Directional Change-Point Detection for Process Control with Multivariate Categorical Data

Jian Li,[1*] Fugee Tsung,[1] 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*

Received 3 October 2011; accepted 20 December 2012
DOI 10.1002/nav.21525
Published online 8 February 2013 in Wiley Online Library (wileyonlinelibrary.com).

Abstract: Most modern processes involve multiple quality characteristics that are all measured on attribute levels, and their overall quality is determined by these characteristics simultaneously. The characteristic factors usually correlate with each other, making multivariate categorical control techniques a must. We study Phase I analysis of multivariate categorical processes (MCPs) to identify the presence of change-points in the reference dataset. A directional change-point detection method based on log-linear models is proposed. The method exploits directional shift information and integrates MCPs into the unified framework of multivariate binomial and multivariate multinomial distributions. A diagnostic scheme for identifying the change-point location and the shift direction is also suggested. Numerical simulations are conducted to demonstrate the detection effectiveness and the diagnostic accuracy. © 2013 Wiley Periodicals, Inc. Naval Research Logistics 60: 160–173, 2013

Keywords: contingency table; generalized likelihood ratio test; log-linear model; multivariate multinomial distribution; statistical process control

1. INTRODUCTION

Nowadays, most processes involve multiple quality characteristics. For instance, a bearing has both an inner diameter and an outer diameter that together determine its usefulness and are assumed to follow a bivariate normal distribution. Here, the observed variables are numerical, and therefore manufacturing bearings represents a multivariate continuous process. Most multivariate statistical process control (SPC) techniques have been devoted to monitoring such processes. We refer to Lowry and Montgomery [13] and Bersimis et al. [2] for nice literature reviews. Conversely, due to intrinsic properties or expensive data collection costs, the continuous values of many quality characteristics are not available. Instead, some attribute levels of them, such as good or bad, may be available, since these categorical values are rough and do not need expensive precise measurements. Such situations have become increasingly common. Examples include products on a production line whose multiple characteristics are classified simply as conforming or nonconforming to their predefined specifications, and multiple indexes in a service flow that are assessed as excellent, acceptable, or unacceptable. Here, the characteristic factors involved have two or more attribute levels, and such data are known as multivariate categorical.

SPC usually consists of two phases. In Phase I, a process dataset is collected and examined to see if any unusual patterns exist. Among others, a change-point is a sustained special cause that remains until some corrective action is taken, and an outlier is an isolated special cause that affects a single sample and then disappears (Hawkins et al. [10]). If there are any change-points or outliers, they are identified and adjusted, resulting in a clean dataset. This dataset is called the in-control (IC) dataset and used for estimating the IC model, which characterizes the process under IC operating conditions. The performance of Phase I analysis is often measured by the power of detecting these abnormalities statistically and the accuracy of identifying them. After this retrospective analysis, quality inspectors will have a good reference IC model. In Phase II, the main task is to construct control charts (Montgomery [16]) and plot the charting statistic of each successive sample online for monitoring the process to see if there is any shift from the IC state.

*Jian Li is currently at the School of Management, Xi'an Jiaotong University, Xi'an, Shaanxi, China

Correspondence to: C. Zou (chlzou@yahoo.com.cn)

Compared to multivariate continuous processes, SPC tools for multivariate categorical processes (MCPs) are relatively rare. Woodall [23] listed many control charts for monitoring attribute data, including the well-known p-chart for binomially distributed variables and the generalized p-chart (Marcucci [15]) for multinomially distributed ones, but most of them are for univariate processes. Among control charts for simultaneously monitoring multiple categorical characteristics, the χ^2-chart proposed by Patel [18], the mnp-chart suggested by Lu et al. [14], and the mp-chart designed by Chiu and Kuo [4] monitor multiple factors all with two levels. It should be noted that these control charts used in Phase II SPC all assume that the reference dataset is IC. However, in reality this is not always the case, and any unusual patterns such as change-points or outliers in the reference dataset could yield an erroneous IC model estimate.

Therefore, Phase I analysis of MCPs is crucial, which, however, still remains a challenge and has not been investigated to the best of our knowledge. We found that the models adopted for establishing the control charts mentioned above are limited in two aspects. First, they can apply only when there are only two attribute levels for each factor. If at least one of the factors has three or more levels, these models become useless. Second, these models focus merely on the marginal sums of each factor, neglecting cross-classifications among multiple factors. So cross-classification probability shifts cannot be detected quickly. To best characterize MCPs and overcome these two drawbacks, we turn to log-linear models (Bishop et al. [3]), which characterize appropriately the association structure among multiple categorical factors. It follows that logarithms of the cross-classification probabilities may be assumed to depend linearly on the levels of multiple factors. Log-linear models resemble multiway analysis of variance (ANOVA) models, in which the expected observations are supposed to be linearly dependent on the levels of several factors.

This article tries to perform Phase I analysis of MCPs by change-point detection. Besides its application in industrial statistics such as SPC, change-point detection is actually useful in many other areas, such as genetic linkage analysis (Siegmund [19]) and array-CGH data analysis (Zhang and Siegmund [25]) in biostatistics. In Phase I SPC, it is assumed that there is already an off-line dataset collected from the process chronologically. We intend to see if there is a time point, at which there is a step change of the process and by which the dataset can be divided into two homogeneous segments. Here, homogeneous means that the data in such a segment come from the same distribution. Note that this change-point detection is an off-line procedure that deals with the beforehand collected reference dataset, rather than an online sequential process that collects data as it proceeds. Sequential change-point detection is similar to Phase II SPC, which at each time point k combines the current kth observation with the earlier ones to see if there is a change-point in the observed dataset up to k (see Lai and Xing [12], Moustakides et al. [17], and Shih et al. [21]). As k increases, this procedure repeats and continues on and on.

Our suggested method is based on log-linear models and applies to the unified framework of multiple factors each with an arbitrary number of attribute levels. Furthermore, to detect change-points as powerfully as possible, the proposed approach also integrates some practical directional shift information. Like multiway ANOVA, it is known that in a log-linear model the cross-classification probabilities are determined by main factor effects and factor interaction effects. At the change-point, shifts may arise in factor effects, which correspond to deviations of some coefficient subvectors in the log-linear model. Such practical knowledge formulated as shift directions should be exploited. A diagnostic scheme is also developed to identify the change-point location and the shift direction upon change-point detection. Monte Carlo simulations are performed to demonstrate the effectiveness of the proposed change-point detection and diagnostic approach.

The rest of this article is organized as follows. The log-linear modeling of MCPs in terms of multivariate binomial and multivariate multinomial distributions is first introduced. Then, the proposed directional change-point detection and diagnostic methodology is described. Finally, numerical simulations investigate the performance of the proposed methodology, and a real-life example demonstrates its implementation. Some derivations can be found in Appendices.

2. MULTIVARIATE CATEGORICAL MODELING

2.1. Multivariate Categorical Processes

Let us first illustrate an MCP with a real manufacturing example. Consider aluminum electrolytic capacitors (AECs) on a production line, the quality of which is mainly evaluated in terms of leakage current (LC), dissipation factor (DF), and capacity (CAP). For an AEC, an electronic device evaluates each of its three characteristics LC, DF, and CAP as conforming or nonconforming automatically at a very high speed. Although obtaining the continuous values of them is possible, with consideration of a huge amount of AECs on the production line, this would be cost prohibitive. So this forms a multivariate categorical process involving three factors each with two attribute levels, giving $2^3 = 8$ level combinations. For each factor, without loss of generality, we denote "conforming" by -1 and "nonconforming" by 1. So the combination $(1, 1, -1)$ represents an AEC with nonconforming LC and DF and conforming CAP.

The AEC example is a trivariate categorical process. Now, we extend it to a general MCP. Imagine p categorical factors $\mathbb{C} = \{C_1, \ldots, C_p\}$, and that each C_i takes a number, say h_i,

of possible levels. The overall cross-classifications among all level combinations of the p factors form a p-way $h_1 \times \cdots \times h_p$ contingency table with $h = \prod_{i=1}^{p} h_i$ cells. In addition, each cell corresponding to a certain level combination of the p factors stores the count under this combination.

Suppose that during a process, observations are collected over a period of time without prior knowledge of their total count. Then, each cell count in the contingency table is subject to an independent Poisson distribution (Bishop et al. [3]). However, for research convenience, we formalize the MCP within the framework of multivariate binomial or multivariate multinomial distributions and then fix the total count N of observations. Conditional on N, the cell counts jointly follow a multinomial distribution. Based on this, for a general p-way contingency table of size $h_1 \times \cdots \times h_p$, denote the probability of an observation falling into the cell corresponding to the level combination $a_1, \ldots, a_p$ by $p_{a_1 \ldots a_p}$ ($a_i = 1, \ldots, h_i$ and $i = 1, \ldots, p$). All of the cell probabilities add up to 1. Denote also the count of observations among a sample of size N in this cell by $n_{a_1 \ldots a_p}$ and the expectation of $n_{a_1 \ldots a_p}$ by $m_{a_1 \ldots a_p} = N p_{a_1 \ldots a_p}$. Then, the joint distribution of the marginal counts $n_{(i)1}, \ldots, n_{(i)h_i}$ with respect to factor C_i ($i = 1, \ldots, p$), where

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

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

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

Actually, $p_{(i)1}, \ldots, p_{(i)h_i}$ are the marginal probabilities of C_i. The joint distribution of the p sets of variables $n_{(i)1}, \ldots, n_{(i)h_i}$ ($i = 1, \ldots, p$), each being a multinomial distribution, is a multivariate multinomial distribution (Johnson et al. [11]). If each factor has only two levels, the distribution simplifies to a multivariate binomial distribution.

2.2. Log-Linear Model

Given the total sample size N, we need to know the relationship between cell probabilities or cell count expectations and their corresponding factor levels. Actually, log-linear models can be used to model this relationship (Bishop et al. [3]). A log-linear model originally relates the logarithms of cell count expectations in a multiway contingency table to factor levels, when there is no restriction on the total sample size and the cell counts are subject to Poisson distributions. Here, since the simple multinomial sampling scheme is adopted with a fixed sample size N, it would be more convenient to focus on the logarithms of cell probabilities, instead of cell count expectations that are the products of size N and cell probabilities.

Take a simple three-way contingency table of size $h_1 \times h_2 \times h_3$ for illustration. The log-linear model describing the relationship between the cell probability $p_{a_1 a_2 a_3}$ ($a_i = 1, \ldots, h_i$ and $i = 1, 2, 3$) in cell(a_1, a_2, a_3) and the factor levels indexed with a_1, a_2, a_3 is

$$\ln p_{a_1 a_2 a_3} = u^{(0)} + u^{(1)}_{a_1} + u^{(2)}_{a_2} + u^{(3)}_{a_3} + u^{(1,2)}_{a_1 a_2} + u^{(1,3)}_{a_1 a_3} + u^{(2,3)}_{a_2 a_3} + u^{(1,2,3)}_{a_1 a_2 a_3},$$

where $u^{(0)}$ is the intercept, $u^{(1)}, u^{(2)}, u^{(3)}$ 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. The cell probabilities satisfy $\sum_{a_1, a_2, a_3} p_{a_1 a_2 a_3} = 1$, and the cell counts $n_{a_1 a_2 a_3}$ jointly follow a multinomial distribution $\mathrm{MN}(N; p_{a_1 a_2 a_3})$ ($a_i = 1, \ldots, h_i$ and $i = 1, 2, 3$). Log-linear models that characterize cell probabilities in a multinomial distribution rather than cell count expectations in independent Poisson distributions, have also appeared in Dahinden et al. [7]. In addition, identifiability requires constraints such as

$$\sum_{a_1} u^{(1)}_{a_1} = \sum_{a_1} u^{(1,2)}_{a_1 a_2} = \sum_{a_1} u^{(1,3)}_{a_1 a_3} = \sum_{a_1} u^{(1,2,3)}_{a_1 a_2 a_3} = 0$$

for factor C_1 along its index a_1. Similarly, such equations exist for factors C_2 and C_3 along their indexes a_2 and a_3, respectively. The dependence among factors should be reflected by the interaction effects among them. For example, $u^{(1,2)}$ represents the dependence between factors C_1 and C_2. Therefore, if a log-linear model involves no interaction effects, the factors are independent of each other.

A log-linear model with separate identifiability constraints is inconvenient to use, but it can be written in another equivalent form, which automatically satisfies these constraints. This is illustrated by a 2×3 contingency table. The identifiability constraints allow setting

$$\begin{array}{lll} u^{(0)} = \beta_0, & u^{(1)}_1 = \beta_1, & u^{(1)}_2 = -\beta_1, \\ u^{(2)}_1 = \beta_2, & u^{(2)}_2 = \beta_3, & u^{(2)}_3 = -\beta_2 - \beta_3, \\ u^{(1,2)}_{1,1} = \beta_4, & u^{(1,2)}_{1,2} = \beta_5, & u^{(1,2)}_{1,3} = -\beta_4 - \beta_5, \\ u^{(1,2)}_{2,1} = -\beta_4, & u^{(1,2)}_{2,2} = -\beta_5, & u^{(1,2)}_{2,3} = \beta_4 + \beta_5. \end{array}$$

Therefore, the probability p_{ij} ($i = 1, 2$; $j = 1, 2, 3$) will be $\ln p_{ij} = \beta_0 + \sum_{k=1}^{5} \beta_k x_k$, and x_k ($k = 1, \ldots, 5$) takes the value 1, 0, or -1 where appropriate. Obviously, β_1 measures the main effect $u^{(1)}$ of factor C_1, $[\beta_2, \beta_3]^T$ measures the main effect $u^{(2)}$ of factor C_2, and $[\beta_4, \beta_5]^T$ measures the interaction effect $u^{(1,2)}$ of factors C_1 and C_2. Clearly, this is a regression form with the cell probabilities as the responses, the factor

levels composing the design matrix, and the β coefficients as the regressors.

Let us now show that this regression form applies to a general multiway contingency table. We shall see that imposed by identifiability constraints, the log-linear model for a p-way contingency table in the original u-term form above can be expressed in a regression form as

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

where $\mathbf{1}$ is a column vector consisting of 1 as all its elements with appropriate dimensions, $\mathbf{p}$ satisfying $\mathbf{1}^T\mathbf{p} = 1$ is the $h \times 1$ probability vector in the h cells of the contingency table, $\mathbf{X}_i$ is an $h \times q_i$ design submatrix corresponding to the ith main or interaction effect and with 1, 0, or -1 as its entries, and $\boldsymbol{\beta}_i$ is the coefficient subvector of size $q_i \times 1$. Clearly, β_0 is a scalar representing the intercept. Let $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_{2^p-1})$ and $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, \ldots, \boldsymbol{\beta}_{2^p-1}^T)^T$, then the design matrix $\widetilde{\mathbf{X}} = (\mathbf{1}, \mathbf{X})$ and the coefficient vector $\widetilde{\boldsymbol{\beta}} = (\beta_0, \boldsymbol{\beta}^T)^T$, and Eq. (1) can be rewritten as

$$\ln \mathbf{p} = \widetilde{\mathbf{X}}\widetilde{\boldsymbol{\beta}} = \mathbf{1}\beta_0 + \mathbf{X}\boldsymbol{\beta}.$$

The log-linear model (1) is at the effect level, and in total there are $2^p - 1$ factor effects. If each factor has only two levels, all of the design submatrixes $\mathbf{X}_i$ will reduce into column vectors, and their corresponding coefficient subvectors $\boldsymbol{\beta}_i$ will all be scalars.

The design matrix $\widetilde{\mathbf{X}}$ guarantees the identifiability constraints, which is in known in advance. Its derivation is a little complex and left in Appendix A. The design submatrixes such as $\mathbf{X}_i$ and their corresponding coefficient subvectors $\boldsymbol{\beta}_i$ are actually arranged following the order from the overall mean β_0, the main effects, up to the highest interaction effect. For example, we consider three factors C_1, C_2, and C_3 with 2, 3, and 3 levels, respectively. The sequence is the overall mean, the main effects C_1, C_2, and C_3, the two-factor interaction effects C_1C_2, C_1C_3, and C_2C_3, and finally the three-factor interaction effect $C_1C_2C_3$. So, we obtain the design matrix $\widetilde{\mathbf{X}}$ and the coefficient vector

$$\begin{aligned}\widetilde{\boldsymbol{\beta}} = [\ & \beta_0 \quad \beta_{(1)} \quad \beta_{(2_1)} \quad \beta_{(2_2)} \quad \beta_{(3_1)} \quad \beta_{(3_2)} \\ & \beta_{(1,2_1)} \quad \beta_{(1,2_2)} \quad \beta_{(1,3_1)} \quad \beta_{(1,3_2)} \quad \beta_{(2_1,3_1)} \quad \beta_{(2_1,3_2)} \\ & \beta_{(2_2,3_1)} \quad \beta_{(2_2,3_2)} \quad \beta_{(1,2_1,3_1)} \quad \beta_{(1,2_1,3_2)} \quad \beta_{(1,2_2,3_1)} \quad \beta_{(1,2_2,3_2)}\ ]^T.\end{aligned}$$

Because of the constraint $\mathbf{1}^T\mathbf{p} = 1$ on the log-linear model (1), given $\boldsymbol{\beta}$, the coefficient β_0 can be determined. Therefore, attention may be paid mainly to $\boldsymbol{\beta}$. Following the above arrangement of coefficients, we see that for example, $\boldsymbol{\beta}_3 = [\beta_{(3_1)}, \beta_{(3_2)}]^T$ measures the main effect of factor C_3, $\boldsymbol{\beta}_4 = [\beta_{(1,2_1)}, \beta_{(1,2_2)}]^T$ measures the two-factor interaction effect of C_1 and C_2, and $\boldsymbol{\beta}_7 = [\beta_{(1,2_1,3_1)}, \beta_{(1,2_1,3_2)}, \beta_{(1,2_2,3_1)}, \beta_{(1,2_2,3_2)}]^T$ measures the three-factor interaction effect of C_1, C_2, and C_3. Clearly, the ith main or interaction effect, the design submatrix $\mathbf{X}_i$, and the coefficient subvector $\boldsymbol{\beta}_i$ $(i = 1, \ldots, 2^p - 1)$ are in one-to-one correspondence. So the probability vector $\mathbf{p}$ is ultimately determined by the magnitudes of the coefficient subvectors $\boldsymbol{\beta}_i$ $(i = 1, \ldots, 2^p - 1)$.

There is another expression of the log-linear model (1), which changes it equivalently from the effect level into the coefficient level and is

$$\ln \mathbf{p} = \mathbf{1}\beta_0 + \sum_{i=1}^{h-1} \mathbf{x}_i \beta_i, \tag{2}$$

where $\mathbf{x}_i$ is the ith column vector of the matrix $\mathbf{X}$ and β_i as a scalar is its corresponding coefficient. For instance, in the example of the contingency table of size $2 \times 3 \times 3$, $\beta_5 = \beta_{(3_2)}$, $\beta_{11} = \beta_{(2_1,3_2)}$, and $\beta_{17} = \beta_{(1,2_2,3_2)}$. In addition, we have $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_{h-1}]$ and $\boldsymbol{\beta} = [\beta_1, \ldots, \beta_{h-1}]^T$. Clearly, there is also correspondence between the ith column vector $\mathbf{x}_i$ and the coefficient β_i $(i = 1, \ldots, h - 1)$.

3. CHANGE-POINT DETECTION IN MCPs

Before introducing the proposed change-point detection method, we need first to state the two-sample multivariate categorical test, which is the basis of the change-point problem.

3.1. Two-Sample Multivariate Categorical Test Problem

In a log-linear model, the marginal distribution of one factor is mainly determined by its main effect, whereas the dependence among multiple factors is represented by their interaction effect. Furthermore, according to the one-to-one correspondence between factor effects and coefficient subvectors, shifts in the marginal distribution of one factor appear in the form of deviations of the coefficient subvector corresponding to its main effect, and shifts in the dependence among multiple factors arise in the form of deviations of the coefficient subvector reflecting their interaction effect. So the practical explanation of shifts in MCPs is defined.

In the systematic framework of multivariate binomial or multivariate multinomial distributions, the two-sample test problem for MCPs is formulated as follows. Given a p-way contingency table with h cells, consider the two multinomial distributions $\mathrm{MN}(N_\mathrm{a}; \mathbf{p}_\mathrm{a})$ and $\mathrm{MN}(N_\mathrm{b}; \mathbf{p}_\mathrm{b})$, where $\mathbf{p}_\mathrm{a}$ and $\mathbf{p}_\mathrm{b}$ are both of size $h \times 1$ and

$$\ln \mathbf{p}_\mathrm{a} = \mathbf{1}\beta_0^{(\mathrm{a})} + \mathbf{X}\boldsymbol{\beta}^{(\mathrm{a})}, \quad \ln \mathbf{p}_\mathrm{a} = \mathbf{1}\beta_0^{(\mathrm{b})} + \mathbf{X}\boldsymbol{\beta}^{(\mathrm{b})}.$$

We want to test the following hypothesis

$$H_0 : \ \mathbf{p}_\mathrm{a} = \mathbf{p}_\mathrm{b} \ \text{ versus } \ H_1 : \ \mathbf{p}_\mathrm{a} \neq \mathbf{p}_\mathrm{b}. \tag{3}$$

Since $\beta_0^{(t)}$ ($t = \mathrm{a}, \mathrm{b}$) can be determined by $\boldsymbol{\beta}^{(t)}$ (as $\mathbf{1}^T\mathbf{p} = 1$), hypothesis (3) is equivalent to

$$H_0 : \ \boldsymbol{\beta}^{(\mathrm{b})} = \boldsymbol{\beta}^{(\mathrm{a})} \ \text{ versus } \ H_1 : \ \boldsymbol{\beta}^{(\mathrm{b})} \neq \boldsymbol{\beta}^{(\mathrm{a})}. \tag{4}$$

Assume that $\mathbf{n}_\mathrm{a}$ and $\mathbf{n}_\mathrm{b}$ are two independent random vectors subject to $\mathrm{MN}(N_\mathrm{a}; \mathbf{p}_\mathrm{a})$ and $\mathrm{MN}(N_\mathrm{b}; \mathbf{p}_\mathrm{b})$, respectively. Then, $\mathbf{1}^T\mathbf{n}_\mathrm{a} = N_\mathrm{a}$ and $\mathbf{1}^T\mathbf{n}_\mathrm{b} = N_\mathrm{b}$. A natural test for hypotheses (3) and (4) may be constructed using the generalized likelihood ratio test (GLRT; Anderson, 2003 [1]). It is easy to see that the maximum likelihood estimates (MLEs) of $\mathbf{p}_\mathrm{a}$ and $\mathbf{p}_\mathrm{b}$ under H_0 in hypothesis (3) are both equal to $(\mathbf{n}_\mathrm{a} + \mathbf{n}_\mathrm{b})/(N_\mathrm{a} + N_\mathrm{b})$, whereas under H_1 they are $\mathbf{n}_\mathrm{a}/N_\mathrm{a}$ and $\mathbf{n}_\mathrm{b}/N_\mathrm{b}$, respectively. So the -2LRT (likelihood ratio test) statistic is

$$\begin{aligned}\theta = {} & 2\mathbf{n}_\mathrm{a}^T \ln\left(\frac{\mathbf{n}_\mathrm{a}}{N_\mathrm{a}}\right) + 2\mathbf{n}_\mathrm{b}^T \ln\left(\frac{\mathbf{n}_\mathrm{b}}{N_\mathrm{b}}\right) \\ & - 2(\mathbf{n}_\mathrm{a} + \mathbf{n}_\mathrm{b})^T \ln\left(\frac{\mathbf{n}_\mathrm{a} + \mathbf{n}_\mathrm{b}}{N_\mathrm{a} + N_\mathrm{b}}\right).\end{aligned} \tag{5}$$

Note that here we consider a saturated log-linear model as expressed in Eqs. (1) and (2), which includes effects of all orders, namely from main effects to the highest p-factor interaction effect. In other words, the considered log-linear model is not a reduced hierarchical one, in which some effects are removed subject to the hierarchical principle (Chritensen [5]). For a saturated log-linear model, the MLE of the cell probability vector is simply the observed cell count vector divided by the total sample size. This is illustrated above for deriving the MLEs of $\mathbf{p}_\mathrm{a}$ and $\mathbf{p}_\mathrm{b}$ under H_0 and H_1. So the existence of these MLEs is clear.

Hypotheses (3) and (4) state that $\boldsymbol{\beta}^{(\mathrm{b})}$ differs from $\boldsymbol{\beta}^{(\mathrm{a})}$ in all coefficients. With the practical definition of shifts, however, it is usually reasonable to assume that in real processes shifts occur in only a few coefficient subvectors or coefficients, reflecting the changes in their corresponding main or interaction effects. Suppose that there is some prior knowledge that a shift arises in only one coefficient in $\boldsymbol{\beta}^{(\mathrm{a})}$, say the ith ($1 \le i \le h-1$) coefficient, by adding an unknown constant δ_i to it. Then, $\widetilde{\boldsymbol{\beta}}^{(\mathrm{a})}$ and $\widetilde{\boldsymbol{\beta}}^{(\mathrm{b})}$ have a relationship such that

$$\begin{cases} \beta_i^{(\mathrm{b})} = \beta_i^{(\mathrm{a})} + \delta_i & \text{for some } i, \text{and } 1 \le i \le h-1, \\ \beta_j^{(\mathrm{b})} = \beta_j^{(\mathrm{a})} & \text{for all } j \neq i, \text{and } 1 \le j \le h-1, \\ \beta_0^{(\mathrm{b})} = \beta_0^{(\mathrm{a})} + \alpha_i, & \end{cases} \tag{6}$$

where $\beta_i^{(t)}$ is the ith element of $\boldsymbol{\beta}^{(t)}$ ($t = \mathrm{a}, \mathrm{b}$) and α_i is also an unknown constant. Here, α_i is actually the increment of $\beta_0^{(\mathrm{a})}$ induced by the constraint $\mathbf{1}^T\mathbf{p} = 1$. Now the hypothesis can be expressed as

$$H_0 : \ \boldsymbol{\beta}^{(\mathrm{b})} = \boldsymbol{\beta}^{(\mathrm{a})} \ \text{ versus } \ H_1 : \ \boldsymbol{\beta}^{(\mathrm{b})} = \boldsymbol{\beta}^{(\mathrm{a})} + \mathbf{d}_i\delta_i, \tag{7}$$

where $\mathbf{d}_i$ is the direction vector of size $(h-1) \times 1$ with 1 at its ith element and 0 elsewhere.

Hypothesis (7) assumes that the location of the only dissimilar coefficient between $\boldsymbol{\beta}^{(\mathrm{a})}$ and $\boldsymbol{\beta}^{(\mathrm{b})}$ is known. Here, we consider a more practical and general case where compared to $\boldsymbol{\beta}^{(\mathrm{a})}$, $\boldsymbol{\beta}^{(\mathrm{b})}$ deviates in only one coefficient, but its location is unknown. So the original alternative hypothesis in Eq. (4) becomes

$$H_1 : \ \boldsymbol{\beta}^{(\mathrm{b})} = \boldsymbol{\beta}^{(\mathrm{a})} + \mathbf{d}_1\delta_1 \ \text{ or } \ldots \text{ or } \ \boldsymbol{\beta}^{(\mathrm{b})} = \boldsymbol{\beta}^{(\mathrm{a})} + \mathbf{d}_{h-1}\delta_{h-1}, \tag{8}$$

where δ_i ($i = 1, \ldots, h-1$) are the unknown shift magnitudes, and the possible shift direction vectors $\mathbf{d}_1, \ldots, \mathbf{d}_{h-1}$, which apply to $\beta_1^{(\mathrm{a})}, \ldots, \beta_{h-1}^{(\mathrm{a})}$, respectively, are defined similarly. In other words, hypothesis (8) states that $\boldsymbol{\beta}^{(\mathrm{b})}$ may differ from $\boldsymbol{\beta}^{(\mathrm{a})}$ in only one unknown coefficient, which may be any of the coefficients from the main factor effects up to the highest p-factor interaction effect. Since the GLRT derived from hypothesis (8) exploits more constructive information about potential shift directions, it should be more powerful than that from hypothesis (4).

Let us now generalize hypothesis (8). Hypothesis (8) considers shifts in effects of all orders, but it is believed that usually deviations containing fewer factors occur more frequently. It follows reasonably that most shifts appear in lower order effects such as main effects and two-factor interaction effects, instead of higher order ones. So we may pay sufficient attention to effects of the first few, say q, orders. This allows focusing detection power on a limited subspace with improved sensitivity. For the log-linear model (2), let the set of coefficient indexes corresponding to the effects of the first q orders be $\mathbb{I}_q$ ($1 \le q \le p$). Then, hypothesis (8) has a coefficient set $\mathbb{I}_p$, and it can further be generalized as

$$H_0 : \ \boldsymbol{\beta}^{(\mathrm{b})} = \boldsymbol{\beta}^{(\mathrm{a})} \ \text{ versus } H_1 : \ \bigcup_{i \in \mathbb{I}_q} \left(\boldsymbol{\beta}^{(\mathrm{b})} = \boldsymbol{\beta}^{(\mathrm{a})} + \mathbf{d}_i\delta_i\right), \quad 1 \le q \le p. \tag{9}$$

If $q = p$, the alternative hypothesis in Eq. (9) is equivalent to hypothesis (8). Since shifts usually occur in low-order effects, it is clear that the larger q is, the less powerful the GLRT based on hypothesis (9) will be. This follows because the alternative hypothesis considers redundant high-order shift directions. If the real shift indeed appears in an effect of the first q orders, the GLRT will certainly be powerful. Even if a shift occurs in an effect of an order higher than q, this change will be reflected to a large extent by the derived -2LRT statistic, and the GLRT will still be effective.

To obtain the -2LRT statistic λ for testing hypothesis (9), we need first to derive the -2LRT statistic λ_i for testing hypothesis (7), where it is assumed that only the coefficient β_i shifts. The MLEs of $\mathbf{p}_\mathrm{a}$ and $\mathbf{p}_\mathrm{b}$ under H_0 of hypothesis

(7) is the same as that in hypothesis (4), but their derivation under H_1 of hypothesis (7) is a little complex, so it is left in Appendix B. Actually, the key step is to get the MLE of δ_i, and from Appendix B it can be seen that the second-order partial derivative of the log-likelihood L with respect to δ_i under H_1 is smaller than or equal to zero. This means the log-likelihood L is concave with respect to δ_i, which guarantees the existence of the MLE of δ_i, hence the existence of the MLEs of $\mathbf{p}_a$ and $\mathbf{p}_b$ under H_1 of hypothesis (7). Given λ_i finalized in Eq. (22) in Appendix B, the -2LRT statistic for testing hypothesis (9) is

$$\lambda = \max_{i \in \mathbb{I}_q} \lambda_i. \tag{10}$$

This section shows how to derive the MLEs of $\mathbf{p}_a$ and $\mathbf{p}_b$ under both H_0 and H_1 of hypotheses (4) and (7). Their existence can also be guaranteed, which further insures that the two-sample multivariate categorical test and therefore the following proposed change-point detection methods can still be performed in a large and sparse contingency table. Here, sparsity means that there are quite a few zero cell counts in the table, which result from either some extremely small cell probabilities or a total sample size that is small relative to the number of cells in the table (Eriksson et al. [8] and Fienberg and Rinaldo [9]).

3.2. Directional Change-Point Detection

Based on the two-sample test with directional shift information, we can build the directional change-point detection method for Phase I analysis of MCPs. The prespecified log-linear model can be summarized as

$$\ln \mathbf{p} = \widetilde{\mathbf{X}}\widetilde{\boldsymbol{\beta}} \quad \text{and} \quad \mathbf{1}^T \mathbf{p} = 1.$$

Denote this model by $F(\widetilde{\mathbf{X}}; \widetilde{\boldsymbol{\beta}})$. In Phase I analysis, it is usually reasonable to assume that the given reference dataset is composed of multiple samples, say M, of size N, and each sample forms an observation vector of size $h \times 1$ subject to the multinomial distribution $\mathrm{MN}(N; \mathbf{p})$. Furthermore, it is also reasonably assumed that the jth multivariate sampling observation vector $\mathbf{n}_j$ is collected over time from the following change-point model

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

where τ is the unknown change-point, and $\widetilde{\boldsymbol{\beta}}^{(a)} \neq \widetilde{\boldsymbol{\beta}}^{(b)}$ are the unknown pre-change and post-change process coefficient vectors, respectively.

To detect the change-point, like Srivastava and Worsley [22] which aimed at detecting a change-point in the mean vector of a multivariate normal distribution, one natural idea is to integrate the GLRT with the binary segmentation procedure. Given the observation vectors $\mathbf{n}_j$ $(j = 1, \ldots, M)$, the change-point τ is unknown and may take any of the values $1, \ldots, M-1$. Therefore, for each possible τ, we pool both the pre-τ and post-τ samples together to form two large samples, respectively, and based on them compare the process coefficient vectors $\widetilde{\boldsymbol{\beta}}^{(a)}$ and $\widetilde{\boldsymbol{\beta}}^{(b)}$ by performing the two-sample test for hypothesis (4) in multivariate categorical distributions. If the maximum of the $M-1$ statistics is large enough, there will be a change-point during the process. To be specific, let

$$\mathbf{n}_k^A = \sum_{j=1}^{k} \mathbf{n}_j, \quad \mathbf{n}_k^B = \sum_{j=k+1}^{M} \mathbf{n}_j,$$
$$N_k^A = kN, \qquad N_k^B = (M-k)N. \tag{11}$$

Actually, $\mathbf{n}_k^A$ and $\mathbf{n}_k^B$ can be regarded as being collected from $\mathrm{MN}(N_k^A; \mathbf{p}_a)$ and $\mathrm{MN}(N_k^B; \mathbf{p}_b)$, respectively, where the probability vectors $\mathbf{p}_t, t = a, b$ are determined by $\widetilde{\boldsymbol{\beta}}^{(t)}$. This results in a two-sample test problem. By replacing $\mathbf{n}_a$, $\mathbf{n}_b$, N_a, and N_b in Eq. (5) by $\mathbf{n}_k^A$, $\mathbf{n}_k^B$, N_k^A, and N_k^B in Eq. (11), respectively, we obtain a -2LRT statistic Θ_k. If the maximum

$$\Theta = \max_{k \in \{1, \ldots, M-1\}} \Theta_k$$

exceeds a critical value, an alarm will sound a change-point. Accordingly, the change-point τ may be estimated as

$$\hat{\tau} = \arg\max_{k \in \{1, \ldots, M-1\}} \Theta_k. \tag{12}$$

The above change-point detection approach is based on the two-sample test without considering directional shift information. In reality, however, practical shifts tend to appear in only a few effects, leading to deviations of only a few coefficient subvectors. In change-point detection, such knowledge about the most likely shift directions should be exploited sufficiently. Similar ideas also appeared in Zou and Tsung [26] and Zou et al. [27], which both dealt with multistage process monitoring and diagnosis by capitalizing on information about shift directions from the first stage to the last one. To this end, we combine the -2LRT statistic in Eq. (10) for testing hypothesis (9) with the foregoing binary segmentation method. This leads to the change-point detection test statistic

$$\Lambda = \max_{i \in \mathbb{I}_q} \Lambda_i = \max_{i \in \mathbb{I}_q} \max_{k \in \{1, \ldots, M-1\}} \Lambda_{i,k},$$

where $\Lambda_{i,k}$ is obtained in a similar fashion to λ_i in Eq. (10) by using $\mathbf{n}_k^A$, $\mathbf{n}_k^B$, N_k^A, and N_k^B in Eq. (11).

There remains an issue about the choice of $\mathbb{I}_q$. Generally, if indeed most shifts arise in lower order effects, the GLRT

for testing hypothesis (9) will be less powerful as more shift directions are included. If q is chosen to be 1, then only shifts in main effects are considered, ignoring correlations between factors. Clearly, this is inappropriate. In practice, most applications pay attention to the first two moments such as means and variances because of their importance in describing a distribution. Moreover, higher moments usually cannot be accurately estimated given a dataset of moderate size. Therefore, we recommend considering only main effects and two-factor interactions and let $\mathbb{I}_q$ be $\mathbb{I}_2$. If the shift indeed occurs in a main or two-factor interaction effect, the GLRT for testing hypothesis (9) with $\mathbb{I}_2$ should definitely be powerful. Further, even if a shift appears in a three-factor or higher order interaction effect, this GLRT indexed by $\mathbb{I}_2$ will still be influenced to a fairly large extent, so it would still be detected powerfully. Therefore, the test statistic for directional change-point detection can be finalized as

$$\Lambda = \max_{i\in\mathbb{I}_2} \Lambda_i = \max_{i\in\mathbb{I}_2} \max_{k\in\{1,\ldots,M-1\}} \Lambda_{i,k} = \max_{k\in\{1,\ldots,M-1\}} \max_{i\in\mathbb{I}_2} \Lambda_{i,k}. \tag{13}$$

Similar to the change-point estimation in Eq. (12), a relevant diagnostic approach could also be developed based on the directional detection method, which aims at identifying both the change-point and the shift direction. However, unlike detection which uses an index subset $\mathbb{I}_2$ and can only tell whether there is a change-point or not, diagnosis needs to recognize the shift direction, which may indeed be in an effect of an order higher than two. Therefore, it is necessary to select a candidate subset of diagnostic index shift directions larger than $\mathbb{I}_2$ in case the real shift directions are left out. The subset for diagnosis should be at least $\mathbb{I}_3$, so that it is still safe even if a shift indeed occurs in a three-factor interaction effect. Conversely, it is believed that shifts in four-factor or higher order interaction effects are rare. In addition, the diagnostic consistency will decrease as more shift directions are included into candidates. Hence, the candidate diagnostic subset of shift directions is suggested as $\mathbb{I}_3$. Accordingly, the change-point τ can be estimated as

$$\hat{\tau} = \arg\max_{k\in\{1,\ldots,M-1\}} \left(\max_{i\in\mathbb{I}_3} \Lambda_{i,k} \right). \tag{14}$$

Based on the estimated change-point $\hat{\tau}$, the shift direction ζ can further be identified as

$$\hat{\zeta} = \arg\max_{i\in\mathbb{I}_3} \Lambda_{i,\hat{\tau}}. \tag{15}$$

It should be emphasized that the performance of such change-point estimation depends on the location of the true change-point. To obtain an accurate estimation, it is usually required that both the pre-τ and post-τ samples are sufficiently large. From the viewpoint of asymptotics, τ needs to satisfy $\tau/M \to \eta \in (0,1)$ as $M \to \infty$, to achieve some sense of consistency (Csörgő and Horváth [6]). If there are too few data on one side of τ, the observations on the other side of τ are almost equivalent to the whole dataset, and the test and detection of the change-point would be rather difficult.

3.3. Approximation of Significance Levels

The derived -2LRT statistics Θ and Λ for change-point detection need to be compared with their critical values. However, since the exact distributions of these statistics seem impossible to find, in this section we suggest a suitable approach to approximate their critical values or equivalently the significance levels.

By noticing $\Lambda = \max_{i\in\mathbb{I}_2} \Lambda_i$ in Eq. (13), the significance level of Λ can be approximated based on that of Λ_i. The significance levels of Θ and Λ_i are in fact approximated in the same way. Here, $\Theta = \max_k \Theta_k$ is the maximal LRT statistic without integrating directional information, and Θ_k is the -2LRT statistic for testing hypothesis (4) where $\widetilde{\boldsymbol{\beta}}^{(\mathrm{b})}$ differs from $\widetilde{\boldsymbol{\beta}}^{(\mathrm{a})}$ in $h-1$ independent parameters. Similarly, $\Lambda_i = \max_k \Lambda_{i,k}$ is the maximal LRT statistic with a given shift direction, and $\Lambda_{i,k}$ is the -2LRT statistic for testing hypothesis (7) where $\widetilde{\boldsymbol{\beta}}^{(\mathrm{b})}$ deviates from $\widetilde{\boldsymbol{\beta}}^{(\mathrm{a})}$ in only one independent parameter. So the difference between Θ and Λ_i lies mainly in the number of changed parameters.

According to Theorem 1.3.2 in Csörgő and Horváth [6], this book recommended an accurate approximation for the p-value of a maximal LRT statistic Z such as Θ or Λ_i, which is

$$\Pr\left(Z^{\frac{1}{2}} > x\right) = \frac{x^d \exp(-x^2/2)}{2^{d/2}\Gamma(d/2)} \left(\ln s - \frac{d}{x^2} \ln s + \frac{4}{x^2} \right) \tag{16}$$

with $d = h-1$ for $Z = \Theta$ and $d = 1$ for $Z = \Lambda_i$ $(i \in \mathbb{I}_2)$, where $s = (1-b_1)(1-b_2)/(b_1 b_2)$ and $b_1 = b_2 = (\ln M)^{\frac{3}{2}}/M$.

Furthermore, based on the approximate p-values of Λ_i, let us now turn to the significance level of $\Lambda = \max_{i\in\mathbb{I}_2} \Lambda_i$. Generally, these Λ_is are correlated, and therefore the analytical form of $\Pr(\Lambda > c)$ seems rather difficult to obtain. One simple and natural idea is to use the classical Bonferroni procedure in terms of rejecting $\Lambda \geq c$ with a desired type I error α if any $\Pr(\Lambda_i > c) \leq \alpha/K$ $(i \in \mathbb{I}_2)$, where K is the cardinality of the subset $\mathbb{I}_2$. If K is small, the Bonferroni procedure would have a false alarm rate close to α. However, it is rather conservative if K is large and the Λ_is are highly correlated. Hence, we recommend using the famous modified Bonferroni procedure (Simes [20]), which is less conservative than the classical one but is still simple to apply.

Table 1. Performance of the modified Bonferroni procedure in approximating the significance levels of Λ.

α	M	$i \in \mathbb{I}_1$, $K = 5$	$i \in \mathbb{I}_2$, $K = 14$	$i \in \mathbb{I}_3$, $K = 21$
	40	0.045	0.036	0.035
0.05	80	0.046	0.044	0.042
	120	0.051	0.046	0.042
	40	0.089	0.077	0.075
0.10	80	0.096	0.084	0.079
	120	0.101	0.089	0.082

In addition, the modified procedure is superior to the classical one when the test statistics involved are highly correlated. The modified Bonferroni procedure for these Λ_i ($i \in \mathbb{I}_2$) are summarized as follows:

Step 1. Calculate Λ_i for each $i \in \mathbb{I}_2$ based on $\Lambda_{i,k}$ ($k = 1, \ldots, M-1$);

Step 2. Use Eq. (16) with $d = 1$ to compute the approximate p-values $\hat{p}_i$ for each Λ_i, $i \in \mathbb{I}_2$;

Step 3. Given a desired type I error α, reject the null hypothesis if $\hat{p}_{(i)} \leq i\alpha/K$ for at least one i, where $\hat{p}_{(1)} \leq \ldots \leq \hat{p}_{(K)}$ are the ordered values of $\hat{p}_1, \ldots, \hat{p}_K$.

Table 1 lists the approximate false alarm rate $\hat{\alpha}$ of Λ in various cases, given a desired type I error α and simulated by applying the modified Bonferroni procedure and Eq. (16). Here, we consider four factors with 2, 2, 2, and 3 levels, as shown in Section 4.2. In addition, for simulation purpose, we also consider shift directions with indexes in the subsets $\mathbb{I}_1$, $\mathbb{I}_2$, and $\mathbb{I}_3$ with their cardinalities $K = 5$, 14, and 21, respectively. The results are based on 5000 replicated simulations. Clearly, the false alarm rate $\hat{\alpha}$ increases as the number K of changed parameters decreases and the number M of samples increases. These simulated $\hat{\alpha}$s are close to their desired value α. Even for the worst case of $K = 21$ and $M = 40$, the method still produces a good approximation.

4. PERFORMANCE ASSESSMENT

In this section, we investigate the powers of the two change-point detection methods, which are based on the two-sample test with and without integrating directional shift information. The investigation is done in both multivariate binomial and multivariate multinomial settings. In the multivariate binomial context, we also compare the two proposed methods with the χ^2-chart developed in Patel [18]. After detection, we test the corresponding diagnostic performance of the two proposed approaches in identifying the change-point τ and the shift direction ζ. Finally, by revisiting the AEC example introduced previously, we illustrate the implementation of the proposed directional change-point detection and diagnostic approaches. All reported results are based on 5000 replicated simulations, and the desired type I error α is set as 0.05 for detection.

4.1. Power Comparison in a Multivariate Binomial Process

Suppose that a production process involves four quality characteristics each with two attribute levels of conforming and nonconforming. This results in a multivariate binomial process with four factors, hence a four-way contingency table of size $2 \times 2 \times 2 \times 2$. Also assume that before the change-point τ, the log-linear model has the coefficient vector

$$\widetilde{\boldsymbol{\beta}}^{(a)} = [\ \beta_0\ \ 0.89\ 0.89\ \ 0.92\ 0.90\ 0.10\ 0.08\ 0.03\ -0.12\ -0.05\ 0.10\ -0.06\ 0.07\ \ 0\ \ 0\ \ 0\ ]^T,$$

where β_0 is the intercept accommodating the constraint $\mathbf{1}^T\mathbf{p} = 1$. Based on $\widetilde{\boldsymbol{\beta}}^{(a)}$, the pre-change probability vector $\mathbf{p}_a$ can be further calculated. Assume that there are $M = 80$ samples of size $N = 600$, and that the change-point τ is set as 30.

The comparison is among the proposed two detection methods and the χ^2-chart approach, and we leave a brief introduction to the application of the χ^2-chart in Appendix C. As mentioned previously, shifts in the marginal distribution of one factor are represented by deviations of its main effect or the corresponding coefficient subvector, and shifts in the dependence among multiple factors are reflected by deviations of their interaction effect or the corresponding coefficient subvector. We suppose that at the change-point τ, only one coefficient, say β_i ($i \in \mathbb{I}_2$), in $\boldsymbol{\beta}^{(a)}$ shifts with an increment δ. In other words, this one-coefficient shift occurs in a main effect or a two-factor interaction effect. Table 2 tabulates the powers of the two change-point detection approaches "with" and "without" the knowledge of shift directions as well as the χ^2-chart. Hereafter, the three methods are represented by "With," "Without," and χ^2 for short in the tables. Note that the approximate false alarm rate $\hat{\alpha}$ of the directional method is 0.040, smaller than the $\hat{\alpha}$ 0.050 of the undirectional one. In spite of this, in the case of one-coefficient shifts such as deviations of $\beta_{(1)}$ corresponding to the main effect of factor C_1, $\beta_{(4)}$, $\beta_{(1,2)}$, $\beta_{(2,3)}$, $\beta_{(2,4)}$, and $\beta_{(3,4)}$, the power of the directional method is uniformly greater than those of the undirectional approach and the χ^2-chart. This is expected, bacause the directional method takes advantage of the information about one-coefficient shifts in the effects of the first two orders, whereas the undirectional approach has to consider all kinds of shifts. In addition, the χ^2-chart does not exploit the pre-change and post-change group information using binary segmentation, and therefore it shows little power.

Table 2. Power comparison for one-coefficient shifts in a multivariate binomial process.

	$\beta_{(1)}$			$\beta_{(4)}$			$\beta_{(1,2)}$		
δ	With	Without	χ^2	With	Without	χ^2	With	Without	χ^2
0.04	0.282	0.161	0.063	0.313	0.172	0.063	0.577	0.324	0.060
0.05	0.463	0.261	0.069	0.519	0.291	0.066	0.822	0.531	0.064
0.06	0.700	0.426	0.077	0.735	0.455	0.076	0.959	0.789	0.092
0.07	0.859	0.595	0.089	0.895	0.653	0.086	0.994	0.924	0.094
0.08	0.955	0.770	0.105	0.971	0.823	0.107	0.999	0.991	0.102
0.10	0.997	0.963	0.144	0.999	0.978	0.154	1.000	1.000	0.139
	$\beta_{(2,3)}$			$\beta_{(2,4)}$			$\beta_{(3,4)}$		
0.04	0.638	0.364	0.063	0.607	0.324	0.066	0.552	0.296	0.064
0.05	0.890	0.638	0.076	0.863	0.581	0.073	0.795	0.523	0.075
0.06	0.982	0.859	0.093	0.974	0.823	0.085	0.950	0.758	0.074
0.07	0.999	0.966	0.113	0.997	0.951	0.099	0.994	0.926	0.091
0.08	1.000	0.995	0.127	1.000	0.992	0.112	1.000	0.984	0.100
0.10	1.000	1.000	0.205	1.000	1.000	0.175	1.000	1.000	0.142

$\hat{\alpha} = 0.040$ for "With," 0.050 for "Without," 0.050 for χ^2-chart.

4.2. Power Comparison in a Multivariate Multinomial Process

Assume that there are four quality characteristics with the first three assessed as conforming or nonconforming and the last one evaluated as excellent, acceptable, or unacceptable. This leads to a multivariate multinomial process with four factors, which can be arranged into a four-way table of size $2 \times 2 \times 2 \times 3$. Also suppose that before the change-point τ, the log-linear model has the coefficient vector

$$\widetilde{\boldsymbol{\beta}}^{(a)} = [\ \begin{matrix} \beta_0 & 0.86 & 0.89 & 0.82 & 0.72 & 0.08 \\ 0.10 & 0.12 & 0.12 & -0.13 & 0.10 & -0.06 \\ 0.07 & 0.16 & -0.14 & 0.13 & -0.10 & -0.08 \\ -0.04 & -0.07 & -0.11 & -0.05 & 0 & 0 \end{matrix}\ ]^T,$$

Table 3. Power comparison for one-coefficient shifts of the first two orders.

	$\beta_{(2)}$		$\beta_{(4_1)}$		$\beta_{(1,3)}$	
δ	With	Without	With	Without	With	Without
0.02	0.121	0.081	0.210	0.107	0.171	0.100
0.03	0.332	0.151	0.555	0.239	0.486	0.221
0.04	0.633	0.314	0.876	0.503	0.847	0.488
0.05	0.888	0.553	0.985	0.809	0.981	0.785
0.06	0.985	0.804	1.000	0.967	0.999	0.950
0.07	0.998	0.944	1.000	0.999	1.000	0.994
	$\beta_{(1,4_2)}$		$\beta_{(2,3)}$		$\beta_{(3,4_1)}$	
0.02	0.130	0.085	0.192	0.102	0.214	0.110
0.03	0.339	0.154	0.532	0.247	0.607	0.273
0.04	0.673	0.327	0.875	0.532	0.910	0.574
0.05	0.908	0.562	0.988	0.832	0.993	0.870
0.06	0.986	0.825	0.999	0.969	1.000	0.984
0.07	0.999	0.953	1.000	0.998	1.000	0.999

$\hat{\alpha} = 0.044$ for "With" and 0.048 for "Without."

where β_0 is the intercept accommodating the constraint $\mathbf{1}^T\mathbf{p} = 1$. Throughout the simulations there are $M = 80$ samples of size $N = 1200$, and the change-point τ is 30. Note that the χ^2-chart method does not apply to this case, so we only compare the proposed two change-point detection approaches.

Let us consider first that at the change-point τ, there is a one-coefficient shift of magnitude δ in β_i of $\boldsymbol{\beta}^{(a)}$ ($i \in \mathbb{I}_2$), which means that this shift occurs in a main or two-factor interaction effect. The comparison results are listed in Table 3 for the directional and undirectional detection approaches. Similar to Table 2 in a multivariate binomial setting, in the case of one-coefficient shifts such as deviations of $\beta_{(2)}$, $\beta_{(4_1)}$, $\beta_{(1,3)}$, $\beta_{(1,4_2)}$, $\beta_{(2,3)}$, and $\beta_{(3,4_1)}$, the directional method outperforms uniformly the undirectional approach in terms of a higher power, which is natural.

Now we turn to the performance of the two methods if at the change-point τ the only shift occurs in a three-factor interaction effect. The results are given in Table 4, where shifts in $\beta_{(1,2,3)}$, $\beta_{(1,3,4_2)}$, and $\beta_{(2,3,4_1)}$ are considered. Since shifts in a third-order effect can still have a strong influence on the

Table 4. Power comparison for one-coefficient shifts of the third order.

	$\beta_{(1,2,3)}$		$\beta_{(1,3,4_2)}$		$\beta_{(2,3,4_1)}$	
δ	With	Without	With	Without	With	Without
0.02	0.157	0.116	0.102	0.075	0.197	0.125
0.03	0.406	0.289	0.229	0.151	0.490	0.315
0.04	0.733	0.623	0.461	0.307	0.817	0.644
0.05	0.939	0.900	0.721	0.571	0.975	0.920
0.06	0.992	0.989	0.907	0.819	0.999	0.992
0.07	1.000	1.000	0.979	0.959	1.000	1.000

$\hat{\alpha} = 0.044$ for "With" and 0.048 for "Without."

Table 5. Power comparison for two-coefficient shifts of the first three orders.

		$\beta_{(1)}$	$\beta_{(4_2)}$	$\beta_{(2)}$	$\beta_{(1,4_1)}$	$\beta_{(3)}$	$\beta_{(2,3)}$	$\beta_{(1,2)}$	$\beta_{(1,3)}$
δ_1	δ_2	With	Without	With	Without	With	Without	With	Without
0.02	0.02	0.180	0.115	0.285	0.162	0.599	0.290	0.589	0.303
0.02	0.04	0.651	0.358	0.915	0.653	0.987	0.839	0.973	0.804
0.04	0.02	0.554	0.335	0.707	0.450	0.957	0.715	0.983	0.833
0.04	0.04	0.799	0.634	0.960	0.867	1.000	0.986	1.000	0.990
		$\beta_{(2,3)}$	$\beta_{(3,4_1)}$	$\beta_{(2,4_1)}$	$\beta_{(2,4_2)}$	$\beta_{(1,2,4_1)}$	$\beta_{(1,2,4_2)}$	$\beta_{(1,3,4_1)}$	$\beta_{(2,3,4_1)}$
0.02	0.02	0.455	0.256	0.532	0.255	0.446	0.263	0.783	0.504
0.02	0.04	0.959	0.787	0.914	0.647	0.852	0.691	0.996	0.970
0.04	0.02	0.935	0.753	0.986	0.832	0.952	0.857	0.996	0.966
0.04	0.04	0.997	0.974	0.999	0.977	0.995	0.984	1.000	0.999

$\hat{\alpha} = 0.044$ for "With" and 0.048 for "Without."

test statistic indexed by $\mathbb{I}_2$, Table 4 shows that the directional method still uniformly outperforms the undirectional one.

We come next to the case of two-coefficient shifts to verify the robustness of the directional method, which specializes in detecting one-coefficient shifts. Here, two coefficients increase by δ_1 and δ_2, for example, $\beta_{(1)} + \delta_1$ and $\beta_{(4_2)} + \delta_2$ simultaneously at the change-point τ. Table 5 demonstrates the powers of the two detection approaches under different combinations of coefficients of the first three orders. We can see that the directional method is still more powerful than the undirectional one, reflecting robustness to shifts in more than one coefficient.

4.3. Diagnostic Performance

The diagnostic performance is investigated in this subsection, under the same parameter settings as in Section 4.2 for a multivariate multinomial process, in which the true change-point τ is set as 30. The change-point τ is estimated based on Eq. (12) for the method without directional shift information and on Eq. (14) for the one with such information. These two schemes are compared in terms of the bias (Bias) as the difference of the average minus the true one, the standard deviation (Std) of change-point estimates $\hat{\tau}$, and the probabilities $\Pr(|\hat{\tau} - \tau| \leq 1)$ and $\Pr(|\hat{\tau} - \tau| \leq 2)$ (denoted by $\Pr_1$ and $\Pr_2$, respectively), which quantify the consistency of change-point identification. Conversely, the shift direction estimate $\hat{\zeta}$ is derived using only Eq. (15) for the directional approach, and to the best of our knowledge, there is currently no method for comparison in recognizing shift directions. Here, the matching probability that the estimated index $\hat{\zeta}$ of the only deviated coefficient β_ζ is indeed the true one ζ, $\Pr(\hat{\zeta} = \zeta)$ denoted by $\Pr_\zeta$, is selected for measuring the accuracy of identifying shift directions. Note that for diagnosis the directional method uses the one-coefficient shift direction index subset $\mathbb{I}_3$ rather than $\mathbb{I}_2$, including coefficients in the main, two-factor interaction, and three-factor interaction effects as the candidate diagnostic shift directions.

Table 6 presents the results for some one-coefficient shifts in the effects of the first three orders, such as $\beta_{(2)}, \beta_{(1,3)}, \beta_{(3,4_1)}$, and $\beta_{(1,3,4_2)}$. It can be seen that the estimator (14) for the directional method is always more consistent than the estimator (12) in recognizing the change-point, giving smaller biases, smaller standard deviations, and higher probabilities $\Pr_1$ and $\Pr_2$. Again, this follows because the estimator (14) exploits fully the directional information. As for the shift direction, the estimator (15) derived from the directional approach performs fairly well in that it has a high matching probability of $\hat{\zeta} = \zeta$, and certainly this probability increases as the shift magnitude δ increases. In addition, according to the practical definition of shifts, estimating the shift direction may provide some insights into MCPs. For example, if the shift direction index is estimated as $\hat{\zeta} = 7$, by the relation $\beta_7 = \beta_{(1,3)}$, we may conclude that the interaction of factors C_1 and C_3 deviates at the change-point.

4.4. Revisiting the AEC Example

Here, we turn to the AEC example introduced previously to show how to implement the proposed directional detection and diagnostic methodology in practice. The AEC example has $p = 3$ factors: LC, DF, and CAP, and each has two levels of conforming and nonconforming. The cross-classification counts with all factor level combinations are stored in a three-way contingency table with $2^3 = 8$ cells. The following log-linear model can appropriately describe the relationship between the cell probabilities and the level combinations:

$$\ln p_{a_1a_2a_3} = \beta_0 + \beta_{(1)}x_1 + \beta_{(2)}x_2 + \beta_{(3)}x_3 + \beta_{(1,2)}x_1x_2 + \beta_{(1,3)}x_1x_3 + \beta_{(2,3)}x_2x_3 + \beta_{(1,2,3)}x_1x_2x_3,$$

where $a_1 = 1, 2; a_2 = 1, 2; a_3 = 1, 2$, all the cell probabilities $p_{a_1a_2a_3}$ sum up to 1, and x_1, x_2, x_3 take 1 or -1 representing the two levels of the three factors LC, DF, and CAP, respectively. There is a reference dataset of the multivariate sampling observations, which involves $M = 120$

Table 6. Diagnostic comparison for one-coefficient shifts of the first three orders.

		Bias		Std		Pr_1		Pr_2		Pr_ζ
β	δ	With	Without	With	Without	With	Without	With	Without	With
	0.03	4.32	6.09	19.5	24.1	0.188	0.105	0.267	0.157	0.522
	0.04	1.70	4.30	13.2	19.9	0.354	0.197	0.461	0.271	0.765
$\beta_{(2)}$	0.05	0.452	2.05	7.58	14.5	0.516	0.320	0.643	0.421	0.917
	0.06	0.205	0.975	4.37	9.74	0.633	0.453	0.761	0.582	0.981
	0.08	0.013	0.083	1.87	3.53	0.787	0.685	0.882	0.803	0.999
	0.10	0.003	0.027	1.17	1.63	0.888	0.830	0.952	0.914	1.000
	0.03	2.89	5.77	16.2	22.4	0.269	0.135	0.361	0.186	0.633
	0.04	1.10	2.75	9.06	16.0	0.469	0.280	0.595	0.376	0.874
$\beta_{(1,3)}$	0.05	0.073	0.968	4.50	10.2	0.622	0.437	0.751	0.564	0.968
	0.06	0.045	0.203	2.75	5.47	0.725	0.593	0.835	0.712	0.992
	0.08	0.025	0.055	1.32	1.99	0.857	0.801	0.937	0.893	1.000
	0.10	−0.005	0.016	0.809	1.01	0.937	0.913	0.976	0.965	1.000
	0.03	2.13	4.75	14.1	20.6	0.309	0.184	0.418	0.254	0.565
	0.04	0.581	2.14	7.17	14.4	0.516	0.343	0.639	0.438	0.797
$\beta_{(3,4_1)}$	0.05	0.126	0.795	3.78	8.10	0.679	0.531	0.791	0.648	0.924
	0.06	0.021	0.159	2.09	4.00	0.775	0.670	0.880	0.785	0.972
	0.08	−0.009	0.002	1.04	1.37	0.899	0.854	0.962	0.933	0.997
	0.10	0.007	0.017	0.634	0.747	0.960	0.945	0.988	0.981	1.000
	0.03	4.32	6.80	19.7	24.4	0.198	0.098	0.269	0.143	0.412
	0.04	1.42	4.27	12.6	20.0	0.357	0.195	0.463	0.269	0.646
$\beta_{(1,3,4_2)}$	0.05	0.732	2.19	7.74	14.2	0.514	0.332	0.636	0.433	0.827
	0.06	0.141	0.878	4.23	9.37	0.648	0.474	0.769	0.595	0.921
	0.08	0.030	0.066	1.82	3.05	0.804	0.716	0.896	0.829	0.988
	0.10	0.014	0.037	1.09	1.48	0.891	0.849	0.953	0.925	0.999

samples of size $N = 500$. Usually, the sample size can be chosen as tens of the number of cells. The observation vector $\mathbf{n}_k$ ($k = 1, \ldots, 120$) is of size 8×1, for instance, $\mathbf{n}_1 = [483, 14, 2, 0, 1, 0, 0, 0]^T$. The elements of $\mathbf{n}_k$ jointly follow a multinomial distribution MN(500; $\mathbf{p}$), where $\mathbf{p}$ is a cell probability vector and may shift at a change-point.

In Phase I, the main task is to check whether there are unusual patterns in the reference dataset. Let us now use the proposed directional detection method to see if a change-point exists. We select the index subset $\mathbb{I}_2$, because shifts tend to occur in main effects and two-factor interaction effects. In addition, the desired type I error α is set as 0.05. By first fixing each $i \in \mathbb{I}_2$, for each $k = 1, \ldots, 119$, based on the grouped data $\sum_{j=1}^{k} \mathbf{n}_j$ and $\sum_{j=k+1}^{120} \mathbf{n}_j$, we use the method given in Appendix B to calculate the -2LRT statistic $\Lambda_{i,k}$ for testing hypothesis (7). Then, for each $i \in \mathbb{I}_2$, we obtain $\Lambda_i = \max_k \Lambda_{i,k}$, which are 2.013, 7.099, 17.33, 5.081, 20.17, and 11.06. By using Eq. (16) with $d = 1$, we calculate the approximate p-values of these Λ_i as 0.900, 0.140, 0.001, 0.323, 0.000, and 0.024, respectively. Arranging these p-values from the minimum to the maximum as $\hat{p}_{(1)}, \ldots, \hat{p}_{(6)}$ ($\mathbb{I}_2$ has a cardinality $K = 6$), we find that $\hat{p}_{(i)} \le i\alpha/6$ holds for $i = 1, 2$. Therefore, there is a change-point in the reference dataset.

Since a change-point exists, the next step is to identify its location and the shift direction, with the assumption that only one coefficient deviates in the effects of the first three orders. Here, the proposed directional diagnostic approach adopts the candidate index subset $\mathbb{I}_3$ for the only potential shift direction. Still, for each $i \in \mathbb{I}_3$ and $k = 1, \ldots, 119$, we calculate the -2LRT statistic $\Lambda_{i,k}$ for testing hypothesis (7). For each k, by calculating $\max_{i \in \mathbb{I}_3} \Lambda_{i,k}$, we see that $\max_{i \in \mathbb{I}_3} \Lambda_{i,48}$ is the maximum among all $\max_{i \in \mathbb{I}_3} \Lambda_{i,k}$ ($k = 1, \ldots, 119$). According to the estimator (14), the change-point is identified as $\hat{\tau} = 48$. For $\Lambda_{i,48}$ ($i \in \mathbb{I}_3$), it turns out that the maximum among them is $\Lambda_{5,48}$. Hence, the shift direction index is recognized as $\hat{\zeta} = 5$ based on the estimator (15), and the shifted coefficient is $\beta_5 = \beta_{(1,3)}$, which represents the interaction effect of factors C_1 and C_3. In the case of the AEC example, this corresponds to a shift in the dependence between the factors LC and CAP.

For comparison, the application to this AEC example of the proposed undirectional change-point method and the χ^2-chart in Patel [18] are also illustrated here. For the undirectional method, obtain $\Theta = \Theta_{51} = 24.94$ as the maximum among Θ_k ($k = 1, \ldots, 119$). By using Eq. (16) with $d = 8 - 1 = 7$, the approximated p-values is further calculated as 0.028, smaller than 0.05. Hence a change-point may exist in the dataset, and the change-point is identified as $\hat{\tau} = 51$ using the estimator (12). Conversely, we construct a χ^2-chart based on the 120 samples in this dataset and get a charting statistic for each sample. The maximum among the

120 charting statistics is 17.56. In the IC state, each charting statistic should follow approximately the $\chi^2(3)$ distribution with d.f. 3, since three factors are involved in the AEC example. The desired type I error α is 0.05, so the control limit of the χ^2-chart may be selected to be $(1-0.05)^{\frac{1}{120}}$th quantile of the $\chi^2(3)$ distribution, which is 18.06 and larger than the maximum among the 120 charting statistics. So the χ^2-chart does not signal a change-point.

5. CONCLUSION

This study has developed a new log-linear directional change-point detection methodology for Phase I analysis of multivariate categorical processes. The method integrates the knowledge of potential shift directions, which are formulated as deviations of coefficient subvectors in log-linear modeling. In addition, a post-signal diagnostic scheme for recognizing the change-point location and the shift direction has been proposed. Both the detection and diagnostic approaches work well in multivariate categorical processes that follow multivariate binomial and multivariate multinomial distributions. Monte Carlo simulations have shown high detection power and good diagnostic consistency.

It should be pointed out that the proposed approach can be readily extended for detecting multiple change-points by using the binary segmentation method recursively (Yao [24]). Moreover, the presence of outliers has seriously adverse effects on the modeling and monitoring of MCPs. Therefore, outlier detection procedures, which aim at identifying any abnormal samples from a dataset, are quite important and deserve future research. In addition, the proposed change-point detection method can be extended to Phase II SPC. In the literature, some authors have developed change-point schemes for online monitoring (see Hawkins et al. [10] and Lai and Xing [12]) without the knowledge of pre- and post-change parameters. In an ongoing effort, we are developing a control scheme that integrates sequential change-point detection and the proposed directional change-point test based on log-linear models, which might be expected to be more powerful in detecting shifts at a certain expense of computational efforts.

APPENDIX A

Derivation of the Design Matrix $\widetilde{\mathbf{X}}$

Here, we take four factors C_1, C_2, C_3, and C_4 with 2, 2, 2, and 3 levels, respectively, for illustration. Let

$$\mathbf{1}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \mathbf{1}_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{J}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix},$$

$$\mathbf{J}_3 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{bmatrix} = \begin{bmatrix} \mathbf{I}_2 \\ -\mathbf{1}_2^T \end{bmatrix}.$$

Note that the column sums of matrixes $\mathbf{J}_2$ and $\mathbf{J}_3$ are all zeros, which assures identifiability. For instance, the design submatrix corresponding to the main effect of C_3 is $\mathbf{1}_2 \otimes \mathbf{1}_2 \otimes \mathbf{J}_2 \otimes \mathbf{1}_3$, where $\otimes$ is Kronecker product operator. The design submatrix corresponding to the two-factor interaction effect of C_2C_4 is $\mathbf{1}_2 \otimes \mathbf{J}_2 \otimes \mathbf{1}_2 \otimes \mathbf{J}_3$, and the design submatrix corresponding to the three-factor interaction effect of $C_1C_3C_4$ is $\mathbf{J}_2 \otimes \mathbf{1}_2 \otimes \mathbf{J}_2 \otimes \mathbf{J}_3$. All the other design submatrixes can be constructed similarly, and they all combine to form $\widetilde{\mathbf{X}}$. In a word, given $\mathbf{1}_2 \otimes \mathbf{1}_2 \otimes \mathbf{1}_2 \otimes \mathbf{1}_3$ with the subscripts corresponding to the number of attribute levels of each factor, the design submatrix corresponding to an effect is obtained by replacing $\mathbf{1}$ with $\mathbf{J}$ with appropriate dimensions at all the positions where the factors are contained in this effect.

APPENDIX B

Derivation of the -2LRT Statistic λ_i for Testing Hypothesis (7)

Under H_1 of Eq. (7), $\widetilde{\boldsymbol{\beta}}^{(a)}$ and $\widetilde{\boldsymbol{\beta}}^{(b)}$ have a relationship as described in Eq. (6). Because of the constraint $\mathbf{1}^T\mathbf{p}_t = 1$ ($t = a, b$), the following equations hold

$$\mathbf{1}^T \exp\left(\mathbf{1}\beta_0^{(a)} + \mathbf{X}\boldsymbol{\beta}^{(a)}\right) = 1,$$
$$\mathbf{1}^T \exp\left(\mathbf{1}\beta_0^{(a)} + \mathbf{X}\boldsymbol{\beta}^{(a)} + \mathbf{1}\alpha_i + \mathbf{x}_i\delta_i\right) = 1,$$

where $\mathbf{x}_i$ is the ith column of $\mathbf{X}$. So $\beta_0^{(a)}$ and α_i can be expressed as

$$\beta_0^{(a)} = -\ln\left(\mathbf{1}^T \exp\left(\mathbf{X}\boldsymbol{\beta}^{(a)}\right)\right),$$
$$\alpha_i = -\ln\left(\mathbf{1}^T \exp\left(\mathbf{1}\beta_0^{(a)} + \mathbf{X}\boldsymbol{\beta}^{(a)} + \mathbf{x}_i\delta_i\right)\right).$$

Again, assume that $\mathbf{n}_a$ and $\mathbf{n}_b$ are two independent random vectors subject to $\mathrm{MN}(N_a; \mathbf{p}_a)$ and $\mathrm{MN}(N_b; \mathbf{p}_b)$, respectively. The log-likelihood function of $\mathbf{p}_a$ and $\mathbf{p}_b$ can be written as

$$\begin{aligned} L &= \mathbf{n}_a^T \ln \mathbf{p}_a + \ln N_a! - \mathbf{1}^T \ln \mathbf{n}_a! + \mathbf{n}_b^T \ln \mathbf{p}_b + \ln N_b! - \mathbf{1}^T \ln \mathbf{n}_b! \\ &= \mathbf{n}_a^T \widetilde{\mathbf{X}}\widetilde{\boldsymbol{\beta}}^{(a)} + \ln N_a! - \mathbf{1}^T \ln \mathbf{n}_a! + \mathbf{n}_b^T \widetilde{\mathbf{X}}\widetilde{\boldsymbol{\beta}}^{(b)} + \ln N_b! - \mathbf{1}^T \ln \mathbf{n}_b! \\ &= (N_a + N_b)\beta_0^{(a)} + (\mathbf{n}_a + \mathbf{n}_b)^T \mathbf{X}\boldsymbol{\beta}^{(a)} + N_b\alpha_i + \mathbf{n}_b^T \mathbf{x}_i\delta_i \\ &\quad + \ln N_a! - \mathbf{1}^T \ln \mathbf{n}_a! + \ln N_b! - \mathbf{1}^T \ln \mathbf{n}_b!. \end{aligned}$$

To determine the MLEs of $\mathbf{p}_a$ and $\mathbf{p}_b$, the first-order partial derivatives of L with respect to both $\boldsymbol{\beta}^{(a)}$ and δ_i are required. After some simplifications, we have

$$\frac{\partial L}{\partial \boldsymbol{\beta}^{(a)}} = \mathbf{X}^T(\mathbf{n}_a + \mathbf{n}_b - N_a\mathbf{p}_a - N_b\mathbf{p}_b),$$
$$\frac{\partial L}{\partial \delta_i} = s(\delta_i) = \mathbf{x}_i^T \mathbf{n}_b - N_b \mathbf{x}_i^T \exp\Big[\mathbf{1}\beta_0^{(a)} + \mathbf{X}\boldsymbol{\beta}^{(a)} + \mathbf{x}_i\delta_i - \mathbf{1}\ln\left(\mathbf{1}^T \exp\left(\mathbf{1}\beta_0^{(a)} + \mathbf{X}\boldsymbol{\beta}^{(a)} + \mathbf{x}_i\delta_i\right)\right)\Big].$$

Let

$$\mathbf{k}(\delta_i) = \exp\Big[\mathbf{1}\beta_0^{(a)} + \mathbf{X}\boldsymbol{\beta}^{(a)} + \mathbf{x}_i\delta_i - \mathbf{1}\ln\Big(\mathbf{1}^T \exp\Big(\mathbf{1}\beta_0^{(a)} + \mathbf{X}\boldsymbol{\beta}^{(a)} + \mathbf{x}_i\delta_i\Big)\Big)\Big].$$

The second-order partial derivative of L with respect to δ_i can be formulated as

$$s'(\delta_i) = \frac{\partial^2 L}{\partial \delta_i^2} = -N_b \mathbf{x}_i^T \left[\mathrm{diag}(\mathbf{k}(\delta_i)) - \mathbf{k}(\delta_i)\mathbf{k}^T(\delta_i)\right]\mathbf{x}_i,$$

where $\mathrm{diag}(\mathbf{a})$ is the diagonal square matrix with the column vector $\mathbf{a}$ as its diagonal elements. Note that $\mathbf{k}(\delta_i)$ actually satisfies $\mathbf{1}^T\mathbf{k}(\delta_i) = 1$, and therefore it can be regarded as the probability vector of a multinomial distribution. In addition, the matrix $\mathrm{diag}(\mathbf{k}(\delta_i)) - \mathbf{k}(\delta_i)\mathbf{k}^T(\delta_i)$ is the covariance matrix of this multinomial distribution, hence semipositive definite. Thus, the second-order partial derivative of L with respect to δ_i

$$\frac{\partial^2 L}{\partial \delta_i^2} \le 0.$$

This means the log-likelihood L is concave with respect to δ_i, which guarantees the existence of the MLE of δ_i, hence the existence of the MLEs of $\mathbf{p}_\mathrm{a}$ and $\mathbf{p}_\mathrm{b}$ under H_1 of hypothesis (7).

It is obvious that

$$s(0) = \mathbf{x}_i^T\mathbf{n}_\mathrm{b} - N_\mathrm{b}\mathbf{x}_i^T\exp\left(\mathbf{1}\beta_0^{(\mathrm{a})} + \mathbf{X}\boldsymbol{\beta}^{(\mathrm{a})}\right) = \mathbf{x}_i^T(\mathbf{n}_\mathrm{b} - N_\mathrm{b}\mathbf{p}_\mathrm{a}),$$
$$s'(0) = -N_\mathrm{b}\mathbf{x}_i^T\mathrm{diag}(\mathbf{p}_\mathrm{a})\mathbf{x}_i + N_\mathrm{b}\mathbf{x}_i^T\mathbf{p}_\mathrm{a}\mathbf{p}_\mathrm{a}^T\mathbf{x}_i = -N_\mathrm{b}\mathbf{x}_i^T\boldsymbol{\Sigma}_\mathrm{a}\mathbf{x}_i,$$

where $\boldsymbol{\Sigma}_\mathrm{a} = \mathrm{diag}(\mathbf{p}_\mathrm{a}) - \mathbf{p}_\mathrm{a}\mathbf{p}_\mathrm{a}^T$. By performing the first-order Taylor expansion of $s(\delta_i)$ at $\delta_i = 0$, we have

$$s(\delta_i) \approx s(0) + s'(0)\delta_i.$$

To obtain the MLE of δ_i, let $s(\delta_i)$ be 0. Therefore, the MLE of δ_i should approximately satisfy

$$\delta_i \approx -\frac{s(0)}{s'(0)} = \frac{1}{N_\mathrm{b}}\left(\mathbf{x}_i^T\boldsymbol{\Sigma}_\mathrm{a}\mathbf{x}_i\right)^{-1}\mathbf{x}_i^T(\mathbf{n}_\mathrm{b} - N_\mathrm{b}\mathbf{p}_\mathrm{a}). \tag{17}$$

Conversely, by performing the first-order Taylor expansion of $\mathbf{p}_\mathrm{b}$ at $\mathbf{p}_\mathrm{a}$ (i.e., $\delta_i = 0$), we have

$$\begin{aligned}\mathbf{p}_\mathrm{b} &= \exp\left(\mathbf{1}\beta_0^{(\mathrm{a})} + \mathbf{X}\boldsymbol{\beta}^{(\mathrm{a})} + \mathbf{1}\alpha_i + \mathbf{x}_i\delta_i\right)\\ &= \exp\left(\mathbf{1}\beta_0^{(\mathrm{a})} + \mathbf{X}\boldsymbol{\beta}^{(\mathrm{a})} - \mathbf{1}\ln\left(\mathbf{1}^T\exp\left(\mathbf{1}\beta_0^{(\mathrm{a})} + \mathbf{X}\boldsymbol{\beta}^{(\mathrm{a})} + \mathbf{x}_i\delta_i\right)\right) + \mathbf{x}_i\delta_i\right)\\ &\approx \exp\left(\mathbf{1}\beta_0^{(\mathrm{a})} + \mathbf{X}\boldsymbol{\beta}^{(\mathrm{a})}\right) + \left(\mathrm{diag}\left(\mathbf{p}_\mathrm{a}y\right)\mathbf{x}_i - \mathbf{p}_\mathrm{a}\mathbf{p}_\mathrm{a}^T\mathbf{x}_i\right)\delta_i\\ &= \mathbf{p}_\mathrm{a} + \boldsymbol{\Sigma}_\mathrm{a}\mathbf{x}_i\delta_i.\end{aligned} \tag{18}$$

Therefore, similar to δ_i, the MLE of $\boldsymbol{\beta}^{(\mathrm{a})}$ should satisfy

$$\frac{\partial L}{\partial \boldsymbol{\beta}^{(\mathrm{a})}} \approx \mathbf{X}^T(\mathbf{n}_\mathrm{a} + \mathbf{n}_\mathrm{b} - N_\mathrm{a}\mathbf{p}_\mathrm{a} - N_\mathrm{b}\mathbf{p}_\mathrm{a} - N_\mathrm{b}\boldsymbol{\Sigma}_\mathrm{a}\mathbf{x}_i\delta_i) = \mathbf{0}. \tag{19}$$

By substituting δ_i in Eq. (17) into Eq. (19), we have

$$\begin{aligned}&\mathbf{X}^T[\mathbf{n}_\mathrm{a} + \mathbf{n}_\mathrm{b} - N_\mathrm{a}\mathbf{p}_\mathrm{a} - N_\mathrm{b}\mathbf{p}_\mathrm{a}\\ &\quad - \boldsymbol{\Sigma}_\mathrm{a}\mathbf{x}_i\left(\mathbf{x}_i^T\boldsymbol{\Sigma}_\mathrm{a}\mathbf{x}_i\right)^{-1}\mathbf{x}_i^T(\mathbf{n}_\mathrm{b} - N_\mathrm{b}\mathbf{p}_\mathrm{a})] = \mathbf{0}.\end{aligned} \tag{20}$$

After some calculations, Eq. (20) plus the constraint $\mathbf{1}^T\mathbf{p}_\mathrm{a} = 1$ leads to

$$\mathbf{p}_\mathrm{a} = \begin{bmatrix}\mathbf{1}^T\\ \mathbf{X}^T(N_\mathrm{a} + N_\mathrm{b}) - \mathbf{X}^T\boldsymbol{\Sigma}_\mathrm{a}\mathbf{A}_i N_\mathrm{b}\end{bmatrix}^{-1}\begin{bmatrix}1\\ \mathbf{X}^T(\mathbf{n}_\mathrm{a} + \mathbf{n}_\mathrm{b}) - \mathbf{X}^T\boldsymbol{\Sigma}_\mathrm{a}\mathbf{A}_i\mathbf{n}_\mathrm{b}\end{bmatrix}, \tag{21}$$

where $\mathbf{A}_i = \mathbf{x}_i(\mathbf{x}_i^T\boldsymbol{\Sigma}_\mathrm{a}\mathbf{x}_i)^{-1}\mathbf{x}_i^T$. Note that Eq. (21) may be regarded as $\mathbf{p}_\mathrm{a} = \mathbf{f}(\mathbf{p}_\mathrm{a})$.

By performing numerical iterations based on Eq. (21) until some stopping criterion is reached, we obtain the MLE $\widehat{\mathbf{p}}_\mathrm{a}^{H_1}$ of $\mathbf{p}_\mathrm{a}$ under H_1 and hence $\widehat{\boldsymbol{\Sigma}}_\mathrm{a}^{H_1}$. Generally, the initial value of $\mathbf{p}_\mathrm{a}$ in the above iteration can be set as $(\mathbf{n}_\mathrm{a} + \mathbf{n}_\mathrm{b})/(N_\mathrm{a} + N_\mathrm{b})$, and our simulations show that this iteration will stop after only a few steps. The MLE $\hat{\delta}_i^{H_1}$ of δ_i can be calculated from Eq. (17). Based on $\widehat{\mathbf{p}}_\mathrm{a}^{H_1}$ and $\hat{\delta}_i^{H_1}$, and according to Eq. (18), the MLE $\widehat{\mathbf{p}}_\mathrm{b}^{H_1}$ of $\mathbf{p}_\mathrm{b}$ under H_1 can also be derived. Under H_0 of Eq. (7), it is known that the MLEs of $\mathbf{p}_\mathrm{a}$ and $\mathbf{p}_\mathrm{b}$ are $\widehat{\mathbf{p}}_\mathrm{a}^{H_0} = \widehat{\mathbf{p}}_\mathrm{b}^{H_0} = (\mathbf{n}_\mathrm{a} + \mathbf{n}_\mathrm{b})/(N_\mathrm{a} + N_\mathrm{b})$. Based on these, the -2LRT statistic λ_i for testing (7) can finally be formulated as

$$\lambda_i = 2\mathbf{n}_\mathrm{a}^T\ln\widehat{\mathbf{p}}_\mathrm{a}^{H_1} + 2\mathbf{n}_\mathrm{b}^T\ln\widehat{\mathbf{p}}_\mathrm{b}^{H_1} - 2(\mathbf{n}_\mathrm{a} + \mathbf{n}_\mathrm{b})^T\ln\left(\frac{\mathbf{n}_\mathrm{a} + \mathbf{n}_\mathrm{b}}{N_\mathrm{a} + N_\mathrm{b}}\right). \tag{22}$$

APPENDIX C

The χ^2-chart Method in Patel [18]

Based on the M samples $\mathbf{n}_j$ $(j = 1, \ldots, M)$ each of size $h \times 1$, we calculate their average $\bar{\mathbf{n}}$ and use $\bar{\mathbf{p}} = \bar{\mathbf{n}}/N$ as the cell probability vector. The covariance matrix is obtained as $\overline{\boldsymbol{\Sigma}} = \mathrm{diag}(\bar{\mathbf{p}}) - \bar{\mathbf{p}}\bar{\mathbf{p}}^T$. Let $\mathbf{X}_P = [\mathbf{X}_1, \ldots, \mathbf{X}_p]^T$, which is the design submatrix corresponding to the main effects of the p factors. In a multivariate binomial setting, $\mathbf{X}_i$ $(i = 1, \ldots, p)$ actually all reduce into column vectors and contain only 1 and -1 as their entries. Further replace all -1s in $\mathbf{X}_P$ by 0, and denote this new matrix by $\mathbf{X}_\mathrm{M}$, that is, $\mathbf{X}_\mathrm{M} = \mathrm{I}$ $(\mathbf{X}_P > 0)$. The charting statistic for the jth sample is calculated as

$$R_j = \frac{1}{N}\left(\mathbf{n}_j - N\bar{\mathbf{p}}\right)^T\mathbf{X}_\mathrm{M}\left(\mathbf{X}_\mathrm{M}^T\overline{\boldsymbol{\Sigma}}\mathbf{X}_\mathrm{M}\right)^{-1}\mathbf{X}_\mathrm{M}^T\left(\mathbf{n}_j - N\bar{\mathbf{p}}\right).$$

The control limit can be selected by simulation to make an false alarm rate of 0.05 when there is no change-point. Then, the charting statistic for each sample is plotted in the chart against the control limit. If at least one charting statistic falls beyond the limit, there may be a change-point.

ACKNOWLEDGMENTS

The authors thank the editor, associate editor, and two anonymous referees for their many helpful comments that have resulted in significant improvements in the article. This work was completed when Li was a PhD student at the Hong Kong University of Science and Technology, whose much help is appreciated and acknowledged. Tsung's research was supported by the RGC Competitive Earmarked Research Grants 620010. Zou's research was supported by the NNSF of China Grants 11001138, 11071128, 11131002, and 11101306, the RFDP of China Grant 20110031110002, the PAPD of Jiangsu Higher Education Institutions, and Foundation for the Author of National Excellent Doctoral Dissertation of PR China H0512101.

REFERENCES

[1] T.W. Anderson, An introduction to multivariate statistical analysis, 3rd ed., New York, 2003.
[2] S. Bersimis, S. Psarakis, and J. Panaretos, Multivariate statistical process control charts: An overview, Qual Reliab Eng Int 23 (2007), 517–543.
[3] Y.M.M. Bishop, S.E. Fienberg, and P.W. Holland, Discrete multivariate analysis, Springer, New York, 2007.
[4] J. Chiu and T. Kuo, Attribute control chart for multivariate poisson distribution, Commun Stat — Theor M 37 (2008), 146–158.
[5] R. Christensen, Log-linear models and logistic regression, Springer, New York, 1997.
[6] M. Csörgő and L. Horváth, Limit theorems in change-point analysis, John Wiley & Sons, New York, 1997.

[7] C. Dahinden, G. Parmigiani, M.C. Emerick, and P. Bühlmann, Penalized likelihood for sparse contingency tables with an application to full-length cDNA libraries, BMC Bioinfomatics 8 (2007), 476.
[8] N. Eriksson, S.E. Fienberg, A. Rinaldo, and S. Sullivant, Polyhedral conditions for the nonexistence of the MLE for hierarchical log-linear models, J Symb Comput 41 (2006), 222–233.
[9] E.S. Fienberg and A. Rinaldo, Maximum likelihood estimation in log-linear models, Ann Statist 40 (2012), 996–1023.
[10] D.M. Hawkins, P. Qiu, and C.W. Kang, The changepoint model for statistical process control, J Qual Technol 35 (2003), 355–366.
[11] N.L. Johnson, S. Kotz, and N. Balakrishnan, Discrete multivariate distributions (1997).
[12] T.L. Lai and H. Xing, Sequential change-point detection when the pre- and post-change parameters are unknown, Sequen Anal 29 (2010), 162–175.
[13] C.A. Lowry and D.C. Montgomery, A review of multivariate control charts, IIE Trans 27 (1995), 800–810.
[14] X.S. Lu, M. Xie, T.N. Goh, and C.D. Lai, Control charts for multivariate attribute processes, Int J Prod Res 36 (1998), 3477–3489.
[15] M. Marcucci, Monitoring multinomial processes, J Qual Technol 17 (1985), 86–91.
[16] D.C. Montgomery, Statistical quality control: A modern introduction, 6th ed., Wiley, New York, 2009.
[17] G.V. Moustakides, A.S. Polunchenko, and A.G. Tartakovsky, A numerical approach to performance analysis of quickest change-point detection procedures, Stat Sinica 21 (2011), 571–596.
[18] H.I. Patel, Quality control methods for multivariate binomial and poisson distributions, Technometrics 15 (1973), 103–112.
[19] D. Siegmund, Model selection in irregular problems: Applications to mapping quantitative trait loci, Biometrika 91 (2004), 785–800.
[20] R.J. Simes, An improved Bonferroni procedure for multiple tests of significance, Biometrika 73 (1986), 751–754.
[21] M. Shih, T.L. Lai, J.F. Heyse, and J. Chen, Sequential generalized likelihood ratio tests for vaccine safety evaluation, Stat Med 29 (2010), 2698–2708.
[22] M.S. Srivastava and K.J. Worsley, Likelihood ratio tests for a change in the multivariate normal mean, J Am Stat Assoc 81 (1986), 199–204.
[23] W.H. Woodall, Control charts based on attribute data: Bibliography and review, J Qual Technol 29 (1997), 172–183.
[24] Y. Yao, Estimating the number of change-points via Schwarz criterion, Stat Probabil Lett 6 (1988), 181–189.
[25] N. Zhang and D. Siegmund, A modified Bayes information criterion with applications to the analysis of comparative genomic hybridization data, Biometrics 63 (2007), 22–32.
[26] C. Zou and F. Tsung, Directional MEWMA schemes for multistage process monitoring and diagnosis, J Qual Technol 40 (2008), 407–427.
[27] C. Zou, F. Tsung, and Y. Liu, A change point approach for phase I analysis in multistage processes, Technometrics 50 (2008), 344–356.