1. Introduction

When a product item is tested, we usually have more information than just pass or fail, e.g., pass or fail low or fail high. When the measurement of each tested product item is recorded as exactly one element in some known set consisting of $k+1$ elements, the data are called either binary for $k = 1$ or polytomous for $k = 2, 3, 4, \ldots$. In the paper, categorical data denote either binary data or polytomous data. See, e.g., Agresti (2002) for the categorical data analysis.

In the Bayesian framework, it is assumed that the prior distribution of the unknown random parameter vector is known. In practice, it is usually a non-trivial task to find an appropriate prior distribution for the unknown random parameter vector. Thus, an empirical Bayes (EB) approach is commonly used instead of a Bayesian approach in the literature. In contrast with the Bayesian inference, the EB inference utilizes the observed data to estimate the unknown hyperparameter vector in the prior distribution and then proceeds to do the standard Bayesian inference as if the estimated prior distribution were the true prior one.

Using an EB approach to monitoring a manufacturing process is not entirely new. For example, utilizing the estimated posterior distributions of the random parameters given the data, Yousry et al. (1991) used the beta-binomial model for manufacturing binary data to judge whether a manufacturing process was in control or not. Recently, to judge wether a manufacturing process was in control or not, Shiau et al. (2004) used the Dirichlet-multinomial model for manufacturing polytomous data utilizing the estimated marginal distributions of the data, and Chen et al. (2004) used the beta-binomial or Dirichlet-multinomial model for manufacturing categorical data utilizing the likelihood ratio (LR) method.

To proceed the discussion, first of all, we briefly introduce the Bayesian inference as follows. In the Bayesian framework, it is assumed that the unknown random parameter vector θ has the known prior probability density function (p.d.f.) or probability mass function (p.m.f.) $\pi(\theta)$ with the known parameter space Θ and that the observed response vector y given θ has the known conditional p.d.f. or p.m.f. $f(y|\theta)$. Then the Bayesian inference is based on the posterior p.d.f. or p.m.f.,

$$p(\theta|y) = \frac{f(y|\theta)\,\pi(\theta)}{\int_\Theta f(y|\theta^*)\,\pi(\theta^*)\,d\theta^*} \text{ or } \frac{f(y|\theta)\,\pi(\theta)}{\sum_{\theta^*\in\Theta} f(y|\theta^*)\,\pi(\theta^*)},$$

of θ given y. When the posterior mean,

$$E(\theta|y) = \frac{\int_\Theta \theta\, f(y|\theta)\,\pi(\theta)\,d\theta}{\int_\Theta f(y|\theta)\,\pi(\theta)\,d\theta} \text{ or } \frac{\sum_{\theta\in\Theta} \theta\, f(y|\theta)\,\pi(\theta)}{\sum_{\theta\in\Theta} f(y|\theta)\,\pi(\theta)},$$

of θ given y exists, it minimizes $E((\tilde{\theta}-\theta)'(\tilde{\theta}-\theta)|y)$ among all estimates $\tilde{\theta}$. In the literature, it is common practice to estimate θ by either $E(\theta|y)$ or the posterior mode, mode$(\theta|y)$, of θ given y. See, e.g., Gelman et al. (2003) for the Bayesian data analysis.

Next, we briefly introduce the EB inference as follows. In the EB framework, it is assumed that the unknown random parameter vector θ has the prior p.d.f. or p.m.f. $\pi(\theta;\alpha)$, a known function of both θ and an unknown hyperparameter vector α, with the known parameter space Θ and that the observed response vector y given θ has the known conditional p.d.f. or p.m.f. $f(y|\theta)$. Then the EB inference is based on the estimated posterior p.d.f. or p.m.f., $p(\theta|y;\alpha)|_{\alpha=\hat{\alpha}}$ ($\equiv p(\theta|y;\hat{\alpha})$), of θ given y, where $\hat{\alpha}$ is some estimate of α. In practice, it is frequent to estimate α by either the maximum likelihood (ML) method or the method of moments. In the literature it is common practice to estimate θ by either the estimated posterior mean, $E(\theta|y;\alpha)|_{\alpha=\hat{\alpha}}$ ($\equiv E(\theta|y;\hat{\alpha})$), of θ given y or the estimated posterior mode, $\text{mode}(\theta|y;\alpha)|_{\alpha=\hat{\alpha}}$ ($\equiv \text{mode}(\theta|y;\hat{\alpha})$), of θ given y. See, e.g., Carlin and Louis (2000) for the EB data analysis.

The remaining of the paper is organized as follows. In Section 2, using the normal-binomial or -multinomial model rather than the beta-binomial or Dirichlet-multinomial model in Chen et al. (2004), the EB inference for manufacturing categorical data is discussed. In Section 3, utilizing the LR method, an EB process monitoring technique for manufacturing categorical data is proposed. A simulation study to demonstrate the proposed methodology is given in Section 4. Finally, some concluding remarks and possible generalizations of the paper are given in Section 5.

2. Empirical Bayes

Suppose that a manufacturing process produces a product that has k possible defect types for some fixed positive integer k. Suppose further that each product item belongs to exactly one of the $k+1$ categories {pass, the first defect type, ..., the kth defect type}. Let t be any positive integer. For $i = 1, \ldots, k$, let p_{it} denote the probability that a product item manufactured at time t possesses the ith defect type. Then $1 -$

$\sum_{i=1}^{k} p_{it}$ ($\equiv p_{0t}$) is the probability that a product item manufactured at time t possesses none of the k defect types. Suppose also that there are in total n_t tested product items manufactured at time t, where n_t is a known positive integer such that $\sup\{n_1, n_2, n_3, \ldots\} < \infty$. For $i = 1, \ldots, k$, let y_{it} denote the number of product items that possess the ith defect type among the n_t tested product items manufactured at time t. Then $n_t - \sum_{i=1}^{k} y_{it}$ ($\equiv y_{0t}$) is the number of product items that possess none of the k defect types among the n_t tested product items manufactured at time t. Set $p_t \equiv (p_{1t}, \ldots, p_{kt})'$, $y_t \equiv (y_{1t}, \ldots, y_{kt})'$, $\mathcal{P}_t \equiv \{p_t : \sum_{i=1}^{k} p_{it} < 1 \text{ and } p_{1t}, \ldots, p_{kt} \in (0, 1)\}$, and $\mathcal{Y}_t \equiv \{y_t : \sum_{i=1}^{k} y_{it} \leq n_t \text{ and } y_{1t}, \ldots, y_{kt} \in \{0, 1, \ldots, n_t\}\}$. Then $\mathcal{P}_t = \mathcal{P}_1$ and the number of elements in $\mathcal{Y}_t$, denoted by $|\mathcal{Y}_t|$, is $(n_t + k)!/(n_t!k!)$. Moreover, suppose that $P(\{p_t \in \mathcal{P}_1\}) = 1$ and that y_t given p_t has either the binomial$(n_t; p_t)$ distribution for $k = 1$ or the multinomial$(n_t; p_t)$ distribution for $k = 2, 3, 4, \ldots$. Let $F_{p_t}, F_{y_t}, F_{y_t|p_t}$, and $F_{p_t|y_t}$ denote, respectively, the prior cumulative distribution function (c.d.f.) of p_t, the marginal c.d.f. of y_t, the conditional c.d.f. of y_t given p_t, and the posterior c.d.f. of p_t given y_t. Set $\theta_t \equiv (\theta_{1t}, \ldots, \theta_{kt})' \equiv (\log(p_{1t}/p_{0t}), \ldots, \log(p_{kt}/p_{0t}))'$. Then $p_t = (\exp(\theta_{1t})/[1+\sum_{j=1}^{k} \exp(\theta_{jt})], \ldots, \exp(\theta_{kt})/[1+\sum_{j=1}^{k} \exp(\theta_{jt})])'$. Let $F_{\theta_t}, F_{y_t|\theta_t}$, and $F_{\theta_t|y_t}$ denote, respectively, the prior c.d.f. of θ_t, the conditional c.d.f. of y_t given θ_t, and the posterior c.d.f. of θ_t given y_t. Then $F_{y_t|\theta_t} = F_{y_t|p_t}$, the conditional p.m.f. of y_t

given θ_t is

$$f(y_t|\theta_t) = 1_{\mathcal{Y}_t}(y_t) \cdot \frac{n_t!}{\prod_{i=0}^k y_{it}!} \cdot \frac{\exp(\theta_t' y_t)}{[1+\sum_{i=1}^k \exp(\theta_{it})]^{n_t}}$$
$$= 1_{\mathcal{Y}_t}(y_t) \cdot \frac{n_t!}{\prod_{i=0}^k y_{it}!} \cdot \prod_{i=0}^k p_{it}^{y_{it}} = f(y_t|p_t), \quad (1)$$

and the marginal p.m.f. of y_t is

$$f(y_t) \equiv 1_{\mathcal{Y}_t}(y_t) \cdot \frac{n_t!}{\prod_{i=0}^k y_{it}!} \cdot \int_{\mathcal{R}^k} \frac{\exp(\theta_t' y_t)}{[1+\sum_{i=1}^k \exp(\theta_{it})]^{n_t}} dF_{\theta_t}(\theta_t)$$
$$= 1_{\mathcal{Y}_t}(y_t) \cdot \frac{n_t!}{\prod_{i=0}^k y_{it}!} \cdot \int_{\mathcal{P}_1} \prod_{i=0}^k p_{it}^{y_{it}} dF_{p_t}(p_t), \quad (2)$$

where $1_{\mathcal{Y}_t}(y_t) \equiv 1$ for $y_t \in \mathcal{Y}_t$ and 0 otherwise.

In the paper, first of all, consider the situation where there are available historical in-control manufacturing categorical data $\{y_1, y_2, \ldots, y_T\}$ for some large positive integer T. In the following, suppose that $(\theta_1', y_1')'$, $(\theta_2', y_2')'$, $\ldots$, $(\theta_T', y_T')'$ are independent random vectors and that $\theta_1, \theta_2, \ldots, \theta_T$ are independent and identically distributed (i.i.d.) normal random vectors with $E(\theta_1) = \mu$ and $Cov(\theta_1) = \Sigma$, i.e., $\theta_t \overset{i.i.d.}{\sim} N(\mu, \Sigma)$ for $t = 1, 2, \ldots, T$, where μ ($\equiv (\mu_1, \ldots, \mu_k)'$) is an unknown $k \times 1$ vector and Σ ($\equiv (\Sigma_{uv})$) is an unknown $k \times k$ positive definite covariance matrix.

Throughout the paper, set $\theta_{1:T} \equiv (\theta_1', \theta_2', \ldots, \theta_T')'$, $y_{1:T} \equiv (y_1', y_2', \ldots, y_T')'$, $\Sigma^{-1} \equiv (\Sigma^{uv})$, and $\alpha \equiv (\mu', \Sigma^{11}, \ldots, \Sigma^{k1}, \Sigma^{22}, \ldots, \Sigma^{k2}, \ldots, \Sigma^{kk})'$, where α is the unknown $k(k+3)/2 \times 1$ hyperparameter vector. Let Φ_α denote the c.d.f. of $N(\mu, \Sigma)$ and let $\mathcal{A}$ denote the set consisting of all possible α's. Then the prior p.d.f. of $\theta_{1:T}$ is

$$\pi(\theta_{1:T}; \alpha) = \prod_{t=1}^T \pi(\theta_t; \alpha) = \prod_{t=1}^T \frac{1}{(2\pi)^{k/2}|\Sigma|^{1/2}} \cdot \exp\left[-\frac{1}{2}(\theta_t - \mu)'\Sigma^{-1}(\theta_t - \mu)\right], \quad (3)$$

where $\pi(\theta_t;\alpha)$ denotes the prior marginal p.d.f. of θ_t. Thus, the posterior p.d.f. of $\theta_{1:T}$ given $y_{1:T}$ is

$$p(\theta_{1:T}|y_{1:T};\alpha) = \prod_{t=1}^{T} p(\theta_t|y_{1:T};\alpha) = \prod_{t=1}^{T} p(\theta_t|y_t;\alpha), \tag{4}$$

the posterior mean of $\theta_{1:T}$ given $y_{1:T}$ is

$$\begin{aligned} & E\left(\theta_{1:T}|y_{1:T};\alpha\right) \\ = & \left([E\left(\theta_1|y_{1:T};\alpha\right)]', [E\left(\theta_2|y_{1:T};\alpha\right)]', \ldots, [E\left(\theta_T|y_{1:T};\alpha\right)]'\right)' \\ = & \left([E\left(\theta_1|y_1;\alpha\right)]', [E\left(\theta_2|y_2;\alpha\right)]', \ldots, [E\left(\theta_T|y_T;\alpha\right)]'\right)', \end{aligned} \tag{5}$$

and the posterior mode of $\theta_{1:T}$ given $y_{1:T}$ is

$$\begin{aligned} & \operatorname{mode}\left(\theta_{1:T}|y_{1:T};\alpha\right) \\ = & \left([\operatorname{mode}\left(\theta_1|y_{1:T};\alpha\right)]', [\operatorname{mode}\left(\theta_2|y_{2:T};\alpha\right)]', \ldots, [\operatorname{mode}\left(\theta_T|y_{1:T};\alpha\right)]'\right)' \\ = & \left([\operatorname{mode}\left(\theta_1|y_1;\alpha\right)]', [\operatorname{mode}\left(\theta_2|y_2;\alpha\right)]', \ldots, [\operatorname{mode}\left(\theta_T|y_T;\alpha\right)]'\right)', \end{aligned} \tag{6}$$

where

$$\begin{aligned} & p(\theta_t|y_{1:T};\alpha) \quad = \quad p(\theta_t|y_t;\alpha) \\ = & \frac{\exp[-(\theta_t-\mu-\Sigma\, y_t)'\Sigma^{-1}(\theta_t-\mu-\Sigma\, y_t)/2]/[1+\sum_{i=1}^{k}\exp(\theta_{it})]^{n_t}}{\int_{\mathcal{R}^k}\exp[-(\theta_t^*-\mu-\Sigma\, y_t)'\Sigma^{-1}(\theta_t^*-\mu-\Sigma\, y_t)/2]/[1+\sum_{i=1}^{k}\exp(\theta_{it}^*)]^{n_t}\, d\theta_t^*}, \end{aligned} \tag{7}$$

$$\begin{aligned} & E(\theta_t|y_{1:T};\alpha) \quad = \quad E(\theta_t|y_t;\alpha) \\ = & \frac{\int_{\mathcal{R}^k}\theta_t\,\exp[-(\theta_t-\mu-\Sigma\, y_t)'\Sigma^{-1}(\theta_t-\mu-\Sigma\, y_t)/2]/[1+\sum_{i=1}^{k}\exp(\theta_{it})]^{n_t}\, d\theta_t}{\int_{\mathcal{R}^k}\exp[-(\theta_t-\mu-\Sigma\, y_t)'\Sigma^{-1}(\theta_t-\mu-\Sigma\, y_t)/2]/[1+\sum_{i=1}^{k}\exp(\theta_{it})]^{n_t}\, d\theta_t}, \end{aligned} \tag{8}$$

and

$$
\begin{aligned}
\text{mode}(\theta_t|y_{1:T};\alpha) &= \text{mode}(\theta_t|y_t;\alpha) \\
= \arg\min_{\theta_t\in\mathcal{R}^k} &\left\{\frac{1}{2}\left(\theta_t-\mu-\Sigma\, y_t\right)'\Sigma^{-1}\left(\theta_t-\mu-\Sigma\, y_t\right)+n_t\,\log\left[1+\sum_{i=1}^{k}\exp(\theta_{it})\right]\right\}
\end{aligned}
\tag{9}
$$

for $t = 1, 2, \ldots, T$.

Let c_T be a prespecified constant in $(0, 1)$, e.g., 0.05, and let $q_{T,1-c_T}(y_{1:T};\alpha) \in (0,\infty)$ such that $P(R_{T,1-c_T}(y_{1:T};\alpha)|y_{1:T};\alpha) = 1 - c_T$, where

$$
R_{T,1-c_T}(y_{1:T};\alpha) \equiv \{\theta_{1:T} : p(\theta_{1:T}|y_{1:T};\alpha) \geq q_{T,1-c_T}(y_{1:T};\alpha)\}. \tag{10}
$$

Then $R_{T,1-c_T}(y_{1:T};\alpha)$ is the size $1-c_T$ highest posterior density (HPD) region for $\theta_{1:T}$ given $y_{1:T}$. Similarly, for $t = 1, 2, \ldots, T$, let $c_{(t)}$ be a prespecified constant in $(0, 1)$, e.g., 0.05, and let $q_{(t),1-c_{(t)}}(y_t;\alpha) \in (0,\infty)$ such that $P(R_{(t),1-c_{(t)}}(y_t;\alpha)|y_t;\alpha) = 1-c_{(t)}$, where

$$
R_{(t),1-c_{(t)}}(y_t;\alpha) \equiv \left\{\theta_t : p(\theta_t|y_t;\alpha) \geq q_{(t),1-c_{(t)}}(y_t;\alpha)\right\}. \tag{11}
$$

Then for $t = 1, 2, \ldots, T$, $R_{(t),1-c_{(t)}}(y_t;\alpha)$ is the size $1-c_{(t)}$ HPD region for θ_t given y_t. For $t = 1, 2, \ldots, T$, as $p(\theta_t|y_{1:T};\alpha) = p(\theta_t|y_t;\alpha)$, $R_{(t),1-c_{(t)}}(y_t;\alpha)$ is also the size $1 - c_{(t)}$ HPD region for θ_t given $y_{1:T}$.

Observe that the marginal p.m.f. of $y_{1:T}$ is

$$\begin{aligned} f(y_{1:T}) &= \prod_{t=1}^{T} f(y_t) \\ &= \prod_{t=1}^{T} 1_{\mathcal{Y}_t}(y_t) \cdot \frac{n_t!}{\prod_{i=0}^{k} y_{it}!} \cdot \int_{\mathcal{R}^k} \frac{\exp(y_t' \theta_t)}{[1+\sum_{i=1}^{k}\exp(\theta_{it})]^{n_t}} \cdot \pi(\theta_t;\alpha)\, d\theta_t \\ &\equiv \prod_{t=1}^{T} 1_{\mathcal{Y}_t}(y_t) \cdot \frac{n_t!}{\prod_{i=0}^{k} y_{it}!} \cdot \int_{\mathcal{R}^k} g_t(\theta_t, y_t)\pi(\theta_t;\alpha)\, d\theta_t \\ &\equiv \prod_{t=1}^{T} 1_{\mathcal{Y}_t}(y_t) \cdot \frac{n_t!}{\prod_{i=0}^{k} y_{it}!} \cdot a_t(\alpha; y_t). \end{aligned} \tag{12}$$

For $t = 1, 2, \ldots, T$, set $\ell(\alpha;\theta_t) \equiv \log[\pi(\theta_t;\alpha)]$, $b_t(\alpha; y_t) \equiv \partial a_t(\alpha; y_t)/\partial\alpha$, and $c_t(\alpha; y_t) \equiv \partial b_t(\alpha; y_t)/\partial\alpha'$. Then

$$b_t(\alpha; y_t) = \int_{\mathcal{R}^k} \frac{\partial \ell(\alpha;\theta_t)}{\partial\alpha} \cdot g_t(\theta_t, y_t) \cdot \pi(\theta_t;\alpha)\, d\theta_t$$

and

$$c_t(\alpha; y_t) = \int_{\mathcal{R}^k} \left[\frac{\partial^2 \ell(\alpha;\theta_t)}{\partial\alpha\partial\alpha'} + \frac{\partial \ell(\alpha;\theta_t)}{\partial\alpha}\frac{\partial \ell(\alpha;\theta_t)}{\partial\alpha'}\right] g_t(\theta_t, y_t) \cdot \pi(\theta_t;\alpha)\, d\theta_t$$

for $t = 1, 2, \ldots, T$. Thus, given $y_{1:T}$, the log-likelihood function for α is

$$\begin{aligned} \log[f(y_{1:T})] &= \prod_{t=1}^{T} \log[f(y_t)] \\ = \sum_{t=1}^{T} \left\{\log(n_t!) - \sum_{i=0}^{k} \log(y_{it}!) + \log[a_t(\alpha; y_t)]\right\} &\equiv \sum_{t=1}^{T} \ell_{(t)}(\alpha) \equiv \ell_T(\alpha), \end{aligned} \tag{13}$$

the score function for α is

$$\frac{\partial \ell_T(\alpha)}{\partial\alpha} = \sum_{t=1}^{T} \frac{b_t(\alpha; y_t)}{a_t(\alpha; y_t)} \equiv \sum_{t=1}^{T} S_{(t)}(\alpha) \equiv S_T(\alpha), \tag{14}$$

the (expected) Fisher information for α is

$$Cov(S_T(\alpha)) = \sum_{t=1}^{T} \sum_{y_t \in \mathcal{Y}_t} S_{(t)}(\alpha) S'_{(t)}(\alpha) f(y_t) \equiv \sum_{t=1}^{T} I_{(t)}(\alpha) \equiv I_T(\alpha), \tag{15}$$

and the observed (Fisher) information for α is

$$-\frac{\partial S_T(\alpha)}{\partial \alpha'} = \sum_{t=1}^{T} \frac{b_t(\alpha; y_t) b_t'(\alpha; y_t) - a_t(\alpha; y_t) c_t(\alpha; y_t)}{a_t^2(\alpha; y_t)} \equiv \sum_{t=1}^{T} J_{(t)}(\alpha) \equiv J_T(\alpha). \quad (16)$$

Note that given $y_{1:T}$, the MLE $\hat{\alpha}_T$ of α solves the score equation $S_T(\alpha) = 0_{k(k+3)/2\times 1}$ for α, i.e., $S_T(\alpha)|_{\alpha=\hat{\alpha}_T}$ ($\equiv S_T(\hat{\alpha}_T)$) $= 0_{k(k+3)/2\times 1}$, where $0_{k(k+3)/2\times 1}$ denotes the $k(k+3)/2 \times 1$ vector $(0, \ldots, 0)^T$. Set $K_T(\alpha) \equiv \sum_{t=1}^{T} S_{(t)}(\alpha) S'_{(t)}(\alpha)$ ($\equiv \sum_{t=1}^{T} K_{(t)}(\alpha)$).

One way to numerically evaluate $\hat{\alpha}_T$ is to utilize the following iterative procedure. First choose a *good* initial value $\hat{\alpha}_T^{(0)}$ for $\hat{\alpha}_T$, e.g., the method-of-moments estimate (MME) of α given in Appendix A, and then iterate the following equations

$$\begin{aligned} \hat{\alpha}_T^{(u+1)} &= \hat{\alpha}_T^{(u)} + \left[M_T^{(u)}(\alpha)\Big|_{\alpha=\hat{\alpha}_T^{(u)}} \right]^{-1} S_T(\alpha)|_{\alpha=\hat{\alpha}_T^{(u)}} \\ &\equiv \hat{\alpha}_T^{(u)} + \left[M_T^{(u)}\left(\hat{\alpha}_T^{(u)}\right)\right]^{-1} S_T\left(\hat{\alpha}_T^{(u)}\right) \end{aligned}$$

for $u = 0, 1, 2, \ldots$ until $\hat{\alpha}_T^{(u)}$ converges to $\hat{\alpha}_T$, where $M_T^{(u)}(\hat{\alpha}_T^{(u)})$ could be any of $I_T(\hat{\alpha}_T^{(u)})$, $J_T(\hat{\alpha}_T^{(u)})$, and $K_T(\hat{\alpha}_T^{(u)})$ for each $u \in \{0, 1, 2, \ldots\}$. When $M_T^{(u)}(\hat{\alpha}_T^{(u)}) = J_T(\hat{\alpha}_T^{(u)})$ for all u's, it is called the Newton-Raphson method. When $M_T^{(u)}(\hat{\alpha}_T^{(u)}) = I_T(\hat{\alpha}_T^{(u)})$ for all u's, it is called the Fisher scoring method. Observe that all of $I_T(\hat{\alpha}_T^{(u)})$'s are positive definite and that all of $K_T(\hat{\alpha}_T^{(u)})$'s are positive semi-definite and generally positive definite. However, $J_T(\hat{\alpha}_T^{(u)})$'s are not necessarily positive semi-definite when the initial value $\hat{\alpha}_T^{(0)}$ is not close to $\hat{\alpha}_T$. As there is no simple closed-form formula for $I_T(\alpha)$ and $|\mathcal{Y}_t|$ is generally much larger than 1 for each $t \in \{1, 2, \ldots, T\}$, it takes too much time to calculate $I_T(\hat{\alpha}_T^{(u)})$ than either $J_T(\hat{\alpha}_T^{(u)})$ or $K_T(\hat{\alpha}_T^{(u)})$ for each $u \in$

$\{0, 1, 2, \ldots\}$. Moreover, as it takes more time to find a *good* initial value for the Newton-Raphson method than the procedure with $M_T^{(u)}(\hat{\alpha}_T^{(u)}) = K_T(\hat{\alpha}_T^{(u)})$ for all u's, a stable and fast method to numerically evaluate $\hat{\alpha}_T$ is suggested as follows. First choose $M_T^{(u)}(\hat{\alpha}_T^{(u)}) = K_T(\hat{\alpha}_T^{(u)})$ for all u's until near convergence, say at $u = u_0$, and then choose $M_T^{(u)}(\hat{\alpha}_T^{(u)}) = J_T(\hat{\alpha}_T^{(u)})$ for $u = u_0+1, u_0+2, u_0+3, \ldots$ until convergence.

Note that $\hat{\alpha}_T = \alpha + O_p(T^{-1/2})$ as $T \to \infty$. Thus, for any fixed $t \in \{1, 2, \ldots, T\}$ and given y_t, all of $p(\theta_t|y_t; \hat{\alpha}_T) - p(\theta_t|y_t; \alpha)$, $E(\theta_t|y_t; \hat{\alpha}_T) - E(\theta_t|y_t; \alpha)$, $\text{mode}(\theta_t|y_t; \hat{\alpha}_T) - \text{mode}(\theta_t|y_t; \alpha)$, $f_T(y_t) - f(y_t)$, and $q_{(t),1-c_{(t)}}(y_t; \hat{\alpha}_T) - q_{(t),1-c_{(t)}}(y_t; \alpha)$ are $O_p(T^{-1/2})$ as $T \to \infty$, where $p(\theta_t|y_t; \hat{\alpha}_T) \equiv p(\theta_t|y_t; \alpha)|_{\alpha=\hat{\alpha}_T}$, $E(\theta_t|y_t; \hat{\alpha}_T) \equiv E(\theta_t|y_t; \alpha)|_{\alpha=\hat{\alpha}_T}$, $\text{mode}(\theta_t|y_t; \hat{\alpha}_T) \equiv \text{mode}(\theta_t|y_t; \alpha)|_{\alpha=\hat{\alpha}_T}$, $f_T(y_t) \equiv f(y_t)|_{\theta \sim \Phi_{\hat{\alpha}_T}}$ with $\Phi_{\hat{\alpha}_T} = \Phi_\alpha|_{\alpha=\hat{\alpha}_T}$, and $q_{(t),1-c_{(t)}}(y_t; \hat{\alpha}_T) \equiv q_{(t),1-c_{(t)}}(y_t; \alpha)|_{\alpha=\hat{\alpha}_T}$. For $t = 1, 2, \ldots, T$, as there is no simple closed-form formula for any of $p(\theta_t|y_t; \hat{\alpha}_T)$, $E(\theta_t|y_t; \hat{\alpha}_T)$, $\text{mode}(\theta_t|y_t; \hat{\alpha}_T)$, $f_T(y_t)$, and $q_{(t),1-c_{(t)}}(y_t; \hat{\alpha}_T)$, all of them might be evaluated numerically. Similarly, as there is no closed formula for $q_{T,1-c_T}(y_{1:T}; \alpha)|_{\alpha=\hat{\alpha}_T}$ ($\equiv q_{T,1-c_T}(y_{1:T}; \hat{\alpha}_T)$), it might also be evaluated numerically. See Appendix B for details.

3. Empirical Bayes process monitoring scheme

In this section, assume that the response vector y_{t_0} is observed at time t_0 and that the random vector $(\theta'_{t_0}, y'_{t_0})'$ is independent of $(\theta'_{1:T}, y'_{1:T})'$, where $t_0 \in \{T+1, T+2, T+3, \ldots\}$ and all of y_{t_0},θ_{t_0}, $\theta_{1:T}$, and $y_{1:T}$ are defined in Section 2. Let γ denote the false alarm rate, i.e., the probability that an out-of-control signal occurs when a manufacturing process is in control, where γ is a prespecified constant in (0,1). In the literature, γ is commonly chosen as $2\Phi(-3)$ (≈ 0.002699796), where Φ denotes the c.d.f. of a standard normal random variable. Set $\mathcal{Y}_{t_0} \equiv \{y_{t_0,1}, y_{t_0,2}, \ldots, y_{t_0,|\mathcal{Y}_{t_0}|}\}$, where both $\mathcal{Y}_{t_0}$ and $|\mathcal{Y}_{t_0}|$ are defined in Section 2. Recall from Section 2 that $F_{\theta_{t_0}}$ denotes the prior c.d.f. of θ_{t_0}. Let $\mathcal{F}(\mathcal{R}^k)$ denote the non-parametric family consisting of all c.d.f.'s on $\mathcal{R}^k$. Note that by assuming that $F_{\theta_{t_0}}$ is the unknown prior c.d.f. of interest in $\mathcal{F}(\mathcal{R}^k)$ rather than in some particular parametric family such as the family of all k-variate normal distributions, we make our process monitoring scheme more general than most of other schemes.

Let $\ell_{(t_0)}(F_{\theta_{t_0}})$ denote the log-likelihood function of $F_{\theta_{t_0}}$ given y_{t_0}. Then $\ell_{(t_0)}(F_{\theta_{t_0}})$ $= \log[f(y_{t_0})]$, where $f(y_{t_0})$ is defined in Section 2. Note that

$$\begin{aligned}
\ell_{(t_0)}\left(F_{\theta_{t_0}}\right) &= \log\left[\int_{\mathcal{P}_1} f(y_{t_0}|p_{t_0})\, dF_{p_{t_0}}(p_{t_0})\right] \\
&\leq \log\left[\int_{\mathcal{P}_1} f(y_{t_0}|p_{t_0})|_{p_{t_0}=y_{t_0}/n_{t_0}}\, dF_{p_{t_0}}(p_{t_0})\right] \\
&= \log\left[f(y_{t_0}|p_{t_0})|_{p_{t_0}=y_{t_0}/n_{t_0}}\right],
\end{aligned}$$

where all of p_{t_0}, $f(y_{t_0}|p_{t_0})$, and $F_{p_{t_0}}$ are defined in Section 2 and $f(y_{t_0}|p_{t_0})$ is maximized if and only if $p_{t_0} = y_{t_0}/n_{t_0}$. Thus,

$$\sup_{F_{\theta_{t_0}} \in \mathcal{F}(\mathcal{R}^k)} \ell_{(t_0)}\left(F_{\theta_{t_0}}\right) = \log\left[f(y_{t_0}|p_{t_0})|_{p_{t_0}=y_{t_0}/n_{t_0}}\right]. \tag{17}$$

When the manufacturing process is in control, the prior c.d.f. of θ_{t_0} is Φ_α, where both α and Φ_α are defined in Section 2. Thus, to monitor the manufacturing process at time t_0, we might be interested in testing the null hypothesis $H_0 : F_{\theta_{t_0}} = \Phi_\alpha$ versus the non-parametric alternative $H_1 : F_{\theta_{t_0}} \neq \Phi_\alpha$.

In the remaining of this section, the hyperparameter vector α is assumed to be known in Subsection 3.1 and unknown in Subsection 3.2, respectively. Note that the reason for utilizing the LR method to monitor a manufacturing process is that the LR test frequently has higher power when the alternative hypothesis is true, which corresponds to good detecting power in the process monitoring when a manufacturing process is out of control.

3.1. Known α

In this subsection, consider the Bayesian situation where the hyperparameter vector α is known. Set $\ell_{(t_0)}(\alpha) \equiv \ell_{(t_0)}(F_{\theta_{t_0}})|_{F_{\theta_{t_0}}=\Phi_\alpha}$. Then the LR statistic for testing the simple null hypothesis $H_0 : F_{\theta_{t_0}} = \Phi_\alpha$ versus the non-parameteric alternative $H_1 : F_{\theta_{t_0}} \neq \Phi_\alpha$ is

$$\begin{aligned} W_{(t_0)}(\alpha) &\equiv 2\left[\sup_{F_{\theta_{t_0}} \in \mathcal{F}(\mathcal{R}^k)} \ell_{(t_0)}\left(F_{\theta_{t_0}}\right) - \ell_{(t_0)}(\alpha)\right] \\ &= 2\left\{\log\left[f(y_{t_0}|p_{t_0})|_{p_{t_0}=y_{t_0}/n_{t_0}}\right] - \ell_{(t_0)}(\alpha)\right\} \\ &= 2\left\{\sum_{i=0}^{k} y_{it_0}\log\left(\frac{y_{it_0}}{n_{t_0}}\right) - \log\left[a_{t_0}(\alpha; y_{t_0})\right]\right\}, \end{aligned} \quad (18)$$

where $0\log(0) \equiv 0$, $a_{t_0}(\alpha; y_{t_0})$ is defined in Section 2, and $P(\{0 < W_{(t_0)}(\alpha) < \infty\}; H_0) = 1$.

The size γ LR test with its corresponding quality control scheme for monitoring the LR statistic $W_{(t_0)}(\alpha)$ could be constructed as follows. For $s = 1, 2, \ldots, |\mathcal{Y}_{t_0}|$, set $W_{(t_0),s}(\alpha) \equiv W_{(t_0)}(\alpha)|_{y_{t_0}=y_{t_0,s}}$. Let $(W_{(t_0),(1)}(\alpha), W_{(t_0),(2)}(\alpha), \ldots, W_{(t_0),(|\mathcal{Y}_{t_0}|)}(\alpha))$ be a permutation of $(W_{(t_0),1}(\alpha), W_{(t_0),2}(\alpha), \ldots, W_{(t_0),|\mathcal{Y}_{t_0}|}(\alpha))$ such that $W_{(t_0),(1)}(\alpha) \leq W_{(t_0),(2)}(\alpha) \leq \ldots \leq W_{(t_0),(|\mathcal{Y}_{t_0}|)}(\alpha)$. As $W_{(t_0)}(\alpha)$ is a discrete random variable, it is nearly impossible to attain the exact false alarm rate γ if a deterministic control limit approach is used. Thus, based on the concept of randomized tests in hypothesis testing, we propose the following randomized control limit approach. To find the randomized upper control limit, we start accumulating $P(\{W_{(t_0)}(\alpha) = W_{(t_0),(|\mathcal{Y}_{t_0}|)}(\alpha)\}; H_0)$ until we reach the first s such that $P(\{W_{(t_0)}(\alpha) \geq W_{(t_0),(s)}(\alpha)\}; H_0)$ $> \gamma$. Denote this s by $m_{(t_0)}(\alpha)$ and set $RUCL_{(t_0)}(\alpha) \equiv W_{(t_0),(m_{(t_0)}(\alpha))}(\alpha)$. If $P(\{W_{(t_0)}(\alpha)$ $> RUCL_{(t_0)}(\alpha)\}; H_0) = \gamma$, which is very unlikely, then there is no need for randomization and set $\gamma_{RUCL,(t_0)}(\alpha) = 0$; otherwise, we need to find the randomization probability $\gamma_{RUCL,(t_0)}(\alpha) \in (0, 1)$.

Specifically, we have

$$m_{(t_0)}(\alpha) = \max\left\{s : P\left(\left\{W_{(t_0)}(\alpha) \geq W_{(t_0),(s)}(\alpha)\right\}; H_0\right) > \gamma\right\}, \quad (19)$$

$$RUCL_{(t_0)}(\alpha) = W_{(t_0),(m_{(t_0)}(\alpha))}(\alpha), \quad (20)$$

and

$$\gamma_{RUCL,(t_0)}(\alpha) = \frac{\gamma - P(\{W_{(t_0)}(\alpha) > RUCL_{(t_0)}(\alpha)\}; H_0)}{P(\{W_{(t_0)}(\alpha) = RUCL_{(t_0)}(\alpha)\}; H_0)}. \quad (21)$$

Finally, the monitoring scheme for the manufacturing process at time t_0 is proposed as follows. If $W_{(t_0)}(\alpha) > RUCL_{(t_0)}(\alpha)$, then we reject H_0 and declare that the process is out of control; if $W_{(t_0)}(\alpha) < RUCL_{(t_0)}(\alpha)$, then we accept H_0 and declare that the process is in control; if $W_{(t_0)}(\alpha) = RUCL_{(t_0)}(\alpha)$, then with probability $\gamma_{RUCL,(t_0)}(\alpha)$ we reject H_0 and declare that the process is out of control or, equivalently, with probability $1 - \gamma_{RUCL,(t_0)}(\alpha)$ we accept H_0 and declare that the process is in control, where the randomization could be done by any random number generator or table.

However, it is possible that $|\mathcal{Y}_{t_0}|$ is very large at time t_0 in a manufacturing process, e.g., $|\mathcal{Y}_{t_0}| = 82,408,626,300$ if $n_{t_0} = 200$ and $k = 6$. In such a situation, it takes too much time to perform the previous size γ LR test. Thus, by a simulation, an approximate size γ LR test with its corresponding quality control scheme for monitoring the LR statistic $W_{(t_0)}(\alpha)$ could be constructed as follows. First generate an i.i.d. sample $\{(\theta^{*'}_{1,1}, y^{*'}_{t_0,1})', (\theta^{*'}_{1,2}, y^{*'}_{t_0,2})', \ldots, (\theta^{*'}_{1,r}, y^{*'}_{t_0,r})'\}$ of size r for some large positive integer r, e.g., $r = 100,000$, such that $\theta^*_{1,s} \sim \Phi_\alpha$ and $y^*_{t_0,s}|\theta^*_{1,s} \sim F_{y_{t_0}|\theta_{t_0}}|_{\theta_{t_0}=\theta^*_{1,s}}$ for $s = 1, 2, \ldots, r$. For $s = 1, 2, \ldots, r$, set $W^*_{(t_0),r,s}(\alpha) \equiv W_{(t_0)}(\alpha)|_{y_{t_0}=y^*_{t_0,s}}$. Let $(W^*_{(t_0),r,(1)}(\alpha), W^*_{(t_0),r,(2)}(\alpha), \ldots, W^*_{(t_0),r,(r)}(\alpha))$ be a permutation of $(W^*_{(t_0),r,1}(\alpha), W^*_{(t_0),r,2}(\alpha), \ldots, W^*_{(t_0),r,r}(\alpha))$ such that $W^*_{(t_0),r,(1)}(\alpha) \leq W^*_{(t_0),r,(2)}(\alpha) \leq \ldots \leq W^*_{(t_0),r,(r)}(\alpha)$. Set $m^*_{(t_0),r}(\alpha) \equiv [r(1 - \gamma)] + 1$ and

$$RUCL^*_{(t_0),r}(\alpha) \equiv W_{(t_0),r,(m^*_{(t_0),r}(\alpha))}(\alpha), \tag{22}$$

where $[r(1-\gamma)]$ denotes the largest integer less than or equal to $r(1-\gamma)$, e.g., $m^*_{(t_0),r}(\alpha) = 99,731$ if $r = 100,000$ and $\gamma = 2\Phi(-3)$. Let $m^*_{L,(t_0),r}(\alpha), m^*_{U,(t_0),r}(\alpha) \in$

$\{1, 2, \ldots, r\}$ such that

$$
\begin{aligned}
W^*_{(t_0),r,(m^*_{L,(t_0),r}(\alpha)-1)}(\alpha) &< W^*_{(t_0),r,(m^*_{L,(t_0),r}(\alpha))}(\alpha) = RUCL^*_{(t_0),r}(\alpha) \\
= W^*_{(t_0),r,(m^*_{U,(t_0),r}(\alpha))}(\alpha) &< W^*_{(t_0),r,(m^*_{U,(t_0),r}(\alpha)+1)}(\alpha),
\end{aligned}
$$

where $W^*_{(t_0),r,(0)}(\alpha) \equiv 0$ and $W^*_{(t_0),r,(r+1)}(\alpha) \equiv \infty$. Set

$$
\gamma^*_{RUCL,(t_0),r}(\alpha) \equiv \frac{\gamma - [r - m^*_{U,(t_0),r}(\alpha)]/r}{[m^*_{U,(t_0),r}(\alpha) - m^*_{L,(t_0),r}(\alpha) + 1]/r} = \frac{r\gamma - r + m^*_{U,(t_0),r}(\alpha)}{m^*_{U,(t_0),r}(\alpha) - m^*_{L,(t_0),r}(\alpha) + 1}. \tag{23}
$$

Finally, the monitoring scheme for the manufacturing process at time t_0 is proposed as follows. If $W_{(t_0)}(\alpha) > RUCL^*_{(t_0),r}(\alpha)$, then we reject H_0 and declare that the process is out of control; if $W_{(t_0)}(\alpha) < RUCL^*_{(t_0),r}(\alpha)$, then we accept H_0 and declare that the process is in control; if $W_{(t_0)}(\alpha) = RUCL^*_{(t_0),r}(\alpha)$, then with probability $\gamma^*_{RUCL,(t_0),r}(\alpha)$ we reject H_0 and declare that the process is out of control or, equivalently, with probability $1 - \gamma^*_{RUCL,(t_0),r}(\alpha)$ we accept H_0 and declare that the process is in control.

Note that under H_0, both $RUCL^*_{(t_0),r}(\alpha) - RUCL_{(t_0)}(\alpha)$ and $\gamma^*_{RUCL,(t_0),r}(\alpha) - \gamma_{RUCL,(t_0)}(\alpha)$ converge to 0 with probability one as $r \to \infty$, where the rate of convergence for the latter is much slower than that for the former as $r \to \infty$. Thus, this test converges to the previous size γ LR test as $r \to \infty$.

3.2. Unknown α

In this subsection, consider the EB situation where the hyperparameter vector α is unknown. Set $y_{1:T,t_0} \equiv (y'_{1:T}, y'_{t_0})'$. Then the LR statistic for testing the parametric null hypothesis $H_0 : F_{\theta_{t_0}} = \Phi_\alpha$ versus the non-parameter alternative $H_1 : F_{\theta_{t_0}} \neq \Phi_\alpha$ is

$$\begin{aligned} W_{T,(t_0)} &\equiv 2\left\{ \sup_{\alpha\in\mathcal{A}, F_{\theta_{t_0}}\in\mathcal{F}(\mathcal{R}^k)} \left[\ell_T(\alpha) + \ell_{(t_0)}\left(F_{\theta_{t_0}}\right)\right] - \sup_{\alpha\in\mathcal{A}} \left[\ell_T(\alpha) + \ell_{(t_0)}\left(\alpha\right)\right] \right\} \\ &= 2\left\{\ell_T(\hat{\alpha}_T) + \log\left[f(y_{t_0}|p_{t_0})|_{p_{t_0}=y_{t_0}/n_{t_0}}\right] - \ell_T\left(\hat{\alpha}_{T,(t_0)}\right) - \ell_{(t_0)}\left(\hat{\alpha}_{T,(t_0)}\right)\right\}, \end{aligned} \tag{24}$$

where all of $\mathcal{A}$, $\ell_T(\alpha)$, and $\hat{\alpha}_T$ are defined in Section 2, $\hat{\alpha}_{T,(t_0)}$ denotes the MLE of α given $y_{1:T,t_0}$ under H_0, and $P(\{0 < W_{T,(t_0)} < \infty\}; H_0) = 1$. For simplicity of notation, set $S_{(t_0)}(\alpha) \equiv \partial\ell_{(t_0)}(\alpha)/\partial\alpha$ and $J_{(t_0)}(\alpha) \equiv -\partial S_{(t_0)}(\alpha)/\partial\alpha'$.

Note that given $y_{1:T,t_0}$, the MLE $\hat{\alpha}_{T,(t_0)}$ of α solves the score equation $S_T(\alpha) + S_{(t_0)}(\alpha) = 0_{k(k+3)/2\times 1}$, i.e., $[S_T(\alpha) + S_{(t_0)}(\alpha)]|_{\alpha=\hat{\alpha}_{T,t_0}}$ $(\equiv S_T(\hat{\alpha}_{T,(t_0)}) + S_{(t_0)}(\hat{\alpha}_{T,(t_0)}))$ $= 0_{k(k+3)/2\times 1}$. One way to numerically evaluate $\hat{\alpha}_{T,(t_0)}$ is to utilize the following Newton-Raphson method. First choose $\hat{\alpha}_T$ as the initial value $\hat{\alpha}^{(0)}_{T,(t_0)}$ for $\hat{\alpha}_{T,(t_0)}$ and then iterate the following equations

$$\hat{\alpha}^{(u+1)}_{T,(t_0)} = \hat{\alpha}^{(u)}_{T,(t_0)} + \left\{ \left[J_T(\alpha) + J_{(t_0)}(\alpha)\right]\Big|_{\alpha=\hat{\alpha}^{(u)}_{T,(t_0)}} \right\}^{-1} \left[S_T(\alpha) + S_{(t_0)}(\alpha)\right]\Big|_{\alpha=\hat{\alpha}^{(u)}_{T,(t_0)}}$$

for $u = 0, 1, 2, \ldots$ until $\hat{\alpha}^{(u)}_{T,(t_0)}$ converges to $\hat{\alpha}_{T,(t_0)}$.

Set $W_{(t_0)}(\hat{\alpha}_T) \equiv W_{(t_0)}(\alpha)|_{\alpha=\hat{\alpha}_T}$ and

$$\hat{W}_{T,(t_0)} \equiv \max\left\{0, W_{(t_0)}\left(\hat{\alpha}_T\right) - S'_{(t_0)}\left(\hat{\alpha}_T\right)\left[J_T\left(\hat{\alpha}_T\right) + J_{(t_0)}\left(\hat{\alpha}_T\right)\right]^{-1} S_{(t_0)}\left(\hat{\alpha}_T\right)\right\}, \tag{25}$$

where $W_{(t_0)}(\alpha)$ is defined in Subsection 3.1, $P(\{0 < W_{(t_0)}(\hat{\alpha}_T) < \infty\}; H_0) = 1$, and $P(\{0 \leq \hat{W}_{T,t_0} < \infty\}; H_0) = 1$. Then under H_0,

$$W_{T,(t_0)} = \hat{W}_{T,(t_0)} + O_p\left(T^{-2}\right) = W_{(t_0)}\left(\hat{\alpha}_T\right) + O_p\left(T^{-1}\right) = W_{(t_0)}\left(\alpha\right) + O_p\left(T^{-1/2}\right) \quad (26)$$

as $T \to \infty$. See Appendix c for details.

In the following, let $\tilde{\alpha}_{T,(t_0)}$ denote either $\hat{\alpha}_T$ or $\hat{\alpha}_{T,(t_0)}$ and let $\tilde{W}_{T,(t_0)}$ denote any of $W_{T,(t_0)}$, $\hat{W}_{T,(t_0)}$, and $W_{(t_0)}(\hat{\alpha}_T)$. Then under H_0, both $\tilde{\alpha}_{T,(t_0)} - \alpha$ and $\tilde{W}_{T,(t_0)} - W_{(t_0)}(\alpha)$ are $O_p(T^{-1/2})$ as $T \to \infty$.

As the hyperparameter vector α is unknown and $(\prod_{t=1}^{T} |\mathcal{Y}_t|)|\mathcal{Y}_{t_0}|$, the number of the elements in $(\prod_{t=1}^{T} \mathcal{Y}_t) \times \mathcal{Y}_{t_0}$, is generally very large, it is nearly impossible to perform the size γ LR test for testing the parameteric null hypothesis $H_0 : F_{\theta_{t_0}} = \Phi_\alpha$ versus the non-parameteric alternative $H_1 : F_{\theta_{t_0}} \neq \Phi_\alpha$. Thus, by a simulation, an approximate size γ LR test with its corresponding quality control scheme for monitoring the (approximate) LR statistic $\tilde{W}_{T,(t_0)}$ could be constructed as follows. For $s = 1, 2, \ldots, |\mathcal{Y}_{t_0}|$, set $\tilde{W}_{T,(t_0),s} \equiv \tilde{W}_{T,(t_0)}|_{y_{t_0}=y_{t_0,s}}$. Let $(\tilde{W}_{T,(t_0),(1)}, \tilde{W}_{T,(t_0),(2)}, \ldots, \tilde{W}_{T,(t_0),(|\mathcal{Y}_{t_0}|)})$ be a permutation of $(\tilde{W}_{T,(t_0),1}, \tilde{W}_{T,(t_0),2}, \ldots, \tilde{W}_{T,(t_0),|\mathcal{Y}_{t_0}|})$ such that $\tilde{W}_{T,(t_0),(1)} \leq \tilde{W}_{T,(t_0),(2)} \leq \ldots \leq \tilde{W}_{T,(t_0),(|\mathcal{Y}_{t_0}|)}$. Set

$$\begin{aligned} \tilde{m}_{U,T,(t_0)} &\equiv \max\left\{s : P\left(\left\{\tilde{W}_{T,(t_0)} \geq \tilde{W}_{T,(t_0),(s)}\right\}\middle| y_{1:T}; H_0\right)|_{\alpha=\tilde{\alpha}_{T,(t_0)}} > \gamma\right\}, \\ R\tilde{U}CL_{T,(t_0)} &\equiv \tilde{W}_{T,(t_0),(\tilde{m}_{U,T,(t_0)})}, \end{aligned} \quad (27)$$

and

$$\tilde{\gamma}_{RUCL,T,(t_0)} \equiv \frac{\gamma - P(\{\tilde{W}_{T,(t_0)} > R\tilde{U}CL_{T,(t_0)}\}|y_{1:T}; H_0)|_{\alpha=\tilde{\alpha}_{T,(t_0)}}}{P(\{\tilde{W}_{T,(t_0)} = R\tilde{U}CL_{T,(t_0)}\}|y_{1:T}; H_0)|_{\alpha=\tilde{\alpha}_{T,(t_0)}}}. \quad (28)$$

Finally, the monitoring scheme for the manufacturing process at time t_0 is proposed as follows. If $\tilde{W}_{T,(t_0)} > R\tilde{U}CL_{T,(t_0)}$, then we reject H_0 and declare that the process is out of control; if $\tilde{W}_{T,(t_0)} < R\tilde{U}CL_{T,(t_0)}$, then we accept H_0 and declare that the process is in control; if $\tilde{W}_{T,(t_0)} = R\tilde{U}CL_{T,(t_0)}$, then with probability $\tilde{\gamma}_{RUCL,T,(t_0)}$ we reject H_0 and declare that the process is out of control or, equivalently, with probability $1 - \tilde{\gamma}_{RUCL,T,(t_0)}$ we accept H_0 and declare that the process is in control.

Note that under H_0, $\tilde{W}_{T,(t_0)} - W_{(t_0)}(\alpha) = O_p(T^{-1/2})$ as $T \to \infty$. Thus, this test approximates the size γ LR test in Subsection 3.1 well for large positive integer T.

Similarly, it is possible that $|\mathcal{Y}_{t_0}|$ is very large at time t_0 in a manufacturing process, e.g., $|\mathcal{Y}_{t_0}| = 82,408,626,300$ if $n_{t_0} = 200$ and $k = 6$. In such a situation, it takes too much time to perform the previous approximate size γ LR test. Thus, by a simulation, an alternative approximate size γ LR test with its corresponding quality control scheme for monitoring the (approximate) LR statistic $\tilde{W}_{T,(t_0)}$ could be constructed as follows. First generate an i.i.d. sample $\{(\tilde{\theta}^{*'}_{1,1}, \tilde{y}^{*'}_{t_0,1})', (\tilde{\theta}^{*'}_{1,2}, \tilde{y}^{*'}_{t_0,2})', \ldots, (\tilde{\theta}^{*'}_{1,r}, \tilde{y}^{*'}_{t_0,r})'\}$ of size r for some large positive integer r, e.g., $r = 100,000$, such that $\tilde{\theta}^*_{1,s} \sim \Phi_\alpha|_{\alpha=\tilde{\alpha}_{T,(t_0)}}$ $(\equiv \Phi_{\tilde{\alpha}_{T,(t_0)}})$ and $\tilde{y}^*_{t_0}|\tilde{\theta}^*_{1,s} \sim F_{y_{t_0}|\theta_{t_0}}|_{\theta_{t_0}=\tilde{\theta}^*_{1,s}}$ for $s = 1, 2, \ldots, r$. For $s = 1, 2, \ldots, r$, set $\tilde{W}^*_{T,(t_0),r,s} \equiv \tilde{W}_{T,(t_0)}|_{y_{t_0}=\tilde{y}^*_{t_0,s}}$. Let $(\tilde{W}^*_{T,(t_0),r,(1)}, \tilde{W}^*_{T,(t_0),r,(2)}, \ldots, \tilde{W}^*_{T,(t_0),r,(r)})$ be a permutation of $(\tilde{W}^*_{T,(t_0),r,1}, \tilde{W}^*_{T,(t_0),r,2}, \ldots, \tilde{W}^*_{T,(t_0),r,r})$ such that $\tilde{W}_{T,(t_0),r,(1)} \leq \tilde{W}_{T,(t_0),r,(2)}$ $\leq \ldots \leq \tilde{W}_{T,(t_0),r,(r)}$. Set $\tilde{m}^*_{T,(t_0),r} \equiv [r(1-\gamma)] + 1$ and

$$R\tilde{U}CL^*_{T,(t_0),r} \equiv \tilde{W}^*_{T,(t_0),(\tilde{m}^*_{T,(t_0),r})}, \tag{29}$$

where $[r(1-\gamma)]$ denotes the largest integer less than or equal to $r(1-\gamma)$, e.g., $\tilde{m}^*_{T,(t_0),r} = 99,731$ if $r = 100,000$ and $\gamma = 2\Phi(-3)$. Let $\tilde{m}^*_{L,T,(t_0),r}(\alpha)$, $\tilde{m}^*_{U,T,(t_0),r}(\alpha) \in \{1, 2, \ldots, r\}$ such that

$$\begin{aligned} \tilde{W}^*_{T,(t_0),(\tilde{m}_{L,T,(t_0),r}-1)} &< \tilde{W}^*_{T,(t_0),(\tilde{m}_{L,T,(t_0),r})} = R\tilde{UCL}^*_{T,(t_0),r} \\ = \tilde{W}^*_{T,(t_0),(\tilde{m}_{U,T,(t_0),r})} &< \tilde{W}^*_{T,(t_0),(\tilde{m}_{U,T,(t_0),r}+1)}, \end{aligned}$$

where $\tilde{W}_{T,(t_0),(0)} \equiv -1$ and $\tilde{W}_{T,(t_0),(r+1)} \equiv \infty$. Set

$$\tilde{\gamma}^*_{RUCL,T,(t_0),r} \equiv \frac{\gamma - (r - \tilde{m}^*_{U,T,(t_0),r})/r}{(\tilde{m}^*_{U,T,(t_0),r} - \tilde{m}^*_{L,T,(t_0),r} + 1)/r} = \frac{r\gamma - r + \tilde{m}^*_{U,T,(t_0),r}}{\tilde{m}^*_{U,T,(t_0),r} - \tilde{m}^*_{L,T,(t_0),r} + 1}. \quad (30)$$

Finally, the monitoring scheme for the manufacturing process at time t_0 is proposed as follows. If $\tilde{W}_{T,(t_0)} > R\tilde{UCL}^*_{T,(t_0),r}$, then we reject H_0 and declare that the process is out of control; if $\tilde{W}_{T,(t_0)} < R\tilde{UCL}^*_{T,(t_0),r}$, then we accept H_0 and declare that the process is in control; if $\tilde{W}_{T,(t_0)} = R\tilde{UCL}^*_{T,(t_0),r}$, then with probability $\tilde{\gamma}^*_{RUCL,T,(t_0),r}$ we reject H_0 and declare that the process is out of control or, equivalently, with probability $1 - \tilde{\gamma}^*_{RUCL,T,(t_0),r}$ we accept H_0 and declare that the process is in control.

Note that both $R\tilde{UCL}^*_{T,(t_0),r} - R\tilde{UCL}_{T,(t_0)}$ and $\tilde{\gamma}^*_{RUCL,T,(t_0),r} - \tilde{\gamma}_{RUCL,T,(t_0)}$ converge to 0 with probability one as $r \to \infty$, where the rate of convergence for the latter is much slower than that for the former as $r \to \infty$. Thus, this test converges to the previous approximate size γ LR test as $r \to \infty$.

4. Simulation study

In order to study the performance of this quality control scheme, we compute the average run length (ARL). The in-control ARL, denoted by ARL_0, is the average number of times to get an out-of-control signal when the manufacturing process is in control. The out-of-control ARL, denoted by ARL_1, is the average number of times to get an out-of-control signal when the manufacturing process is out of control. If the false alarm rate is γ, then $\text{ARL}_0 = 1/\gamma$, e.g., $\text{ARL}_0 \approx 370.3983$ if $\gamma = 2\Phi(-3)$.

To calculate the ARL_1, first of all, find both the randomized upper control limit $RUCL_{(t_0)}(\alpha)$ and the randomization probability $\gamma_{RUCL,(t_0)}(\alpha)$. Next, simulate an i.i.d. sample $\{(\theta^{*'}_{1,1}, y^{*'}_{t_0,1})', \ldots, (\theta^{*'}_{1,r}, y^{*'}_{t_0,r})'\}$ of size r for some large positive integer r, e.g., $r = 100,000$, such that $\theta^*_{1,s} \sim F_{\theta_{t_0}}$ and $y^*_{t_0,s}|\theta^*_{1,s} \sim F_{y_{t_0}|\theta_{t_0}}|_{\theta_{t_0}=\theta^*_{1,s}}$ for $s = 1, 2, \ldots, r$, where $F_{\theta_{t_0}} \neq \Phi_\alpha$. For $s = 1, 2, \ldots, r$, set $W^*_{(t_0),r,s}(\alpha) \equiv W_{(t_0)}(\alpha)|_{y_{t_0}=y^*_{t_0},s}$. Let $(W^*_{(t_0),r,(1)}(\alpha), W^*_{(t_0),r,(2)}(\alpha), \ldots, W^*_{(t_0),r,(r)}(\alpha))$ be a permutation of $(W^*_{(t_0),r,1}(\alpha), W^*_{(t_0),r,2}(\alpha), \ldots, W^*_{(t_0),r,r}(\alpha))$ such that $W^*_{(t_0),r,(1)}(\alpha) \leq W^*_{(t_0),r,(2)}(\alpha) \leq \ldots \leq W^*_{(t_0),r,(r)}(\alpha)$. Let $m^*_{L,(t_0),r}(\alpha)$, $m^*_{U,(t_0),r}(\alpha) \in \{1, 2, \ldots, r\}$ such that

$$W^*_{(t_0),r,(m^*_{L,(t_0),r}-1)}(\alpha) \quad < \quad RUCL_{(t_0)}(\alpha) \quad \leq \quad W^*_{(t_0),r,(m^*_{U,(t_0),r}+1)}(\alpha)$$

and

$$W^*_{(t_0),r,(m^*_{U,(t_0),r})}(\alpha) \;\leq\; RUCL_{(t_0)}(\alpha) \;<\; W^*_{(t_0),r,(m^*_{U,(t_0),r}+1)}(\alpha),$$

where $W^*_{(t_0),r,(0)}(\alpha) \equiv 0$ and $W^*_{(t_0),r,(r+1)}(\alpha) \equiv \infty$. Set

$$P_{out,(t_0),r}(\alpha) \equiv \frac{r - m^*_{U,(t_0),r}(\alpha) + \gamma[m^*_{U,(t_0),r}(\alpha) - m^*_{L,(t_0),r}(\alpha) + 1]}{r}.$$

Then $P_{out,(t_0),r}(\alpha)$ converges to the out-of-control probability as $r \to \infty$ when the manufacturing process is out of control. Thus, $1/P_{out,(t_0),r}(\alpha) \to ARL_1$ as $r \to \infty$.

Just for the illustration purpose, the following tables give the randomized upper control limit $RUCL_{(t_0)}(\alpha)$ and its randomization probability $\gamma_{RUCL,(t_0)}(\alpha)$ for various α's and n_{t_0}'s for $k = 2$.

Set $p_1^{(0)} \equiv (p_{11}^{(0)}, \ldots, p_{k1}^{(0)})'$, $p_{1,L}^{(0)} \equiv (p_{11,L}^{(0)}, \ldots, p_{k1,L}^{(0)})'$, $p_{1,U}^{(0)} \equiv (p_{11,U}^{(0)}, \ldots, p_{k1,U}^{(0)})'$, $p_{01}^{(0)} \equiv 1 - \sum_{i=1}^{k} p_{i1}^{(0)}$, $\mu \equiv (\log(p_{11}^{(0)}/p_{01}^{(0)}), \ldots, \log(p_{k1}^{(0)}/p_{01}^{(0)}))'$, $\sigma_i \equiv [\log(p_{i1,U}^{(0)}/p_{01,L}^{(0)}) - \log(p_{i1,L}^{(0)}/p_{01,U}^{(0)})]/2$ for $i = 1, \ldots, k$, and $\Sigma_{uv} \equiv \rho_{uv}\sigma_u\sigma_v$ for $u, v = 1, \ldots, k$, where $p_1^{(0)} \in \mathcal{P}_1$, $0 < p_{i1,L}^{(0)} < p_{i1}^{(0)} < p_{i1,U}^{(0)} < 1$, $-1 < \rho_{uv} < 1$, and $\Sigma \equiv (\Sigma_{uv})$ is a positive definite matrix.

Table 1: $p_{(01)}^{(0)}$, $p_{01,L}^{(0)}$, $p_{01,U}^{(0)}$, and ρ

	$p_{01}^{(0)}$	$p_{01,L}^{(0)}$	$p_{01,U}^{(0)}$	ρ
Case 1	0.85	0.80	0.90	0.3
Case 2	0.80	0.75	0.85	0.3
Case 3	0.70	0.65	0.75	0.3
Case 4	0.60	0.55	0.65	0.3
Case 5	0.50	0.45	0.55	0.3

Table 2: $p_1^{(0)}$, $p_{1,L}^{(0)}$, and $p_{1,U}^{(0)}$

	$p_1^{(0)}$	$p_{1,L}^{(0)} = (p_{11,L}^{(0)}, p_{21,L}^{(0)})'$	$p_{1,U}^{(0)} = (p_{11,U}^{(0)}, p_{21,U}^{(0)})'$
Case 1	$(0.10, 0.05)'$	$(0.05, 0.025)'$	$(0.15, 0.075)'$
Case 2	$(0.15, 0.05)'$	$(0.05, 0.025)'$	$(0.20, 0.075)'$
Case 3	$(0.20, 0.10)'$	$(0.15, 0.075)'$	$(0.25, 0.125)'$
Case 4	$(0.30, 0.10)'$	$(0.20, 0.075)'$	$(0.35, 0.125)'$
Case 5	$(0.30, 0.20)'$	$(0.20, 0.150)'$	$(0.35, 0.250)'$

Table 3: α

	α
Case 1	$(-2.1401, -2.8332, 2.9708, -0.8912, 2.9708)'$
Case 2	$(-1.6740, -2.7726, 1.9241, -0.7129, 2.9350)'$
Case 3	$(-1.2528, -1.9459, 10.279, -3.0838, 10.279)'$
Case 4	$(-0.6931, -1.7918, 8.3242, -2.6770, 9.5656)'$
Case 5	$(-0.5108, -0.9163, 7.6044, -2.4378, 8.6831)'$

Table 4: $RUCL_{(t_0)}(\alpha)$

	$n_{t_0} = 20$	$n_{t_0} = 30$	$n_{t_0} = 50$	$n_{t_0} = 100$
Case 1	11.1625	12.3359	12.9654	14.3988
Case 2	12.1689	12.5600	13.3028	14.8388
Case 3	12.6104	12.6891	12.9089	13.5552
Case 4	12.7874	12.7396	13.2475	13.8787
Case 5	13.1051	13.2070	13.5308	14.1464

Table 5: $\gamma_{RUCL,(t_0)}(\alpha)$

	$n_{t_0} = 20$	$n_{t_0} = 30$	$n_{t_0} = 50$	$n_{t_0} = 100$
Case 1	0.0705	0.7295	0.3479	0.5085
Case 2	0.3745	0.4376	0.7363	0.6736
Case 3	0.6094	0.8320	0.3054	0.5559
Case 4	0.8804	0.4088	0.2596	0.1560
Case 5	0.9361	0.5732	0.4771	0.8717

5. Conclusions and possible generalizations

In the paper, the normal-binomial or -multinomial model rather than the beta-binomial or Dirichlet-multinomial model in Chen et al. (2004) is used for manufacturing categorical data. Then the EB inference for manufacturing categorical data is discussed. Finally, utilizing the LR method, an EB process monitoring technique for manufacturing categorical data is proposed to monitor the process whether it is in control or not.

In the paper, we assume that the transformed random parameter vectors are i.i.d. normal when the manufacturing process is in control. However, in practice they are usually correlated and stationary rather than independent. What we want to do next is to consider the correlated and stationary normal case, e.g., the vector autoregressive (VAR) models, or the vector moving average (VMA) models, or the vector autoregressive moving average (VARMA) models.

Appendix A

For $t = 1, 2, \ldots, T$, set

$$\begin{aligned} \hat{\theta}_t^{(0)} &\equiv \left(\log\left(\frac{y_{1t}+1/2}{y_{0t}+1/2}\right), \ldots, \log\left(\frac{y_{kt}+1/2}{y_{0t}+1/2}\right) \right)', \\ \hat{\mu}_T^{(0)} &\equiv \frac{1}{T}\sum_{t=1}^{T} \hat{\theta}_t^{(0)}, \end{aligned}$$

and

$$\hat{\Sigma}_T^{(0)} \equiv \frac{1}{T-1}\sum_{t=1}^{T} \left(\hat{\theta}_t^{(0)} - \hat{\mu}_T^{(0)}\right)\left(\hat{\theta}_t^{(0)} - \hat{\mu}_T^{(0)}\right)'.$$

Then $\hat{\alpha}_T^{(0)} \equiv \alpha|_{\mu=\hat{\mu}_T^{(0)}, \Sigma=\hat{\Sigma}_T^{(0)}}$ is an MME of α given $y_{1:T}$

Appendix B

One way to numerically evaluate all of $p(\theta_t|y_t;\hat{\alpha}_T)$, $E(\theta_t|y_t;\hat{\alpha}_T)$, and $f_T(y_t)$ for $t=1,2,\ldots,T$ is to perform the following simulation. Let $t\in\{1,2,\ldots,T\}$ be fixed. First simulate an i.i.d. sample $\{\theta^*_{1,1},\theta^*_{1,2},\ldots,\theta^*_{1,r}\}$ of size r from $\Phi_{\hat{\alpha}_T}$ for some large positive integer r, e.g., $r=100{,}000$. Set $\theta^*_{1,s}\equiv(\theta^*_{11,s},\ldots,\theta^*_{k1,s})'$ for $s=1,2,\ldots,r$. Next, compute

$$
\begin{aligned}
a^*_{t,r}(\hat{\alpha}_T;y_t) &\equiv \frac{1}{r}\sum_{s=1}^{r}\frac{\exp(y_t'\,\theta^*_{1,s})}{[1+\sum_{i=1}^{k}\exp(\theta^*_{i1,s})]^{n_t}},\\
p^*_r(\theta_t|y_t;\hat{\alpha}_T) &\equiv \frac{\exp(y_t'\,\theta_t)}{[1+\sum_{i=1}^{k}\exp(\theta_{it})]^{n_t}}\cdot\frac{\exp[-(\theta_t-\hat{\mu}_T)'\,\hat{\Sigma}_T^{-1}\,(\theta_t-\hat{\mu}_T)/2]}{(2\pi)^{k/2}\,|\hat{\Sigma}_T|^{1/2}\,a^*_{t,r}(\hat{\alpha}_T;y_t)},\\
E^*_r(\theta_t|y_t;\hat{\alpha}_T) &\equiv \frac{1}{a^*_{t,r}(\hat{\alpha}_T;y_t)}\cdot\frac{1}{r}\sum_{s=1}^{r}\frac{\exp(y_t'\,\theta^*_{1,s})}{[1+\sum_{i=1}^{k}\exp(\theta^*_{i1,s})]^{n_t}}\cdot\theta^*_{1,s},
\end{aligned}
$$

and

$$
f^*_{T,r}(y_t)\equiv 1_{\mathcal{Y}_t}(y_t)\cdot\frac{n_t!}{\prod_{i=0}^{k}y_{it}!}\cdot a^*_{t,r}(\hat{\alpha}_T;y_t)
$$

for $t=1,2,\ldots,T$, where $\hat{\mu}_T\equiv\mu|_{\alpha=\hat{\alpha}_T}$ and $\hat{\Sigma}_T\equiv\Sigma|_{\alpha=\hat{\alpha}_T}$. Then for $t=1,2,\ldots,T$, all of $p^*_r(\theta_t|y_t;\hat{\alpha}_T)-p(\theta_t|y_t;\hat{\alpha}_T)$, $E^*_r(\theta_t|y_t;\hat{\alpha}_T)-E(\theta_t|y_t;\hat{\alpha}_T)$, and $f^*_{T,r}(y_t)-\hat{f}_T(y_t)$ converge to 0 with probability one as $r\to\infty$.

A quicker way to numerically evaluate all of $p(\theta_t|y_t;\hat{\alpha}_T)$, $E(\theta_t|y_t;\hat{\alpha}_T)$, and $\hat{f}_T(y_t)$ for $t=1,2,\ldots,T$ is to use the following multivariate Gauss-Hermite integration method. Compute

$$
\begin{aligned}
\tilde{a}_{t,m}(\hat{\alpha}_T; y_t) &\equiv \pi^{-k/2} \sum_{u_1=1}^{m_1} w_{u_1}^{(1)} \cdots \sum_{u_k=1}^{m_k} w_{u_k}^{(k)} \cdot \frac{\exp(y_t' \tilde{\theta}_u)}{[1 + \sum_{j=1}^{k} \exp(\tilde{\theta}_{u_j})]^{n_t}}, \\
\tilde{p}_m(\theta_t | y_t; \hat{\alpha}_T) &\equiv \frac{\exp(y_t' \theta_t)}{[1 + \sum_{i=1}^{k} \exp(\theta_{it})]^{n_t}} \cdot \frac{\exp[-(\theta_t - \hat{\mu}_T)' \hat{\Sigma}_T^{-1} (\theta_t - \hat{\mu}_T)/2]}{(2\pi)^{k/2} |\hat{\Sigma}_T|^{1/2} \tilde{a}_{t,m}(\hat{\alpha}_T; y_t)}, \\
\tilde{E}_m(\theta_t | y_t; \hat{\alpha}_T) &\equiv \frac{\pi^{-k/2}}{\tilde{a}_{t,m}(\hat{\alpha}_T; y_t)} \cdot \sum_{u_1=1}^{m_1} w_{u_1}^{(1)} \cdots \sum_{u_k=1}^{m_k} w_{u_k}^{(k)} \cdot \frac{\exp(y_t' \tilde{\theta}_u)}{[1 + \sum_{j=1}^{k} \exp(\tilde{\theta}_{u_j})]^{n_t}} \cdot \tilde{\theta}_u,
\end{aligned}
$$

and

$$
\tilde{\hat{f}}_m(y_t) \equiv 1_{\mathcal{Y}_t}(y_t) \cdot \frac{n_t!}{\prod_{i=0}^{k} y_{it}!} \cdot \tilde{a}_{t,m}(\hat{\alpha}_T; y_t)
$$

for $t = 1, 2, \ldots, T$, where $m \equiv (m_1, \ldots, m_k)$, $u \equiv (u_1, \ldots, u_k)$, $\tilde{\theta}_u \equiv (\tilde{\theta}_{u_1}, \ldots, \tilde{\theta}_{u_k})' \equiv \sqrt{2} \hat{\Sigma}_T^{1/2} x_u + \hat{\mu}_T$, $x_u \equiv (x_{u_1}^{(1)}, \ldots, x_{u_k}^{(k)})'$, $x_{u_j}^{(j)}$ denotes the u_jth zero of the Hermite polynomial with degree m_j, and $w_{u_j}^{(j)}$ denotes the corresponding weight for $x_{u_j}^{(j)}$. Then for $t = 1, 2, \ldots, T$, all of $\tilde{p}_m(\theta_t | y_t; \hat{\alpha}_T) - p(\theta_t | y_t; \hat{\alpha})$, $\tilde{E}_m(\theta_t | y_t; \hat{\alpha}_T) - E(\theta_t | y_t; \hat{\alpha}_T)$, and $\tilde{\hat{f}}_{T,m}(y_t) - \hat{f}_T(y_t)$ converge to 0 as $\min\{m_1, \ldots, m_k\} \to \infty$.

One way to numerically evaluate $q_{T,1-c_T}(y_{1:T}; \hat{\alpha}_T)$ is to perform the following simulation. First simulate an i.i.d. sample $\{\tilde{\theta}_{1:T,1}, \tilde{\theta}_{1:T,2}, \ldots, \tilde{\theta}_{1:T,r}\}$ of size r from $p(\theta_{1:T} | y_{1:T}; \alpha)|_{\alpha = \hat{\alpha}_T}$ $(\equiv p(\theta_{1:T} | y_{1:T}; \hat{\alpha}_T))$ by the rejection method for some large positive integer r, e.g., $r = 100,000$. For $s = 1, 2, \ldots, r$, set $\tilde{\theta}_{1:T,s} \equiv (\tilde{\theta}_{1,s}', \tilde{\theta}_{2,s}', \ldots, \tilde{\theta}_{T,s}')'$ where $\tilde{\theta}_{t,s} \equiv (\tilde{\theta}_{1t,s}, \ldots, \tilde{\theta}_{kt,s})'$ for $t = 1, 2, \ldots, T$. Compute $p(\theta_{1:T} | y_{1:T}; \hat{\alpha}_T)|_{\theta_{1:T} = \tilde{\theta}_{1:T,s}}$ $(\equiv p(\tilde{\theta}_{1:T,s} | y_{1:T}; \hat{\alpha}_T))$ for $s = 1, 2, \ldots, r$. Let $(\tilde{\theta}_{1:T,(1)}, \tilde{\theta}_{1:T,(2)}, \ldots, \tilde{\theta}_{1:T,(r)})$ be a permutation of $(\tilde{\theta}_{1:T,1}, \tilde{\theta}_{1:T,2}, \ldots, \tilde{\theta}_{1:T,r})$ such that $p(\tilde{\theta}_{1:T,(1)} | y_{1:T}; \hat{\alpha}_T) \leq p(\tilde{\theta}_{1:T,(2)} | y_{1:T}; \hat{\alpha}_T) \leq \ldots \leq p(\tilde{\theta}_{1:T,(r)} | y_{1:T}; \hat{\alpha}_T)$. Set $\tilde{q}_{T,1-c_T,r}(y_{1:T}; \hat{\alpha}_T) \equiv p(\tilde{\theta}_{1:T,([rc])} | y_{1:T}; \hat{\alpha}_T)$, where $[rc]$ denotes the largest integer less than or equal to rc, e.g., $[rc] = 5,000$ if $r = 100,000$ and $c = 0.05$. Then $\tilde{q}_{T,1-c_T,r}(y_{1:T};\hat{\alpha}_T) \to q_{T,1-c_T}(y_{1:T};\hat{\alpha}_T)$ with probability one as $r \to \infty$.

Similarly, one way to numerically evaluate $q_{(t),1-c_{(t)}}(y_t;\hat{\alpha}_T)$ for $t = 1,2,\ldots,T$ is to perform the following simulation. Let $t \in \{1,2,\ldots,T\}$ be fixed. First simulate an i.i.d. sample $\{\tilde{\theta}_{t,1},\tilde{\theta}_{t,2},\ldots,\tilde{\theta}_{t,r}\}$ of size r from $p(\theta_t|y_t;\hat{\alpha}_T)$ by the rejection method for some large positive integer r, e.g., $r = 100,000$. For $s = 1,2,\ldots,r$, set $\tilde{\theta}_{t,s} \equiv (\tilde{\theta}_{1t,s},\ldots,\tilde{\theta}_{kt,s})'$. Compute $p(\theta_t|y_t;\alpha)|_{\theta_t=\tilde{\theta}_{t,s},\alpha=\hat{\alpha}_T}$ $(\equiv p(\tilde{\theta}_{t,s}|y_t;\hat{\alpha}_T))$ for $s = 1,2,\ldots,r$. Let $(\tilde{\theta}_{t,(1)},\ \tilde{\theta}_{t,(2)},\ \ldots,\tilde{\theta}_{t,(r)})$ be a permutation of $(\tilde{\theta}_{t,1},\ \tilde{\theta}_{t,2},\ldots,\tilde{\theta}_{t,r})$ such that $p(\tilde{\theta}_{t,(1)}|y_t;\hat{\alpha}_T) \leq p(\tilde{\theta}_{t,(2)}|y_t;\hat{\alpha}_T) \leq \ldots \leq p(\tilde{\theta}_{t,(r)}|y_t;\hat{\alpha}_T)$. Set $\tilde{q}_{(t),1-c_{(t)},r}(y_t;\hat{\alpha}_T) \equiv p(\tilde{\theta}_{t,([rc])}|y_t;\hat{\alpha}_T)$, where $[rc]$ denotes the largest integer less than or equal to rc, e.g., $[rc] = 5,000$ if $r = 100,000$ and $c = 0.05$. Then $\tilde{q}_{(t),1-c_{(t)},r}(y_t;\hat{\alpha}_T) \to q_{(t),1-c_{(t)}}(y_t;\hat{\alpha}_T)$ with probability one as $r \to \infty$.

Appendix C

Under H_0, it follows from the Taylor series expansion that

$$\begin{aligned} 0 &= S_T\left(\hat{\alpha}_{T,(t_0)}\right) + S_{(t_0)}\left(\hat{\alpha}_{T,(t_0)}\right) \\ &\approx S_{(t_0)}\left(\hat{\alpha}_T\right) - \left[J_T\left(\hat{\alpha}_T\right) + J_{(t_0)}\left(\hat{\alpha}_T\right)\right]\left(\hat{\alpha}_{T,(t_0)} - \hat{\alpha}_T\right) \end{aligned}$$

as $T \to \infty$, which implies that

$$\hat{\alpha}_{T,(t_0)} - \hat{\alpha}_T \approx \left[J_T\left(\hat{\alpha}_T\right) + J_{(t_0)}\left(\hat{\alpha}_T\right)\right]^{-1} S_{(t_0)}\left(\hat{\alpha}_T\right)$$

as $T \to \infty$. Thus, under H_0,

$$\hat{\alpha}_{T,(t_0)} = \hat{\alpha}_T + O_p(T^{-1}) = \alpha + O_p(T^{-1/2})$$

as $T \to \infty$. Observe that under H_0,

$$\begin{aligned}
& \ell_T\left(\hat{\alpha}_{T,(t_0)}\right) + \ell_{(t_0)}\left(\hat{\alpha}_{T,(t_0)}\right) - \ell_T\left(\hat{\alpha}_T\right) \\
= \; & \ell_{(t_0)}\left(\hat{\alpha}_T\right) + S'_{(t_0)}\left(\hat{\alpha}_T\right)\left(\hat{\alpha}_{T,(t_0)} - \hat{\alpha}_T\right) \\
& -\frac{1}{2}\left(\hat{\alpha}_{T,(t_0)} - \hat{\alpha}_T\right)'\left[J_T\left(\hat{\alpha}_T\right) + J_{(t_0)}\left(\hat{\alpha}_T\right)\right]\left(\hat{\alpha}_{T,(t_0)} - \hat{\alpha}_T\right) + O_p\left(T^{-2}\right) \\
= \; & \ell_{(t_0)}\left(\hat{\alpha}_T\right) + \frac{1}{2} S'_{(t_0)}\left(\hat{\alpha}_T\right)\left[J_T\left(\hat{\alpha}_T\right) + J_{(t_0)}\left(\hat{\alpha}_T\right)\right]^{-1} S_{(t_0)}\left(\hat{\alpha}_T\right) + O_p\left(T^{-2}\right) \\
= \; & \ell_{(t_0)}\left(\hat{\alpha}_T\right) + O_p\left(T^{-1}\right) \\
= \; & \ell_{(t_0)}\left(\alpha\right) + O_p\left(T^{-1/2}\right)
\end{aligned}$$

as $T \to \infty$. Thus, equation (26) holds.

References

1. Agresti, A. (2002). *Categorical Data Analysis*, 2nd ed. John Wiley & Sons, New York.

2. Carlin, B. P. and Louis, T. A. (2000). *Bayes and Empirical Bayes Methods for Data Analysis*, 2nd ed. Chapman & Hall/CRC, Boca Raton.

3. Chen, C.-R., Shiau, J.-J. H., Liao, H.-H., and Feltz, C. J. (2004). An empirical Bayes process monitoring technique for categorical data. Technical Report, Institute of Statistics, National Chiao Tung University, Hsinchu, Taiwan.

4. Feltz, C. J. and Shiau, J.-J. H. (2001). Statitical process monitoring using an empirical Bayes multivariate process control chart. *Quality and Reliability Engineering International*, **17**, 119-124.

5. Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2003). *Bayesian Data Analysis*, 2nd ed. Chapman & Hall/CRC, Boca Raton.

6. Shiau, J.-J. H., Chen, C.-R., and Feltz, C. J. (2004). An empirical Bayes process monitoring technique for polytomous data (to appear). *Quality and Reliability Engineering International*, **20**.

7. Sturm, G. W., Feltz, C. J., and Yousry, M. A. (1991). An empirical Bayes strategy for analysing manufacturing data in real time. *Quality and Reliability Engineering International*, **7**, 159-167.

8. Yousry, M. A., Sturm, G. W., Feltz, C. J., and Noorossana, R. (1991). Process monitoring in real time: empirical Bayes approach - discrete case. *Quality and Reliability Engineering International*, **7**, 123-132.

Table 6: The sample standard error vector(SD), and the sample mean squared error vector(MSE) of $\hat{\alpha}_{MM}$'s, with different α's, T's, and $n_1 = n_2 = \ldots = n_T$.

	$(-2.1401, -2.8332, 2.9708, -0.8912, 2.9708)'$
$T = 100$, $n_1 = 20$, SD	$(0.0886, 0.0842, 0.0967, 0.0757, 0.0689)'$
$T = 100$, $n_1 = 20$, MSE	$(0.0109, 0.0520, 2.6332, 0.4141, 2.1658)'$
$T = 100$, $n_1 = 40$, SD	$(0.8118, 0.0868, 0.1101, 0.0727, 0.0903)'$
$T = 100$, $n_1 = 40$, MSE	$(0.0066, 0.0101, 1.9750, 0.4037, 2.4135)'$
$T = 200$, $n_1 = 20$, SD	$(0.0630, 0.0598, 0.0685, 0.0531, 0.0176)'$
$T = 200$, $n_1 = 20$, MSE	$(0.0070, 0.0484, 2.6694, 0.4033, 2.2012)'$
	$(-1.6740, -2.7726, 1.9241, -0.7129, 2.9350)'$
$T = 100$, $n_1 = 20$, SD	$(0.0951, 0.0847, 0.1305, 0.0849, 0.0721)'$
$T = 100$, $n_1 = 20$, MSE	$(0.2514, 0.08473.2384, 0.3662, 2.1598)'$
$T = 100$, $n_1 = 40$, SD	$(0.0899, 0.0863, 0.1320, 0.0816, 0.0923)'$
$T = 100$, $n_1 = 40$, MSE	$(0.2281, 0.0222, 2.6874, 0.3777, 2.4238)'$
$T = 200$, $n_1 = 20$, SD	$(0.0679, 0.0597, 0.0916, 0.0592, 0.0499)'$
$T = 200$, $n_1 = 20$, MSE	$(0.2478, 0.0817, 3.2858, 0.3599, 2.1909)'$
	$(-1.2528, -1.9459, 10.2792, -3.0838, 10.2792)'$
$T = 100$, $n_1 = 20$, SD	$(0.0679, 0.0780, 0.0844, 0.0547, 0.0851)'$
$T = 100$, $n_1 = 20$, MSE	$(0.7904, 0.8418, 0.5768, 0.3032, 1.6477)'$
$T = 100$, $n_1 = 40$, SD	$(0.05261, 0.0660, 0.0510, 0.0365, 0.0814)'$
$T = 100$, $n_1 = 40$, MSE	$(0.7888, 0.7871, 1.2448, 0.1902, 0.4421)'$
$T = 200$, $n_1 = 20$, SD	$(0.0481, 0.0555, 0.0603, 0.392, 0.0603)'$
$T = 200$, $n_1 = 20$, MSE	$(0.7877, 0.8378, 0.5479, 0.2822, 1.6795)'$

Figure 1: The histograms of $\hat{\mu}_{1,MM}$'s for 10,000 samples and $\alpha = (-2.14, -2.83, 2.97, -0.89, 2.97)'$ where $n_1 = n_2 = \ldots = n_T$.

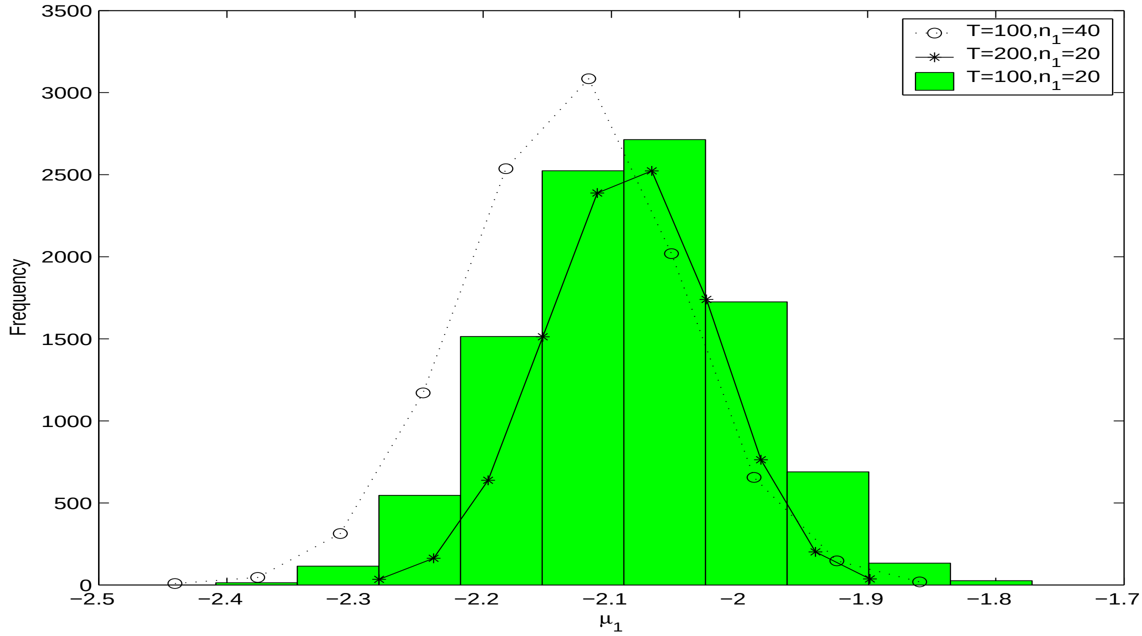

Figure 2: The histograms of $\hat{\mu}_{2,MM}$'s for 10,000 samples and $\alpha = (-2.14, -2.83, 2.97, -0.89, 2.97)'$ where $n_1 = n_2 = \ldots = n_T$.

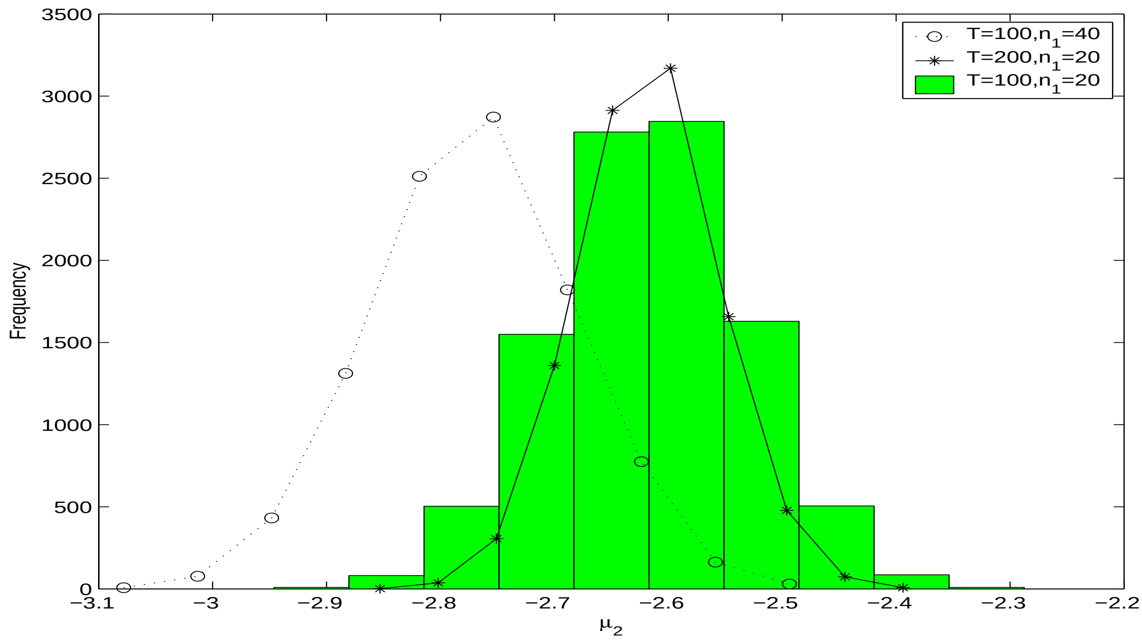

Figure 3: The histograms of $\hat{\rho}_{MM}$'s for $10,000$ samples and $\alpha = (-2.14, -2.83, 2.97, -0.89, 2.97)'$ where $n_1 = n_2 = \ldots = n_T$ and $\rho = 0.3$.

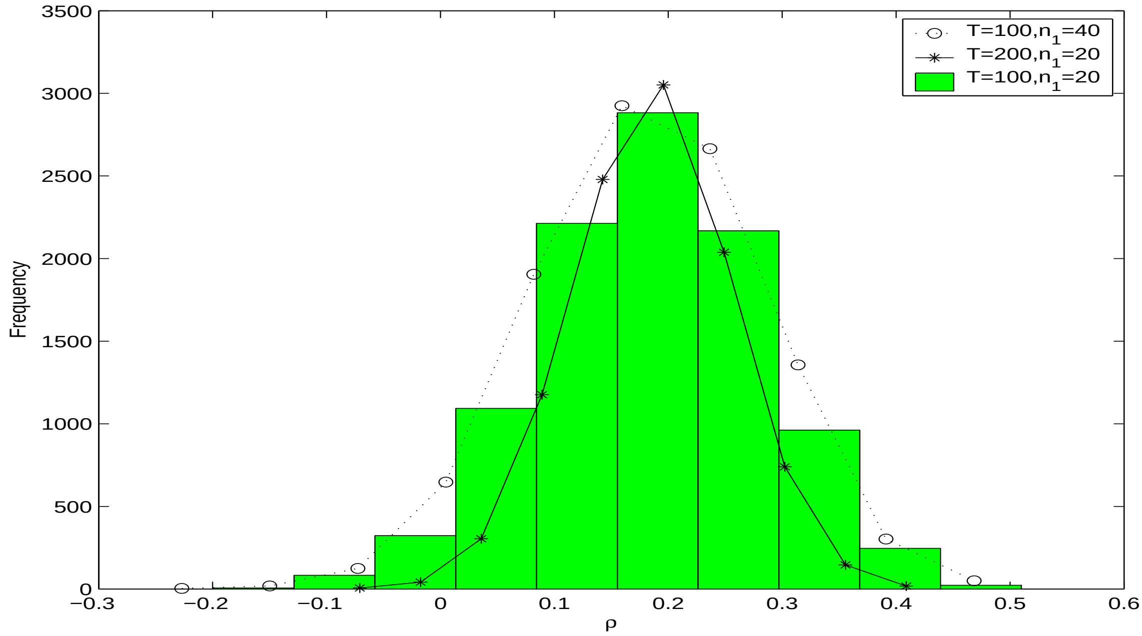

Figure 4: The histograms of $\log[(1+\hat{\rho}_{MM})/(1-\hat{\rho}_{MM})]$'s for $10,000$ samples and $\alpha = (-2.14, -2.83, 2.97, -0.89, 2.97)'$ where $n_1 = n_2 = \ldots = n_T$ and $\log[(1+\rho)/(1-\rho)] = 0.619$.

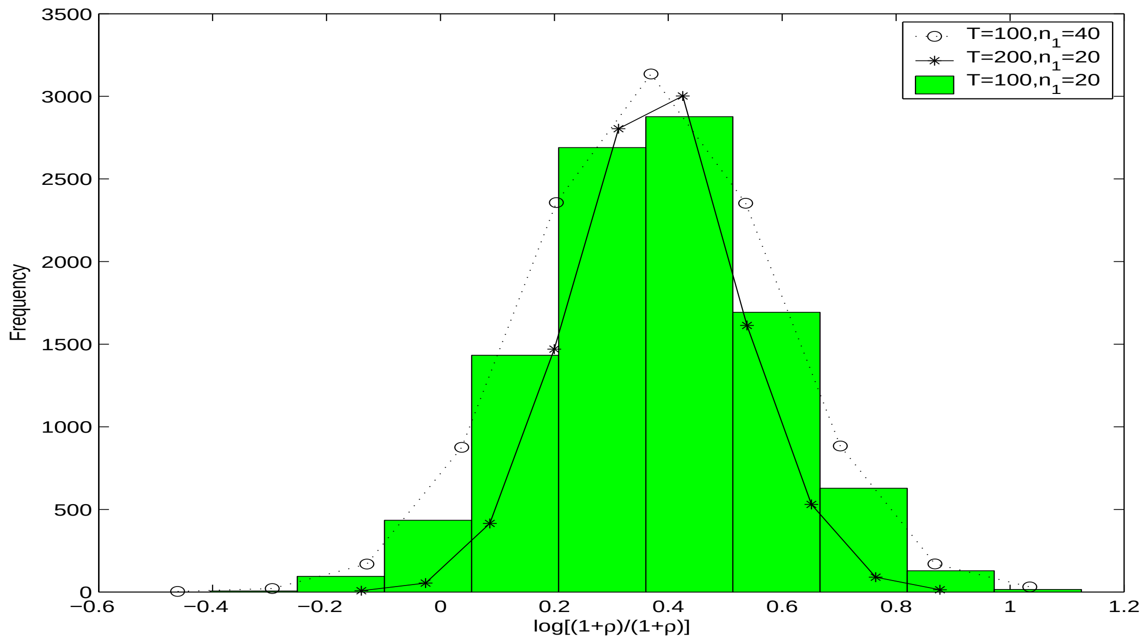

Figure 5: The histograms of $\hat{\sigma}_{1,MM}$'s for 10,000 samples and $\alpha = (-2.14, -2.83, 2.97, -0.89, 2.97)'$ where $n_1 = n_2 = \ldots = n_T$ and $\sigma_1 = 0.608$.

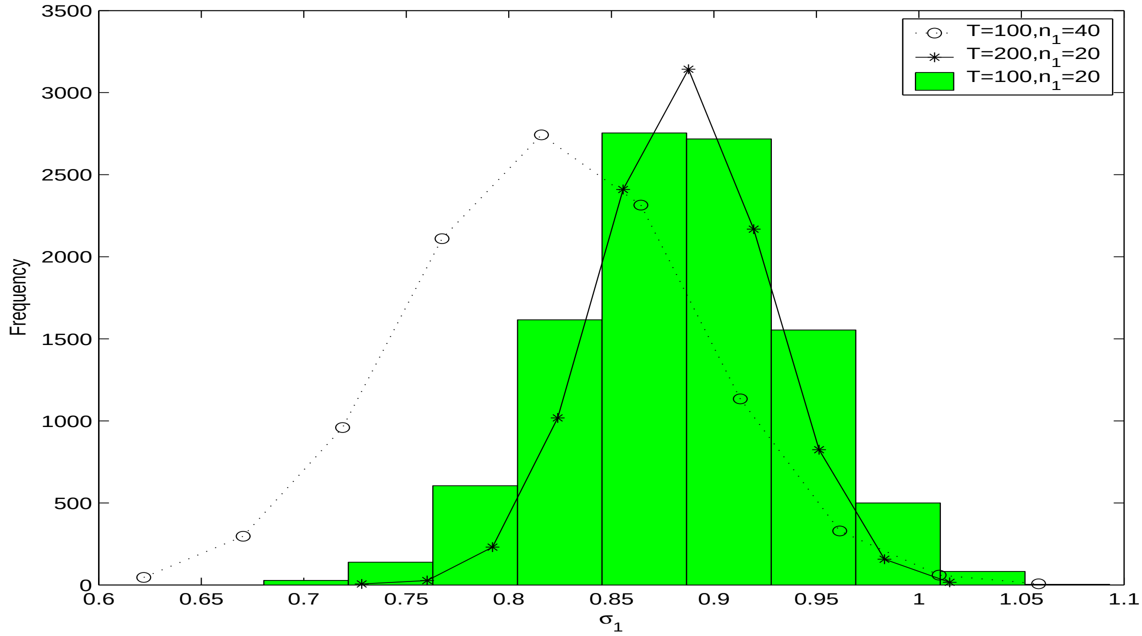

Figure 6: The histograms of $\log(\sigma_{1,\hat{M}M})$'s for 10,000 samples and $\alpha = (-2.14, -2.83, 2.97, -0.89, 2.97)'$ where $n_1 = n_2 = \ldots = n_T$ and $\log(\sigma_1) = -0.497$.

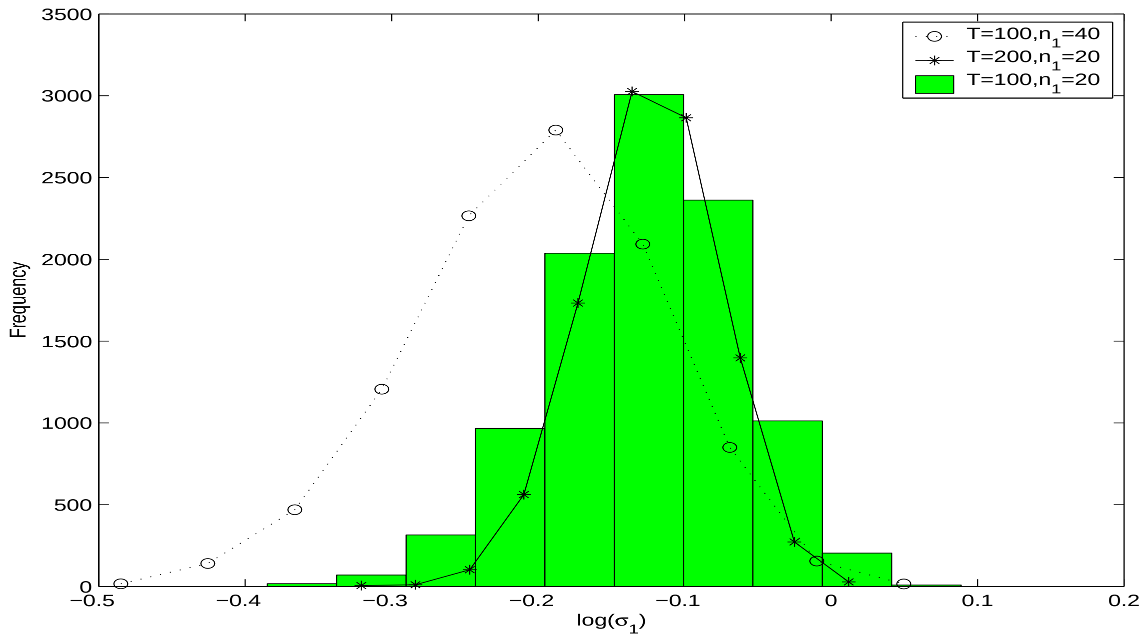

Figure 7: The histograms of $\hat{\sigma}_{2,MM}$'s for 10,000 samples and $\alpha = (-2.14, -2.83, 2.97, -0.89, 2.97)'$ where $n_1 = n_2 = \ldots = n_T$ and $\sigma_2 = 0.608$.

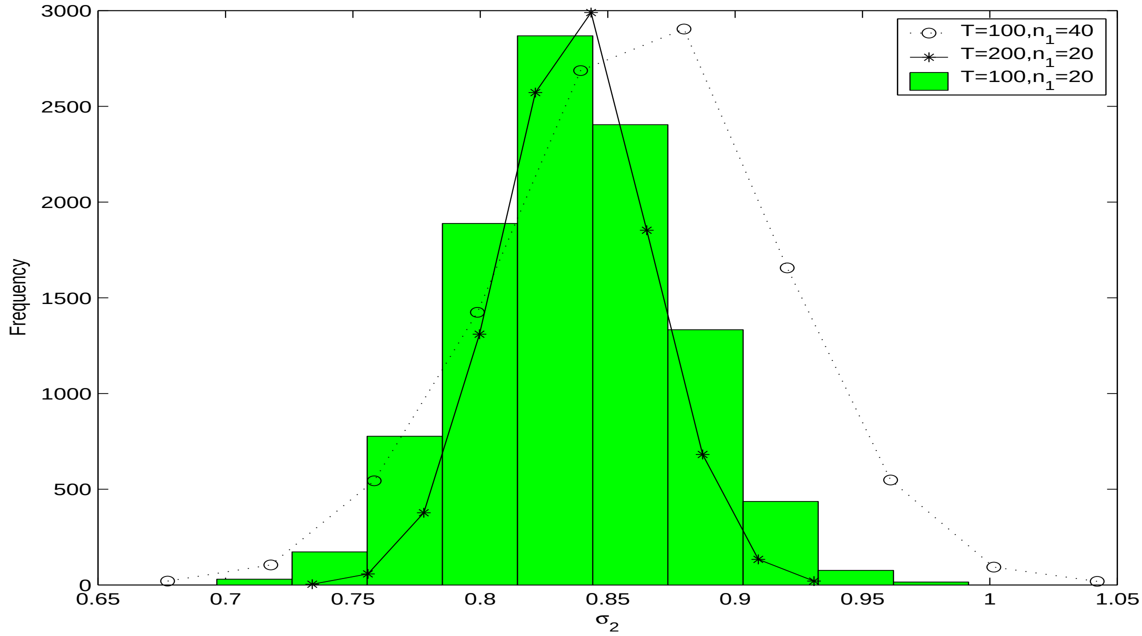

Figure 8: The histograms of $\log(\sigma_{2,\hat{M}M})$'s for 10,000 samples and $\alpha = (-2.14, -2.83, 2.97, -0.89, 2.97)'$ where $n_1 = n_2 = \ldots = n_T$ and $\log(\sigma_2) = -0.497$.

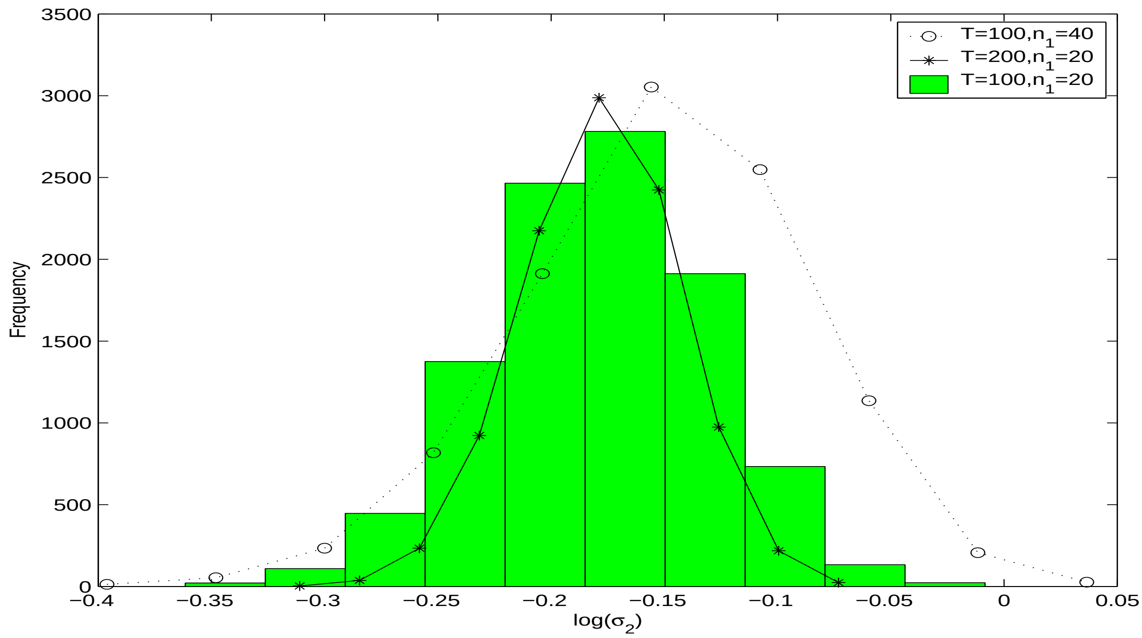

Figure 9: The Q-Q plots of $\hat{\alpha}_{MM}$'s for $10,000$ samples and $\alpha = (-2.14, -2.83, 2.97, -0.89, 2.97)'$, where $n_1 = n_2 = \ldots = n_T = 20$ and $T = 100$.

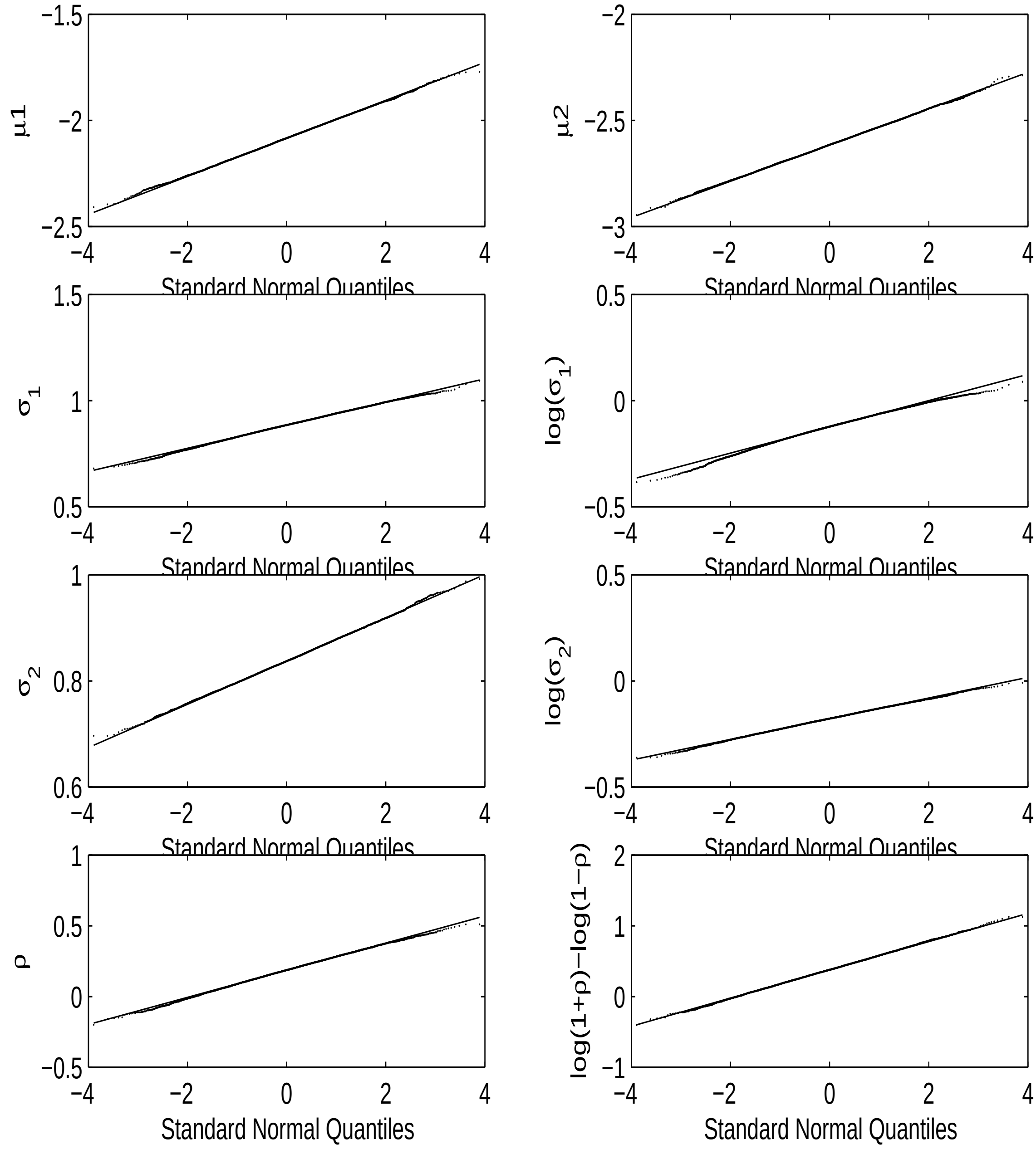

Figure 10: The Q-Q plots of $\hat{\alpha}_{MM}$'s for 10,000 samples and $\alpha = (-2.14, -2.83, 2.97, -0.89, 2.97)'$, where $n_1 = n_2 = \ldots = n_T = 40$ and $T = 100$.

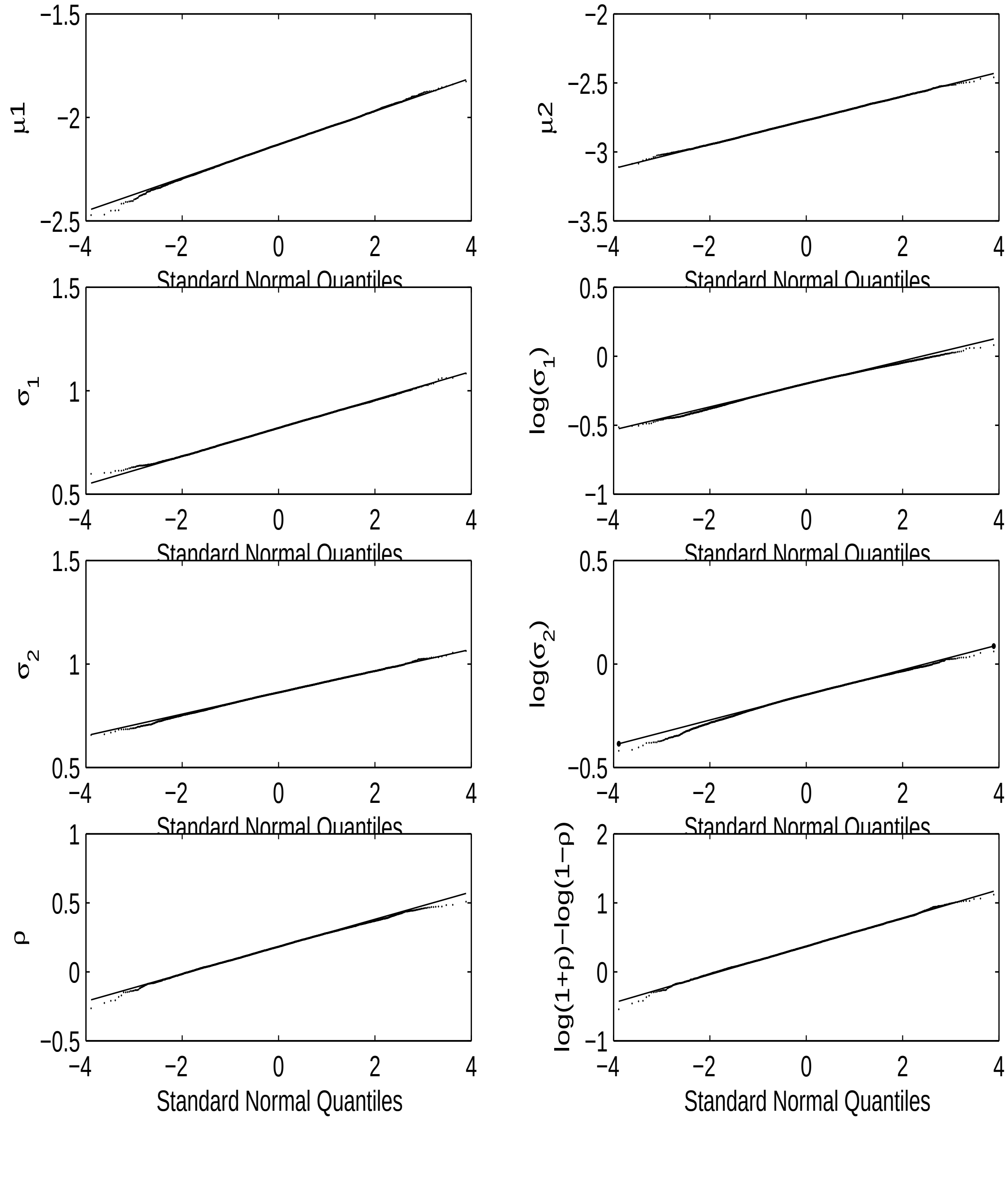

Figure 11: The Q-Q plots of $\hat{\alpha}_{MM}$'s for 10,000 samples and $\alpha = (-2.14, -2.83, 2.97, -0.89, 2.97)'$, where $n_1 = n_2 = \ldots = n_T = 20$ and $T = 200$.

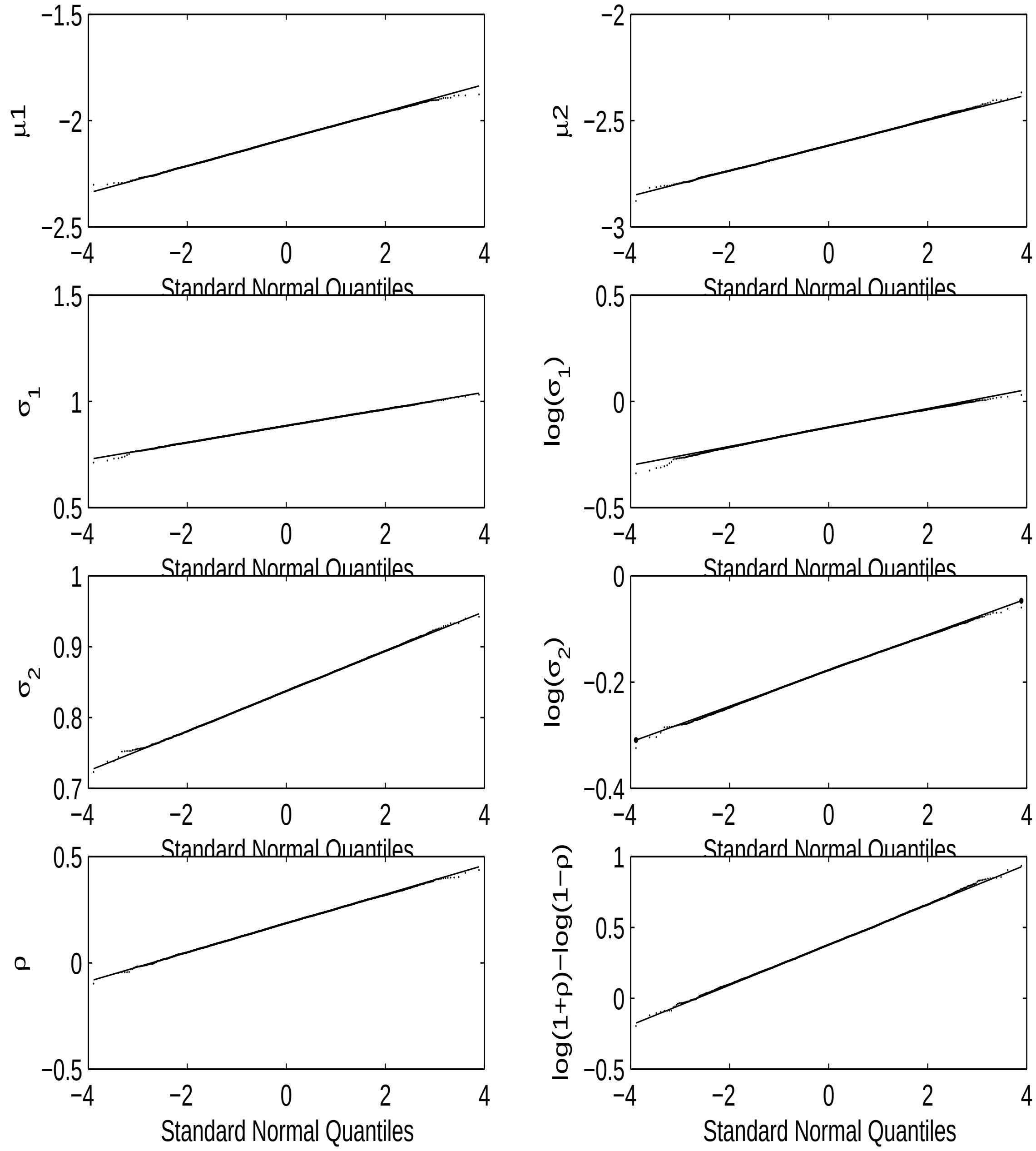

Figure 12: The control chart of $W_{(t_0)}(\alpha)$'s with true distribution is Φ_α for $\alpha = (-2.14, -2.83, 2.97, -0.89, 2.97)'$ and $n_t = 20$ and the generation distribution is Φ_{α_1} for $\alpha_1 = (-0.51, -0.92, 7.60, -2.44, 8.68)'$.

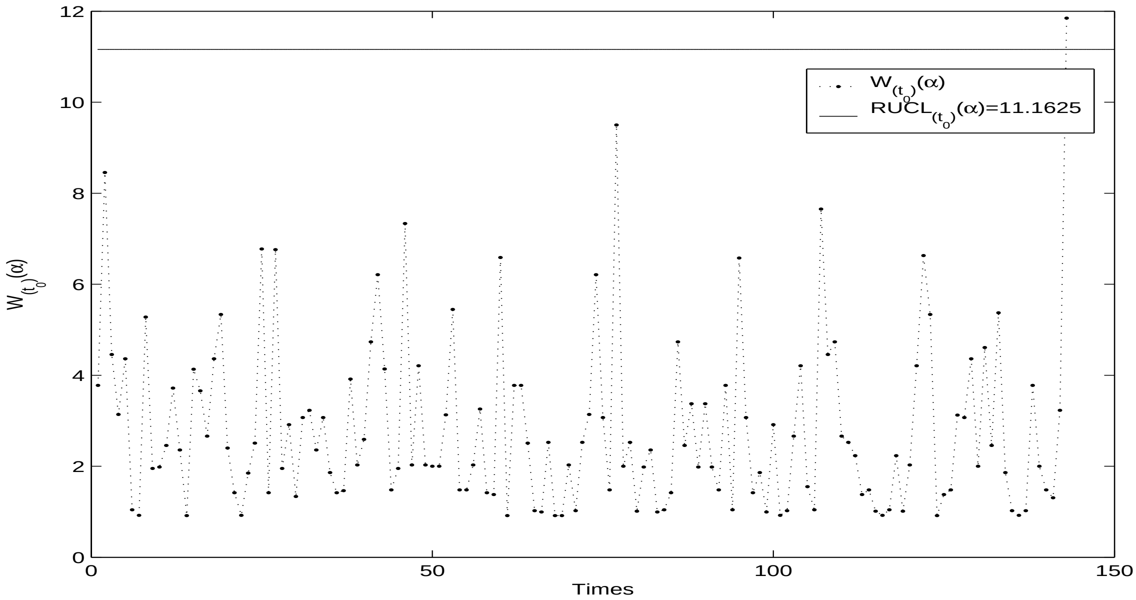

Figure 13: The control chart of $W_{(t_0)}(\alpha)$'s with true distribution is Φ_α for $\alpha = (-2.14, -2.83, 2.97, -0.89, 2.97)'$ and $n_t = 20$ and the generation distribution is Dirichlet(7,2,1).

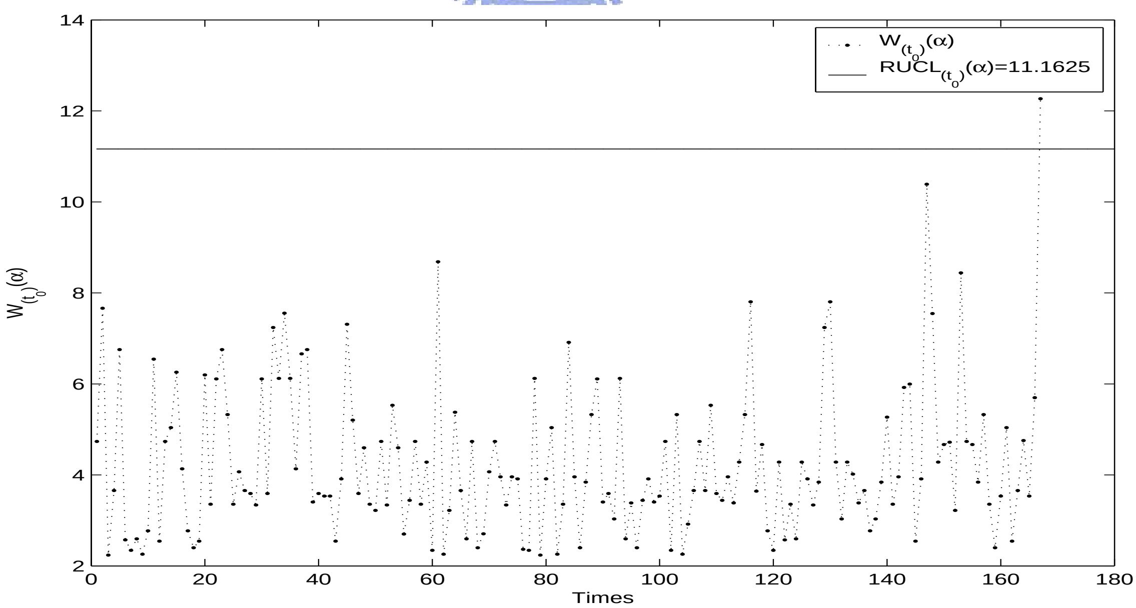

Figure 14: The control chart of $W_{(t_0)}(\alpha)$'s with true distribution is Φ_α for $\alpha = (-2.14, -2.83, 2.97, -0.89, 2.97)'$ and $n_t = 20$ and the generation distribution is also Φ_α.