binomial GLR control chart for monitoring a proportion. *Journal of Quality Technology*, **44**, 192–208.

Johnson, N.L., Kotz, S. and Balakrishnan, N. (1997) *Discrete Multivariate Distributions*, Wiley, New York.

Kaminsky, F.C., Benneyan, J.C., Davis, R.D. and Burke, R.J. (1992) Statistical control charts based on a geometric distribution. *Journal of Quality Technology*, **24**, 63–69.

Li, J., Tsung, F. and Zou, C. (2012) Directional control schemes for multivariate categorical processes. *Journal of Quality Technology*, **44**, 136–154.

Lowry, C.A. and Montgomery, D.C. (1995) A review of multivariate control charts. *IIE Transactions*, **27**, 800–810.

Lowry, C.A., Woodall, W.H., Champ, C.W. and Rigdon, S.E. (1992) A multivariate exponentially weighted moving average control chart. *Technometrics*, **34**, 46–53.

Lu, X.S., Xie, M., Goh, T.N. and Lai, C.D. (1998) Control charts for multivariate attribute processes. *International Journal of Production Research*, **36**, 3477–3489.

Lucas, J.M. and Saccucci, M.S. (1990) Exponentially weighted moving average control schemes: properties and enhancements. *Technometrics*, **32**, 1–29.

Marcucci, M. (1985) Monitoring multinomial processes. *Journal of Quality Technology*, **17**, 86–91.

McCool, J.I. and Joyner-Motley, T. (1998) Control charts applicable when the fraction nonconforming is small. *Journal of Quality Technology*, **30**, 240–247.

McCullagh, P. and Nelder, J.A. (1989) *Generalized Linear Models*, 2nd edition, Chapman & Hall/CRC.

Nelson, L.S. (1994) A control chart for parts-per-million nonconforming items. *Journal of Quality Technology*, **26**, 239–240.

Niaki, S.T.A. and Abbasi, B. (2007) On the monitoring of multi-attribute high-quality production processes. *Metrika*, **66**, 373–388.

Patel, H.I. (1973) Quality control methods for multivariate binomial and poisson distributions. *Technometrics*, **15**, 103–112.

Qiu, P. (2008) Distribution-free multivariate process control based on log-linear modeling. *IIE Transactions*, **40**, 664–677.

Ryan, A.G., Wells, L.J. and Woodall, W.H. (2011) Methods for monitoring multiple proportions when inspecting continuously. *Journal of Quality Technology*, **43**, 237–248.

Serfling, R.J. (1980) *Approximation Theorems of Mathematical Statistics*, John Wiley & Sons, New York, NY.

Topalidou, E. and Psarakis, S. (2009) Review of multinomial and multiattribute quality control charts. *Quality and Reliability Engineering International*, **25**, 773–804.

Weiss, C.H. (2012) Continuously monitoring categorical processes. *Quality Technology & Quantitative Management*, **9**, 171–188.

Woodall, W.H. and Ncube, M.M. (1985) Multivariate CUSUM quality control procedures. *Technometrics*, **27**, 285–292.

Woodall, W.H. (1997) Control charts based on attribute data: bibliography and review. *Journal of Quality Technology*, **29**, 172–183.

Yang, Z., Xie, M., Kuralmani, V. and Tsui, K.L. (2002) On the performance of geometric charts with estimated control limits. *Journal of Quality Technology*, **34**, 448–458.

Yashchin, E. (2012) On detection of changes in categorical data. *Quality Technology & Quantitative Management*, **9**, 79–96.

Zou, C., Jiang, W. and Tsung, F. (2011) A LASSO-based SPC diagnostic framework for multivariate statistical process control. *Technometrics*, **53**, 297–309.

Zou, C., Tsung, F. and Wang, Z. (2007) Monitoring general linear profiles using multivariate EWMA schemes. *Technometrics*, **49**, 395–408.