國立交通大學

統計學研究所

碩士論文

A PROCESS MONITORING TECHNIQUE FOR CATEGORICAL DATA USING A PARAMETRIC TWO-COMPONENTS MIXTURE PRIOR FAMILY

研 究 生：余家慧

指導教授：陳志榮 教授

中華民國九十五年六月

類別資料混合先驗分配之經驗貝氏製程監控技術

A PROCESS MONITORING TECHNIQUE FOR CATEGORICAL DATA USING A PARAMETRIC TWO-COMPONENTS MIXTURE PRIOR FAMILY

研 究 生 ：余家慧　　Student：Chia-Hui Yu

指導教授 ： 陳志榮　博士　　Advisor： Dr. Chih-Rung Chen

國 立 交 通 大 學

統計學研究所

碩 士 論 文

A Thesis

Submitted to Institute of Statistics

College of Science

Nation Chiao Tung University

in partial Fulfillment of the Requirements

for the degree of Master

in

Statistics

June 2006

Hsinchu, Taiwan

中華民國九十五年六月

類別資料混合先驗分配之經驗貝氏製程監控技術

研究生：余家慧　　　　　　　　　指導教授：陳志榮 教授

國立交通大學統計學研究所

摘 要

在此篇論文中，首先對在製程控制下的先驗分配提出一個由兩個成份組成的混合先驗有母數族。然後在可以得到製程控制下所產生的某些類別資料時，提出一個經驗貝氏的方法。接著提出一個例子來解釋此經驗貝氏模型。為了建構模型，我們討論此經驗貝氏模型之配適度和簡化。利用概似比的方法，提出貝氏和經驗貝氏製程監控技術來作為本篇論文的主要目的。最後藉由平均連串長度來研究此製程監控技術的表現。

關鍵字: 經驗貝氏; 製程監控; 類別資料; 混合先驗; beta-二項式; Dirichlet-多項式; 變換-常態-二項式; 變換-常態-多項式; 管制圖; 品質管制

A PROCESS MONITORING TECHNIQUE FOR CATEGORICAL DATA USING A TWO-COMPONENTS MIXTURE PRIOR PARAMETRIC FAMILY

Student : Chia-Hui Yu Advisor : Dr. Chih-Rung Chen

Institute of Statistics

National Chiao Tung University

Abstract

In the paper, first, a two-components mixture prior parametric family for the in-control prior distribution is proposed in a manufacturing process. Then an empirical Bayes approach is proposed when there are available in-control categorical data generated from the manufacturing process. As an illustration, an example of the proposed empirical Bayes model is introduced. For the purpose of model building, the goodness of fit and the simplification of the proposed model are discussed. Utilizing the likelihood ratio method, both Bayesian and empirical Bayes monitoring techniques are proposed as the main purpose of the paper. Finally, the performance of the proposed process monitoring scheme is studied in terms of the average run length to show the robustness of the methodology.

Key words: Empirical Bayes; Process monitoring; Categorical data; Mixture prior; Beta-binomial; Dirichlet-multinomial; Transformed-normal-binomial; Transformed-normal-multinomial; Control chart; Quality control

誌謝

在交大統計學研究所兩年的日子裡，很感謝陪著我一路走來的師長與同學。非常感謝所上老師們的教導，以及同學們在學業上所給予我的協助，使我在研究所的生活既豐富又充實‧很感謝我的指導教授陳志榮老師，指引我探索論文相關領域並且指導我完成論文‧也很感謝口試委員許文郁老師，洪志真老師，黃榮臣老師對於我的論文所提供寶貴的指導與建議。

生活上也很感激我的家人和朋友們在精神上的支持與鼓勵，使我能順利的完成學業。謹以此論文給我的家人、師長、同學和朋友，表達內心最誠摯的感激！

余家慧　　謹誌于
國立交通大學統計學研究所
中華民國九十五年八月

Contents

1 INTRODUCTION

In a manufacturing process, suppose that a product has k possible types of defects for some known positive integer k. For each tested product item, the result could be recorded as exactly one of the following $k+1$ disjoint categories: $\{$the first defect type, $\ldots$, the kth defect type, pass$\}$. Such data are called either binary for $k = 1$ or polytomous for $k \geq 2$. In the paper, categorical data denote either binary or polytomous data. See, e.g., McCullagh and Nelder (1989, Chapters 4 and 5) or Agresti (2002) for a review of the categorical data analysis.

In a Bayesian framework, the prior distribution of the unobserved random parameters is pre-specified explicitly, i.e., it does not depend on the observed data. However, it is usually a non-trivial task for practitioners to pre-specify an appropriate prior distribution of the random parameters. Thus, an empirical Bayes approach is commonly used instead.

In an empirical Bayes framework, there exist some unknown hyperparameters in the prior distribution of the unobserved random parameters. Then the marginal distribution of the observed data is utilized to estimate the hyperparameters. Finally, a Bayesian inference is made for the random parameters by treating the estimated prior distribution as the prior distribution. Since the estimated prior distribution does depend on the observed data, an empirical Bayes inference is not a Bayesian inference.

There are some research works utilizing the empirical Bayes model to monitor the categorical data generated in a manufacturing process. For example, Yousry *et al.* (1991) used the beta-binomial empirical Bayes model to monitor the binary data and utilized the method of moments for estimation of the hyperparameters. Recently, Shiau *et al.* (2005) used the Dirichlet-multinomial empirical Bayes model to monitor the polytomous data and utilized both the pseudo maximum likelihood method and the method of moments for estimation of the hyperparameters. Chen *et al.* (2004) used the beta-binomial/Dirichlet-multinomial empirical Bayes model to monitor the categorical data and utilized the maximum likelihood method for estimation of the hyperparameters. Similarly, Chen *et al.* (2005) used the transformed-normal-binomial/multinomial empirical Bayes model to monitor the categorical data and utilized the maximum likelihood method for estimation of the hyperparameters. Chen and Liu (2005) developed a model selection technique between two empirical Bayes models for the categorical data.

To proceed the discussion, we give a brief description on the Bayesian inference as follows: In a Bayesian framework, the prior distribution of the unobserved random parameter vector $\boldsymbol{\theta}$ has an explicitly pre-specified prior probability density function (p.d.f.) or probability mass function (p.m.f.) $\pi(\boldsymbol{\theta})$ and that the response vector $\mathbf{y}$ given $\boldsymbol{\theta}$ has a known conditional p.d.f. or p.m.f. $f(\mathbf{y}|\boldsymbol{\theta})$, where the function $\pi(\cdot)$ does not depend on $\mathbf{y}$. Then the Bayesian inference is based on the posterior p.d.f. or p.m.f. $p(\boldsymbol{\theta}|\mathbf{y})$ of $\boldsymbol{\theta}$ given $\mathbf{y}$, where

$$p(\boldsymbol{\theta}|\mathbf{y}) \quad \propto \quad f(\mathbf{y}|\boldsymbol{\theta})\,\pi(\boldsymbol{\theta}).$$

In the Bayesian terminology, $\pi(\boldsymbol{\theta})$, $f(\mathbf{y}|\boldsymbol{\theta})$, and $p(\boldsymbol{\theta}|\mathbf{y})$ are also called the prior likelihood, the likelihood, and the posterior likelihood of $\boldsymbol{\theta}$, respectively. In the literature, it is common practice to estimate $\boldsymbol{\theta}$ by the posterior mean $E(\boldsymbol{\theta}|\mathbf{y})$ of $\boldsymbol{\theta}$

given $\mathbf{y}$, where

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

with $P(\{\boldsymbol{\theta} \in \Theta\}) = 1$. An alternative estimator of $\boldsymbol{\theta}$ is the posterior mode $\text{mode}(\boldsymbol{\theta}|\mathbf{y})$ of $\boldsymbol{\theta}$ given $\mathbf{y}$, where

$$\text{mode}(\boldsymbol{\theta}|\mathbf{y}) \quad = \quad \arg\sup_{\boldsymbol{\theta}\in\Theta} p(\boldsymbol{\theta}|\mathbf{y}) \quad = \quad \arg\sup_{\boldsymbol{\theta}\in\Theta} f(\mathbf{y}|\boldsymbol{\theta})\, \pi(\boldsymbol{\theta}).$$

See, e.g., Gelman *et al.* (2004) for a review of the Bayesian data analysis.

Next, we give a brief description on the empirical Bayes inference as follows: In an empirical Bayes framework, the unobserved random parameter vector $\boldsymbol{\theta}$ has a prior p.d.f. or p.m.f. $\pi(\boldsymbol{\theta};\boldsymbol{\lambda})$ for some unknown hyperparameter vector $\boldsymbol{\lambda}$ and that the response vector $\mathbf{y}$ given $\boldsymbol{\theta}$ has a known conditional p.d.f. or p.m.f. $f(\mathbf{y}|\boldsymbol{\theta})$. An empirical Bayes inference is simply a Bayesian inference discussed above with $\pi(\boldsymbol{\theta})$ being replaced by $\pi(\boldsymbol{\theta};\boldsymbol{\lambda})|_{\boldsymbol{\lambda}=\hat{\boldsymbol{\lambda}}(\mathbf{y})}$ $(\equiv \pi(\boldsymbol{\theta};\hat{\boldsymbol{\lambda}}(\mathbf{y})))$, where $\hat{\boldsymbol{\lambda}}(\mathbf{y})$ is an estimator of $\boldsymbol{\lambda}$. Then an empirical Bayes inference is based on the estimated posterior p.d.f. or p.m.f. $p(\boldsymbol{\theta}|\mathbf{y};\boldsymbol{\lambda})|_{\boldsymbol{\lambda}=\hat{\boldsymbol{\lambda}}(\mathbf{y})}$ $(\equiv p(\boldsymbol{\theta}|\mathbf{y};\hat{\boldsymbol{\lambda}}(\mathbf{y})))$ of $\boldsymbol{\theta}$ given $\mathbf{y}$, where

$$p(\boldsymbol{\theta}|\mathbf{y};\boldsymbol{\lambda}) \quad \propto \quad f(\mathbf{y}|\boldsymbol{\theta})\, \pi(\boldsymbol{\theta};\boldsymbol{\lambda}).$$

In practice, either the maximum likelihood estimator or a method-of-moments estimator of $\boldsymbol{\lambda}$ is usually used as $\hat{\boldsymbol{\lambda}}(\mathbf{y})$ in an empirical Bayes inference. Similarly, it is common practice to estimate $\boldsymbol{\theta}$ by the estimated posterior mean $E(\boldsymbol{\theta}|\mathbf{y};\boldsymbol{\lambda})|_{\boldsymbol{\lambda}=\hat{\boldsymbol{\lambda}}(\mathbf{y})}$

($\equiv E(\boldsymbol{\theta}|\mathbf{y};\hat{\boldsymbol{\lambda}}(\mathbf{y}))$) of $\boldsymbol{\theta}$ given $\mathbf{y}$, where

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

with $P(\{\boldsymbol{\theta}\in\Theta\};\boldsymbol{\lambda}) = 1$. An alternative estimator of $\boldsymbol{\theta}$ is the estimated posterior mode $\text{mode}(\boldsymbol{\theta}|\mathbf{y};\boldsymbol{\lambda})|_{\boldsymbol{\lambda}=\hat{\boldsymbol{\lambda}}(\mathbf{y})}$ ($\equiv \text{mode}(\boldsymbol{\theta}|\mathbf{y};\hat{\boldsymbol{\lambda}}(\mathbf{y}))$) of $\boldsymbol{\theta}$ given $\mathbf{y}$, where

$$\text{mode}(\boldsymbol{\theta}|\mathbf{y};\boldsymbol{\lambda}) = \arg\sup_{\boldsymbol{\theta}\in\Theta} p(\boldsymbol{\theta}|\mathbf{y};\boldsymbol{\lambda}) = \arg\sup_{\boldsymbol{\theta}\in\Theta} f(\mathbf{y}|\boldsymbol{\theta})\,\pi(\boldsymbol{\theta};\boldsymbol{\lambda}).$$

See, e.g., Carlin and Louis (2000) for a review of the empirical Bayes data analysis.

The remaining parts of the paper is organized as follows. In Section 2, a two-components mixture prior parametric family for the in-control prior distribution is proposed in a manufacturing process. In Section 3, an empirical Bayes approach is proposed when there are available in-control categorical data generated from the manufacturing process. An example of the proposed empirical Bayes model is introduced in Section 4. The goodness of fit and the simplification of the proposed model are discussed in Sections 5 and 6, respectively. Utilizing the likelihood ratio method, both Bayesian and empirical Bayes monitoring techniques are proposed in Section 7. The performance of the proposed process monitoring scheme is studied in terms of the average run length in Section 8. Some concluding remarks are given in the final section.

2 A TWO-COMPONENTS MIXTURE PRIOR PARAMETRIC FAMILY

Assume that a product item is classified as one of the following $k+1$ disjoint categories: {the first defect type, ..., the kth defect type, pass}, where k is a known positive integer. Let t be any positive integer. For $i \in \{1, \ldots, k\}$, let θ_{it} denote the probability that a product item manufactured at time t has the ith defect type. Then $1 - \sum_{i=1}^{k} \theta_{it}$ ($\equiv \theta_{k+1,t}$) is the probability that a product item manufactured at time t passes the test. Set $\boldsymbol{\theta}_t \equiv (\theta_{1t}, \ldots, \theta_{kt})^T$ and $\Theta \equiv \{\boldsymbol{\theta}_t\colon \theta_{1t}, \ldots, \theta_{kt} > 0 \text{ and } \sum_{i=1}^{k} \theta_{it} < 1\}$. In the paper, $\boldsymbol{\theta}_t$ is called the (unobserved) random parameter vector at time t. Let $F_{\boldsymbol{\theta}_t}$ denote the prior cumulative distribution function (c.d.f.) of $\boldsymbol{\theta}_t$. For simplicity of notation, set $\mathcal{R}^m \equiv (-\infty, \infty)^m$ for any positive integer m.

Throughout the paper, the manufacturing process is said to be in control at time t if and only if $F_{\boldsymbol{\theta}_t} = F$, where F is an unknown in-control prior c.d.f. on Θ with p.d.f. $\pi(\boldsymbol{\theta}_t)$. In other words, the manufacturing process is said to be out of control at time t if and only if $F_{\boldsymbol{\theta}_t} \neq F$.

For $u \in \{1, 2\}$, let $\{F_{u,\boldsymbol{\lambda}_u}\colon \boldsymbol{\lambda}_u \in \Lambda_u\}$ denote the uth component prior parametric family, where $\boldsymbol{\lambda}_u$ is a $q_u \times 1$ hyperparameter vector for some known positive integer q_u, each $F_{u,\boldsymbol{\lambda}_u}$ is a known prior c.d.f. on Θ with p.d.f. $\pi_u(\boldsymbol{\theta}_t; \boldsymbol{\lambda}_u)$, and Λ_u is a known open subset of $\mathcal{R}^{q_u}$. Assume that $\partial^2 \pi_u(\boldsymbol{\theta}_t; \boldsymbol{\lambda}_u)/\partial \boldsymbol{\lambda}_u \partial \boldsymbol{\lambda}_u^T$ exists for each $\boldsymbol{\theta}_t \in \Theta$, $\boldsymbol{\lambda}_u \in \Lambda_u$, and $u \in \{1, 2\}$. Let $\{F_{\boldsymbol{\lambda}}\colon \boldsymbol{\lambda} \in \Lambda\}$ denote the two-components mixture prior parametric family, where $\boldsymbol{\lambda}$ ($\equiv (\omega, \boldsymbol{\lambda}_1^T, \boldsymbol{\lambda}_2^T)^T$) is a $(1 + q_1 + q_2) \times 1$ ($\equiv q \times 1$) hyperparameter vector, each $F_{\boldsymbol{\lambda}}$ is a known prior c.d.f. on Θ with p.d.f.

$$\pi(\boldsymbol{\theta}_t;\boldsymbol{\lambda}) \quad \equiv \quad \frac{\exp(\omega)}{1+\exp(\omega)}\cdot\pi_1(\boldsymbol{\theta}_t;\boldsymbol{\lambda}_1)+\frac{1}{1+\exp(\omega)}\cdot\pi_2(\boldsymbol{\theta}_t;\boldsymbol{\lambda}_2), \tag{1}$$

and $\Lambda \equiv [-\infty,\infty]\times\Lambda_1\times\Lambda_2$. Assume that the two-components mixture prior parametric family is identifiable, i.e., $F_{\boldsymbol{\lambda}^1} \neq F_{\boldsymbol{\lambda}^2}$ if $\boldsymbol{\lambda}^1 \neq \boldsymbol{\lambda}^2$ with $\boldsymbol{\lambda}^1,\boldsymbol{\lambda}^2 \in \Lambda$. When $\omega=\infty$, the two-components mixture prior parametric family is simplified to the first component prior parametric family with $\pi(\boldsymbol{\theta}_t;\boldsymbol{\lambda})=\pi_1(\boldsymbol{\theta}_t;\boldsymbol{\lambda}_1)$. When $\omega=-\infty$, the two-components mixture prior parametric family is simplified to the second component prior parametric family with $\pi(\boldsymbol{\theta}_t;\boldsymbol{\lambda})=\pi_2(\boldsymbol{\theta}_t;\boldsymbol{\lambda}_2)$. See, e.g., McLachlan and Peel (2000).

For any $\boldsymbol{\lambda}\in\Lambda$, the Kullback-Leibler divergence between the *in-control* prior c.d.f. F and the prior c.d.f. $F_{\boldsymbol{\lambda}}$ is defined as

$$d(F,F_{\boldsymbol{\lambda}}) \quad \equiv \quad \int_\Theta \log\left[\frac{\pi(\boldsymbol{\theta}_t)}{\pi(\boldsymbol{\theta}_t;\boldsymbol{\lambda})}\right] dF(\boldsymbol{\theta}_t) \quad \equiv \quad d(\boldsymbol{\lambda}). \tag{2}$$

By the Jensen inequality,

$$\begin{aligned} d(\boldsymbol{\lambda}) \quad &= \quad \int_\Theta -\log\left[\frac{\pi(\boldsymbol{\theta}_t;\boldsymbol{\lambda})}{\pi(\boldsymbol{\theta}_t)}\right] dF(\boldsymbol{\theta}_t) \quad \geq \quad -\log\left[\int_\Theta \frac{\pi(\boldsymbol{\theta}_t;\boldsymbol{\lambda})}{\pi(\boldsymbol{\theta}_t)}\cdot\pi(\boldsymbol{\theta}_t)\,d\boldsymbol{\theta}_t\right] \\ &= \quad -\log\left[\int_{\{\boldsymbol{\theta}_t:\pi(\boldsymbol{\theta}_t)>0\}} \pi(\boldsymbol{\theta}_t;\boldsymbol{\lambda})\,d\boldsymbol{\theta}_t\right] \quad \geq \quad -\log\left[\int_\Theta \pi(\boldsymbol{\theta}_t;\boldsymbol{\lambda})\,d\boldsymbol{\theta}_t\right] \quad = \quad 0 \end{aligned}$$

for $\boldsymbol{\lambda}\in\Lambda$, where $d(\boldsymbol{\lambda})=0$ if and only if $F_{\boldsymbol{\lambda}}=F$.

Assume that all of the following conditions hold: For $\boldsymbol{\lambda}\in(-\infty,\infty)\times\Lambda_1\times\Lambda_2$

($\equiv \Lambda^o$), $\partial^2 d(\boldsymbol{\lambda})/\partial\boldsymbol{\lambda}\partial\boldsymbol{\lambda}^T$ exists,

$$\frac{\partial d(\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}} = \int_\Theta \frac{\partial}{\partial \boldsymbol{\lambda}} \log\left[\frac{\pi(\boldsymbol{\theta}_t)}{\pi(\boldsymbol{\theta}_t;\boldsymbol{\lambda})}\right] dF(\boldsymbol{\theta}_t) \equiv -\bar{S}(\boldsymbol{\lambda}),$$

and

$$\frac{\partial^2 d(\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}\partial\boldsymbol{\lambda}^T} = \int_\Theta \frac{\partial^2}{\partial \boldsymbol{\lambda}\partial\boldsymbol{\lambda}^T} \log\left[\frac{\pi(\boldsymbol{\theta}_t)}{\pi(\boldsymbol{\theta}_t;\boldsymbol{\lambda})}\right] dF(\boldsymbol{\theta}_t) \equiv \bar{J}(\boldsymbol{\lambda}).$$

Assume that there exists a unique $\boldsymbol{\lambda}^0 \in \Lambda^o$ such that

$$\boldsymbol{\lambda}^0 = \arg\inf_{\boldsymbol{\lambda}\in\Lambda} d(\boldsymbol{\lambda}). \tag{3}$$

Then $\bar{S}(\boldsymbol{\lambda}^0) = 0_{q\times 1}$. Observe that, for $\boldsymbol{\lambda} \in \Lambda^o$,

$$\begin{aligned}\bar{S}(\boldsymbol{\lambda}) &= \int_\Theta \frac{\partial \pi(\boldsymbol{\theta}_t;\boldsymbol{\lambda})/\partial\boldsymbol{\lambda}}{\pi(\boldsymbol{\theta}_t;\boldsymbol{\lambda})} dF(\boldsymbol{\theta}_t) \equiv \int_\Theta \frac{\pi_{\boldsymbol{\lambda}}(\boldsymbol{\theta}_t;\boldsymbol{\lambda})}{\pi(\boldsymbol{\theta}_t;\boldsymbol{\lambda})} dF(\boldsymbol{\theta}_t) \\ &\equiv \int_\Theta S(\boldsymbol{\lambda};\boldsymbol{\theta}_t)\, dF(\boldsymbol{\theta}_t) \equiv E(S(\boldsymbol{\lambda};\boldsymbol{\theta}_t);F)\end{aligned} \tag{4}$$

and

$$\begin{aligned}\bar{J}(\boldsymbol{\lambda}) &= \int_\Theta -\frac{\partial S(\boldsymbol{\lambda};\boldsymbol{\theta}_t)}{\partial \boldsymbol{\lambda}^T} dF(\boldsymbol{\theta}_t) \\ &= \int_\Theta \left\{-\frac{\partial^2\pi(\boldsymbol{\theta}_t;\boldsymbol{\lambda})/\partial\boldsymbol{\lambda}\partial\boldsymbol{\lambda}^T}{\pi(\boldsymbol{\theta}_t;\boldsymbol{\lambda})} + \frac{\pi_{\boldsymbol{\lambda}}(\boldsymbol{\theta}_t;\boldsymbol{\lambda})\,\pi_{\boldsymbol{\lambda}}^T(\boldsymbol{\theta}_t;\boldsymbol{\lambda})}{[\pi(\boldsymbol{\theta}_t;\boldsymbol{\lambda})]^2}\right\} dF(\boldsymbol{\theta}_t) \\ &\equiv \int_\Theta \left\{-\frac{\pi_{\boldsymbol{\lambda}\boldsymbol{\lambda}^T}(\boldsymbol{\theta}_t;\boldsymbol{\lambda})}{\pi(\boldsymbol{\theta}_t;\boldsymbol{\lambda})} + \frac{\pi_{\boldsymbol{\lambda}}(\boldsymbol{\theta}_t;\boldsymbol{\lambda})\,\pi_{\boldsymbol{\lambda}}^T(\boldsymbol{\theta}_t;\boldsymbol{\lambda})}{[\pi(\boldsymbol{\theta}_t;\boldsymbol{\lambda})]^2}\right\} dF(\boldsymbol{\theta}_t) \\ &\equiv \int_\Theta J(\boldsymbol{\lambda};\boldsymbol{\theta}_t)\, dF(\boldsymbol{\theta}_t) \equiv E(J(\boldsymbol{\lambda};\boldsymbol{\theta}_t);F).\end{aligned} \tag{5}$$

For $\boldsymbol{\lambda} \in \Lambda^{\text{o}}$, set

$$\begin{aligned}\bar{K}(\boldsymbol{\lambda}) &\equiv \int_{\Theta} S(\boldsymbol{\lambda};\boldsymbol{\theta}_t)\, S^T(\boldsymbol{\lambda};\boldsymbol{\theta}_t)\, dF(\boldsymbol{\theta}_t) = \int_{\Theta} \frac{\pi_{\boldsymbol{\lambda}}(\boldsymbol{\theta}_t;\boldsymbol{\lambda})\,\pi^T_{\boldsymbol{\lambda}}(\boldsymbol{\theta}_t;\boldsymbol{\lambda})}{[\pi(\boldsymbol{\theta}_t;\boldsymbol{\lambda})]^2}\, dF(\boldsymbol{\theta}_t) \\ &\equiv \int_{\Theta} K(\boldsymbol{\lambda};\boldsymbol{\theta}_t)\, dF(\boldsymbol{\theta}_t) \equiv E(K(\boldsymbol{\lambda};\boldsymbol{\theta}_t);F). \end{aligned} \tag{6}$$

Then $\bar{K}(\boldsymbol{\lambda})$ is a positive definite covariance matrix for $\boldsymbol{\lambda} \in \Lambda^{\text{o}}$. For $\boldsymbol{\lambda}_u \in \Lambda_u$ and $u \in \{1,2\}$, set $\pi_{u,\boldsymbol{\lambda}_u}(\boldsymbol{\theta}_t;\boldsymbol{\lambda}_u) \equiv \partial\pi_u(\boldsymbol{\theta}_t;\boldsymbol{\lambda}_u)/\partial\boldsymbol{\lambda}_u$ and $\pi_{u,\boldsymbol{\lambda}_u\boldsymbol{\lambda}_u^T}(\boldsymbol{\theta}_t;\boldsymbol{\lambda}_u) \equiv \partial^2\pi_u(\boldsymbol{\theta}_t;\boldsymbol{\lambda}_u)/\partial\boldsymbol{\lambda}_u\partial\boldsymbol{\lambda}_u^T$.

Observe that, for $\boldsymbol{\lambda} \in \Lambda^{\text{o}}$,

$$\begin{aligned}
\frac{\partial\pi(\boldsymbol{\theta}_t;\boldsymbol{\lambda})}{\partial\omega} &= \frac{\exp(\omega)}{[1+\exp(\omega)]^2}\cdot[\pi_1(\boldsymbol{\theta}_t;\boldsymbol{\lambda}_1)-\pi_2(\boldsymbol{\theta}_t;\boldsymbol{\lambda}_2)] \equiv \pi_\omega(\boldsymbol{\theta}_t;\boldsymbol{\lambda}),\\
\frac{\partial\pi(\boldsymbol{\theta}_t;\boldsymbol{\lambda})}{\partial\boldsymbol{\lambda}_1} &= \frac{\exp(\omega)}{1+\exp(\omega)}\cdot\pi_{1,\boldsymbol{\lambda}_1}(\boldsymbol{\theta}_t;\boldsymbol{\lambda}_1) \equiv \pi_{\boldsymbol{\lambda}_1}(\boldsymbol{\theta}_t;\boldsymbol{\lambda}),\\
\frac{\partial\pi(\boldsymbol{\theta}_t;\boldsymbol{\lambda})}{\partial\boldsymbol{\lambda}_2} &= \frac{1}{1+\exp(\omega)}\cdot\pi_{2,\boldsymbol{\lambda}_2}(\boldsymbol{\theta}_t;\boldsymbol{\lambda}_2) \equiv \pi_{\boldsymbol{\lambda}_2}(\boldsymbol{\theta}_t;\boldsymbol{\lambda}),\\
\frac{\partial^2\pi(\boldsymbol{\theta}_t;\boldsymbol{\lambda})}{\partial\omega^2} &= \frac{\exp(\omega)\,[1-\exp(\omega)]}{[1+\exp(\omega)]^3}\cdot[\pi_1(\boldsymbol{\theta}_t;\boldsymbol{\lambda}_1)-\pi_2(\boldsymbol{\theta}_t;\boldsymbol{\lambda}_2)] \equiv \pi_{\omega\omega}(\boldsymbol{\theta}_t;\boldsymbol{\lambda}),\\
\frac{\partial^2\pi(\boldsymbol{\theta}_t;\boldsymbol{\lambda})}{\partial\boldsymbol{\lambda}_1\partial\boldsymbol{\lambda}_1^T} &= \frac{\exp(\omega)}{1+\exp(\omega)}\cdot\pi_{1,\boldsymbol{\lambda}_1\boldsymbol{\lambda}_1^T}(\boldsymbol{\theta}_t;\boldsymbol{\lambda}_1) \equiv \pi_{\boldsymbol{\lambda}_1\boldsymbol{\lambda}_1^T}(\boldsymbol{\theta}_t;\boldsymbol{\lambda}),\\
\frac{\partial^2\pi(\boldsymbol{\theta}_t;\boldsymbol{\lambda})}{\partial\boldsymbol{\lambda}_2\partial\boldsymbol{\lambda}_2^T} &= \frac{1}{1+\exp(\omega)}\cdot\pi_{2,\boldsymbol{\lambda}_2\boldsymbol{\lambda}_2^T}(\boldsymbol{\theta}_t;\boldsymbol{\lambda}_2) \equiv \pi_{\boldsymbol{\lambda}_2\boldsymbol{\lambda}_2^T}(\boldsymbol{\theta}_t;\boldsymbol{\lambda}),\\
\frac{\partial^2\pi(\boldsymbol{\theta}_t;\boldsymbol{\lambda})}{\partial\omega\partial\boldsymbol{\lambda}_1^T} &= \frac{\exp(\omega)}{[1+\exp(\omega)]^2}\cdot\pi^T_{1,\boldsymbol{\lambda}_1}(\boldsymbol{\theta}_t;\boldsymbol{\lambda}_1) \equiv \pi_{\omega\boldsymbol{\lambda}_1^T}(\boldsymbol{\theta}_t;\boldsymbol{\lambda}),\\
\frac{\partial^2\pi(\boldsymbol{\theta}_t;\boldsymbol{\lambda})}{\partial\omega\partial\boldsymbol{\lambda}_2^T} &= -\frac{\exp(\omega)}{[1+\exp(\omega)]^2}\cdot\pi^T_{2,\boldsymbol{\lambda}_2}(\boldsymbol{\theta}_t;\boldsymbol{\lambda}_2) \equiv \pi_{\omega\boldsymbol{\lambda}_2^T}(\boldsymbol{\theta}_t;\boldsymbol{\lambda}),
\end{aligned}$$

and

$$\frac{\partial^2\pi(\boldsymbol{\theta}_t;\boldsymbol{\lambda})}{\partial\boldsymbol{\lambda}_1\partial\boldsymbol{\lambda}_2^T} = 0_{q_1\times q_2} \equiv \pi_{\boldsymbol{\lambda}_1\boldsymbol{\lambda}_2^T}(\boldsymbol{\theta}_t;\boldsymbol{\lambda}),$$

where $0_{q_1\times q_2}$ denotes the $q_1\times q_2$ matrix with all components being 0.

One way to evaluate $\boldsymbol{\lambda}^0$ is to iterate the following procedure until $\boldsymbol{\lambda}^{(v)}$ converges to $\boldsymbol{\lambda}^0$: First choose a *good* initial value $\boldsymbol{\lambda}^{(0)} \in \Lambda^{\mathrm{o}}$ for $\boldsymbol{\lambda}^0$. Next, set

$$\boldsymbol{\lambda}^{*(v+1)} \equiv \boldsymbol{\lambda}^{(v)} + \bar{J}^{-1}\left(\boldsymbol{\lambda}^{(v)}\right) \bar{S}\left(\boldsymbol{\lambda}^{(v)}\right) \tag{7}$$

when $\boldsymbol{\lambda}^{(v)}$ is defined for $v \in \{0, 1, 2, \ldots\}$. If $\boldsymbol{\lambda}^{*(v+1)} \in \Lambda^{\mathrm{o}}$ and $d(\boldsymbol{\lambda}^{*(v+1)}) \leq d(\boldsymbol{\lambda}^{(v)})$, set $\boldsymbol{\lambda}^{(v+1)} \equiv \boldsymbol{\lambda}^{*(v+1)}$; otherwise, set

$$\boldsymbol{\lambda}^{*(u,v+1)} \equiv \boldsymbol{\lambda}^{(v)} + \frac{1}{2^u} \cdot \bar{K}^{-1}\left(\boldsymbol{\lambda}^{(v)}\right) \bar{S}\left(\boldsymbol{\lambda}^{(v)}\right) \tag{8}$$

for $u \in \{0, 1, 2, \ldots\}$ and set $\boldsymbol{\lambda}^{(v+1)} \equiv \boldsymbol{\lambda}^{*(m^*_{v+1}, v+1)}$, where $m^*_{v+1} \equiv \min\{u\colon\ u \in \{0, 1, 2, \ldots\},\ \boldsymbol{\lambda}^{*(u,v+1)} \in \Lambda^{\mathrm{o}},\ \boldsymbol{\lambda}^{*(u+1,v+1)} \in \Lambda^{\mathrm{o}},$ and $d(\boldsymbol{\lambda}^{*(u,v+1)}) < \min\{d(\boldsymbol{\lambda}^{(v)}), d(\boldsymbol{\lambda}^{*(u+1,v+1)})\}\}$.

Note that, by the Taylor series expansion, we obtain

$$\begin{aligned} d\left(\boldsymbol{\lambda}^{*(u,v+1)}\right) &= d\left(\boldsymbol{\lambda}^{(v)}\right) - \bar{S}^T\left(\boldsymbol{\lambda}^{(v)}\right)\left(\boldsymbol{\lambda}^{*(u,v+1)} - \boldsymbol{\lambda}^{(v)}\right) + \cdots \\ &= d\left(\boldsymbol{\lambda}^{(v)}\right) - \frac{1}{2^u} \cdot \bar{S}^T\left(\boldsymbol{\lambda}^{(v)}\right) \bar{K}^{-1}\left(\boldsymbol{\lambda}^{(v)}\right) \bar{S}\left(\boldsymbol{\lambda}^{(v)}\right) + O\left(\frac{1}{2^{2u}}\right) \end{aligned}$$

as $u \to \infty$ for any fixed non-negative integer v. Since $\bar{S}^T(\boldsymbol{\lambda}^{(v)})\bar{K}^{-1}(\boldsymbol{\lambda}^{(v)})\bar{S}(\boldsymbol{\lambda}^{(v)}) > 0$ for any fixed non-negative integer v, $d(\boldsymbol{\lambda}^{*(u,v+1)})$ is a strictly increasing function of u for large u with limit $d(\boldsymbol{\lambda}^{(v)})$, which implies that m^*_{v+1} is well-defined. Thus, $d(\boldsymbol{\lambda}^{(v)})$ is a decreasing function of v, i.e., $d(\boldsymbol{\lambda}^{(0)}) \geq d(\boldsymbol{\lambda}^{(1)}) \geq d(\boldsymbol{\lambda}^{(2)}) \geq \ldots$.

When any of $d(\boldsymbol{\lambda})$, $\bar{S}(\boldsymbol{\lambda})$, $\bar{J}(\boldsymbol{\lambda})$, and $\bar{K}(\boldsymbol{\lambda})$ does not have a closed-form formula, we may first simulate an independent and identically distributed (i.i.d.) sample $\{\boldsymbol{\theta}_t^{(1)}, \ldots, \boldsymbol{\theta}_t^{(R)}\}$ of size R, e.g., $R = 50\,000$, from the in-control prior c.d.f. F and then numerically evaluate $d(\boldsymbol{\lambda})$, $\bar{S}(\boldsymbol{\lambda})$, $\bar{J}(\boldsymbol{\lambda})$, and $\bar{K}(\boldsymbol{\lambda})$ by

$$\hat{d}(\boldsymbol{\lambda}) \equiv \frac{1}{R} \cdot \sum_{r=1}^{R} \log \left[\frac{\pi(\boldsymbol{\theta}_t)}{\pi(\boldsymbol{\theta}_t; \boldsymbol{\lambda})} \right] \Bigg|_{\boldsymbol{\theta}_t = \boldsymbol{\theta}_t^{(r)}}, \tag{9}$$

$$\hat{\bar{S}}(\boldsymbol{\lambda}) \equiv \frac{1}{R} \cdot \sum_{r=1}^{R} \frac{\pi_{\boldsymbol{\lambda}}(\boldsymbol{\theta}_t; \boldsymbol{\lambda})}{\pi(\boldsymbol{\theta}_t; \boldsymbol{\lambda})} \Bigg|_{\boldsymbol{\theta}_t = \boldsymbol{\theta}_t^{(r)}}, \tag{10}$$

$$\hat{\bar{J}}(\boldsymbol{\lambda}) \equiv \frac{1}{R} \cdot \sum_{r=1}^{R} \left\{ -\frac{\pi_{\boldsymbol{\lambda}\boldsymbol{\lambda}^T}(\boldsymbol{\theta}_t; \boldsymbol{\lambda})}{\pi(\boldsymbol{\theta}_t; \boldsymbol{\lambda})} + \frac{\pi_{\boldsymbol{\lambda}}(\boldsymbol{\theta}_t; \boldsymbol{\lambda})\, \pi_{\boldsymbol{\lambda}}^T(\boldsymbol{\theta}_t; \boldsymbol{\lambda})}{[\pi(\boldsymbol{\theta}_t; \boldsymbol{\lambda})]^2} \right\} \Bigg|_{\boldsymbol{\theta}_t = \boldsymbol{\theta}_t^{(r)}}, \tag{11}$$

and

$$\hat{\bar{K}}(\boldsymbol{\lambda}) \equiv \frac{1}{R} \cdot \sum_{r=1}^{R} \frac{\pi_{\boldsymbol{\lambda}}(\boldsymbol{\theta}_t; \boldsymbol{\lambda})\, \pi_{\boldsymbol{\lambda}}^T(\boldsymbol{\theta}_t; \boldsymbol{\lambda})}{[\pi(\boldsymbol{\theta}_t; \boldsymbol{\lambda})]^2} \Bigg|_{\boldsymbol{\theta}_t = \boldsymbol{\theta}_t^{(r)}}, \tag{12}$$

respectively.

3 AN EMPIRICAL BAYES APPROACH

Let t be any positive integer. Suppose that there are n_t tested product items manufactured at time t, where n_t is a known positive integer. For $i \in \{1, \ldots, k\}$, let y_{it} denote the number of the tested product items which have the ith defect type among the n_t tested product items manufactured at time t. Then $n_t - \sum_{i=1}^{k} y_{it}$ ($\equiv y_{k+1,t}$) is the number of the tested product items which pass the test among the n_t tested product items manufactured at time t. Set $\mathbf{y}_t \equiv (y_{1t}, \ldots, y_{kt})^T$ and $\mathcal{Y}_{n_t} \equiv \{\mathbf{y}_t\colon\ y_{1t}, \ldots, y_{kt} \in \{0, 1, \ldots, n_t\}$ and $\sum_{i=1}^{k} y_{it} \leq n_t\}$. In the paper, $\mathbf{y}_t$ is called the (observed) response vector at time t.

At each time t, assume that the response vector $\mathbf{y}_t$ given the random parameter vector $\boldsymbol{\theta}_t$ is distributed as either the conditional binomial$(n_t; \boldsymbol{\theta}_t)$ distribution for $k = 1$ or the conditional multinomial$(n_t; \boldsymbol{\theta}_t)$ distribution for $k \geq 2$, denoted by

$\mathbf{y}_t|\boldsymbol{\theta}_t \sim \text{binomial}(n_t; \boldsymbol{\theta}_t)$ for $k = 1$ or $\text{multinomial}(n_t; \boldsymbol{\theta}_t)$ for $k \geq 2$. Let $F_{\mathbf{y}_t|\boldsymbol{\theta}_t}$ denote the conditional c.d.f. of $\mathbf{y}_t$ given $\boldsymbol{\theta}_t$ with p.m.f.

$$f(\mathbf{y}_t|\boldsymbol{\theta}_t) = 1_{\mathcal{Y}_{n_t}}(\mathbf{y}_t) \cdot \frac{n_t!}{\prod_{i=1}^{k+1} y_{it}!} \cdot \prod_{i=1}^{k+1} \theta_{it}^{y_{it}}, \tag{13}$$

where $1_{\mathcal{Y}_{n_t}}(\mathbf{y}_t) = 1$ for $\mathbf{y}_t \in \mathcal{Y}_{n_t}$ and 0 otherwise. Let $F_{\mathbf{y}_t}$ denote the marginal c.d.f. of $\mathbf{y}_t$ with p.m.f.

$$f(\mathbf{y}_t; F_{\boldsymbol{\theta}_t}) = 1_{\mathcal{Y}_{n_t}}(\mathbf{y}_t) \cdot \frac{n_t!}{\prod_{i=1}^{k+1} y_{it}!} \cdot \int_{\Theta} \prod_{i=1}^{k+1} \theta_{it}^{y_{it}} \, dF_{\boldsymbol{\theta}_t}(\boldsymbol{\theta}_t). \tag{14}$$

For $\boldsymbol{\lambda}_u \in \Lambda_u$ and $u \in \{1, 2\}$, let $F_{\mathbf{y}_t;u,\boldsymbol{\lambda}_u}$ denote the marginal c.d.f. of $\mathbf{y}_t$ with p.m.f.

$$\begin{aligned} f_u(\mathbf{y}_t; \boldsymbol{\lambda}_u) &= 1_{\mathcal{Y}_{n_t}}(\mathbf{y}_t) \cdot \frac{n_t!}{\prod_{i=1}^{k+1} y_{it}!} \cdot \int_{\Theta} \prod_{i=1}^{k+1} \theta_{it}^{y_{it}} \, dF_{u,\boldsymbol{\lambda}_u}(\boldsymbol{\theta}_t) \\ &= 1_{\mathcal{Y}_{n_t}}(\mathbf{y}_t) \cdot \frac{n_t!}{\prod_{i=1}^{k+1} y_{it}!} \cdot \int_{\Theta} \left(\prod_{i=1}^{k+1} \theta_{it}^{y_{it}} \right) \cdot \pi_u(\boldsymbol{\theta}_t; \boldsymbol{\lambda}_u) \, d\boldsymbol{\theta}_t \end{aligned} \tag{15}$$

when $\boldsymbol{\theta}_t$ is distributed as the prior c.d.f. $F_{u,\boldsymbol{\lambda}_u}$, denoted by $\boldsymbol{\theta}_t \sim F_{u,\boldsymbol{\lambda}_u}$ and $\mathbf{y}_t \sim F_{\mathbf{y}_t;u,\boldsymbol{\lambda}_u}$. For $\boldsymbol{\lambda} \in \Lambda$, let $F_{\mathbf{y}_t;\boldsymbol{\lambda}}$ denote the marginal c.d.f. of $\mathbf{y}_t$ with p.m.f.

$$f(\mathbf{y}_t; \boldsymbol{\lambda}) = \frac{\exp(\omega)}{1 + \exp(\omega)} \cdot f_1(\mathbf{y}_t; \boldsymbol{\lambda}_1) + \frac{1}{1 + \exp(\omega)} \cdot f_2(\mathbf{y}_t; \boldsymbol{\lambda}_2) \tag{16}$$

when $\boldsymbol{\theta}_t$ is distributed as the prior c.d.f. $F_{\boldsymbol{\lambda}}$, denoted by $\boldsymbol{\theta}_t \sim F_{\boldsymbol{\lambda}}$ and $\mathbf{y}_t \sim F_{\mathbf{y}_t;\boldsymbol{\lambda}}$.

For $\boldsymbol{\lambda}_u \in \Lambda_u$ and $u \in \{1, 2\}$, assume that

$$\begin{aligned} \frac{\partial f_u(\mathbf{y}_t; \boldsymbol{\lambda}_u)}{\partial \boldsymbol{\lambda}_u} &= 1_{\mathcal{Y}_{n_t}}(\mathbf{y}_t) \cdot \frac{n_t!}{\prod_{i=1}^{k+1} y_{it}!} \cdot \int_\Theta \frac{\partial}{\partial \boldsymbol{\lambda}_u} \left[\left(\prod_{i=1}^{k+1} \theta_{it}^{y_{it}} \right) \cdot \pi_u(\boldsymbol{\theta}_t; \boldsymbol{\lambda}_u) \right] d\boldsymbol{\theta}_t \\ &\equiv f_{u,\boldsymbol{\lambda}_u}(\mathbf{y}_t; \boldsymbol{\lambda}_u) \end{aligned}$$

and

$$\begin{aligned} \frac{\partial^2 f_u(\mathbf{y}_t; \boldsymbol{\lambda}_u)}{\partial \boldsymbol{\lambda}_u \partial \boldsymbol{\lambda}_u^T} &= 1_{\mathcal{Y}_{n_t}}(\mathbf{y}_t) \cdot \frac{n_t!}{\prod_{i=1}^{k+1} y_{it}!} \cdot \int_\Theta \frac{\partial^2}{\partial \boldsymbol{\lambda}_u \partial \boldsymbol{\lambda}_u^T} \left[\left(\prod_{i=1}^{k+1} \theta_{it}^{y_{it}} \right) \cdot \pi_u(\boldsymbol{\theta}_t; \boldsymbol{\lambda}_u) \right] d\boldsymbol{\theta}_t \\ &\equiv f_{u,\boldsymbol{\lambda}_u \boldsymbol{\lambda}_u^T}(\mathbf{y}_t; \boldsymbol{\lambda}_u). \end{aligned}$$

Then, for $\boldsymbol{\lambda}_u \in \Lambda_u$ and $u \in \{1, 2\}$,

$$f_{u,\boldsymbol{\lambda}_u}(\mathbf{y}_t; \boldsymbol{\lambda}_u) = 1_{\mathcal{Y}_{n_t}}(\mathbf{y}_t) \cdot \frac{n_t!}{\prod_{i=1}^{k+1} y_{it}!} \cdot \int_\Theta \left(\prod_{i=1}^{k+1} \theta_{it}^{y_{it}} \right) \cdot \frac{\pi_{u,\boldsymbol{\lambda}_u}(\boldsymbol{\theta}_t; \boldsymbol{\lambda}_u)}{\pi_u(\boldsymbol{\theta}_t; \boldsymbol{\lambda}_u)} \, dF_{u,\boldsymbol{\lambda}_u}(\boldsymbol{\theta}_t) \quad (17)$$

and

$$f_{u,\boldsymbol{\lambda}_u \boldsymbol{\lambda}_u^T}(\mathbf{y}_t; \boldsymbol{\lambda}_u) = 1_{\mathcal{Y}_{n_t}}(\mathbf{y}_t) \cdot \frac{n_t!}{\prod_{i=1}^{k+1} y_{it}!} \cdot \int_\Theta \left(\prod_{i=1}^{k+1} \theta_{it}^{y_{it}} \right) \cdot \frac{\pi_{u,\boldsymbol{\lambda}_u \boldsymbol{\lambda}_u^T}(\boldsymbol{\theta}_t; \boldsymbol{\lambda}_u)}{\pi_u(\boldsymbol{\theta}_t; \boldsymbol{\lambda}_u)} \, dF_{u,\boldsymbol{\lambda}_u}(\boldsymbol{\theta}_t). \quad (18)$$

In the paper, it is assumed that the in-control prior c.d.f. $F = F_{\boldsymbol{\lambda}^0}$ for some unique $\boldsymbol{\lambda}^0 \in \Lambda$. Then $d(\boldsymbol{\lambda}^0) = 0$. Assume that there are available historical *in-control* response vectors $\{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_T\}$ generated in the manufacturing process for some known large positive integer T, where $(\boldsymbol{\theta}_1^T, \mathbf{y}_1^T)^T, (\boldsymbol{\theta}_2^T, \mathbf{y}_2^T)^T, \ldots, (\boldsymbol{\theta}_T^T, \mathbf{y}_T^T)^T$ are independent $2k \times 1$ random vectors. Set $\boldsymbol{\theta} \equiv (\boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T, \ldots, \boldsymbol{\theta}_T^T)^T$, $\mathbf{y} \equiv (\mathbf{y}_1^T, \mathbf{y}_2^T, \ldots, \mathbf{y}_T^T)^T$, and $\mathcal{Y} \equiv \mathcal{Y}_{n_1} \times \mathcal{Y}_{n_2} \times \cdots \times \mathcal{Y}_{n_T}$, where $\boldsymbol{\theta}$ and $\mathbf{y}$ are, respectively, called the historical in-control (unobserved) random vector and the historical in-control (observed) response vector in the paper. Let $F_{\mathbf{y};\boldsymbol{\lambda}^0}$ denote the marginal c.d.f. of $\mathbf{y}$ with p.m.f.

$$f\left(\mathbf{y};\boldsymbol{\lambda}^0\right) = \prod_{t=1}^{T} f\left(\mathbf{y}_t;\boldsymbol{\lambda}^0\right). \tag{19}$$

Given the historical in-control response vector $\mathbf{y}$, the log-likelihood function for $\boldsymbol{\lambda}$ is

$$\begin{aligned}
\ell(\boldsymbol{\lambda};\mathbf{y}) &\equiv \log[f(\mathbf{y};\boldsymbol{\lambda})] \equiv \log\left[\prod_{t=1}^{T} f(\mathbf{y}_t;\boldsymbol{\lambda})\right] \\
&= \sum_{t=1}^{T} \log\left[\frac{\exp(\omega)}{1+\exp(\omega)}\cdot f_1(\mathbf{y}_t;\boldsymbol{\lambda}_1) + \frac{1}{1+\exp(\omega)}\cdot f_2(\mathbf{y}_t;\boldsymbol{\lambda}_2)\right] \\
&\equiv \sum_{t=1}^{T} \log\left\{\frac{\exp(\omega)}{1+\exp(\omega)}\cdot \exp[\ell_1(\boldsymbol{\lambda}_1;\mathbf{y}_t)] + \frac{1}{1+\exp(\omega)}\cdot \exp[\ell_2(\boldsymbol{\lambda}_2;\mathbf{y}_t)]\right\} \\
&\equiv \sum_{t=1}^{T} \ell(\boldsymbol{\lambda};\mathbf{y}_t),
\end{aligned} \tag{20}$$

the score function for $\boldsymbol{\lambda}$ is

$$\begin{aligned}
S(\boldsymbol{\lambda};\mathbf{y}) &\equiv \frac{\partial \ell(\boldsymbol{\lambda};\mathbf{y})}{\partial \boldsymbol{\lambda}} = \sum_{t=1}^{T} \frac{\partial \ell(\boldsymbol{\lambda};\mathbf{y}_t)}{\partial \boldsymbol{\lambda}} = \sum_{t=1}^{T} \frac{\partial f(\mathbf{y}_t;\boldsymbol{\lambda})/\partial \boldsymbol{\lambda}}{f(\mathbf{y}_t;\boldsymbol{\lambda})} \\
&\equiv \sum_{t=1}^{T} \frac{f_{\boldsymbol{\lambda}}(\mathbf{y}_t;\boldsymbol{\lambda})}{f(\mathbf{y}_t;\boldsymbol{\lambda})} \equiv \sum_{t=1}^{T} S(\boldsymbol{\lambda};\mathbf{y}_t),
\end{aligned} \tag{21}$$

and the observed (Fisher) information for $\boldsymbol{\lambda}$ is

$$\begin{aligned}
J(\boldsymbol{\lambda};\mathbf{y}) &\equiv -\frac{\partial S(\boldsymbol{\lambda};\mathbf{y})}{\partial \boldsymbol{\lambda}^T} = \sum_{t=1}^{T} -\frac{\partial S(\boldsymbol{\lambda};\mathbf{y}_t)}{\partial \boldsymbol{\lambda}^T} \\
&= \sum_{t=1}^{T} \left\{ \frac{f_{\boldsymbol{\lambda}}(\mathbf{y}_t;\boldsymbol{\lambda})\, f_{\boldsymbol{\lambda}}^T(\mathbf{y}_t;\boldsymbol{\lambda})}{[f(\mathbf{y}_t;\boldsymbol{\lambda})]^2} - \frac{\partial^2 f(\mathbf{y}_t;\boldsymbol{\lambda})/\partial\boldsymbol{\lambda}\partial\boldsymbol{\lambda}^T}{f(\mathbf{y}_t;\boldsymbol{\lambda})} \right\} \\
&\equiv \sum_{t=1}^{T} \left\{ \frac{f_{\boldsymbol{\lambda}}(\mathbf{y}_t;\boldsymbol{\lambda})\, f_{\boldsymbol{\lambda}}^T(\mathbf{y}_t;\boldsymbol{\lambda})}{[f(\mathbf{y}_t;\boldsymbol{\lambda})]^2} - \frac{f_{\boldsymbol{\lambda}\boldsymbol{\lambda}^T}(\mathbf{y}_t;\boldsymbol{\lambda})}{f(\mathbf{y}_t;\boldsymbol{\lambda})} \right\} \\
&\equiv \sum_{t=1}^{T} J(\boldsymbol{\lambda};\mathbf{y}_t). \qquad (22)
\end{aligned}$$

For $\boldsymbol{\lambda} \in \Lambda^{\mathrm{o}}$, set

$$\begin{aligned}
K(\boldsymbol{\lambda};\mathbf{y}) &\equiv \sum_{t=1}^{T} S(\boldsymbol{\lambda};\mathbf{y}_t)\, S^T(\boldsymbol{\lambda};\mathbf{y}_t) = \sum_{t=1}^{T} \frac{f_{\boldsymbol{\lambda}}(\mathbf{y}_t;\boldsymbol{\lambda})\, f_{\boldsymbol{\lambda}^T}(\mathbf{y}_t;\boldsymbol{\lambda})}{[f(\mathbf{y}_t;\boldsymbol{\lambda})]^2} \\
&\equiv \sum_{t=1}^{T} K(\boldsymbol{\lambda};\mathbf{y}_t). \qquad (23)
\end{aligned}$$

Then $K(\boldsymbol{\lambda};\mathbf{y})$ is a non-negative definite covariance matrix for $\boldsymbol{\lambda} \in \Lambda^{\mathrm{o}}$ and $\mathbf{y} \in \mathcal{Y}$. For large T, $K(\boldsymbol{\lambda};\mathbf{y})$ is in general a positive definite covariance matrix for $\boldsymbol{\lambda} \in \Lambda^{\mathrm{o}}$ and $\mathbf{y} \in \mathcal{Y}$.

Observe that, for $\boldsymbol{\lambda} \in \Lambda^o$,

$$\begin{aligned}
\frac{\partial f(\mathbf{y}_t;\boldsymbol{\lambda})}{\partial \omega} &= \frac{\exp(\omega)}{[1+\exp(\omega)]^2} \cdot [f_1(\mathbf{y}_t;\boldsymbol{\lambda}_1) - f_2(\mathbf{y}_t;\boldsymbol{\lambda}_2)] \equiv f_\omega(\mathbf{y}_t;\boldsymbol{\lambda}), \\
\frac{\partial f(\mathbf{y}_t;\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}_1} &= \frac{\exp(\omega)}{1+\exp(\omega)} \cdot f_{1,\boldsymbol{\lambda}_1}(\mathbf{y}_t;\boldsymbol{\lambda}_1) \equiv f_{\boldsymbol{\lambda}_1}(\mathbf{y}_t;\boldsymbol{\lambda}), \\
\frac{\partial f(\mathbf{y}_t;\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}_2} &= \frac{1}{1+\exp(\omega)} \cdot f_{2,\boldsymbol{\lambda}_2}(\mathbf{y}_t;\boldsymbol{\lambda}_2) \equiv f_{\boldsymbol{\lambda}_2}(\mathbf{y}_t;\boldsymbol{\lambda}), \\
\frac{\partial^2 f(\mathbf{y}_t;\boldsymbol{\lambda})}{\partial \omega^2} &= \frac{\exp(\omega)\,[1-\exp(\omega)]}{[1+\exp(\omega)]^3} \cdot [f_1(\mathbf{y}_t;\boldsymbol{\lambda}_1) - f_2(\mathbf{y}_t;\boldsymbol{\lambda}_2)] \equiv f_{\omega\omega}(\mathbf{y}_t;\boldsymbol{\lambda}), \\
\frac{\partial^2 f(\mathbf{y}_t;\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}_1 \partial \boldsymbol{\lambda}_1^T} &= \frac{\exp(\omega)}{1+\exp(\omega)} \cdot f_{1,\boldsymbol{\lambda}_1\boldsymbol{\lambda}_1^T}(\mathbf{y}_t;\boldsymbol{\lambda}_1) \equiv f_{\boldsymbol{\lambda}_1\boldsymbol{\lambda}_1^T}(\mathbf{y}_t;\boldsymbol{\lambda}), \\
\frac{\partial^2 f(\mathbf{y}_t;\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}_2 \partial \boldsymbol{\lambda}_2^T} &= \frac{1}{1+\exp(\omega)} \cdot f_{2,\boldsymbol{\lambda}_2\boldsymbol{\lambda}_2^T}(\mathbf{y}_t;\boldsymbol{\lambda}_2) \equiv f_{\boldsymbol{\lambda}_2\boldsymbol{\lambda}_2^T}(\mathbf{y}_t;\boldsymbol{\lambda}), \\
\frac{\partial^2 f(\mathbf{y}_t;\boldsymbol{\lambda})}{\partial \omega \partial \boldsymbol{\lambda}_1^T} &= \frac{\exp(\omega)}{[1+\exp(\omega)]^2} \cdot f_{1,\boldsymbol{\lambda}_1}^T(\mathbf{y}_t;\boldsymbol{\lambda}_1) \equiv f_{\omega\boldsymbol{\lambda}_1^T}(\mathbf{y}_t;\boldsymbol{\lambda}), \\
\frac{\partial^2 f(\mathbf{y}_t;\boldsymbol{\lambda})}{\partial \omega \partial \boldsymbol{\lambda}_2^T} &= -\frac{\exp(\omega)}{[1+\exp(\omega)]^2} \cdot f_{2,\boldsymbol{\lambda}_2}^T(\mathbf{y}_t;\boldsymbol{\lambda}_2) \equiv f_{\omega\boldsymbol{\lambda}_2^T}(\mathbf{y}_t;\boldsymbol{\lambda}),
\end{aligned}$$

and

$$\frac{\partial^2 f(\mathbf{y}_t;\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}_1 \partial \boldsymbol{\lambda}_2^T} = 0_{q_1\times q_2} \equiv f_{\boldsymbol{\lambda}_1\boldsymbol{\lambda}_2^T}(\mathbf{y}_t;\boldsymbol{\lambda}).$$

The maximum likelihood estimator (MLE) $\hat{\boldsymbol{\lambda}}(\mathbf{y})$ $(\equiv \hat{\boldsymbol{\lambda}})$ of $\boldsymbol{\lambda}$ solves the score equation $S(\boldsymbol{\lambda};\mathbf{y}) = 0_{q\times 1}$ for $\boldsymbol{\lambda}$. That is, $S(\boldsymbol{\lambda};\mathbf{y})|_{\boldsymbol{\lambda}=\hat{\boldsymbol{\lambda}}}$ $(\equiv S(\hat{\boldsymbol{\lambda}};\mathbf{y})) = 0_{q\times 1}$.

One way to evaluate $\hat{\boldsymbol{\lambda}}$ is to iterate the following procedure until $\boldsymbol{\lambda}^{(v)}$ converges to $\hat{\boldsymbol{\lambda}}$: First choose a *good* initial value $\boldsymbol{\lambda}^{(0)} \in \Lambda^o$ for $\hat{\boldsymbol{\lambda}}$. Next, set

$$\boldsymbol{\lambda}^{*(v+1)} \equiv \boldsymbol{\lambda}^{(v)} + J^{-1}\left(\boldsymbol{\lambda}^{(v)};\mathbf{y}\right) S\left(\boldsymbol{\lambda}^{(v)};\mathbf{y}\right) \tag{24}$$

when $\boldsymbol{\lambda}^{(v)}$ is defined for $v \in \{0,1,2,\ldots\}$. If $\boldsymbol{\lambda}^{*(v+1)} \in \Lambda^o$ and $\ell(\boldsymbol{\lambda}^{*(v+1)};\mathbf{y}) \geq$

$\ell(\boldsymbol{\lambda}^{(v)};\mathbf{y})$, set $\boldsymbol{\lambda}^{(v+1)} \equiv \boldsymbol{\lambda}^{*(v+1)}$; otherwise, set

$$\boldsymbol{\lambda}^{*(u,v+1)} \equiv \boldsymbol{\lambda}^{(v)} + \frac{1}{2^u} \cdot K^{-1}\left(\boldsymbol{\lambda}^{(v)};\mathbf{y}\right) S\left(\boldsymbol{\lambda}^{(v)};\mathbf{y}\right) \tag{25}$$

for $u \in \{0,1,2,\ldots\}$ and set $\boldsymbol{\lambda}^{(v+1)} \equiv \boldsymbol{\lambda}^{*(m^*_{v+1},v+1)}$, where $m^*_{v+1} \equiv \min\{u\colon\ u \in \{0,1,2,\ldots\},\ \boldsymbol{\lambda}^{*(u,v+1)} \in \Lambda^{\mathrm{o}},\ \boldsymbol{\lambda}^{*(u+1,v+1)} \in \Lambda^{\mathrm{o}}$, and $\ell(\boldsymbol{\lambda}^{*(u,v+1)};\mathbf{y}) < \min\{\ell(\boldsymbol{\lambda}^{(v)};\mathbf{y}), \ell(\boldsymbol{\lambda}^{*(u+1,v+1)};\mathbf{y})\}\}$.

Note that, by the Taylor series expansion, we obtain

$$\begin{aligned}
& \ell\left(\boldsymbol{\lambda}^{*(u,v+1)};\mathbf{y}\right) \\
= {} & \ell\left(\boldsymbol{\lambda}^{(v)};\mathbf{y}\right) + S^T\left(\boldsymbol{\lambda}^{(v)};\mathbf{y}\right)\left(\boldsymbol{\lambda}^{*(u,v+1)} - \boldsymbol{\lambda}^{(v)}\right) + \cdots \\
= {} & \ell\left(\boldsymbol{\lambda}^{(v)};\mathbf{y}\right) + \frac{1}{2^u} \cdot S^T\left(\boldsymbol{\lambda}^{(v)};\mathbf{y}\right) K^{-1}\left(\boldsymbol{\lambda}^{(v)};\mathbf{y}\right) S\left(\boldsymbol{\lambda}^{(v)};\mathbf{y}\right) + O\left(\frac{1}{2^{2\,u}}\right)
\end{aligned}$$

as $u \to \infty$ for any non-negative integer v. Since $S^T(\boldsymbol{\lambda}^{(v)};\mathbf{y})K^{-1}(\boldsymbol{\lambda}^{(v)};\mathbf{y})S(\boldsymbol{\lambda}^{(v)};\mathbf{y}) > 0$ for any fixed non-negative integer v, $\ell(\boldsymbol{\lambda}^{*(u,v+1)};\mathbf{y})$ is a strictly decreasing function of u for large u with limit $\ell(\boldsymbol{\lambda}^{(v)};\mathbf{y})$, which implies that m^*_{v+1} is well-defined. Thus, $\ell(\boldsymbol{\lambda}^{(v)};\mathbf{y})$ is an increasing function of v, i.e., $\ell(\boldsymbol{\lambda}^{(0)};\mathbf{y}) \le \ell(\boldsymbol{\lambda}^{(1)};\mathbf{y}) \le \ell(\boldsymbol{\lambda}^{(2)};\mathbf{y}) \le \ldots$.

When any of $\ell(\boldsymbol{\lambda};\mathbf{y})$, $S(\boldsymbol{\lambda};\mathbf{y})$, $J(\boldsymbol{\lambda};\mathbf{y})$, and $K(\boldsymbol{\lambda};\mathbf{y})$ does not have a closed-form formula, we may numerically evaluate any of them as follows: First, for $u \in \{1,2\}$, simulate an i.i.d. sample $\{\boldsymbol{\theta}_1^{(u,1)},\ldots,\boldsymbol{\theta}_1^{(u,R)}\}$ of size R, e.g., $R = 50\,000$, from the prior c.d.f. $F_{u,\boldsymbol{\lambda}_u}$. Secondly, for $u \in \{1,2\}$, numerically evaluate $f_u(\mathbf{y}_t;\boldsymbol{\lambda}_u)$, $f_{u,\boldsymbol{\lambda}_u}(\mathbf{y}_t;\boldsymbol{\lambda}_u)$, and $f_{u,\boldsymbol{\lambda}_u\boldsymbol{\lambda}_u^T}(\mathbf{y}_t;\boldsymbol{\lambda}_u)$ by

$$\hat{f}_u(\mathbf{y}_t;\boldsymbol{\lambda}_u) \equiv 1_{y_{n_t}}(\mathbf{y}_t) \cdot \frac{n_t!}{\prod_{i=1}^{k+1} y_{it}!} \cdot \frac{1}{R} \cdot \sum_{r=1}^{R}\left(\prod_{i=1}^{k+1}\theta_{it}^{y_{it}}\right)\Bigg|_{\boldsymbol{\theta}_t=\boldsymbol{\theta}_1^{(u,r)}}, \tag{26}$$

$$
\begin{aligned}
&\hat{f}_{u,\boldsymbol{\lambda}_u}(\mathbf{y}_t;\boldsymbol{\lambda}_u) \\
\equiv\ & 1_{\mathcal{Y}_{n_t}}(\mathbf{y}_t)\cdot\frac{n_t!}{\prod_{i=1}^{k+1} y_{it}!}\cdot\frac{1}{R}\cdot\sum_{r=1}^{R}\left[\left(\prod_{i=1}^{k+1}\theta_{it}^{y_{it}}\right)\cdot\frac{\pi_{u,\boldsymbol{\lambda}_u}(\boldsymbol{\theta}_t;\boldsymbol{\lambda}_u)}{\pi_u(\boldsymbol{\theta}_t;\boldsymbol{\lambda}_u)}\right]\Bigg|_{\boldsymbol{\theta}_t=\boldsymbol{\theta}_1^{(u,r)}},
\end{aligned}
\tag{27}
$$

and

$$
\begin{aligned}
&\hat{f}_{u,\boldsymbol{\lambda}_u\boldsymbol{\lambda}_u^T}(\mathbf{y}_t;\boldsymbol{\lambda}_u) \\
\equiv\ & 1_{\mathcal{Y}_{n_t}}(\mathbf{y}_t)\cdot\frac{n_t!}{\prod_{i=1}^{k+1} y_{it}!}\cdot\frac{1}{R}\cdot\sum_{r=1}^{R}\left[\left(\prod_{i=1}^{k+1}\theta_{it}^{y_{it}}\right)\cdot\frac{\pi_{u,\boldsymbol{\lambda}_u\boldsymbol{\lambda}_u^T}(\boldsymbol{\theta}_t;\boldsymbol{\lambda}_u)}{\pi_u(\boldsymbol{\theta}_t;\boldsymbol{\lambda}_u)}\right]\Bigg|_{\boldsymbol{\theta}_t=\boldsymbol{\theta}_1^{(u,r)}},
\end{aligned}
\tag{28}
$$

respectively. As all of $f(\mathbf{y}_t;\boldsymbol{\lambda})$, $f_{\boldsymbol{\lambda}}(\mathbf{y}_t;\boldsymbol{\lambda})$, and $f_{\boldsymbol{\lambda}\boldsymbol{\lambda}^T}(\mathbf{y}_t;\boldsymbol{\lambda})$ have closed-form formulae of ω, $f_u(\mathbf{y}_t;\boldsymbol{\lambda}_u)$, $f_{u,\boldsymbol{\lambda}_u}(\mathbf{y}_t;\boldsymbol{\lambda}_u)$, and $f_{u,\boldsymbol{\lambda}_u\boldsymbol{\lambda}_u^T}(\mathbf{y}_t;\boldsymbol{\lambda}_u)$ for $u\in\{1,2\}$, thirdly, numerically evaluate $f(\mathbf{y}_t;\boldsymbol{\lambda})$, $f_{\boldsymbol{\lambda}}(\mathbf{y}_t;\boldsymbol{\lambda})$, and $f_{\boldsymbol{\lambda}\boldsymbol{\lambda}^T}(\mathbf{y}_t;\boldsymbol{\lambda})$ by $\hat{f}(\mathbf{y}_t;\boldsymbol{\lambda})$, $\hat{f}_{\boldsymbol{\lambda}}(\mathbf{y}_t;\boldsymbol{\lambda})$, and $\hat{f}_{\boldsymbol{\lambda}\boldsymbol{\lambda}^T}(\mathbf{y}_t;\boldsymbol{\lambda})$, respectively, utilizing their closed-form formulae. Finally, numerically evaluate $\ell(\boldsymbol{\lambda};\mathbf{y})$, $S(\boldsymbol{\lambda};\mathbf{y})$, $J(\boldsymbol{\lambda};\mathbf{y})$, and $K(\boldsymbol{\lambda};\mathbf{y})$ by

$$
\hat{\ell}(\boldsymbol{\lambda};\mathbf{y}) \equiv \sum_{t=1}^{T}\log\left[\hat{f}(\mathbf{y}_t;\boldsymbol{\lambda})\right], \tag{29}
$$

$$
\hat{S}(\boldsymbol{\lambda};\mathbf{y}) = \sum_{t=1}^{T}\frac{\hat{f}_{\boldsymbol{\lambda}}(\mathbf{y}_t;\boldsymbol{\lambda})}{\hat{f}(\mathbf{y}_t;\boldsymbol{\lambda})}, \tag{30}
$$

$$
\hat{J}(\boldsymbol{\lambda};\mathbf{y}) = \sum_{t=1}^{T}\left\{\frac{\hat{f}_{\boldsymbol{\lambda}}(\mathbf{y}_t;\boldsymbol{\lambda})\,\hat{f}_{\boldsymbol{\lambda}}^T(\mathbf{y}_t;\boldsymbol{\lambda})}{[\hat{f}(\mathbf{y}_t;\boldsymbol{\lambda})]^2}-\frac{\hat{f}_{\boldsymbol{\lambda}\boldsymbol{\lambda}^T}(\mathbf{y}_t;\boldsymbol{\lambda})}{\hat{f}(\mathbf{y}_t;\boldsymbol{\lambda})}\right\}, \tag{31}
$$

and

$$
\hat{K}(\boldsymbol{\lambda};\mathbf{y}) = \sum_{t=1}^{T}\frac{\hat{f}_{\boldsymbol{\lambda}}(\mathbf{y}_t;\boldsymbol{\lambda})\,\hat{f}_{\boldsymbol{\lambda}}^T(\mathbf{y}_t;\boldsymbol{\lambda})}{[\hat{f}(\mathbf{y}_t;\boldsymbol{\lambda})]^2}, \tag{32}
$$

respectively.

4 AN EXAMPLE

For illustration of the proposed methodology, the first component prior parametric family is chosen as the family of all beta/Dirichlet distributions because it is a conjugate family of binomial/multinomial distributions. The second component prior parametric family is chosen as the family of all transformed normal distributions (defined below) because it is a rich family of distributions, offering important distribution shapes that cannot be achieved within the family of all beta/Dirichlet distributions. See, e.g., O'Hagan and Forster (2004, Chapter 12).

4.1 The First Component Prior Parametric Family

Let the first component prior parametric family $\{F_{1,\boldsymbol{\lambda}_1}\colon\ \boldsymbol{\lambda}_1 \in \Lambda_1\}$ denote the family of all beta/Dirichlet distributions, where $\boldsymbol{\lambda}_1 \equiv (\lambda_{11}, \ldots \lambda_{1,k+1})^T$ $(\equiv (\lambda_{11}, \ldots \lambda_{1q_1})^T)$, $\Lambda_1 \equiv \mathcal{R}^{k+1}$, and $F_{1,\boldsymbol{\lambda}_1}$ has p.d.f.

$$
\begin{aligned}
\pi_1(\boldsymbol{\theta}_t; \boldsymbol{\lambda}_1) &= 1_\Theta(\boldsymbol{\theta}_t) \cdot \frac{\Gamma[\sum_{i=1}^{k+1} \exp(\lambda_{1i})]}{\prod_{i=1}^{k+1} \Gamma[\exp(\lambda_{1i})]} \cdot \prod_{i=1}^{k+1} \theta_{it}^{\exp(\lambda_{1i})-1} \\
&\equiv 1_\Theta(\boldsymbol{\theta}_t) \cdot \frac{\Gamma[\exp(\lambda_{10})]}{\prod_{i=1}^{k+1} \Gamma[\exp(\lambda_{1i})]} \cdot \prod_{i=1}^{k+1} \theta_{it}^{\exp(\lambda_{1i})-1} \qquad (33)
\end{aligned}
$$

with $1_\Theta(\boldsymbol{\theta}_t) = 1$ for $\boldsymbol{\theta}_t \in \Theta$ and 0 otherwise. Since $\{F_{1,\boldsymbol{\lambda}_1}\colon\ \boldsymbol{\lambda}_1 \in \Lambda_1\}$ is chosen as a conjugate family of binomial/multinomial distributions, all of $f_1(\mathbf{y}_t; \boldsymbol{\lambda}_1)$, $f_{1,\boldsymbol{\lambda}_1}(\mathbf{y}_t; \boldsymbol{\lambda}_1)$, and $f_{1,\boldsymbol{\lambda}_1\boldsymbol{\lambda}_1^T}(\mathbf{y}_t; \boldsymbol{\lambda}_1)$ have closed-form formulae for $\boldsymbol{\lambda}_1 \in \Lambda_1$ as follows:

For $\boldsymbol{\lambda}_1 \in \Lambda_1$, it follows from Johnson *et al.* (1997, pages 80 and 81) that

$$
\begin{aligned}
& f_1(\mathbf{y}_t; \boldsymbol{\lambda}_1) \\
= \; & 1_{\mathcal{Y}_{n_t}}(\mathbf{y}_t) \cdot \exp\left\{ \sum_{j=0}^{n_t-1} \log\left[\frac{j+1}{\exp(\lambda_{10})+j}\right] - \sum_{i=1}^{k+1} \sum_{j=0}^{y_{it}-1} \log\left[\frac{j+1}{\exp(\lambda_{1i})+j}\right] \right\}.
\end{aligned} \tag{34}
$$

For $\boldsymbol{\lambda}_1 \in \Lambda_1$, set $\exp(\boldsymbol{\lambda}_1) \equiv (\exp(\lambda_{11}), \ldots, \exp(\lambda_{1,k+1}))^T$. Then, for $\boldsymbol{\lambda}_1 \in \Lambda_1$ and $\mathbf{y}_t \in \mathcal{Y}_{n_t}$, we have

$$
\begin{aligned}
S_1(\boldsymbol{\lambda}_1; \mathbf{y}_t) \; \equiv \; \frac{\partial}{\partial \boldsymbol{\lambda}_1} \ell_1(\boldsymbol{\lambda}_1; \mathbf{y}_t) \; &= \; \frac{f_{1,\boldsymbol{\lambda}_1}(\mathbf{y}_t; \boldsymbol{\lambda}_1)}{f_1(\mathbf{y}_t; \boldsymbol{\lambda}_1)} \\
&= \left(\sum_{j=0}^{y_{1t}-1} \frac{\exp(\lambda_{11})}{\exp(\lambda_{11})+j}, \ldots, \sum_{j=0}^{y_{k+1,t}-1} \frac{\exp(\lambda_{1,k+1})}{\exp(\lambda_{1,k+1})+j} \right)^T \\
&\quad - \left[\sum_{j=0}^{n_t-1} \frac{1}{\exp(\lambda_{10})+j} \right] \cdot \exp(\boldsymbol{\lambda}_1)
\end{aligned}
$$

and

$$
\begin{aligned}
J_1(\boldsymbol{\lambda}_1; \mathbf{y}_t) \; \equiv \; -\frac{\partial}{\partial \boldsymbol{\lambda}_1^T} S_1(\boldsymbol{\lambda}_1; \mathbf{y}_t) \; &= \; S_1(\boldsymbol{\lambda}_1; \mathbf{y}_t)\, S_1^T(\boldsymbol{\lambda}_1; \mathbf{y}_t) - \frac{f_{1,\boldsymbol{\lambda}_1\boldsymbol{\lambda}_1^T}(\mathbf{y}_t; \boldsymbol{\lambda}_1)}{f_1(\mathbf{y}_t; \boldsymbol{\lambda}_1)} \\
&= \operatorname{diag}\left\{ \sum_{j=0}^{y_{1t}-1} \frac{j \cdot \exp(\lambda_{11})}{[\exp(\lambda_{11})+j]^2}, \ldots, \sum_{j=0}^{y_{k+1,t}-1} \frac{j \cdot \exp(\lambda_{1,k+1})}{[\exp(\lambda_{1,k+1})+j]^2} \right\} \\
&\quad + \left[\sum_{j=0}^{n_t-1} \frac{1}{\exp(\lambda_{10})+j} \right] \cdot \operatorname{diag}\{\exp(\lambda_{11}), \ldots, \exp(\lambda_{1,k+1})\} \\
&\quad - \left\{ \sum_{j=0}^{n_t-1} \frac{1}{[\exp(\lambda_{10})+j]^2} \right\} \cdot \exp(\boldsymbol{\lambda}_1)\, \exp(\boldsymbol{\lambda}_1)^T.
\end{aligned}
$$

Thus, for $\mathbf{y}_t \in \mathcal{Y}_{n_t}$ and $\boldsymbol{\lambda}_1 \in \Lambda_1$,

$$f_{1,\boldsymbol{\lambda}_1}(\mathbf{y}_t; \boldsymbol{\lambda}_1) \quad = \quad f_1(\mathbf{y}_t; \boldsymbol{\lambda}_1) \cdot S_1(\boldsymbol{\lambda}_1; \mathbf{y}_t) \tag{35}$$

and

$$f_{1,\boldsymbol{\lambda}_1\boldsymbol{\lambda}_1^T}(\mathbf{y}_t; \boldsymbol{\lambda}_1) \quad = \quad f_1(\mathbf{y}_t; \boldsymbol{\lambda}_1) \cdot \left[S_1(\boldsymbol{\lambda}_1; \mathbf{y}_t)\, S_1^T(\boldsymbol{\lambda}_1; \mathbf{y}_t) - J_1(\boldsymbol{\lambda}_1; \mathbf{y}_t) \right]. \tag{36}$$

4.2 The Second Component Prior Parametric Family

Let the second component prior parametric family $\{F_{2,\boldsymbol{\lambda}_2}\colon\ \boldsymbol{\lambda}_2 \in \Lambda_2\}$ denote the family of all transformed normal distributions defined as follows: Set $(\log(\theta_{1t}/\theta_{k+1,t}), \ldots, \log(\theta_{kt}/\theta_{k+1,t}))^T \equiv \boldsymbol{\eta}_t\ (\equiv (\eta_{1t}, \ldots, \eta_{kt})^T)$. Then, for $i \in \{1, \ldots, k\}$, $\theta_{it} = \exp(\eta_{it})/[1 + \sum_{i'=1}^{k} \exp(\eta_{i't})]$. Let $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denote the k-variate normal distribution with mean vector $\boldsymbol{\mu}\ (\equiv (\mu_1, \ldots, \mu_k)^T) \in \mathcal{R}^k$ and $k \times k$ positive definite covariance matrix $\boldsymbol{\Sigma}\ (\equiv (\Sigma_{ii'}))$. Set $\boldsymbol{\Sigma}^{-1} \equiv (\Sigma^{ii'})$ and $\mathbf{R} \equiv (\Sigma^{ii'}/\sqrt{\Sigma^{ii}\,\Sigma^{i'i'}})$ $(\equiv (\rho_{ii'}))$. Then

$$\boldsymbol{\Sigma}^{-1} \quad = \quad \operatorname{diag}\left\{\sqrt{\Sigma^{11}}, \ldots, \sqrt{\Sigma^{kk}}\right\} \mathbf{R} \operatorname{diag}\left\{\sqrt{\Sigma^{11}}, \ldots, \sqrt{\Sigma^{kk}}\right\}.$$

Set

$$\begin{aligned}\boldsymbol{\lambda}_2 \quad &\equiv \quad \left(\boldsymbol{\mu}^T, \log\left(\Sigma^{11}\right), \ldots, \log\left(\Sigma^{kk}\right), \log\left(\frac{1+\rho_{12}}{1-\rho_{12}}\right), \ldots, \log\left(\frac{1+\rho_{1k}}{1-\rho_{1k}}\right),\right.\\ &\qquad \left.\ldots, \log\left(\frac{1+\rho_{k-1,k}}{1-\rho_{k-1,k}}\right)\right)^T \\ &\equiv \quad (\lambda_{21}, \ldots, \lambda_{2,k(k+3)/2})^T \quad \equiv \quad (\lambda_{21}, \ldots, \lambda_{2q_2})^T \quad \in \quad \Lambda_2,\end{aligned}$$

where $\Lambda_2 \equiv \{\boldsymbol{\lambda}_2\colon\ \boldsymbol{\mu} \in \mathcal{R}^k$ and $\boldsymbol{\Sigma}$ is a $k \times k$ positive definite covariance matrix$\}$. Then Λ_2 is an open subset of $\mathcal{R}^{q_2}$. For $\boldsymbol{\lambda}_2 \in \Lambda_2$, let $\phi(\cdot;\boldsymbol{\lambda}_2)$ and $\Phi_{\boldsymbol{\lambda}_2}$ denote, respectively, the p.d.f. and the c.d.f. of the $N(\boldsymbol{\mu},\boldsymbol{\Sigma})$ distribution. For $\boldsymbol{\lambda}_2 \in \Lambda_2$, set $\partial\phi(\cdot;\boldsymbol{\lambda}_2)/\partial\boldsymbol{\lambda}_2 \equiv \phi_{\boldsymbol{\lambda}_2}(\cdot;\boldsymbol{\lambda}_2)$ and $\partial^2\phi(\cdot;\boldsymbol{\lambda}_2)/\partial\boldsymbol{\lambda}_2\partial\boldsymbol{\lambda}_2^T \equiv \phi_{\boldsymbol{\lambda}_2\boldsymbol{\lambda}_2^T}(\cdot;\boldsymbol{\lambda}_2)$. When $\boldsymbol{\eta}_t \sim N(\boldsymbol{\mu},\boldsymbol{\Sigma})$, we say that $\boldsymbol{\theta}_t$ is distributed as the transformed-normal$(\boldsymbol{\mu},\boldsymbol{\Sigma})$ distribution, denoted by $\boldsymbol{\theta}_t \sim F_{2,\boldsymbol{\lambda}_2}$, with p.d.f.

$$\begin{aligned}\pi_2(\boldsymbol{\theta}_t;\boldsymbol{\lambda}_2) &= \phi(\boldsymbol{\eta}_t;\boldsymbol{\lambda}_2)\cdot\left|\det\left(\frac{\partial\boldsymbol{\eta}_t}{\partial\boldsymbol{\theta}_t^T}\right)\right| \\ &= 1_\Theta(\boldsymbol{\theta}_t)\cdot\frac{1}{(2\pi)^{k/2}\,|\boldsymbol{\Sigma}|^{1/2}\prod_{i=1}^{k+1}\theta_{it}}\cdot\exp\left[-\frac{1}{2}\cdot(\boldsymbol{\eta}_t-\boldsymbol{\mu})^T\,\boldsymbol{\Sigma}^{-1}\,(\boldsymbol{\eta}_t-\boldsymbol{\mu})\right], \qquad (37)\end{aligned}$$

where

$$\frac{\partial\boldsymbol{\eta}_t}{\partial\boldsymbol{\theta}_t^T} = \operatorname{diag}\left\{\frac{1}{\theta_{1t}},\ldots,\frac{1}{\theta_{kt}}\right\}+\frac{1}{\theta_{k+1,t}}\cdot 1_{k\times 1}\,1_{k\times 1}^T$$

with $1_{k\times 1}$ denoting the $k\times 1$ vector $(1,\ldots,1)^T$.

For $\boldsymbol{\lambda}_2 \in \Lambda_2$, we obtain

$$\begin{aligned}f_2(\mathbf{y}_t;\boldsymbol{\lambda}_2) &= 1_{\mathcal{Y}_{n_t}}(\mathbf{y}_t)\cdot\frac{n_t!}{\prod_{i=1}^{k+1}y_{it}!}\cdot\int_\Theta\prod_{i=1}^{k+1}\theta_{it}^{y_{it}}\,dF_{2,\boldsymbol{\lambda}_2}(\boldsymbol{\theta}_t) \\ &= 1_{\mathcal{Y}_{n_t}}(\mathbf{y}_t)\cdot\frac{n_t!}{\prod_{i=1}^{k+1}y_{it}!}\cdot\int_{\mathcal{R}^k}\frac{\exp(\mathbf{y}_t^T\,\boldsymbol{\eta}_t)}{[1+\sum_{i=1}^k\exp(\eta_{it})]^{n_t}}\,d\Phi_{\boldsymbol{\lambda}_2}(\boldsymbol{\eta}_t), \qquad (38)\end{aligned}$$

$$\begin{aligned}&f_{2,\boldsymbol{\lambda}_2}(\mathbf{y}_t;\boldsymbol{\lambda}_2) \\ &= 1_{\mathcal{Y}_{n_t}}(\mathbf{y}_t)\cdot\frac{n_t!}{\prod_{i=1}^{k+1}y_{it}!}\cdot\int_\Theta\left(\prod_{i=1}^{k+1}\theta_{it}^{y_{it}}\right)\cdot\frac{\pi_{2,\boldsymbol{\lambda}_2}(\boldsymbol{\theta}_t;\boldsymbol{\lambda}_2)}{\pi_2(\boldsymbol{\theta}_t;\boldsymbol{\lambda}_2)}\,dF_{2,\boldsymbol{\lambda}_2}(\boldsymbol{\theta}_t) \\ &= 1_{\mathcal{Y}_{n_t}}(\mathbf{y}_t)\cdot\frac{n_t!}{\prod_{i=1}^{k+1}y_{it}!}\cdot\int_{\mathcal{R}^k}\frac{\exp(\mathbf{y}_t^T\,\boldsymbol{\eta}_t)}{[1+\sum_{i=1}^k\exp(\eta_{it})]^{n_t}}\cdot\frac{\phi_{\boldsymbol{\lambda}_2}(\boldsymbol{\eta}_t;\boldsymbol{\lambda}_2)}{\phi(\boldsymbol{\eta}_t;\boldsymbol{\lambda}_2)}\,d\Phi_{\boldsymbol{\lambda}_2}(\boldsymbol{\eta}_t), \qquad (39)\end{aligned}$$

and

$$
\begin{aligned}
& f_{2,\boldsymbol{\lambda}_2\boldsymbol{\lambda}_2^T}(\mathbf{y}_t;\boldsymbol{\lambda}_2) \\
= \; & 1_{\mathcal{Y}_{n_t}}(\mathbf{y}_t)\cdot\frac{n_t!}{\prod_{i=1}^{k+1} y_{it}!}\cdot\int_{\Theta}\left(\prod_{i=1}^{k+1}\theta_{it}^{y_{it}}\right)\cdot\frac{\pi_{2,\boldsymbol{\lambda}_2\boldsymbol{\lambda}_2^T}(\boldsymbol{\theta}_t;\boldsymbol{\lambda}_2)}{\pi_2(\boldsymbol{\theta}_t;\boldsymbol{\lambda}_2)}\,dF_{2,\boldsymbol{\lambda}_2}(\boldsymbol{\theta}_t) \\
= \; & 1_{\mathcal{Y}_{n_t}}(\mathbf{y}_t)\cdot\frac{n_t!}{\prod_{i=1}^{k+1} y_{it}!}\cdot\int_{\mathcal{R}^k}\frac{\exp(\mathbf{y}_t^T\boldsymbol{\eta}_t)}{[1+\sum_{i=1}^k\exp(\eta_{it})]^{n_t}}\cdot\frac{\phi_{\boldsymbol{\lambda}_2\boldsymbol{\lambda}_2^T}(\boldsymbol{\eta}_t;\boldsymbol{\lambda}_2)}{\phi(\boldsymbol{\eta}_t;\boldsymbol{\lambda}_2)}\,d\Phi_{\boldsymbol{\lambda}_2}(\boldsymbol{\eta}_t).
\end{aligned}
\tag{40}
$$

Note that, for $\boldsymbol{\theta}_t\in\Theta$ and $\boldsymbol{\lambda}_2\in\Lambda_2$,

$$
\frac{\pi_{2,\boldsymbol{\lambda}_2}(\boldsymbol{\theta}_t;\boldsymbol{\lambda}_2)}{\pi_2(\boldsymbol{\theta}_t;\boldsymbol{\lambda}_2)} = \frac{\phi_{\boldsymbol{\lambda}_2}(\boldsymbol{\eta}_t;\boldsymbol{\lambda}_2)}{\phi(\boldsymbol{\eta}_t;\boldsymbol{\lambda}_2)} = -\frac{1}{2}\cdot\frac{\partial}{\partial\boldsymbol{\lambda}_2}(\boldsymbol{\eta}_t-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{\eta}_t-\boldsymbol{\mu}) \tag{41}
$$

and

$$
\begin{aligned}
\frac{\pi_{2,\boldsymbol{\lambda}_2\boldsymbol{\lambda}_2^T}(\boldsymbol{\theta}_t;\boldsymbol{\lambda}_2)}{\pi_2(\boldsymbol{\theta}_t;\boldsymbol{\lambda}_2)} &= \frac{\phi_{\boldsymbol{\lambda}_2\boldsymbol{\lambda}_2^T}(\boldsymbol{\eta}_t;\boldsymbol{\lambda}_2)}{\phi(\boldsymbol{\eta}_t;\boldsymbol{\lambda}_2)} \\
&= -\frac{1}{2}\cdot\frac{\partial^2}{\partial\boldsymbol{\lambda}_2\partial\boldsymbol{\lambda}_2^T}(\boldsymbol{\eta}_t-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{\eta}_t-\boldsymbol{\mu})+\frac{\phi_{\boldsymbol{\lambda}_2}(\boldsymbol{\eta}_t;\boldsymbol{\lambda}_2)\,\phi_{\boldsymbol{\lambda}_2}^T(\boldsymbol{\eta}_t;\boldsymbol{\lambda}_2)}{[\phi(\boldsymbol{\eta}_t;\boldsymbol{\lambda}_2)]^2}.
\end{aligned}
\tag{42}
$$

Since none of $f_2(\mathbf{y}_t;\boldsymbol{\lambda}_2)$, $f_{2,\boldsymbol{\lambda}_2}(\mathbf{y}_t;\boldsymbol{\lambda}_2)$, and $f_{2,\boldsymbol{\lambda}_2\boldsymbol{\lambda}_2^T}(\mathbf{y}_t;\boldsymbol{\lambda}_2)$ has a closed-form formula, we may numerically evaluate all of them as follows: First simulate an i.i.d. sample $\{\boldsymbol{\theta}_1^{(2,1)},\ldots,\boldsymbol{\theta}_1^{(2,R)}\}$ of size R, e.g., $R=50\,000$, from the prior c.d.f. $F_{2,\boldsymbol{\lambda}_2}$ and then numerically evaluate $f_2(\mathbf{y}_t;\boldsymbol{\lambda}_2)$, $f_{2,\boldsymbol{\lambda}_2}(\mathbf{y}_t;\boldsymbol{\lambda}_2)$, and $f_{2,\boldsymbol{\lambda}_2\boldsymbol{\lambda}_2^T}(\mathbf{y}_t;\boldsymbol{\lambda}_2)$ by

$$
\hat{f}_2(\mathbf{y}_t;\boldsymbol{\lambda}_2) \equiv 1_{\mathcal{Y}_{n_t}}(\mathbf{y}_t)\cdot\frac{n_t!}{\prod_{i=1}^{k+1} y_{it}!}\cdot\frac{1}{R}\cdot\sum_{r=1}^{R}\left(\prod_{i=1}^{k+1}\theta_{it}^{y_{it}}\right)\Bigg|_{\boldsymbol{\theta}_t=\boldsymbol{\theta}_1^{(2,r)}}, \tag{43}
$$

$$\hat{f}_{2,\boldsymbol{\lambda}_2}(\mathbf{y}_t;\boldsymbol{\lambda}_2) \equiv 1_{\mathcal{Y}_{n_t}}(\mathbf{y}_t)\cdot\frac{n_t!}{\prod_{i=1}^{k+1} y_{it}!}\cdot\frac{1}{R}\cdot\sum_{r=1}^{R}\left[\left(\prod_{i=1}^{k+1}\theta_{it}^{y_{it}}\right)\cdot\frac{\pi_{2,\boldsymbol{\lambda}_2}(\boldsymbol{\theta}_t;\boldsymbol{\lambda}_2)}{\pi_2(\boldsymbol{\theta}_t;\boldsymbol{\lambda}_2)}\right]\Bigg|_{\boldsymbol{\theta}_t=\boldsymbol{\theta}_1^{(2,r)}}, \quad (44)$$

and

$$\hat{f}_{2,\boldsymbol{\lambda}_2\boldsymbol{\lambda}_2^T}(\mathbf{y}_t;\boldsymbol{\lambda}_2) \equiv 1_{\mathcal{Y}_{n_t}}(\mathbf{y}_t)\cdot\frac{n_t!}{\prod_{i=1}^{k+1} y_{it}!}\cdot\frac{1}{R}\cdot\sum_{r=1}^{R}\left[\left(\prod_{i=1}^{k+1}\theta_{it}^{y_{it}}\right)\cdot\frac{\pi_{2,\boldsymbol{\lambda}_2\boldsymbol{\lambda}_2^T}(\boldsymbol{\theta}_t;\boldsymbol{\lambda}_2)}{\pi_2(\boldsymbol{\theta}_t;\boldsymbol{\lambda}_2)}\right]\Bigg|_{\boldsymbol{\theta}_t=\boldsymbol{\theta}_1^{(2,r)}}, \quad (45)$$

respectively.

An alternative way to numerically evaluate all of $f_2(\mathbf{y}_t;\boldsymbol{\lambda}_2)$, $f_{2,\boldsymbol{\lambda}_2}(\mathbf{y}_t;\boldsymbol{\lambda}_2)$, and $f_{2,\boldsymbol{\lambda}_2\boldsymbol{\lambda}_2^T}(\mathbf{y}_t;\boldsymbol{\lambda}_2)$ is to utilize the multivariate Gauss-Hermite quadrature, e.g., see Fahrmeir and Tutz (2001, pages 447-449). All of nodes and weights of the Hermite polynomial of 32 degrees are shown in the appendix for the multivariate Gauss-Hermite quadrature.

In the paper, a simulation study is conducted for the following four cases where

$$F = F_{\boldsymbol{\lambda}^0} = \frac{\exp(\omega^0)}{1+\exp(\omega^0)}\cdot F_{1,\boldsymbol{\lambda}_1^0} + \frac{1}{1+\exp(\omega^0)}\cdot F_{2,\ \boldsymbol{\lambda}_2^0},$$

$k = 1$, $T = 300$, and $n_1 = \ldots = n_T = 300$ with $\boldsymbol{\lambda}^0 \equiv (\omega^0, \boldsymbol{\lambda}_1^{0T}, \boldsymbol{\lambda}_2^{0T})^T$ ($\equiv (\omega^0, \lambda_{11}^0, \lambda_{12}^0, \lambda_{21}^0, \lambda_{22}^0)^T$).

Case 1: $\boldsymbol{\lambda}^0 = (\log(1/5), \log(85), \log(15), -0.716, \log[1/(0.214)^2])^T$. In particular, $\exp(\omega^0)/[1+\exp(\omega^0)] = 1/6$, $F_{1,\boldsymbol{\lambda}_1^0}$ is the beta$(85,15)$ distribution, and $F_{2,\boldsymbol{\lambda}_2^0}$ is the transformed-normal$(-0.716, (0.214)^2)$ distribution.

Case 2: $\boldsymbol{\lambda}^0 = (\log(1), \log(80), \log(20), -0.410, \log[1/(0.205)^2])^T$. In particular, $\exp(\omega^0)/[1+\exp(\omega^0)] = 1/2$, $F_{1,\boldsymbol{\lambda}_1^0}$ is the beta$(80,20)$ distribution,and $F_{2,\boldsymbol{\lambda}_2^0}$ is the transformed-normal$(-0.410, (0.205)^2)$ distribution.

Case 3: $\boldsymbol{\lambda}^0 = (\log(1), \log(60), \log(40), -1.405, \log[1/(0.253)^2])^T$. In particular, $\exp(\omega^0)/[1+\exp(\omega^0)] = 1/2$, $F_{1,\boldsymbol{\lambda}_1^0}$ is the beta$(60, 40)$ distribution, and $F_{2,\boldsymbol{\lambda}_2^0}$ is the transformed-normal$(-1.405, (0.253)^2)$ distribution.

Case 4: $\boldsymbol{\lambda}^0 = (\log(5), \log(73), \log(27), -0.203, \log[1/(0.202)^2])^T$. In particular, $\exp(\omega^0)/[1+\exp(\omega^0)] = 5/6$, $F_{1,\boldsymbol{\lambda}_1^0}$ is the beta$(73, 27)$ distribution, and $F_{2,\boldsymbol{\lambda}_2^0}$ is the transformed-normal$(-0.203, (0.202)^2)$ distribution.

5 GOODNESS OF FIT

In this section, the goodness of fit of the proposed model for a set of available historical *in-control* response vectors, $\{\mathbf{y}_1, \ldots, \mathbf{y}_T\}$, generated in a manufacturing process is discussed. Recall that $\boldsymbol{\theta} \equiv (\boldsymbol{\theta}_1^T, \ldots, \boldsymbol{\theta}_T^T)^T$, $\mathbf{y} \equiv (\mathbf{y}_1^T, \ldots, \mathbf{y}_T^T)^T$, $\mathcal{Y} \equiv \mathcal{Y}_{n_1} \times \cdots \times \mathcal{Y}_{n_T}$, and F is the in-control prior c.d.f.

Consider the null hypothesis H_0: $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_T \overset{i.i.d.}{\sim} F \in \{F_{\boldsymbol{\lambda}}\colon \boldsymbol{\lambda} \in \Lambda\}$ versus the alternative H_1: $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_T \overset{i.i.d.}{\sim} F \notin \{F_{\boldsymbol{\lambda}}\colon \boldsymbol{\lambda} \in \Lambda\}$. Let $\mathcal{F}(\Theta)$ denote the set of all prior c.d.f.'s on Θ and let $\ell(F; \mathbf{y})$ denote the log-likelihood function of F given $\mathbf{y}$. Then

$$\ell(F; \mathbf{y}) \equiv \log\left[\prod_{t=1}^{T} f(\mathbf{y}_t; F)\right] = \sum_{t=1}^{T} \log[f(\mathbf{y}_t; F)] \equiv \sum_{t=1}^{T} \ell(F; \mathbf{y}_t),$$

where

$$f(\mathbf{y}_t; F) = \int_{\Theta} f(\mathbf{y}_t | \boldsymbol{\theta}_t)\, dF(\boldsymbol{\theta}_t).$$

Let $W_T(\mathbf{y})$ denote the corresponding likelihood ratio (LR) statistic given $\mathbf{y}$.

Then

$$W_T(\mathbf{y}) \equiv 2\left[\sup_{F\in\mathcal{F}(\Theta)} \ell(F;\mathbf{y}) - \sup_{\boldsymbol{\lambda}\in\Lambda} \ell(\,\boldsymbol{\lambda};\mathbf{y})\right] \equiv 2\left[\ell\left(\hat{F};\mathbf{y}\right) - \ell\left(\hat{\boldsymbol{\lambda}};\mathbf{y}\right)\right], \quad (46)$$

where $\hat{F}$ is the non-parametric MLE of F given $\mathbf{y}$ under H_1 and $\hat{\boldsymbol{\lambda}}$ is the parametric MLE of $\boldsymbol{\lambda}$ under H_0. Since it takes too much time to calculate the critical point for performing the LR test, an alternative goodness-of-fit test is proposed in the paper as follows:

Note that the empirical prior c.d.f. $\tilde{F}$ with p.m.f. $T^{-1}\cdot\sum_{t=1}^{T} 1_{\{\boldsymbol{\theta}_t\}}$ converges to F in distribution as $T\to\infty$ and that, for $t\in\{1,\ldots,T\}$, the MLE $\mathbf{y}_t/n_t$ of $\boldsymbol{\theta}_t$ given $\mathbf{y}_t$ converges to $\boldsymbol{\theta}_t$ as $n_t\to\infty$. Since $\boldsymbol{\theta}_1,\ldots,\boldsymbol{\theta}_T$ are unobserved, the empirical prior c.d.f. $\tilde{F}$ is unavailable. Thus, we utilize the estimated empirical prior c.d.f. F^* with p.m.f. $T^{-1}\cdot\sum_{t=1}^{T} 1_{\{\mathbf{y}_t/n_t\}}$ to estimate F. When all of $n_1,\ldots,n_T$, and T tend to ∞, F^* converges to F in distribution.

In the paper, consider the goodness-of-fit statistic

$$W_T^*(\mathbf{y}) \equiv 2\left[\ell(F;\mathbf{y})|_{F=F^*} - \ell\left(\hat{\boldsymbol{\lambda}};\mathbf{y}\right)\right] \equiv 2\left[\ell(F^*;\mathbf{y}) - \ell\left(\hat{\boldsymbol{\lambda}};\mathbf{y}\right)\right]. \quad (47)$$

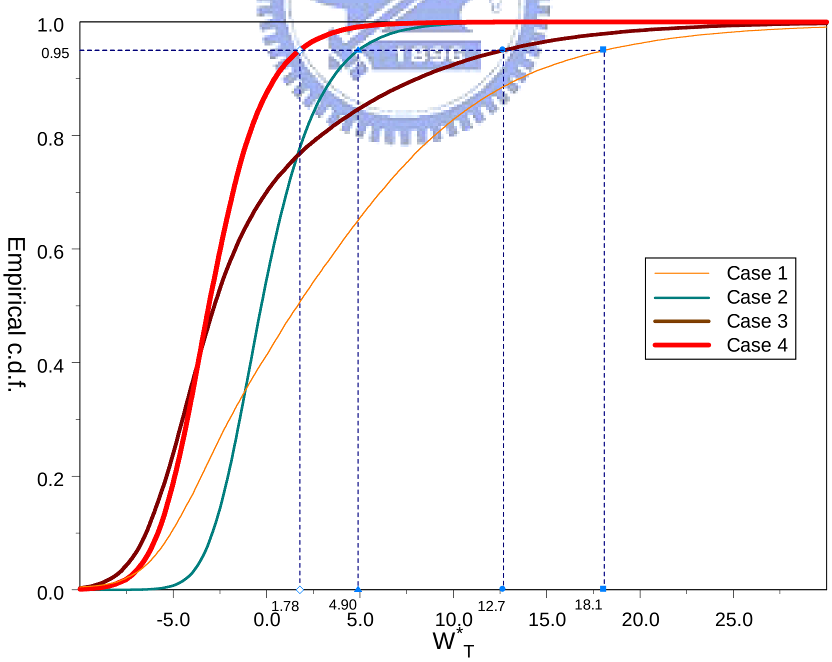
One way to calculate the critical point for performing the goodness-of-fit test is as follows: First simulate an i.i.d. sample $\{\mathbf{y}^{(1)},\ldots,\mathbf{y}^{(R)}\}$, e.g., $R=50\,000$, from the estimated in-control marginal c.d.f. $F_{\mathbf{y};\boldsymbol{\lambda}^0}|_{\boldsymbol{\lambda}^0=\hat{\boldsymbol{\lambda}}}$ ($\equiv F_{\mathbf{y};\hat{\boldsymbol{\lambda}}}$). Let $(\mathbf{y}_{(1)},\ldots,\mathbf{y}_{(R)})$ be a permutation of $(\mathbf{y}^{(1)},\ldots,\mathbf{y}^{(R)})$ such that $W_T^*(\mathbf{y}_{(1)})\le\ldots\le W_T^*(\mathbf{y}_{(R)})$. Let α be a known constant with $0<\alpha<1$, e.g., 0.05. An approximate size $1-\alpha$ goodness-of-fit test is to reject H_0 if and only if $W_T^*(\mathbf{y}) > W_T^*(\mathbf{y}_{([R\,(1-\alpha)])})$, where $[R\,(1-\alpha)]$ is the largest integer less than or equal to $R\,(1-\alpha)$.

The corresponding values of $W_T^*(\mathbf{y}_{([R\,(1-\alpha)])})$'s for Cases 1-4 in Section 4 are shown in Table 1, where $k = 1$, $T = 300$, $n_1 = \ldots = n_T = 300$, $R = 50\,000$, and $\alpha = 0.05$. And the empirical c.d.f.'s of $W_T^*(\mathbf{y})$'s for Cases 1-4 in Section 4 are shown in Figures 1, where $k = 1$, $T = 300$, $n_1 = \ldots = n_T = 300$, $R = 50\,000$.

Table 1: The values of $W_T^*(\mathbf{y}_{([R\,(1-\alpha)])})$'s for Cases 1-4, where $k = 1$, $T = 300$, $n_1 = \ldots = n_T = n_t = 300$, $R = 50\,000$, and $\alpha = 0.05$.

	Case 1	Case 2	Case 3	Case 4
$W_T^*(\mathbf{y}_{([R\,(1-\alpha)])})$	18.1	4.90	12.7	1.78

Figure 1: The empirical c.d.f.'s of W_T^*'s for Case 1-4, where $k = 1$, $T = 300$, $n_1 = \ldots = n_T = n_t = 300$, and $R = 50\,000$.



6 SIMPLIFICATION

In this section, the simplification of the two-components mixture prior parametric family to either the first or the second component prior parametric family is discussed if the null hypothesis of the previous goodness-of-fit test is not rejected.

Let $u \in \{1, 2\}$ be fixed. Consider the null hypothesis H_{u0}: $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_T \overset{i.i.d.}{\sim} F \in \{F_{u,\boldsymbol{\lambda}_u}\colon \boldsymbol{\lambda}_u \in \Lambda_u\}$ versus the alternative H_{u1}: $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_T \overset{i.i.d.}{\sim} F \in \{F_{\boldsymbol{\lambda}}\colon \boldsymbol{\lambda} \in \Lambda\}$. Let $W_{u,T}(\mathbf{y})$ denote the LR statistic given $\mathbf{y}$, where

$$\begin{aligned} W_{u,T}(\mathbf{y}) &\equiv 2\left[\ell\left(\hat{\boldsymbol{\lambda}};\mathbf{y}\right) - \sup_{\boldsymbol{\lambda}_u\in\Lambda_u}\sum_{t=1}^{T}\ell_u(\boldsymbol{\lambda}_u;\mathbf{y}_t)\right] \\ \equiv 2\left[\ell\left(\hat{\boldsymbol{\lambda}};\mathbf{y}\right) - \sup_{\boldsymbol{\lambda}_u\in\Lambda_u}\ell_u(\boldsymbol{\lambda}_u;\mathbf{y})\right] &\equiv 2\left[\ell\left(\hat{\boldsymbol{\lambda}};\mathbf{y}\right) - \ell_u\left(\hat{\boldsymbol{\lambda}}_u;\mathbf{y}\right)\right] \end{aligned} \tag{48}$$

with $\hat{\boldsymbol{\lambda}}_u$ denoting the MLE of $\boldsymbol{\lambda}_u$ given $\mathbf{y}$ under the uth component prior parametric family.

One way to calculate the critical point for performing the LR test is as follows: First simulate $\{\mathbf{y}^{(u,1)}, \ldots, \mathbf{y}^{(u,R)}\}$, e.g., $R = 50\,000$, from the estimated in-control marginal c.d.f. $F_{\mathbf{y};u,\boldsymbol{\lambda}_u^0}|_{\boldsymbol{\lambda}_u^0=\hat{\boldsymbol{\lambda}}_u}$ $(\equiv F_{\mathbf{y};u,\hat{\boldsymbol{\lambda}}_u})$. Let $(\mathbf{y}^{(u)}_{(1)}, \ldots, \mathbf{y}^{(u)}_{(R)})$ be a permutation of $(\mathbf{y}^{(u,1)}, \ldots, \mathbf{y}^{(u,R)})$ such that $W_{u,T}(\mathbf{y}^{(u)}_{(1)}) \leq \ldots \leq W_{u,T}(\mathbf{y}^{(u)}_{(R)})$. Let α be a known constant with $0 < \alpha < 1$, e.g., 0.05. An approximate size $1-\alpha$ LR test is to reject H_{u0} if and only if $W_{u,T}(\mathbf{y}) > W_{u,T}(\mathbf{y}^{(u)}_{([R\,(1-\alpha)])})$, where $[R\,(1-\alpha)]$ is the largest integer less than or equal to $R\,(1-\alpha)$.

When both H_{10} and H_{20} are rejected, the proposed two-components mixture prior parametric family for the in-control prior distribution is selected. The corresponding monitoring technique is developed in the following section.

When H_{10} is not rejected but H_{20} is rejected, the first component prior parametric family for the in-control prior distribution is selected. The corresponding monitoring technique is developed in Chen *et al.* (2004).

When H_{10} is rejected but H_{20} is not rejected, the second component prior parametric family for the in-control prior distribution is selected. The corresponding monitoring technique is developed in Chen *et al.* (2005).

When neither H_{10} nor H_{20} is rejected, the model selection technique developed in Chen and Liu (2005) can be utilized. The corresponding monitoring technique is developed in either Chen *et al.* (2004) or Chen *et al.* (2005).

The corresponding values of $W_{u,T}(\mathbf{y}^{(u)}_{([R\,(1-\alpha)])})$'s for Cases 1-4 in Section 4 are shown in Table 2, where $u \in \{1, 2\}$, $k = 1$, $T = 300$, $n_1 = \ldots = n_T = 300$, $R = 50\,000$, and $\alpha = 0.05$. And the empirical c.d.f.'s of $W_{1,T}(\mathbf{y})$'s and $W_{2,T}(\mathbf{y})$'s for Cases 1-4 in Section 4 are shown in Figures 2 and 3, where $k = 1$, $T = 300$, $n_1 = \ldots = n_T = 300$, $R = 50\,000$.

Table 2: The values of $W_{u,T}(\mathbf{y}^{(u)}_{([R\,(1-\alpha)])})$'s for Cases 1-4, where $u \in \{1, 2\}$, $k = 1$, $T = 300$, $n_1 = \ldots = n_T = n_t = 300$, $R = 50\,000$, and $\alpha = 0.05$.

	Case 1	Case 2	Case 3	Case 4
$W_{1,T}(\mathbf{y}^{(1)}_{([R\,(1-\alpha)])})$	2.146	1.762	0.566	1.284
$W_{2,T}(\mathbf{y}^{(2)}_{([R\,(1-\alpha)])})$	1.035	0.653	1.789	0.335

Figure 2: The empirical c.d.f.'s of $W_{1,T}$'s for Case 1-4, where $k = 1$, $T = 300$, $n_1 = \ldots = n_T = n_t = 300$, $R = 50\,000$.

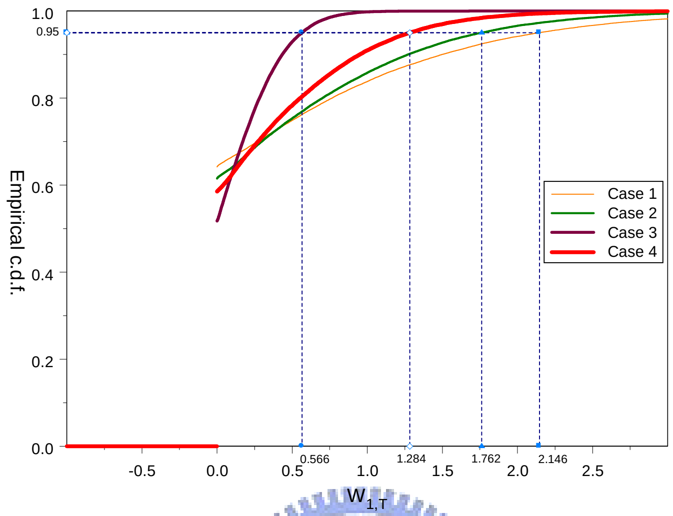
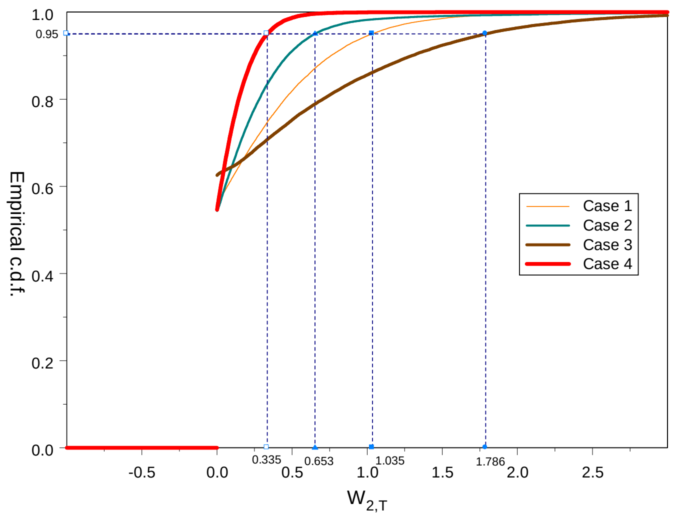


Figure 3: The empirical c.d.f.'s of $W_{2,T}$'s for Case 1-4, where $k = 1$, $T = 300$, $n_1 = \ldots = n_T = n_t = 300$, $R = 50\,000$.



7 A PROCESS MONITORING SCHEME

Let P_{in} denote the false-alarm rate, i.e., the probability that an out-of-control signal occurs when the manufacturing process is in control. Conventionally, P_{in} is taken to be $2\,\Phi(-3)$ ($\approx 0.002\,699\,8$), where Φ is the c.d.f. of the standard normal distribution. In this section, utilizing the LR method, a Bayesian (or an empirical Bayes) monitoring scheme for the manufacturing process is proposed when $F = F_{\boldsymbol{\lambda}^0} \in \{F_{\boldsymbol{\lambda}}\colon \boldsymbol{\lambda} \in \Lambda\}$ for some known (or unknown) $\boldsymbol{\lambda}^0 \in \Lambda$. The main reason for using the LR test is that it often has a higher power than other tests when the alternative hypothesis is true, which corresponds to a better detecting power in monitoring the process when the process is out of control.

In order to monitor the manufacturing process at time t ($> T$), suppose that the response vector $\mathbf{y}_t$ is observed. Then we are interested in testing whether or not the manufacturing process is in control at time t. Recall that $F_{\boldsymbol{\theta}_t}$ is the prior c.d.f. of $\boldsymbol{\theta}_t$ and that $\mathcal{F}(\Theta)$ is the set of all c.d.f.'s on Θ.

7.1 A BAYESIAN MONITORING SCHEME

In this subsection, consider the case where $F = F_{\boldsymbol{\lambda}^0} \in \{F_{\boldsymbol{\lambda}}\colon \boldsymbol{\lambda} \in \Lambda\}$ for some known $\boldsymbol{\lambda}^0 \in \Lambda$. To monitor the manufacturing process at time t, the null hypothesis H_0: $F_{\boldsymbol{\theta}_t} = F_{\boldsymbol{\lambda}^0}$ versus the alternative H_1: $F_{\boldsymbol{\theta}_t} \neq F_{\boldsymbol{\lambda}^0}$, i.e., $F_{\boldsymbol{\theta}_t} \in \mathcal{F}(\Theta)\backslash\{F_{\boldsymbol{\lambda}^0}\}$, is tested.

List all the elements of the sample space $\mathcal{Y}_{n_t}$ of $\mathbf{y}_t$ by $\{\mathbf{y}_t^{(1)}, \ldots, \mathbf{y}_t^{(|\mathcal{Y}_{n_t}|)}\}$, where $|\mathcal{Y}_{n_t}|$ ($= (n_t + k)!/(n_t!k!)$) is the number of elements in $\mathcal{Y}_{n_t}$. Regard $F_{\boldsymbol{\theta}_t}$ as the unknown parameter of interest in $\mathcal{F}(\Theta)$. Then the unknown parameter of interest is non-parametric. Let $\ell(F_{\boldsymbol{\theta}_t}; \mathbf{y}_t)$ ($\equiv \log[f(\mathbf{y}_t; F_{\boldsymbol{\theta}_t})]$) denote the log-likelihood function of F_{θ_t} given $\mathbf{y}_t$. Note that

$$\begin{aligned}\ell(F_{\boldsymbol{\theta}_t};\mathbf{y}_t) &= \log\left[\int_\Theta f(\mathbf{y}_t|\boldsymbol{\theta}_t)\,dF_{\boldsymbol{\theta}_t}(\boldsymbol{\theta}_t)\right] \leq \log\left[\int_\Theta \sup_{\boldsymbol{\theta}_t\in\Theta} f(\mathbf{y}_t|\boldsymbol{\theta}_t)\,dF_{\boldsymbol{\theta}_t}(\boldsymbol{\theta}_t)\right]\\ &= \log\left[\int_\Theta f(\mathbf{y}_t|\boldsymbol{\theta}_t)|_{\boldsymbol{\theta}_t=\mathbf{y}_t/n_t}\,dF_{\boldsymbol{\theta}_t}(\boldsymbol{\theta}_t)\right] = \log\left[f(\mathbf{y}_t|\boldsymbol{\theta}_t)|_{\boldsymbol{\theta}_t=\mathbf{y}_t/n_t}\right],\end{aligned}$$

where the binomial/multinomial likelihood $f(\mathbf{y}_t|\boldsymbol{\theta}_t)$ for $\boldsymbol{\theta}_t$ given $\mathbf{y}_t$ attains its maximum at $\boldsymbol{\theta}_t = \mathbf{y}_t/n_t$. Thus, the MLE $\hat{F}_{\boldsymbol{\theta}_t}$ of $F_{\boldsymbol{\theta}_t}$ given $\mathbf{y}_t$ has p.m.f. $1_{\{\mathbf{y}_t/n_t\}}$ and

$$\sup_{F_{\boldsymbol{\theta}_t}\in\mathcal{F}(\Theta)} \ell(F_{\boldsymbol{\theta}_t};\mathbf{y}_t) = \ell(F_{\boldsymbol{\theta}_t};\mathbf{y}_t)|_{F_{\boldsymbol{\theta}_t}=\hat{F}_{\boldsymbol{\theta}_t}} \equiv \ell\left(\hat{F}_{\boldsymbol{\theta}_t};\mathbf{y}_t\right) = \log\left[f(\mathbf{y}_t|\boldsymbol{\theta}_t)|_{\boldsymbol{\theta}_t=\mathbf{y}_t/n_t}\right].$$

Let $W_{t,\boldsymbol{\lambda}^0}(\mathbf{y}_t)$ denote the corresponding LR statistic, where

$$W_{t,\boldsymbol{\lambda}^0}(\mathbf{y}_t) = 2\left\{\log\left[f(\mathbf{y}_t|\boldsymbol{\theta}_t)|_{\boldsymbol{\theta}_t=\mathbf{y}_t/n_t}\right] - \ell\left(\boldsymbol{\lambda}^0;\mathbf{y}_t\right)\right\} \tag{49}$$

with $P(\{0 < W_{t,\boldsymbol{\lambda}^0}(\mathbf{y}_t) < \infty\}; F_{\mathbf{y}_t;\boldsymbol{\lambda}^0}) = 1$.

The size P_{in} LR test and a control chart of monitoring the LR statistic $W_{t,\boldsymbol{\lambda}^0}(\mathbf{y}_t)$ can be constructed as follows: Let $(\mathbf{y}_{t,(1)},\ldots,\mathbf{y}_{t,(|\mathcal{Y}_{n_t}|)})$ be a permutation of $(\mathbf{y}_t^{(1)},\ldots,\mathbf{y}_t^{(|\mathcal{Y}_{n_t}|)})$ such that $W_{t,\boldsymbol{\lambda}^0}(\mathbf{y}_{t,(1)}) \leq \ldots \leq W_{t,\boldsymbol{\lambda}^0}(\mathbf{y}_{t,(|\mathcal{Y}_{n_t}|)})$. Note that $W_{t,\boldsymbol{\lambda}^0}(\mathbf{y}_t)$ is a discrete random variable. If a deterministic upper control limit is used, a pre-specified false-alarm rate P_{in} ($\in (0,1)$), e.g., $2\,\Phi(-3)$, is nearly impossible to attain. However, there is no problem to attain any pre-specified false-alarm rate based on the concept of a randomized-upper-control-limit approach proposed in Shiau *et al.* (2005). To find the randomized upper control limit ($\equiv RUCL_{\boldsymbol{\lambda}^0}$), we start accumulating the right tail probability from $W_{t,\boldsymbol{\lambda}^0}(\mathbf{y}_{t,(|\mathcal{Y}_{n_t}|)})$ until we reach the first r

such that $P(\{W_{t,\boldsymbol{\lambda}^0}(\mathbf{y}_t) \geq W_{t,\boldsymbol{\lambda}^0}(\mathbf{y}_{t,(r)})\}; F_{\mathbf{y}_t;\boldsymbol{\lambda}^0}) \geq P_{in}$. Denote this r by $m_{\boldsymbol{\lambda}^0}$, i.e.,

$$m_{\boldsymbol{\lambda}^0} = \max\left\{r\colon P\left(\left\{W_{t,\boldsymbol{\lambda}^0}(\mathbf{y}_t) \geq W_{t,\boldsymbol{\lambda}^0}\left(\mathbf{y}_{t,(r)}\right)\right\}; F_{\mathbf{y}_t;\boldsymbol{\lambda}^0}\right) \geq P_{in}\right\}. \quad (50)$$

If $P(\{W_{t,\boldsymbol{\lambda}^0}(\mathbf{y}_t) \geq W_{t,\boldsymbol{\lambda}^0}(\mathbf{y}_{t,(m_{\boldsymbol{\lambda}^0})})\}; F_{\mathbf{y}_t;\boldsymbol{\lambda}^0}) = P_{in}$, which is nearly impossible, then there is no need for randomization and $W_{t,\boldsymbol{\lambda}^0}(\mathbf{y}_{t,(m_{\boldsymbol{\lambda}^0})})$ is the upper control limit ($\equiv UCL_{\boldsymbol{\lambda}^0}$). If $P(\{W_{t,\boldsymbol{\lambda}^0}(\mathbf{y}_t) \geq W_{t,\boldsymbol{\lambda}^0}(\mathbf{y}_{t,(m_{\boldsymbol{\lambda}^0})})\}; F_{\mathbf{y}_t;\boldsymbol{\lambda}^0}) > P_{in}$, then $W_{t,\boldsymbol{\lambda}^0}(\mathbf{y}_{t,(m_{\boldsymbol{\lambda}^0})}) = RUCL_{\boldsymbol{\lambda}^0}$. Note that there may be more than one $\mathbf{y}_{t,(r)}$ such that $W_{t,\boldsymbol{\lambda}^0}(\mathbf{y}_{t,(r)}) = RUCL_{\boldsymbol{\lambda}^0}$. Let $m_{\boldsymbol{\lambda}^0,L}$, $m_{\boldsymbol{\lambda}^0,U} \in \{1, \ldots, |\mathcal{Y}_{n_t}|\}$ such that

$$\begin{aligned} W_{t,\boldsymbol{\lambda}^0}\left(\mathbf{y}_{t,(m_{\boldsymbol{\lambda}^0,L-1})}\right) &< W_{t,\boldsymbol{\lambda}^0}\left(\mathbf{y}_{t,(m_{\boldsymbol{\lambda}^0,L})}\right) = RUCL_{\boldsymbol{\lambda}^0} = W_{t,\boldsymbol{\lambda}^0}\left(\mathbf{y}_{t,(m_{\boldsymbol{\lambda}^0,U})}\right) \\ &< W_{t,\boldsymbol{\lambda}^0}\left(\mathbf{y}_{t,(m_{\boldsymbol{\lambda}^0,U+1})}\right), \end{aligned}$$

where $W_{t,\boldsymbol{\lambda}^0}(\mathbf{y}_{t,(0)}) \equiv 0$ and $W_{t,\boldsymbol{\lambda}^0}(\mathbf{y}_{t,(|\mathcal{Y}_{n_t}|+1)}) \equiv \infty$. Then the randomization is done by signaling an out-of-control alarm with probability

$$\begin{aligned} P_{in,\boldsymbol{\lambda}^0,RUCL} &= \frac{P_{in} - P(\{W_{t,\boldsymbol{\lambda}^0}(\mathbf{y}_t) > RUCL_{\boldsymbol{\lambda}^0}\}; F_{\mathbf{y}_t;\boldsymbol{\lambda}^0})}{P(\{W_{t,\boldsymbol{\lambda}^0}(\mathbf{y}_t) = RUCL_{\boldsymbol{\lambda}^0}\}; F_{\mathbf{y}_t;\boldsymbol{\lambda}^0})} \\ &= \frac{P_{in} - \sum_{r=m_{\boldsymbol{\lambda}^0,U}+1}^{|\mathcal{Y}_{n_t}|} P(\{W_{t,\boldsymbol{\lambda}^0}(\mathbf{y}_t) = W_{t,\boldsymbol{\lambda}^0}(\mathbf{y}_{t,(r)})\}; F_{\mathbf{y}_t;\boldsymbol{\lambda}^0})}{\sum_{r=m_{\boldsymbol{\lambda}^0,L}}^{m_{\boldsymbol{\lambda}^0,U}} P(\{W_{t,\boldsymbol{\lambda}^0}(\mathbf{y}_t) = W_{t,\boldsymbol{\lambda}^0}(\mathbf{y}_{t,(r)})\}; F_{\mathbf{y}_t;\boldsymbol{\lambda}^0})}. \quad (51) \end{aligned}$$

This leads to

$$\begin{aligned} P_{in} &= P\left(\left\{W_{t,\boldsymbol{\lambda}^0}(\mathbf{y}_t) > RUCL_{\boldsymbol{\lambda}^0}\right\}; F_{\mathbf{y}_t;\boldsymbol{\lambda}^0}\right) \\ &\quad + P_{in,\boldsymbol{\lambda}^0,RUCL} \cdot P\left(\left\{W_{t,\boldsymbol{\lambda}^0}(\mathbf{y}_t) = RUCL_{\boldsymbol{\lambda}^0}\right\}; F_{\mathbf{y}_t;\boldsymbol{\lambda}^0}\right) \end{aligned}$$

and $0 < P_{in,\boldsymbol{\lambda}^0,RUCL} \leq 1$. When $P_{in,\boldsymbol{\lambda}^0,RUCL} = 1$, there is no need for randomization.

The monitoring scheme is as follows: If $W_{t,\boldsymbol{\lambda}^0}(\mathbf{y}_t) > RUCL_{\boldsymbol{\lambda}^0}$, then the null hypothesis H_0: $F_{\boldsymbol{\theta}_t} = F_{\boldsymbol{\lambda}^0}$ is rejected and the manufacturing process is declared to be out of control at time t; if $W_{t,\boldsymbol{\lambda}^0}(\mathbf{y}_t) < RUCL_{\boldsymbol{\lambda}^0}$, then the null hypothesis H_0: $F_{\boldsymbol{\theta}_t} = F_{\boldsymbol{\lambda}^0}$ is not rejected and the manufacturing process is declared to be in control at time t; if $W_{t,\boldsymbol{\lambda}^0}(\mathbf{y}_t) = RUCL_{\boldsymbol{\lambda}^0}$, then, with probability $P_{in,\boldsymbol{\lambda}^0,RUCL}$, the null hypothesis H_0: $F_{\boldsymbol{\theta}_t} = F_{\boldsymbol{\lambda}^0}$ is rejected and the manufacturing process is declared to be out of control at time t.

The corresponding values of $RUCL_{\boldsymbol{\lambda}^0}$'s and $P_{in,\boldsymbol{\lambda}^0,RUCL}$'s for Cases 1-4 in Section 4 are shown in Table 3, where $k = 1$, $T = 300$, $n_1 = \ldots = n_T = n_t = 300$, and $P_{in} = 2\,\Phi(-3)$ ($\approx 0.002\,699\,8$).

Table 3: The values of $RUCL_{\boldsymbol{\lambda}^0}$'s and $P_{in,\boldsymbol{\lambda}^0,RUCL}$'s for Cases 1-4 in Section 4, where $k = 1$, $T = 300$, $n_1 = \ldots = n_T = n_t = 300$, and $P_{in} = 2\,\Phi(-3)$ ($\approx$ $0.002\,699\,8$).

	Case 1	Case 2	Case 3	Case 4
$RUCL_{\boldsymbol{\lambda}^0}$	10.5	10.4	10.5	10.6
$P_{in,\boldsymbol{\lambda}^0,RUCL}$	0.547	0.537	0.765	0.120

7.2 AN EMPIRICAL BAYES MONITORING SCHEME

In this subsection, consider the case where $F = F_{\boldsymbol{\lambda}^0} \in \{F_{\boldsymbol{\lambda}}\colon \boldsymbol{\lambda} \in \Lambda\}$ for some unknown $\boldsymbol{\lambda}^0 \in \Lambda$. To monitor the manufacturing process at time t ($> T$), the null hypothesis H_0: $F_{\boldsymbol{\theta}_t} = F_{\boldsymbol{\lambda}^0}$ versus the alternative H_1: $F_{\boldsymbol{\theta}_t} \neq F_{\boldsymbol{\lambda}^0}$ is tested.

The LR statistic $W_{t,\boldsymbol{\lambda}^0}(\mathbf{y}_t)$ proposed in Section 7.1 for known $\boldsymbol{\lambda}^0$ can be estimated by $W_{t,\boldsymbol{\lambda}^0}(\mathbf{y}_t)|_{\boldsymbol{\lambda}^0=\hat{\boldsymbol{\lambda}}}$ $(\equiv W_{t,\hat{\boldsymbol{\lambda}}}(\mathbf{y}_t))$, where $\hat{\boldsymbol{\lambda}}$ is the MLE of $\boldsymbol{\lambda}$ given $\mathbf{y}$ and

$$W_{t,\hat{\boldsymbol{\lambda}}}(\mathbf{y}_t) \;=\; 2\left\{\log\left[f(\mathbf{y}_t|\boldsymbol{\theta}_t)|_{\boldsymbol{\theta}_t=\mathbf{y}_t/n_t}\right] - \ell\left(\hat{\boldsymbol{\lambda}};\mathbf{y}_t\right)\right\} \qquad (52)$$

with $P(\{0 < W_{t,\hat{\boldsymbol{\lambda}}}(\mathbf{y}_t) < \infty\}; F_{\mathbf{y}_t;\boldsymbol{\lambda}^0}) = 1$. Note that $\hat{\boldsymbol{\lambda}} = \boldsymbol{\lambda}^0 + O_p(1/\sqrt{T})$ as $T \to \infty$, which implies that $W_{t,\hat{\boldsymbol{\lambda}}}(\mathbf{y}_t) = W_{t,\boldsymbol{\lambda}^0}(\mathbf{y}_t) + O_p(1/\sqrt{T})$ as $T \to \infty$.

An empirical Bayes monitoring scheme can be constructed by replacing the unknown $\boldsymbol{\lambda}^0$ by $\hat{\boldsymbol{\lambda}}$ in the Bayesian monitoring scheme described in Section 7.1, where $RUCL_{\boldsymbol{\lambda}^0}$ and $P_{in,\boldsymbol{\lambda}^0,RUCL}$ are estimated by $RUCL_{\boldsymbol{\lambda}^0}|_{\boldsymbol{\lambda}^0=\hat{\boldsymbol{\lambda}}}$ $(\equiv RUCL_{\hat{\boldsymbol{\lambda}}})$ and $P_{in,\boldsymbol{\lambda}^0,RUCL}|_{\boldsymbol{\lambda}^0=\hat{\boldsymbol{\lambda}}}$ $(\equiv P_{in,\hat{\boldsymbol{\lambda}},RUCL})$, respectively.

To see how the additional estimation error resulting from the empirical Bayes approach affects the performance of the monitoring scheme, the Kullback-Leibler divergence $d(F_{\boldsymbol{\lambda}^0}, F_{\hat{\boldsymbol{\lambda}}})$ between $F_{\boldsymbol{\lambda}^0}$ and $F_{\hat{\boldsymbol{\lambda}}}$ can be used as a measure of how close $F_{\hat{\boldsymbol{\lambda}}}$ is to $F_{\boldsymbol{\lambda}^0}$, where $F_{\hat{\boldsymbol{\lambda}}} \equiv F_{\boldsymbol{\lambda}}|_{\boldsymbol{\lambda}=\hat{\boldsymbol{\lambda}}}$ and

$$\begin{aligned} d\left(F_{\boldsymbol{\lambda}^0}, F_{\hat{\boldsymbol{\lambda}}}\right) &\equiv \int_{\Theta} \log\left[\frac{\pi(\boldsymbol{\theta}_t;\boldsymbol{\lambda})|_{\boldsymbol{\lambda}=\boldsymbol{\lambda}^0}}{\pi(\boldsymbol{\theta}_t;\boldsymbol{\lambda})|_{\boldsymbol{\lambda}=\hat{\boldsymbol{\lambda}}}}\right] \cdot \pi(\boldsymbol{\theta}_t;\boldsymbol{\lambda})|_{\boldsymbol{\lambda}=\boldsymbol{\lambda}^0}\, d\boldsymbol{\theta}_t \\ &\equiv \int_{\Theta} \log\left[\frac{\pi(\boldsymbol{\theta}_t;\boldsymbol{\lambda}^0)}{\pi(\boldsymbol{\theta}_t;\hat{\boldsymbol{\lambda}})}\right] \cdot \pi(\boldsymbol{\theta}_t;\boldsymbol{\lambda}^0)\, d\boldsymbol{\theta}_t. \qquad (53) \end{aligned}$$

When there is no closed-form formula for $d(F_{\boldsymbol{\lambda}^0}, F_{\hat{\boldsymbol{\lambda}}})$, it can be numerically evaluated as follows: First simulate an i.i.d. sample $\{\boldsymbol{\theta}_1^{(1)}, \ldots, \boldsymbol{\theta}_1^{(R_1)}\}$ of size R_1, e.g., $R_1 = 50\,000$, from the in-control prior c.d.f. $F_{\boldsymbol{\lambda}^0}$ and then numerically evaluate $d(F_{\boldsymbol{\lambda}^0}, F_{\hat{\boldsymbol{\lambda}}})$ by

$$\hat{d}\,(F_{\boldsymbol{\lambda}^0}, F_{\hat{\boldsymbol{\lambda}}}) \equiv \frac{1}{R_1} \cdot \sum_{r=1}^{R_1} \log \left[\frac{\pi(\boldsymbol{\theta}_t; \boldsymbol{\lambda}^0)}{\pi(\boldsymbol{\theta}_t; \hat{\boldsymbol{\lambda}})} \right] \Bigg|_{\boldsymbol{\theta}_t = \boldsymbol{\theta}_1^{(r)}} . \tag{54}$$

One way to numerically evaluate the τ quantile of $d(F_{\boldsymbol{\lambda}^0}, F_{\hat{\boldsymbol{\lambda}}})$ for $\tau \in (0,1)$ is as follows: First simulate an i.i.d. sample $\{\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(R)}\}$, e.g., $R = 50\,000$, from the in-control marginal c.d.f. $F_{\mathbf{y};\boldsymbol{\lambda}^0}$. Let $(\mathbf{y}_{(1)}, \ldots, \mathbf{y}_{(R)})$ be a permutation of $(\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(R)})$ such that $\hat{d}(F_{\boldsymbol{\lambda}^0}, F_{\hat{\boldsymbol{\lambda}}(\mathbf{y}_{(1)})}) \leq \ldots \leq \hat{d}(F_{\boldsymbol{\lambda}^0}, F_{\hat{\boldsymbol{\lambda}}(\mathbf{y}_{(R)})})$. Then the τ quantile of $d(F_{\boldsymbol{\lambda}^0}, F_{\hat{\boldsymbol{\lambda}}})$ can be estimated by $\hat{d}(F_{\boldsymbol{\lambda}^0}, F_{\hat{\boldsymbol{\lambda}}(\mathbf{y}_{([R\,\tau])})})$, where $[R\tau]$ is the largest integer less than or equal to $R\,\tau$.

The corresponding values of $RUCL_{\hat{\boldsymbol{\lambda}}(\mathbf{y}_{([R\,\tau])})}$'s and $P_{in,\hat{\boldsymbol{\lambda}}(\mathbf{y}_{([R\,\tau])}),RUCL}$'s for Cases 1-4 in Section 4 are shown in Table 4, where $k = 1$, $T = 300$, $n_1 = \ldots = n_T = n_t = 300$, $R = R_1 = 50\,000$, and $\tau \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$.

Table 4: $RUCL_{\hat{\boldsymbol{\lambda}}(\mathbf{y}_{([R_2\,\tau])})}$'s and $P_{in,\hat{\boldsymbol{\lambda}}(\mathbf{y}_{([R_2\,\tau])}),RUCL}$'s for Case 1-4, where $k = 1$, $T = 300$, $n_1 = n_2 = \ldots = n_T = n_t = 300$, $R = R_1 = 50\,000$, and $\tau \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$.

Case 1	$\tau = 0.1$	$\tau = 0.3$	$\tau = 0.5$	$\tau = 0.7$	$\tau = 0.9$
$RUCL_{\hat{\boldsymbol{\lambda}}(\mathbf{y}_{([R\,\tau])})}$	10.6	10.5	10.9	10.2	10.5
$P_{in,\hat{\boldsymbol{\lambda}}(\mathbf{y}_{([R\,\tau])}),RUCL}$	0.257	0.0513	0.552	0.00126	0.261

Case 2	$\tau = 0.1$	$\tau = 0.3$	$\tau = 0.5$	$\tau = 0.7$	$\tau = 0.9$
$RUCL_{\hat{\boldsymbol{\lambda}}(\mathbf{y}_{([R\,\tau])})}$	10.6	10.6	10.4	10.6	10.4
$P_{in,\hat{\boldsymbol{\lambda}}(\mathbf{y}_{([R\tau])}),RUCL}$	0.954	0.807	0.216	0.566	0.587

Case 3	$\tau = 0.1$	$\tau = 0.3$	$\tau = 0.5$	$\tau = 0.7$	$\tau = 0.9$
$RUCL_{\hat{\boldsymbol{\lambda}}(\mathbf{y}_{([R\tau])})}$	10.5	10.3	10.3	10.5	10.6
$P_{in,\hat{\boldsymbol{\lambda}}(\mathbf{y}_{([R\tau])}),RUCL}$	0.688	0.0511	0.433	0.840	0.602

Case 4	$\tau = 0.1$	$\tau = 0.3$	$\tau = 0.5$	$\tau = 0.7$	$\tau = 0.9$
$RUCL_{\hat{\boldsymbol{\lambda}}(\mathbf{y}_{([R\tau])})}$	10.5	10.4	10.7	10.5	10.3
$P_{in,\hat{\boldsymbol{\lambda}}(\mathbf{y}_{([R\tau])}),RUCL}$	0.346	0.506	0.353	0.449	0.553

From Tables 3 and 4, it is easily seen that all of $RUCL_{\hat{\boldsymbol{\lambda}}(\mathbf{y}_{([R\tau])})}$'s are close to $RUCL_{\boldsymbol{\lambda}^0}$, but $P_{in,\hat{\boldsymbol{\lambda}}(\mathbf{y}_{([R\tau])}),RUCL}$'s are not necessarily close to $P_{in,\boldsymbol{\lambda}^0,RUCL}$ for Cases 1-4.

8 AVERAGE RUN LENGTH BEHAVIOR

In this section, the performance of the proposed process monitoring scheme is studied in terms of the average run length. Let ARL_0 denote the average run length for an out-of-control signal to occur when the manufacturing process is in control. Recall that P_{in} is the false-alarm rate, i.e., the probability that an out-of-control signal occurs when the manufacturing process is in control. Then $ARL_0 = 1/P_{in}$. When $P_{in} = 2\,\Phi(-3)$ ($\approx 0.002\,699\,8$), $ARL_0 = 1/[2\,\Phi(-3)]$ (≈ 370.40). Let ARL_1 denote the average run length for an out-of-control signal to occur when the manufacturing process is out of control. Let P_{out} denote the correct-alarm rate, i.e., the probability that an out-of-control signal occurs when the manufacturing process is out of control. Similarly, $ARL_1 = 1/P_{out}$.

8.1 A BAYESIAN APPROACH

In this subsection, consider the case where $F = F_{\boldsymbol{\lambda}^0} \in \{F_{\boldsymbol{\lambda}}\colon \boldsymbol{\lambda} \in \Lambda\}$ for some known $\boldsymbol{\lambda}^0 \in \Lambda$. To monitor the manufacturing process at time t, the monitoring scheme proposed in Section 7.1 is used for the null hypothesis H_0: $F_{\boldsymbol{\theta}_t} = F_{\boldsymbol{\lambda}^0}$ versus the alternative H_1: $F_{\boldsymbol{\theta}_t} \neq F_{\boldsymbol{\lambda}^0}$.

Set $ARL_0 \equiv ARL_{0,\boldsymbol{\lambda}^0}$, $P_{in} \equiv P_{in,\boldsymbol{\lambda}^0}$, $ARL_1 \equiv ARL_{1,\boldsymbol{\lambda}^0,F_{\boldsymbol{\theta}_t}}$, and $P_{out} \equiv P_{out,\boldsymbol{\lambda}^0,F_{\boldsymbol{\theta}_t}}$. When $P_{in,\boldsymbol{\lambda}^0}$ is pre-specified to be $2\,\Phi(-3)$ ($\approx 0.002\,699\,8$),

$$ARL_{0,\boldsymbol{\lambda}^0} = \frac{1}{P_{in,\boldsymbol{\lambda}^0}} = \frac{1}{2\,\Phi(-3)} \approx 370.40. \tag{55}$$

When $F_{\boldsymbol{\theta}_t} \neq F_{\boldsymbol{\lambda}^0}$,

$$\begin{aligned} P_{out,\boldsymbol{\lambda}^0,F_{\boldsymbol{\theta}_t}} &= P\left(\{W_{t,\boldsymbol{\lambda}^0}(\mathbf{y}_t) > RUCL_{\boldsymbol{\lambda}^0}\}\,; F_{\mathbf{y}_t}\right) \\ &\quad + P_{in,\boldsymbol{\lambda}^0,RUCL} \cdot P\left(\{W_{t,\boldsymbol{\lambda}^0}(\mathbf{y}_t) = RUCL_{\boldsymbol{\lambda}^0}\}\,; F_{\mathbf{y}_t}\right), \end{aligned} \tag{56}$$

where all of $W_{t,\boldsymbol{\lambda}^0}(\mathbf{y}_t)$, $RUCL_{\boldsymbol{\lambda}^0}$, and $P_{in,\boldsymbol{\lambda}^0,RUCL}$ are defined in Section 7.1 and $F_{\mathbf{y}_t}$ is defined in Section 3.

The corresponding values of $P_{out,\boldsymbol{\lambda}^0,F_{\boldsymbol{\lambda}^i}}$'s, and $ARL_{1,\boldsymbol{\lambda}^0,F_{\boldsymbol{\lambda}^i}}$'s for Cases 1-4 in Section 4 are shown in Tables 5 and 6, where $k = 1$, $T = 300$, $n_1 = \ldots = n_T = n_t = 300$, and $i \in \{1, 2, 3\}$.

Case 1: $\boldsymbol{\lambda}^1 = (\log(1/4), \log(80), \log(20), -2.210, \log[1/(0.210)^2])^T$, $\boldsymbol{\lambda}^2 = (\log(1/9), \log(90), \log(10), -1.552, \log[1/(0.220)^2])^T$, and $\boldsymbol{\lambda}^3 = (\log(4/21), \log(80), \log(20), -0.503, \log[1/(0.216)^2])^T$.

Case 2: $\boldsymbol{\lambda}^1 = (\log(9/11), \log(85), \log(15), -0.510, \log[1/(0.210)^2])^T$, $\boldsymbol{\lambda}^2 = (\log(11/9), \log(72), \log(18), -2.030, \log[1/(0.210)^2])^T$, and $\boldsymbol{\lambda}^3 = (\log(14/11), \log(80),$

$\log(20), -0.203, \log[1/(0.202)^2])^T$.

Case 3: $\boldsymbol{\lambda}^1 = (\log(2/3), \log(65), \log(35), -0.110, \log[1/(0.210)^2])^T$, $\boldsymbol{\lambda}^2 = (\log(1), \log(70), \log(30), -2.005, \log[1/(0.253)^2])^T$, and $\boldsymbol{\lambda}^3 = (\log(3/2), \log(60), \log(40), -0.203, \log[1/(0.202)^2])^T$.

Case 4: $\boldsymbol{\lambda}^1 = (\log(4), \log(70), \log(20), -1.510, \log[1/(0.210)^2])^T$, $\boldsymbol{\lambda}^2 = (\log(3), \log(88), \log(22), -1.203, \log[1/(0.220)^2])^T$, and $\boldsymbol{\lambda}^3 = (\log(83/17), \log(80), \log(20), -1.203, \log[1/(0.041)^2])^T$.

Table 5: The values of $P_{out,\boldsymbol{\lambda}^0,F_{\boldsymbol{\lambda}^i}}$'s for Cases 1-4, where $k = 1$, $T = 300$, $n_1 = \ldots = n_T = n_t = 300$, and $i \in \{1, 2, 3\}$.

	Case 1	Case 2	Case 3	Case 4
$P_{out,\boldsymbol{\lambda}^0,F_{\boldsymbol{\lambda}^1}}$	0.0568	0.024	0.0487	0.0434
$P_{out,\boldsymbol{\lambda}^0,F_{\boldsymbol{\lambda}^2}}$	0.0153	0.0652	0.0929	0.0492
$P_{out,\boldsymbol{\lambda}^0,F_{\boldsymbol{\lambda}^3}}$	0.0169	0.0140	0.0135	0.0590

Table 6: The values of $ARL_{1,\boldsymbol{\lambda}^0,F_{\boldsymbol{\lambda}^i}}$'s for Cases 1-4, where $k = 1$, $T = 300$, $n_1 = \ldots = n_T = n_t = 300$, and $i \in \{1, 2, 3\}$.

	Case 1	Case 2	Case 3	Case 4
$ARL_{1,\boldsymbol{\lambda}^0,F_{\boldsymbol{\lambda}^1}}$	17.6	41.2	20.5	23.1
$ARL_{1,\boldsymbol{\lambda}^0,F_{\boldsymbol{\lambda}^2}}$	65.3	15.3	10.8	20.3
$ARL_{1,\boldsymbol{\lambda}^0,F_{\boldsymbol{\lambda}^3}}$	59.1	71.6	74.2	16.9

8.2 AN EMPIRICAL BAYES APPROACH

In this subsection, consider the case where $F = F_{\boldsymbol{\lambda}^0} \in \{F_{\boldsymbol{\lambda}}\colon \boldsymbol{\lambda} \in \Lambda\}$ for some unknown $\boldsymbol{\lambda}^0 \in \Lambda$. To monitor the manufacturing process at time t, the monitoring scheme proposed in Section 7.2 is used for the null hypothesis H_0: $F_{\boldsymbol{\theta}_t} = F_{\boldsymbol{\lambda}^0}$ versus the alternative H_1: $F_{\boldsymbol{\theta}_t} \neq F_{\boldsymbol{\lambda}^0}$.

Set $ARL_0 \equiv ARL_{0,\boldsymbol{\lambda}^0,\hat{\boldsymbol{\lambda}}}$, $P_{in} \equiv P_{in,\boldsymbol{\lambda}^0,\hat{\boldsymbol{\lambda}}}$, $ARL_1 \equiv ARL_{1,\boldsymbol{\lambda}^0,\hat{\boldsymbol{\lambda}},F_{\boldsymbol{\theta}_t}}$, and $P_{out} \equiv P_{out,\boldsymbol{\lambda}^0,\hat{\boldsymbol{\lambda}},F_{\boldsymbol{\theta}_t}}$. When $\hat{\boldsymbol{\lambda}} = \boldsymbol{\lambda}^0$, which is nearly impossible, we have $P_{in,\boldsymbol{\lambda}^0,\hat{\boldsymbol{\lambda}}} = P_{in,\boldsymbol{\lambda}^0}$ and $P_{out,\boldsymbol{\lambda}^0,\hat{\boldsymbol{\lambda}},F_{\boldsymbol{\theta}_t}} = P_{out,\boldsymbol{\lambda}^0,F_{\boldsymbol{\theta}_t}}$, where both $P_{in,\boldsymbol{\lambda}^0}$ and $P_{out,\boldsymbol{\lambda}^0,F_{\boldsymbol{\theta}_t}}$ are defined in Section 8.1. When $\hat{\boldsymbol{\lambda}} \neq \boldsymbol{\lambda}^0$, we have

$$\begin{aligned} P_{in,\boldsymbol{\lambda}^0,\hat{\boldsymbol{\lambda}}} &= P\left(\left\{W_{t,\hat{\boldsymbol{\lambda}}}(\mathbf{y}_t) > RUCL_{\hat{\boldsymbol{\lambda}}}\right\}; F_{\mathbf{y}_t;\boldsymbol{\lambda}^0}\right) \\ &\quad + P_{in,\hat{\boldsymbol{\lambda}},RUCL} \cdot P\left(\left\{W_{t,\hat{\boldsymbol{\lambda}}}(\mathbf{y}_t) = RUCL_{\hat{\boldsymbol{\lambda}}}\right\}; F_{\mathbf{y}_t;\boldsymbol{\lambda}^0}\right) \end{aligned} \tag{57}$$

and

$$\begin{aligned} P_{out,\boldsymbol{\lambda}^0,\hat{\boldsymbol{\lambda}},F_{\boldsymbol{\theta}_t}} &= P\left(\left\{W_{t,\hat{\boldsymbol{\lambda}}}(\mathbf{y}_t) > RUCL_{\hat{\boldsymbol{\lambda}}}\right\}; F_{\mathbf{y}_t}\right) \\ &\quad + P_{in,\hat{\boldsymbol{\lambda}},RUCL} \cdot P\left(\left\{W_{t,\hat{\boldsymbol{\lambda}}}(\mathbf{y}_t) = RUCL_{\hat{\boldsymbol{\lambda}}}\right\}; F_{\mathbf{y}_t}\right), \end{aligned} \tag{58}$$

where all of $W_{t,\hat{\boldsymbol{\lambda}}}(\mathbf{y}_t)$, $RUCL_{\hat{\boldsymbol{\lambda}}}$, and $P_{in,\hat{\boldsymbol{\lambda}},RUCL}$ are defined in Section 7.2 and both $F_{\mathbf{y}_t;\boldsymbol{\lambda}^0}$ and $F_{\mathbf{y}_t}$ are defined in Section 3.

To see how the additional estimation error resulting from the empirical Bayes approach affects the performance of the average run length, the Kullback-Leibler divergence $d(F_{\boldsymbol{\lambda}^0}, F_{\hat{\boldsymbol{\lambda}}})$ between $F_{\boldsymbol{\lambda}^0}$ and $F_{\hat{\boldsymbol{\lambda}}}$ defined in Section 7.2 can be used as a measure of how close $F_{\hat{\boldsymbol{\lambda}}}$ is to $F_{\boldsymbol{\lambda}^0}$. See Section 7.2 for details.

The corresponding values of $P_{in,\boldsymbol{\lambda}^0,\hat{\boldsymbol{\lambda}}(\mathbf{y}_{([R\tau])})}$'s and $ARL_{0,\boldsymbol{\lambda}^0,\hat{\boldsymbol{\lambda}}(\mathbf{y}_{([R\tau])})}$'s for Cases

1-4 in Section 4 are shown in Table 7, where $k = 1$, $T = 300$, $n_1 = \ldots = n_T = n_t = 300$, $R = R_1 = 50\,000$, and $\tau \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$.

Table 7: The values of $P_{in,\boldsymbol{\lambda}^0,\hat{\boldsymbol{\lambda}}(\mathbf{y}_{([R_2\,\tau])})}$'s and $ARL_{0,\boldsymbol{\lambda}^0,\hat{\boldsymbol{\lambda}}(\mathbf{y}_{([R\tau])})}$'s for Cases 1-4, where $k = 1$, $T = 300$, $n_1 = \ldots = n_T = n_t = 300$, $R = R_1 = 50\,000$, and $\tau \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$.

Case 1	$\tau = 0.1$	$\tau = 0.3$	$\tau = 0.5$	$\tau = 0.7$	$\tau = 0.9$
$P_{in,\boldsymbol{\lambda}^0,\hat{\boldsymbol{\lambda}}(\mathbf{y}_{([R\,\tau])})}$	0.00169	0.00334	0.00123	0.00374	0.00132
$ARL_{0,\boldsymbol{\lambda}^0,\hat{\boldsymbol{\lambda}}(\mathbf{y}_{([R\,\tau])})}$	590.7	299.8	812.9	267.5	759.5

Case 2	$\tau = 0.1$	$\tau = 0.3$	$\tau = 0.5$	$\tau = 0.7$	$\tau = 0.9$
$P_{in,\boldsymbol{\lambda}^0,\hat{\boldsymbol{\lambda}}(\mathbf{y}_{([R\,\tau])})}$	0.00281	0.00320	0.00204	0.00332	0.00372
$ARL_{0,\boldsymbol{\lambda}^0,\hat{\boldsymbol{\lambda}}(\mathbf{y}_{([R\,\tau])})}$	356	313	490	301	269

Case 3	$\tau = 0.1$	$\tau = 0.3$	$\tau = 0.5$	$\tau = 0.7$	$\tau = 0.9$
$P_{in,\boldsymbol{\lambda}^0,\hat{\boldsymbol{\lambda}}(\mathbf{y}_{([R\,\tau])})}$	0.00324	0.00202	0.00373	0.00247	0.00146
$ARL_{0,\boldsymbol{\lambda}^0,\hat{\boldsymbol{\lambda}}(\mathbf{y}_{([R\,\tau])})}$	309	495	268	406	685

Case 4	$\tau = 0.1$	$\tau = 0.3$	$\tau = 0.5$	$\tau = 0.7$	$\tau = 0.9$
$P_{in,\boldsymbol{\lambda}^0,\hat{\boldsymbol{\lambda}}(\mathbf{y}_{([R\,\tau])})}$	0.00282	0.00445	0.00162	0.00279	0.00569
$ARL_{0,\boldsymbol{\lambda}^0,\hat{\boldsymbol{\lambda}}(\mathbf{y}_{([R\,\tau])})}$	355.1	225.0	615.6	358.9	175.7

The corresponding values of $P_{out,\boldsymbol{\lambda}^0,\hat{\boldsymbol{\lambda}}(\mathbf{y}_{([R\,\tau])}),F_{\boldsymbol{\lambda}^i}}$'s and $ARL_{1,\boldsymbol{\lambda}^0,\hat{\boldsymbol{\lambda}}(\mathbf{y}_{([R\tau])}),F_{\boldsymbol{\lambda}^i}}$'s for Cases 1-4 in Section 4 are shown in Tables 8-10, where $k = 1$, $T = 300$, $n_1 = \ldots = n_T = n_t = 300$, $R = R_1 = 50\,000$, $\tau \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$, and $i \in \{1, 2, 3\}$.

Table 8: The values of $P_{out,\boldsymbol{\lambda}^0,\hat{\boldsymbol{\lambda}}(\mathbf{y}_{([R\,\tau])}),F_{\boldsymbol{\lambda}^1}}$'s and $ARL_{1,\boldsymbol{\lambda}^0,\hat{\boldsymbol{\lambda}}(\mathbf{y}_{([R\,\tau])}),F_{\boldsymbol{\lambda}^1}}$'s for Cases

1-4, where $k = 1$, $T = 300$, $n_1 = \ldots = n_T = n_t = 300$, $R = R_1 = 50\,000$, and $\tau \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$.

Case 1	$\tau = 0.1$	$\tau = 0.3$	$\tau = 0.5$	$\tau = 0.7$	$\tau = 0.9$
$P_{out,\boldsymbol{\lambda}^0,\hat{\boldsymbol{\lambda}}(\mathbf{y}_{([R\tau])}),F_{\boldsymbol{\lambda}^1}}$	0.0409	0.0763	0.0192	0.0999	0.0318
$ARL_{1,\boldsymbol{\lambda}^0,\hat{\boldsymbol{\lambda}}(\mathbf{y}_{([R\tau])}),F_{\boldsymbol{\lambda}^1}}$	24.4	13.1	52.0	10.0	31.4

Case 2	$\tau = 0.1$	$\tau = 0.3$	$\tau = 0.5$	$\tau = 0.7$	$\tau = 0.9$
$P_{out,\boldsymbol{\lambda}^0,\hat{\boldsymbol{\lambda}}(\mathbf{y}_{([R\tau])}),F_{\boldsymbol{\lambda}^1}}$	0.0296	0.0288	0.0130	0.0199	0.0401
$ARL_{1,\boldsymbol{\lambda}^0,\hat{\boldsymbol{\lambda}}(\mathbf{y}_{([R\tau])}),F_{\boldsymbol{\lambda}^1}}$	33.7	34.8	77.0	50.2	25.0

Case 3	$\tau = 0.1$	$\tau = 0.3$	$\tau = 0.5$	$\tau = 0.7$	$\tau = 0.9$
$P_{out,\boldsymbol{\lambda}^0,\hat{\boldsymbol{\lambda}}(\mathbf{y}_{([R\tau])}),F_{\boldsymbol{\lambda}^1}}$	0.0647	0.0327	0.0365	0.0606	0.0451
$ARL_{1,\boldsymbol{\lambda}^0,\hat{\boldsymbol{\lambda}}(\mathbf{y}_{([R\tau])}),F_{\boldsymbol{\lambda}^1}}$	15.5	30.5	27.4	16.5	22.2

Case 4	$\tau = 0.1$	$\tau = 0.3$	$\tau = 0.5$	$\tau = 0.7$	$\tau = 0.9$
$P_{out,\boldsymbol{\lambda}^0,\hat{\boldsymbol{\lambda}}(\mathbf{y}_{([R\tau])}),F_{\boldsymbol{\lambda}^1}}$	0.0463	0.0520	0.0190	0.0520	0.0268
$ARL_{1,\boldsymbol{\lambda}^0,\hat{\boldsymbol{\lambda}}(\mathbf{y}_{([R\tau])}),F_{\boldsymbol{\lambda}^1}}$	21.6	19.2	52.6	19.2	37.3

Table 9: The values of $P_{out,\boldsymbol{\lambda}^0,\hat{\boldsymbol{\lambda}}(\mathbf{y}_{([R\tau])}),F_{\boldsymbol{\lambda}^2}}$'s and $ARL_{1,\boldsymbol{\lambda}^0,\hat{\boldsymbol{\lambda}}(\mathbf{y}_{([R\tau])}),F_{\boldsymbol{\lambda}^2}}$'s for Cases 1-4, where $k = 1$, $T = 300$, $n_1 = \ldots = n_T = n_t = 300$, $R = R_1 = 50\,000$, and $\tau \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$.

Case 1	$\tau = 0.1$	$\tau = 0.3$	$\tau = 0.5$	$\tau = 0.7$	$\tau = 0.9$
$P_{out,\boldsymbol{\lambda}^0,\hat{\boldsymbol{\lambda}}(\mathbf{y}_{([R\tau])}),F_{\boldsymbol{\lambda}^2}}$	0.0126	0.0184	0.00802	0.0218	0.0108
$ARL_{1,\boldsymbol{\lambda}^0,\hat{\boldsymbol{\lambda}}(\mathbf{y}_{([R\tau])}),F_{\boldsymbol{\lambda}^2}}$	79.5	54.4	125	46.0	92.8

Case 2	$\tau = 0.1$	$\tau = 0.3$	$\tau = 0.5$	$\tau = 0.7$	$\tau = 0.9$
$P_{out,\boldsymbol{\lambda}^0,\hat{\boldsymbol{\lambda}}(\mathbf{y}_{([R\tau])}),F_{\boldsymbol{\lambda}^2}}$	0.0808	0.0778	0.0323	0.0515	0.110
$ARL_{1,\boldsymbol{\lambda}^0,\hat{\boldsymbol{\lambda}}(\mathbf{y}_{([R\tau])}),F_{\boldsymbol{\lambda}^2}}$	12.4	12.9	31.0	19.4	9.11

Case 3	$\tau = 0.1$	$\tau = 0.3$	$\tau = 0.5$	$\tau = 0.7$	$\tau = 0.9$
$P_{out,\boldsymbol{\lambda}^0,\hat{\boldsymbol{\lambda}}(\mathbf{y}_{([R\tau])}),F_{\boldsymbol{\lambda}^2}}$	0.0776	0.0938	0.136	0.0497	0.0371
$ARL_{1,\boldsymbol{\lambda}^0,\hat{\boldsymbol{\lambda}}(\mathbf{y}_{([R\tau])}),F_{\boldsymbol{\lambda}^2}}$	12.9	10.7	7.36	20.1	26.9

Case 4	$\tau = 0.1$	$\tau = 0.3$	$\tau = 0.5$	$\tau = 0.7$	$\tau = 0.9$
$P_{out,\boldsymbol{\lambda}^0,\hat{\boldsymbol{\lambda}}(\mathbf{y}_{([R\tau])}),F_{\boldsymbol{\lambda}^2}}$	0.0525	0.0586	0.0223	0.0586	0.0310
$ARL_{1,\boldsymbol{\lambda}^0,\hat{\boldsymbol{\lambda}}(\mathbf{y}_{([R\tau])}),F_{\boldsymbol{\lambda}^2}}$	19.1	17.1	44.8	17.1	32.2

Table 10: The values of $P_{out,\boldsymbol{\lambda}^0,\hat{\boldsymbol{\lambda}}(\mathbf{y}_{([R\tau])}),F_{\boldsymbol{\lambda}^3}}$'s and $ARL_{1,\boldsymbol{\lambda}^0,\hat{\boldsymbol{\lambda}}(\mathbf{y}_{([R\tau])}),F_{\boldsymbol{\lambda}^3}}$'s for Cases 1-4, where $k = 1$, $T = 300$, $n_1 = \ldots = n_T = n_t = 300$, $R = R_1 = 50\,000$, and $\tau \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$.

Case 1	$\tau = 0.1$	$\tau = 0.3$	$\tau = 0.5$	$\tau = 0.7$	$\tau = 0.9$
$P_{out,\boldsymbol{\lambda}^0,\hat{\boldsymbol{\lambda}}(\mathbf{y}_{([R\tau])}),F_{\boldsymbol{\lambda}^3}}$	0.0108	0.0181	0.0113	0.0157	0.00902
$ARL_{1,\boldsymbol{\lambda}^0,\hat{\boldsymbol{\lambda}}(\mathbf{y}_{([R\tau])}),F_{\boldsymbol{\lambda}^3}}$	92.6	55.4	88.9	63.5	110.9

Case 2	$\tau = 0.1$	$\tau = 0.3$	$\tau = 0.5$	$\tau = 0.7$	$\tau = 0.9$
$P_{out,\boldsymbol{\lambda}^0,\hat{\boldsymbol{\lambda}}(\mathbf{y}_{([R\tau])}),F_{\boldsymbol{\lambda}^3}}$	0.0121	0.0152	0.0139	0.0195	0.0120
$ARL_{1,\boldsymbol{\lambda}^0,\hat{\boldsymbol{\lambda}}(\mathbf{y}_{([R\tau])}),F_{\boldsymbol{\lambda}^3}}$	82.8	65.9	72.1	51.3	83.2

Case 3	$\tau = 0.1$	$\tau = 0.3$	$\tau = 0.5$	$\tau = 0.7$	$\tau = 0.9$
$P_{out,\boldsymbol{\lambda}^0,\hat{\boldsymbol{\lambda}}(\mathbf{y}_{([R\,\tau])}),F_{\boldsymbol{\lambda}^3}}$	0.0199	0.00786	0.00911	0.0182	0.0121
$ARL_{1,\boldsymbol{\lambda}^0,\hat{\boldsymbol{\lambda}}(\mathbf{y}_{([R\tau])}),F_{\boldsymbol{\lambda}^3}}$	50.3	127	109	55.1	82.6

Case 4	$\tau = 0.1$	$\tau = 0.3$	$\tau = 0.5$	$\tau = 0.7$	$\tau = 0.9$
$P_{out,\boldsymbol{\lambda}^0,\hat{\boldsymbol{\lambda}}(\mathbf{y}_{([R\,\tau])}),F_{\boldsymbol{\lambda}^3}}$	0.0627	0.0695	0.0282	0.0695	0.0383
$ARL_{1,\boldsymbol{\lambda}^0,\hat{\boldsymbol{\lambda}}(\mathbf{y}_{([R\tau])}),F_{\boldsymbol{\lambda}^3}}$	16.0	14.4	35.5	14.4	26.1

From Tables 8-10, there is no pattern that the values of $P_{in,\boldsymbol{\lambda}^0,\hat{\boldsymbol{\lambda}}}$ and $P_{out,\boldsymbol{\lambda}^0,\hat{\boldsymbol{\lambda}},F_{\boldsymbol{\lambda}^i}}$ for Case 1-Case 4 increase as the Kullback-Leibler divergence $d(F_{\boldsymbol{\lambda}^0}, F_{\hat{\boldsymbol{\lambda}}})$ increases, where $k = 1$, $T = 300$, $n_1 = \ldots = n_T = n_t = 300$, $R = R_1 = 50\,000$, and $i \in \{1, 2, 3\}$.

9 CONCLUSIONS

In the paper, first, a two-components mixture prior parametric family for the in-control prior distribution is proposed in a manufacturing process. Then an empirical Bayes approach is proposed when there are available in-control categorical data generated from the manufacturing process. As an illustration, an example of the proposed empirical Bayes model is introduced. For the purpose of model building, the goodness of fit and the simplification of the proposed model are discussed. Utilizing the likelihood ratio method, both Bayesian and empirical Bayes monitoring techniques are proposed as the main purpose of the paper. Finally, the performance of the proposed process monitoring scheme is studied in terms of the average run length to show the robustness of the methodology.

APPENDIX

All of nodes and weights of the Hermite polynomial of 32 degrees are shown in the following table. This table is obtained from the following website:

http://www.efunda.com/math/num_integration/findgausshermite.cfm

No. i	abscissas x_i	weights w_i
1	-7.12581390983	$7.31067642754 \times 10^{-23}$
2	-6.40949814928	$9.23173653482 \times 10^{-19}$
3	-5.81222594946	$1.19734401957 \times 10^{-15}$
4	-5.27555098664	$4.21501019491 \times 10^{-13}$
5	-4.77716450334	$5.93329148347 \times 10^{-11}$
6	-4.30554795347	$4.09883215841 \times 10^{-9}$
7	-3.85375548542	$1.57416779440 \times 10^{-7}$
8	-3.41716749282	$3.65058512533 \times 10^{-6}$
9	-2.99249082501	$5.41658405999 \times 10^{-5}$
10	-2.57724953773	$5.36268365495 \times 10^{-4}$
11	-2.16949918361	$3.65489032677 \times 10^{-3}$
12	-1.76765410946	$1.75534288315 \times 10^{-2}$
13	-1.37037641095	$6.04581309559 \times 10^{-2}$
14	-0.97650046359	$1.51269734077 \times 10^{-1}$
15	-0.58497876544	$2.77458142303 \times 10^{-1}$
16	-0.19484074157	$3.75238352593 \times 10^{-1}$

No. i	abscissas x_i	weights w_i
17	0.19484074157	$3.75238352593 \times 10^{-1}$
18	0.58497876544	$2.77458142303 \times 10^{-1}$
19	0.97650046359	$1.51269734077 \times 10^{-1}$
20	1.37037641095	$6.04581309559 \times 10^{-2}$
21	1.76765410946	$1.75534288315 \times 10^{-2}$
22	2.16949918361	$3.65489032677 \times 10^{-3}$
23	2.57724953773	$5.36268365495 \times 10^{-4}$
24	2.99249082501	$5.41658405999 \times 10^{-5}$
25	3.41716749282	$3.65058512533 \times 10^{-6}$
26	3.85375548542	$1.57416779440 \times 10^{-7}$
27	4.30554795347	$4.09883215841 \times 10^{-9}$
28	4.77716450334	$5.93329148347 \times 10^{-11}$
29	5.27555098664	$4.21501019491 \times 10^{-13}$
30	5.81222594946	$1.19734401957 \times 10^{-15}$
31	6.40949814928	$9.23173653482 \times 10^{-19}$
32	7.12581390983	$7.31067642754 \times 10^{-23}$

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. and Liu, C.-Y. (2005). A Model Selection Technique Between Two Empirical Bayes Models for Categorical Data. Technical Report, Institute of Statistics, National Chiao Tung University, Hsinchu, Taiwan.

4. Chen, C.-R., Shiau, J.-J. H., Liao, H.-H., and Feltz, C. J. (2004). A process monitoring technique for categorical data under the beta-binomial/Dirichlet-multinomial empirical Bayes model. Technical Report, Institute of Statistics, National Chiao Tung University, Hsinchu, Taiwan.

5. Chen, C.-R., Shiau, J.-J. H., Lin, T.-Y., and Feltz, C. J. (2005). A process monitoring technique for categorical data under the transformed-normal-binomial/multinomial empirical Bayes model. Technical Report, Institute of Statistics, National Chiao Tung University, Hsinchu, Taiwan.

6. Fahrmeir, A. and Tutz, G. (2001). *Multivariate Statistical Modelling Based on Generalized Linear Models*, 2nd ed. Springer-Verlag, New York.

7. Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2004). *Bayesian Data Analysis*, 2nd ed. Chapman & Hall/CRC, Boca Raton.

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

9. McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*, 2nd ed. Chapman and Hall, London.

10. McLachlan, G. J. and Peel, D. (2000). Finite Mixture Models. John Wiley & Sons, New York.

11. O'Hagan, A. and Forster, J. (2004). *Kendall's Advanced Theory of Statistics, Volume 2B: Bayesian Inference*, 2nd ed. Arnold, London.

12. Shiau, J.-J. H., Chen, C.-R., and Feltz, C. J. (2005). An empirical Bayes process monitoring technique for polytomous data. *Quality and Reliability Engineering International*, **21**, 13-28.

13. 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.