QUALITY AND RELIABILITY ENGINEERING INTERNATIONAL
Qual. Reliab. Engng. Int. 2005; **21**:13–28
Published online 15 November 2004 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/qre.604

Research

An Empirical Bayes Process Monitoring Technique for Polytomous Data

Jyh-Jen H. Shiau[1], Chih-Rung Chen[1] and Carol J. Feltz[2,*,†]
[1]*Institute of Statistics, National Chiao Tung University, Hsinchu, Taiwan 300*
[2]*Division of Statistics, Northern Illinois University, DeKalb, IL 60115, U.S.A.*

When a product item is tested, usually one has more information than just pass or fail. Often there are categories of failure modes. The purpose of this paper is to develop a method to monitor the fractions of the tested items falling into different categories of pass/fail modes. Using the multinomial model with Dirichlet prior, we describe the theory underlying an empirical Bayes approach to monitoring polytomous data generated in manufacturing processes. A pseudo maximum likelihood estimator (PMLE) and the method-of-moments estimator (MME) of the hyperparameters of the prior distribution are considered and compared by a simulation study. It is found that the PMLE performs slightly better than the MME. A monitoring scheme based on the marginal distributions of the observed pass/fail fractions is proposed. The average run length behavior of the proposed monitoring scheme is investigated. Finally, an example to illustrate the use of the technique is given. Copyright © 2004 John Wiley & Sons, Ltd.

KEY WORDS: empirical Bayes; multinomial; Dirichlet; polytomous; attribute data; control charts; quality control

1. INTRODUCTION

When a product item is tested, usually one has more information than just pass or fail. Often there are categories of failures. For instance, a product may have several categories of failure modes. Another example might be that the measurements of the tested items are recorded as fail low, pass, or fail high. When the measurements are recorded only as pass or fail, we say that the measurements are binary. When the measurements are recorded as $k + 1$ possible values for some known $k \in \{2, 3, \ldots\}$, we say that the measurements are polytomous. (For a good reference for binary or polytomous data, see McCullagh and Nelder[1], Chapters 4 and 5.) Although many authors have investigated methods for monitoring the non-conforming fraction for binary data (see Chapter 6 of Montgomery[2] for a good literature review), few have developed methods for polytomous data. One set of authors, Voss *et al.*[3], have investigated the multinomial model to monitor the quality of bacterial colony counting procedures for polytomous data.

*Correspondence to: Carol J. Feltz, Division of Statistics, Northern Illinois University, DeKalb, IL 60115, U.S.A.
†E-mail: feltz@math.niu.edu

Contract/grant sponsor: National Science Council of the Republic of China; contract/grant number: NSC91-2118-M-009-007

Received 18 April 2003
Revised 9 June 2003

Applying a Bayesian methodology to polytomous data has been investigated by several researchers. One of the first, Lindley[4], developed the Bayesian analysis of contingency tables for polytomous data using the multinomial model with an improper prior distribution. Lindley's approach was classic Bayesian, where the prior distribution was specified in advance. In a more recent research, Walley[5] used the multinomial model with an imprecise Dirichlet prior to model polytomous data in cases where no prior information was available. He applied his method to clinical trial data to identify the efficacy of a treatment. Nair *et al.*[6] used a Bayesian approach to analyze a mixture of Poisson data in manufacturing.

In this paper, the fractions of tested products falling into different categories of pass/fail modes are modeled by the multinomial distribution with the Dirichlet prior distribution that has unknown hyperparameters. In contrast with the methods mentioned above, we utilize the previous process data to estimate the hyperparameters. This is an empirical Bayes method. We then propose a scheme to monitor the fractions of the tested product items, and identify if the process has changed over time.

The empirical Bayes approach for monitoring process data is not entirely new. Sturm *et al.*[7] developed an empirical Bayes technique to monitor a process with variable data. To monitor a process of attribute (binary) data, Yousry *et al.*[8] used a binomial model with a beta prior to monitor yield and defect data. These techniques were found to be very useful in industrial settings. The empirical Bayes approach explicitly allows the process parameters to change over time. By allowing the process to change over time, we can estimate the amount of change inherent in the process. A measure of the process change is captured in the process variation. This paper extends the work in Yousry *et al.*[8] to the monitoring of processes with polytomous data. In addition, in this paper, we investigate two different methods for estimating the hyperparameters: the method of moments and the pseudo maximum likelihood method introduced by Gong and Samaniego[9].

The paper is organized as follows. In Section 2, we describe the Bayesian model considered in this paper for polytomous data. In Section 3, an empirical Bayes approach is proposed by estimating the hyperparameters of the prior distribution. In Section 4, we develop a new monitoring scheme based on the estimated marginal distributions of the observed proportions. Due to the discreteness nature of the observed proportions, a randomized-control-limits scheme is proposed to achieve the usual desirable false alarm rate. The average run length of the proposed scheme is studied for various situations in Section 5. In Section 6, a numerical example is presented to illustrate the use and the effectiveness of the proposed empirical Bayes process monitoring scheme. In Section 7, we conclude the paper with a brief summary.

2. A BAYESIAN MODEL FOR POLYTOMOUS DATA

Consider a manufacturing process producing a product that has k different types of defects for some positive integer k. For simplicity, it is assumed that one product item cannot have more than one defect type. Let p_{it} be the probability of a product item having the ith defect type at time t for $i = 1, \dots, k$. Then the yield probability can be defined as $p_{0t} = 1 - \sum_{i=1}^{k} p_{it}$, the probability of no defects at time t.

Let the yield variable x_{0t} be the number of items that do not have any of the k defects out of n_t randomly chosen product items tested at time t. Here n_t is a positive integer. Let the ith defect-type variable x_{it} be the number of tested items that are of the ith defect type at time t for $i = 1, \dots, k$. Then $x_{0t} = n_t - \sum_{i=1}^{k} x_{it}$.

Assume that the observed random vector $\mathbf{x}_t = (x_{0t}, x_{1t}, \dots, x_{kt})'$ is distributed as Multinomial$(n_t; \mathbf{p}_t)$, where $\mathbf{p}_t = (p_{0t}, p_{1t}, \dots, p_{kt})'$ with $0 \le p_{0t}, p_{1t}, \dots, p_{kt} \le 1$ and $\sum_{i=0}^{k} p_{it} = 1$. Then the conditional probability mass function (p.m.f.) of $\mathbf{x}_t$ given $\mathbf{p}_t$ is

$$f(\mathbf{x}_t|\mathbf{p}_t) = \frac{n_t!}{x_{0t}!x_{1t}! \cdots x_{kt}!} p_{0t}^{x_{0t}} p_{1t}^{x_{1t}} \cdots p_{kt}^{x_{kt}} \tag{1}$$

for $x_{0t}, x_{1t}, \dots, x_{kt} \in \{0, 1, \dots, n_t\}$ and $\sum_{i=0}^{k} x_{it} = n_t$.

Given $\mathbf{p}_t$, the conditional *sampling* mean and variance of x_{it}/n_t and the conditional *sampling* covariance of x_{it}/n_t and x_{jt}/n_t are

$$E(x_{it}/n_t|\mathbf{p}_t) = p_{it}$$
$$\mathrm{Var}(x_{it}/n_t|\mathbf{p}_t) = p_{it}(1-p_{it})/n_t$$
$$\mathrm{Cov}(x_{it}/n_t, x_{jt}/n_t|\mathbf{p}_t) = -p_{it}p_{jt}/n_t$$

respectively, for $i \neq j$ and $i, j = 0, 1, \ldots, k$. (See, for example, p. 323 of Carlin and Louis[10].) Using concise matrix notation, the conditional *sampling* mean and variance–covariance matrix of $\mathbf{x}_t/n_t$ given $\mathbf{p}_t$ are

$$E(\mathbf{x}_t/n_t|\mathbf{p}_t) = \mathbf{p}_t$$
$$\mathrm{Cov}(\mathbf{x}_t/n_t|\mathbf{p}_t) = (\mathrm{diag}\{\mathbf{p}_t\} - \mathbf{p}_t\mathbf{p}_t')/n_t$$

where $\mathrm{diag}\{\mathbf{p}_t\} \equiv \mathrm{diag}\{p_{0t}, p_{1t}, \ldots, p_{kt}\}$.

In the Bayesian framework, the process parameters p_{it} are assumed to be random. We have found that the variability of the process parameters is a good characteristic for modeling factory processes. In the classical Bayesian approach, it is usually assumed that $\mathbf{p}_t$ has a prespecified Dirichlet($\boldsymbol{\alpha}$) prior distribution, where $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \ldots, \alpha_k)'$ is a known vector with $\alpha_0, \alpha_1, \ldots, \alpha_k > 0$. Let $\alpha_s = \sum_{i=0}^{k} \alpha_i$, the sum of all α_i. Then the prior probability density function (p.d.f.) of $\mathbf{p}_t$ is

$$f(\mathbf{p}_t; \boldsymbol{\alpha}) = \frac{\Gamma(\alpha_s)}{\Gamma(\alpha_0)\Gamma(\alpha_1)\cdots\Gamma(\alpha_k)} p_{0t}^{\alpha_0-1} p_{1t}^{\alpha_1-1} \cdots p_{kt}^{\alpha_k-1} \tag{2}$$

for $0 \leq p_{0t}, p_{1t}, \ldots, p_{kt} \leq 1$ and $\sum_{i=0}^{k} p_{it} = 1$.

Let $\boldsymbol{\alpha}^*(= (\alpha_0^*, \alpha_1^*, \ldots, \alpha_k^*)') = \boldsymbol{\alpha}/\alpha_s$. The *process* mean and variance–covariance matrix of $\mathbf{p}_t$ are

$$E(\mathbf{p}_t) = \boldsymbol{\alpha}^*$$
$$\mathrm{Cov}(\mathbf{p}_t) = (\mathrm{diag}\{\boldsymbol{\alpha}^*\} - \boldsymbol{\alpha}^*\boldsymbol{\alpha}^{*\prime})/(\alpha_s + 1)$$

respectively. (See, for example, p. 327 of Carlin and Louis[10].)

The variation of $\mathbf{p}_t$ will be referred to as the process variation while the variation of $\mathbf{x}_t/n_t$ given $\mathbf{p}_t$ will be referred to as the sampling variation.

By the double expectation method, the marginal mean of $\mathbf{x}_t/n_t$ can be obtained as follows:

$$E(\mathbf{x}_t/n_t) = E[E(\mathbf{x}_t/n_t|\mathbf{p}_t)] = E(\mathbf{p}_t) = \boldsymbol{\alpha}^* \tag{3}$$

By some simple algebra, the marginal variance–covariance matrix of $\mathbf{x}_t/n_t$ can be found to be

$$\begin{aligned}\mathrm{Cov}(\mathbf{x}_t/n_t) &= E[\mathrm{Cov}(\mathbf{x}_t/n_t|\mathbf{p}_t)] + \mathrm{Cov}[E(\mathbf{x}_t/n_t|\mathbf{p}_t)] \\ &= E(\mathrm{diag}\{\mathbf{p}_t\} - \mathbf{p}_t\mathbf{p}_t')/n_t + \mathrm{Cov}(\mathbf{p}_t) \\ &= \alpha_s(\mathrm{diag}\{\boldsymbol{\alpha}^*\} - \boldsymbol{\alpha}^*\boldsymbol{\alpha}^{*\prime})/[n_t(\alpha_s+1)] + (\mathrm{diag}\{\boldsymbol{\alpha}^*\} - \boldsymbol{\alpha}^*\boldsymbol{\alpha}^{*\prime})/(\alpha_s+1) \\ &= (\alpha_s + n_t)(\mathrm{diag}\{\boldsymbol{\alpha}^*\} - \boldsymbol{\alpha}^*\boldsymbol{\alpha}^{*\prime})/[n_t(\alpha_s+1)]\end{aligned} \tag{4}$$

Thus, it is seen that the marginal variance–covariance matrix of $\mathbf{x}_t/n_t$ is decomposed into the sum of the *sampling* variance–covariance matrix and the *process* variance–covariance matrix.

In Bayesian terminology, the Dirichlet distribution is the conjugate prior distribution for the multinomial model given in (1). Choosing a conjugate prior distribution has an advantage that the corresponding posterior distribution follows the same parametric form as the prior distribution. The Dirichlet($\boldsymbol{\alpha}$) distribution provides great flexibility for modeling $\mathbf{p}_t$ because, with different values of $\boldsymbol{\alpha}$, it has a wide range of shapes. It is well known (see, for example, p. 76 of Gelman *et al.*[11]) that the posterior distribution of $\mathbf{p}_t$ given $\mathbf{x}_t$ is the

Dirichlet($\boldsymbol{\alpha} + \mathbf{x}_t$) distribution

$$f(\mathbf{p}_t|\mathbf{x}_t, \boldsymbol{\alpha}) = \frac{\Gamma(\alpha_s + n_t)}{\Gamma(\alpha_0 + x_{0t})\Gamma(\alpha_1 + x_{1t}) \cdots \Gamma(\alpha_k + x_{kt})} p_{0t}^{\alpha_0 + x_{0t} - 1} p_{1t}^{\alpha_1 + x_{1t} - 1} \cdots p_{kt}^{\alpha_k + x_{kt} - 1} \tag{5}$$

for $0 \le p_{0t}, p_{1t}, \ldots, p_{kt} \le 1$ and $\sum_{i=0}^{k} p_{it} = 1$.

Often, the problem of interest is to estimate $\mathbf{p}_t$, the yield probability along with all defect-type probabilities at time t. In the Bayesian approach we usually estimate $\mathbf{p}_t$ by the posterior mean of $\mathbf{p}_t$ given $\mathbf{x}_t$, which is the Bayes estimator for the quadratic error loss. The posterior mean and variance–covariance matrix of $\mathbf{p}_t$ given $\mathbf{x}_t$ are

$$\begin{aligned} E(\mathbf{p}_t|\mathbf{x}_t) &= (\boldsymbol{\alpha} + \mathbf{x}_t)/(\alpha_s + n_t) \equiv \widetilde{\mathbf{p}}_t(\boldsymbol{\alpha}) \\ \mathrm{Cov}(\mathbf{p}_t|\mathbf{x}_t) &= [\mathrm{diag}\{\widetilde{\mathbf{p}}_t(\boldsymbol{\alpha})\} - \widetilde{\mathbf{p}}_t(\boldsymbol{\alpha})\widetilde{\mathbf{p}}_t'(\boldsymbol{\alpha})]/(\alpha_s + n_t + 1) \end{aligned} \tag{6}$$

respectively.

Note that the posterior mean of $\mathbf{p}_t$ given $\mathbf{x}_t$ can be rewritten as

$$\widetilde{\mathbf{p}}_t(\boldsymbol{\alpha}) = w_t\boldsymbol{\alpha}^* + (1 - w_t)\frac{\mathbf{x}_t}{n_t}$$

where $w_t = \alpha_s/(\alpha_s + n_t)$. This indicates that the posterior mean $\widetilde{\mathbf{p}}_t(\boldsymbol{\alpha})$ is a weighted average of prior mean $\boldsymbol{\alpha}^*$ and observed proportions $\mathbf{x}_t/n_t$. A large weight w_t, indicating more weight on the prior information, pulls the posterior mean towards the prior mean, while a small weight w_t, indicating more weight on data, pulls the posterior mean towards the observed proportions.

3. AN EMPIRICAL BAYES APPROACH FOR POLYTOMOUS DATA

In this section, the fractions of tested product items falling into $k + 1$ categories of pass or failure modes are modeled using the multinomial model and the Dirichlet($\boldsymbol{\alpha}$) prior distribution, where $\boldsymbol{\alpha}$ is an unknown hyperparameter vector. Using an empirical Bayes approach, we let data speak for themselves in estimating $\boldsymbol{\alpha}$. With the components of the model in place, we now focus on using the process data to estimate $\boldsymbol{\alpha}$.

First of all, we derive the marginal distribution of $\mathbf{x}_t$. The marginal p.m.f. of $\mathbf{x}_t$ can be found by

$$\begin{aligned} f(\mathbf{x}_t; \boldsymbol{\alpha}) &= \frac{f(\mathbf{x}_t, \mathbf{p}_t; \boldsymbol{\alpha})}{f(\mathbf{p}_t|\mathbf{x}_t; \boldsymbol{\alpha})} = \frac{f(\mathbf{x}_t|\mathbf{p}_t)f(\mathbf{p}_t; \boldsymbol{\alpha})}{f(\mathbf{p}_t|\mathbf{x}_t; \boldsymbol{\alpha})} \\ &= \frac{n_t!}{x_{0t}!x_{1t}! \cdots x_{kt}!} \frac{\Gamma(\alpha_s)}{\Gamma(\alpha_0)\Gamma(\alpha_1) \cdots \Gamma(\alpha_k)} \frac{\Gamma(\alpha_0 + x_{0t})\Gamma(\alpha_1 + x_{1t}) \cdots \Gamma(\alpha_k + x_{kt})}{\Gamma(\alpha_s + n_t)} \\ &= \frac{n_t!}{\prod_{j=1}^{n_t}(\alpha_s + j - 1)} \prod_{i=0}^{k} \frac{\prod_{j=1}^{x_{it}}(\alpha_i + j - 1)}{x_{it}!} \\ &= \exp\left[\sum_{j=1}^{n_t} \log\left(\frac{j}{\alpha_s + j - 1}\right) - \sum_{i=0}^{k}\sum_{j=1}^{x_{it}} \log\left(\frac{j}{\alpha_i + j - 1}\right)\right] \end{aligned} \tag{7}$$

for $x_{0t}, x_{1t}, \ldots, x_{kt} \in \{0, 1, \ldots, n_t\}$ and $\sum_{i=0}^{k} x_{it} = n_t$.

This shows that the marginal distribution of $\mathbf{x}_t$ is the multivariate Pólya-Eggenberger distribution or the Dirichlet-compound multinomial distribution (see, for example, p. 80 of Johnson *et al.*[12]) with parameters n_t and $\boldsymbol{\alpha}$. In particular, the marginal distribution of x_{it} is the Pólya distribution with parameters n_t, α_i, and α_s for each $i = 0, 1, \ldots, k$.

Let $\mathbf{x}_1, \ldots, \mathbf{x}_T$ be independent observations such that $\mathbf{x}_i$ has marginal p.m.f. (7) for $i = 1, \ldots, T$. Note that, by (3), $\mathbf{x}_t/n_t$ is an unbiased estimator of $\boldsymbol{\alpha}^*$ for $t = 1, \ldots, T$. Thus it is natural to estimate $\boldsymbol{\alpha}^*$ by the weighted average of all $\mathbf{x}_t/n_t$ with weights $n_t/\sum_{t'=1}^{T} n_{t'}$. Then we can estimate $\boldsymbol{\alpha}^*$ by

$$\widehat{\boldsymbol{\alpha}}^* = \frac{\sum_{t=1}^{T} \mathbf{x}_t}{\sum_{t=1}^{T} n_t} \tag{8}$$

By (3) and (4),

$$E\left(\widehat{\boldsymbol{\alpha}}^*\right) = \boldsymbol{\alpha}^* \tag{9}$$

$$\mathrm{Cov}(\widehat{\boldsymbol{\alpha}}^*) = \frac{\sum_{t=1}^{T} n_t(\alpha_s + n_t)}{(\alpha_s + 1)(\sum_{t=1}^{T} n_t)^2}(\mathrm{diag}\{\boldsymbol{\alpha}^*\} - \boldsymbol{\alpha}^*\boldsymbol{\alpha}^{*\prime}) \tag{10}$$

Under regularity conditions, $\widehat{\boldsymbol{\alpha}}^*$ is a strongly consistent estimator of $\boldsymbol{\alpha}^*$ (i.e. $\widehat{\boldsymbol{\alpha}}^*$ converges to $\boldsymbol{\alpha}^*$ almost surely) and

$$\frac{(\alpha_s + 1)^{1/2} \sum_{t=1}^{T} n_t}{[\sum_{t=1}^{T} n_t(\alpha_s + n_t)]^{1/2}} \left(\widehat{\boldsymbol{\alpha}}^* - \boldsymbol{\alpha}^*\right) \xrightarrow{d} N\left(\mathbf{0}, \mathrm{diag}\{\boldsymbol{\alpha}^*\} - \boldsymbol{\alpha}^*\boldsymbol{\alpha}^{*\prime}\right) \tag{11}$$

as $T \to \infty$, where $\xrightarrow{d}$ means convergence in distribution and $N(\mathbf{0}, \mathrm{diag}\{\boldsymbol{\alpha}^*\} - \boldsymbol{\alpha}^*\boldsymbol{\alpha}^{*\prime})$ denotes the multivariate normal distribution with mean $\mathbf{0}$ and variance–covariance matrix $\mathrm{diag}\{\boldsymbol{\alpha}^*\} - \boldsymbol{\alpha}^*\boldsymbol{\alpha}^{*\prime}$.

With (8), if we have an estimate $\widehat{\alpha}_s$ of α_s, then we have an estimate $\widehat{\alpha}_s\widehat{\boldsymbol{\alpha}}^*$ of $\boldsymbol{\alpha}$. By treating α_s as the only hyperparameter to be estimated in the model, we consider the following two methods for estimating α_s: (i) the method of moments and (ii) the pseudo maximum likelihood method.

We first describe the method of moments. By (4), the marginal variance of x_{it}/n_t can be rewritten as

$$\sigma_{it}^2 \equiv \mathrm{Var}(x_{it}/n_t) = \frac{\alpha_i^*(1-\alpha_i^*)}{n_t}\frac{\alpha_s + n_t}{\alpha_s + 1}$$

for $i = 0, 1, \ldots, k$ and $t = 1, \ldots, T$, which implies that

$$(\alpha_s + 1)\sum_{t=1}^{T} n_t \sum_{i=0}^{k} \sigma_{it}^2 = \left(T\alpha_s + \sum_{t=1}^{T} n_t\right)\sum_{i=0}^{k} \alpha_i^*(1-\alpha_i^*) \tag{12}$$

Solving (12) for α_s, we have

$$\alpha_s = \frac{\sum_{t=1}^{T} n_t \sum_{i=0}^{k} \alpha_i^*(1-\alpha_i^*) - \sum_{t=1}^{T} n_t \sum_{i=0}^{k} \sigma_{it}^2}{\sum_{t=1}^{T} n_t \sum_{i=0}^{k} \sigma_{it}^2 - T\sum_{i=0}^{k} \alpha_i^*(1-\alpha_i^*)} \tag{13}$$

By (3) and (9), $\sum_{t=1}^{T} n_t \sum_{i=0}^{k} \sigma_{it}^2$ can be estimated by

$$\sum_{t=1}^{T} n_t \sum_{i=0}^{k}(x_{it}/n_t - \widehat{\alpha}_i^*)^2$$

Thus, we can estimate α_s by the following method-of-moments estimator (MME):

$$\widehat{\alpha}_{s,\mathrm{MM}} = \frac{\sum_{t=1}^{T} n_t \sum_{i=0}^{k} \widehat{\alpha}_i^*(1-\widehat{\alpha}_i^*) - \sum_{t=1}^{T} n_t \sum_{i=0}^{k}(x_{it}/n_t - \widehat{\alpha}_i^*)^2}{\sum_{t=1}^{T} n_t \sum_{i=0}^{k}(x_{it}/n_t - \widehat{\alpha}_i^*)^2 - T\sum_{i=0}^{k} \widehat{\alpha}_i^*(1-\widehat{\alpha}_i^*)} \tag{14}$$

Under regularity conditions, $\widehat{\alpha}_{s,\mathrm{MM}}$ is a weakly consistent estimator of α_s (i.e. $\widehat{\alpha}_{s,\mathrm{MM}}$ converges to α_s in probability) and is asymptotically normally distributed as $T \to \infty$.

The second method for estimating α_s to be described is the pseudo maximum likelihood method introduced by Gong and Samaniego[9]. Again, by plugging $\widehat{\boldsymbol{\alpha}}^*$ into $\boldsymbol{\alpha}^*$, α_s is the only hyperparameter under consideration. By (7), the pseudo-likelihood function for α_s can be defined as

$$L_{\mathrm{P}}(\alpha_s; \mathbf{x}_1, \ldots, \mathbf{x}_T) \equiv \prod_{t=1}^{T} f(\mathbf{x}_t; \alpha)|_{\boldsymbol{\alpha}^*=\widehat{\boldsymbol{\alpha}}^*}$$
$$= \exp\left\{\sum_{t=1}^{T}\left[\sum_{j=1}^{n_t}\log\left(\frac{j}{\alpha_s+j-1}\right) - \sum_{i=0}^{k}\sum_{j=1}^{x_{it}}\log\left(\frac{j}{\widehat{\alpha}_i^*\alpha_s+j-1}\right)\right]\right\} \tag{15}$$

Then the pseudo-log-likelihood function for α_s is

$$\ell_{\mathrm{P}}(\alpha_s; \mathbf{x}_1, \ldots, \mathbf{x}_T) \equiv \log[L_{\mathrm{P}}(\alpha_s; \mathbf{x}_1, \ldots, \mathbf{x}_T)]$$
$$= \sum_{t=1}^{T}\left[\sum_{j=1}^{n_t}\log\left(\frac{j}{\alpha_s+j-1}\right) - \sum_{i=0}^{k}\sum_{j=1}^{x_{it}}\log\left(\frac{j}{\widehat{\alpha}_i^*\alpha_s+j-1}\right)\right]$$

the pseudo-score function for α_s is

$$s_{\mathrm{P}}(\alpha_s; \mathbf{x}_1, \ldots, \mathbf{x}_T) \equiv \frac{\partial \ell_{\mathrm{P}}(\alpha_s; \mathbf{x}_1, \ldots, \mathbf{x}_T)}{\partial \alpha_s}$$
$$= \sum_{t=1}^{T}\left\{\left[\sum_{i=0}^{k}\widehat{\alpha}_i^*\sum_{j=1}^{x_{it}}\frac{1}{\widehat{\alpha}_i^*\alpha_s+j-1}\right] - \sum_{j=1}^{n_t}\frac{1}{\alpha_s+j-1}\right\}$$
$$\equiv \sum_{t=1}^{T} s_{\mathrm{P},t}(\alpha_s; \mathbf{x}_1, \ldots, \mathbf{x}_T)$$

and the pseudo-observed information for α_s is

$$j_{\mathrm{P}}(\alpha_s; \mathbf{x}_1, \ldots, \mathbf{x}_T) \equiv -\frac{\partial^2 \ell_{\mathrm{P}}(\alpha_s; \mathbf{x}_1, \ldots, \mathbf{x}_T)}{\partial \alpha_s^2}$$
$$= \sum_{t=1}^{T}\left\{\left[\sum_{i=0}^{k}\widehat{\alpha}_i^{*2}\sum_{j=1}^{x_{it}}\frac{1}{(\widehat{\alpha}_i^*\alpha_s+j-1)^2}\right] - \sum_{j=1}^{n_t}\frac{1}{(\alpha_s+j-1)^2}\right\}$$

Utilizing the Newton–Raphson method, the pseudo maximum likelihood estimator (PMLE) $\widehat{\alpha}_{s,\mathrm{PML}}$ of α_s can be obtained as follows. First choose a good initial value $\widehat{\alpha}_{s,\mathrm{PML}}^{(0)}$ of α_s, for example, the method-of-moment estimate $\widehat{\alpha}_{s,\mathrm{MM}}$ given in (14). Then iterate the following equation

$$\widehat{\alpha}_{s,\mathrm{PML}}^{(u+1)} = \widehat{\alpha}_{s,\mathrm{PML}}^{(u)} + \frac{s_{\mathrm{P}}(\widehat{\alpha}_{s,\mathrm{PML}}^{(u)}; \mathbf{x}_1, \ldots, \mathbf{x}_T)}{j_{\mathrm{P}}(\widehat{\alpha}_{s,\mathrm{PML}}^{(u)}; \mathbf{x}_1, \ldots, \mathbf{x}_T)}$$

for $u = 0, 1, 2, \ldots$ until convergence. Under regularity conditions, $\widehat{\alpha}_{s,\mathrm{PML}}$ is a weakly consistent estimator of α_s and asymptotically normally distributed as $T \to \infty$.

Plugging $\widehat{\boldsymbol{\alpha}} = \widehat{\alpha}_s\widehat{\boldsymbol{\alpha}}^*$, with $\widehat{\alpha}_s = \widehat{\alpha}_{s,\mathrm{MM}}$ or $\widehat{\alpha}_{s,\mathrm{PML}}$, into the posterior mean (6), we get an empirical Bayes estimator of $\mathbf{p}_t$:

$$\widehat{\mathbf{p}}_t = \frac{\widehat{\alpha}_s\widehat{\boldsymbol{\alpha}}^* + \mathbf{x}_t}{\widehat{\alpha}_s + n_t} \tag{16}$$

which can be rewritten as

$$\widehat{\mathbf{p}}_t = \frac{\widehat{\alpha}_s}{\widehat{\alpha}_s + n_t}\widehat{\boldsymbol{\alpha}}^* + \frac{n_t}{\widehat{\alpha}_s + n_t}\frac{\mathbf{x}_t}{n_t} \tag{17}$$

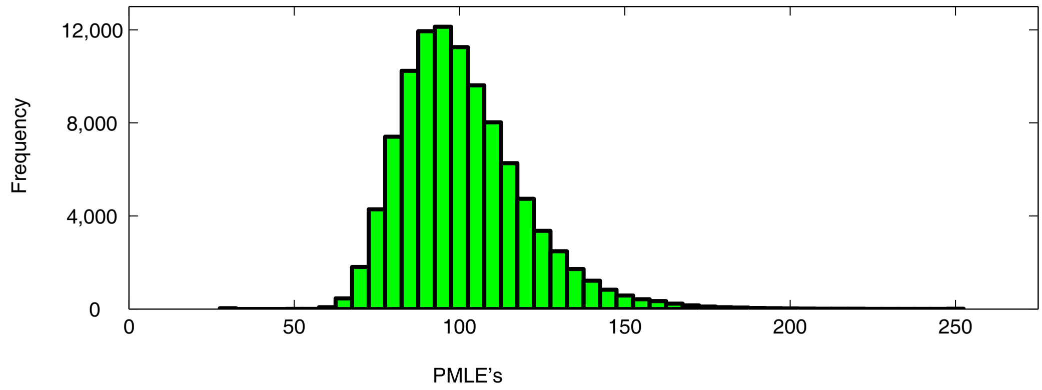

Figure 1. The histogram of $\widehat{\alpha}_{s,\mathrm{PML}}$ for 100 000 samples with $n_t = 50$, $T = 300$, and $\boldsymbol{\alpha} = (70, 20, 10)$

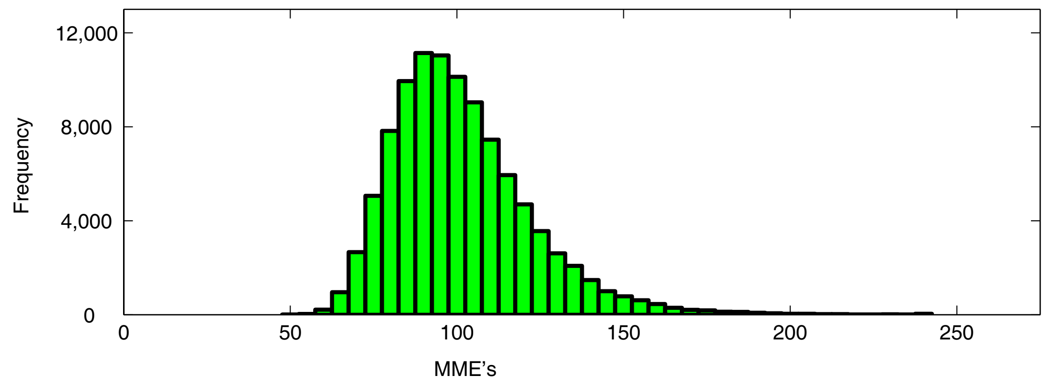

Figure 2. The histogram of $\widehat{\alpha}_{s,\mathrm{MM}}$ for 100 000 samples with $n_t = 50$, $T = 300$, and $\boldsymbol{\alpha} = (70, 20, 10)$

a weighted average of the process mean estimator $\widehat{\boldsymbol{\alpha}}^*$ and observed proportions $\mathbf{x}_t/n_t$. Increasing $\widehat{\alpha}_s$ puts more weight on prior mean estimator $\widehat{\boldsymbol{\alpha}}^*$.

We conduct a simulation study to compare the performance of both PMLE and MME of α_s. For simplicity, consider $k = 2$. For each $t = 1, \ldots, 300$, we first generate $\mathbf{p}_t$ from the Dirichlet($\boldsymbol{\alpha}$) prior distribution and then generate $\mathbf{x}_t$ from the multinomial(n_t; $\mathbf{p}_t$) distribution, where $\boldsymbol{\alpha} = (70, 20, 10)'$ and $n_t = 50$. Thus, we obtain a sample $\{\mathbf{x}_1, \ldots, \mathbf{x}_{300}\}$ of size 300. Compute both PMLE and MME of α_s for this sample. Repeat the above procedure independently for 100 000 times to obtain 100 000 PMLE and 100 000 MMEs.

Figures 1 and 2 present the histograms of these PMLEs and MMEs, respectively. The 100 000 PMLEs have sample mean 100.92, sample standard deviation 18.96, and sample mean squared error 360.38, while the corresponding 100 000 MME have sample mean 101.33, sample standard deviation 21.09, and sample mean squared error 446.70. Furthermore, among these 100 000 samples, there are 57 619 samples for which the PMLE of α_s is closer to the true α_s than the corresponding MME. Thus, the PMLE performs slightly better than the MME for this particular example. Some other examples are tried and results are similar. Thus, in the following, α_s is estimated by the PMLE.

Now with all components of the proposed empirical Bayes approach in place, we are ready to describe a monitoring scheme for polytomous data considered in this paper.

4. A MONITORING SCHEME FOR POLYTOMOUS DATA

Control limits of a control chart are often obtained by finding a region of extreme points for which the coverage probability is a prespecified false alarm rate γ. The most popular value for γ is the probability that a standard normal random variable exceeds 3, i.e. $\gamma = 0.002\,6998$. For a manufacturing process of polytomous data described above, we use the marginal distributions of the observed proportions to construct control limits.

Recall that $\mathbf{x}_t$ has a Dirichlet-compound multinomial distribution with p.m.f. given in (7). However, it is difficult to find a reasonable out-of-control region for such a multivariate discrete distribution. Thus, for simplicity and better interpretation, we consider monitoring the non-conformity of each component of the observed proportions $\mathbf{x}_t/n_t$. Since the marginal distribution of x_{it} is a Pólya distribution with parameters n_t, α_i, and α_s for each $i = 0, 1, \ldots, k$, the marginal p.m.f. of x_{it} is

$$f(x_{it}; \alpha_i, \alpha_s) = \exp\left[\sum_{j=1}^{n_t} \log\left(\frac{j}{\alpha_s + j - 1}\right) - \sum_{j=1}^{x_{it}} \log\left(\frac{j}{\alpha_i + j - 1}\right) - \sum_{j=1}^{n_t - x_{it}} \log\left(\frac{j}{\alpha_s - \alpha_i + j - 1}\right)\right] \tag{18}$$

for $x_{it} \in \{0, 1, \ldots, n_t\}$.

We remark here that there are different opinions on the choice of the false alarm rate for each component of $\mathbf{x}_t/n_t$. A well-known conservative choice is $\gamma/(k+1)$. This scheme is conservative because the overall false alarm rate for all $k+1$ control charts is most likely smaller than γ. The false alarm rates for different components of $\mathbf{x}_t/n_t$ are parameters to choose when designing a monitoring scheme. We shall not discuss this issue further in this paper.

Let $i \in \{0, 1, \ldots, k\}$ be fixed. A control chart for monitoring the non-conformity of the observed proportion x_{it}/n_t can be constructed as follows. Observe that x_{it}/n_t is a discrete random variable. If the deterministic control limits are used, the conventional out-of-control probability γ is almost impossible to attain. And it is found that the actual out-of-control probability varies quite a bit for various processes.

Based on the concept of the randomized test in hypothesis testing, we propose a randomized-control-limits approach. Note that the marginal distribution of x_{it}/n_t is skewed unless $\alpha_i = \alpha_s/2$. There are many possible ways to split the out-of-control probability γ into the two tails when the process is in control. The simplest way is the equal split. That is, $\gamma/2$ for each tail.

Now to find the lower control limit (LCL), from (18), we start accumulating the tail probability from 0 until we reach the first l such that $\sum_{x_{it}=0}^{l} f(x_{it}; \alpha_i, \alpha_s) \geq \gamma/2$. If the equality holds, which is very unlikely, then there is no need for randomization and LCL $= l/n_t$. If the equality does not hold, which means that $\sum_{x_{it}=0}^{l-1} f(x_{it}; \alpha_i, \alpha_s) < \gamma/2$ and $\sum_{x_{it}=0}^{l} f(x_{it}; \alpha_i, \alpha_s) > \gamma/2$, then l/n_t is the randomized lower control limit (RLCL).

The randomization is done by signaling out-of-control with probability

$$\gamma_{\mathrm{RLCL}}(\alpha_i, \alpha_s) = \frac{\gamma/2 - \sum_{x_{it}=0}^{l-1} f(x_{it}; \alpha_i, \alpha_s)}{f(l; \alpha_i, \alpha_s)} \tag{19}$$

when $x_{it} = l$ is observed. A randomized upper control limit (RUCL) can be obtained similarly. For charting, in addition to the two randomized control limits, we use the median of the marginal distribution of x_{it}/n_t as the center line of the control chart.

For example, for the equal-split method, Figure 3 presents the marginal p.m.f. of x_{it} for a process with $n_t = 50$, $\alpha_i = 10$, and $\alpha_s = 100$. Let $\gamma = 0.002\,6998$. Since $f(0; 10, 100) = 0.014\,272$, $f(15; 10, 100) = 0.000\,860\,32$, and $f(16; 10, 100) + \cdots + f(50; 10, 100) = 0.000\,644\,96$, it is found that RLCL $= 0$ and RUCL $= 15/50 = 0.3$, with $\gamma_{\mathrm{RLCL}} = 0.001\,3499/0.014\,272 = 0.094\,582$ and $\gamma_{\mathrm{RUCL}} = (0.001\,3499 - 0.000\,644\,96)/0.000\,860\,32 = 0.819\,39$.

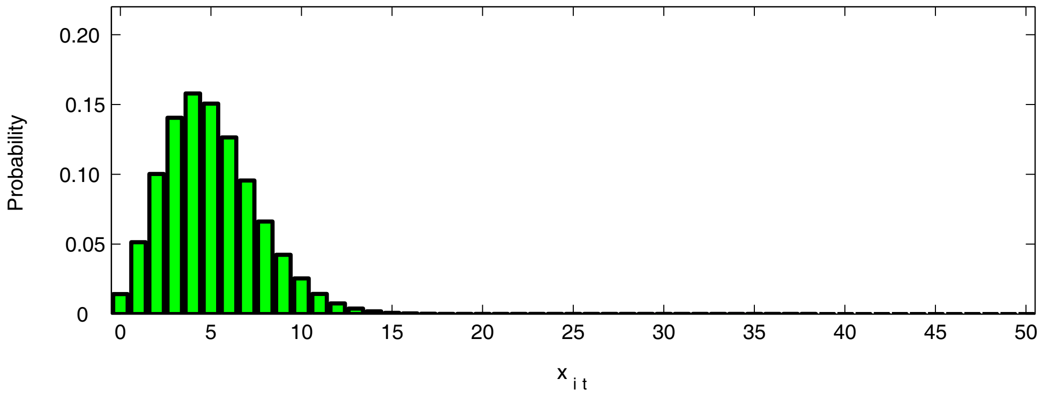

Figure 3. The marginal p.m.f. of x_{it} with $n_t = 50$ and $(\alpha_i^*, \alpha_s) = (0.1, 100)$

5. THE AVERAGE RUN LENGTH OF THE CONTROL SCHEME

To study the performance of this control scheme, we compute the average run length (ARL) for in-control and several out-of-control situations. The ARL is the average number of observations required until an out-of-control alarm is signaled. The in-control ARL, usually denoted by ARL_0, is $\text{ARL}_0 = 1/\gamma$ $(= 370.40$ if $\gamma = 0.002\,6998)$ by the control chart design.

To compute the out-of-control ARL, usually denoted by ARL_1, we need to specify an out-of-control situation, say, x_{it} has a Pólya distribution with parameters n_t, $\tilde{\alpha}_i$, and $\tilde{\alpha}_s$ such that $\tilde{\alpha}_i \neq \alpha_i$ and/or $\tilde{\alpha}_s \neq \alpha_s$. The out-of-control probability is

$$P_{\text{out}}(\tilde{\alpha}_i, \tilde{\alpha}_s) = \sum_{x_{it}=0}^{n_t\cdot\text{RLCL}-1} f(x_{it}; \tilde{\alpha}_i, \tilde{\alpha}_s) + \gamma_{\text{RLCL}} \cdot f(n_t \cdot \text{RLCL}; \tilde{\alpha}_i, \tilde{\alpha}_s)$$
$$+ \gamma_{\text{RUCL}} \cdot f(n_t \cdot \text{RUCL}; \tilde{\alpha}_i, \tilde{\alpha}_s) + \sum_{x_{it}=n_t\cdot\text{RUCL}+1}^{n_t} f(x_{it}; \tilde{\alpha}_i, \tilde{\alpha}_s)$$

Then $\text{ARL}_1(\tilde{\alpha}_i, \tilde{\alpha}_s) = 1/P_{\text{out}}(\tilde{\alpha}_i, \tilde{\alpha}_s)$. For the case that $(\alpha_i^*, \alpha_s) = (0.1, 100)$ (i.e. $(\alpha_i, \alpha_s) = (10, 100)$), Table I gives the ARL_0 and ARL_1 for various out-of-control $(\tilde{\alpha}_i^*, \tilde{\alpha}_s)$'s for the cases that $n_t = 50$, 100, and 200, respectively. For Table I, $\alpha_i^* = 0.1$, ARL_0 is given when $\tilde{\alpha}_i^*$ is also equal to 0.1. We repeat the above study for three more cases: (1) $(\alpha_i^*, \alpha_s) = (0.05, 100)$, (2) $(\alpha_i^*, \alpha_s) = (0.15, 100)$, and (3) $(\alpha_i^*, \alpha_s) = (0.5, 100)$. The results are given in Tables II–IV. For each of these tables, ARL_0 is given when $\tilde{\alpha}_i^* = \alpha_i^*$. It is seen that the ARL_1 decrease fairly fast as $\tilde{\alpha}_i^*$ gets further away from α_i^*, and as n_t increases. Also, when α_i^* gets larger (up to 0.5), ARL_1 gets smaller for the $\tilde{\alpha}_i^*$ having the same relative deviation from α_i^* (i.e. $(\tilde{\alpha}_i^* - \alpha_i^*)/\alpha_i^*$). In other words, for the less skewed marginal distribution of x_{it}/n_t, the detecting power is better. An interesting situation occurs in Table II, when $\tilde{\alpha}_i^*$ equals 0.04, and $n_t = 50$. In this situation, the ARL_1 is larger than ARL_0. This is because $\tilde{\alpha}_i^* = 0.04$ is too close to the true value $\alpha_u^* (= 0.05)$ and the sample size 50 is not large enough to have good power in detecting the difference. The above ARLs are computed assuming that the α_i and α_s are known.

In practice, we have to estimate α_i and α_s from the process data. Denote their estimates by $\widehat{\alpha}_i$ and $\widehat{\alpha}_s$, respectively. To investigate the effect of the estimation error of $\widehat{\alpha}_i$ and $\widehat{\alpha}_s$ on ARLs, the results of another simulation study are presented as follows.

The 100 000 samples described in Section 3 are used for this study. For each $i = 0, 1, \ldots, k$, we order these 100 000 samples by the size of $|\widehat{\alpha}_i^* - \alpha_i^*|$ and then pick the 10 000th, 50 000th, and 90 000th ordered samples (i.e. the 10th, 50th, and 90th percentiles in terms of the deviation from the true α_i^* among the 100 000 simulated samples) to construct the control charts. As an example, Table V gives the ARLs of

Table I. ARL_0 and ARL_1 values are listed for various $\tilde{\alpha}_i^*$ and n_t with $(\alpha_i^*, \alpha_s) = (0.1, 100)$ and $\tilde{\alpha}_s = 100$. ARL_0 is given when $\tilde{\alpha}_i^* = 0.1$

$\tilde{\alpha}_i^*$	$n_t = 50$, $\alpha_i^* = 0.1$ RLCL = 0 $\gamma_{\mathrm{RLCL}} = 0.094\,582$ RUCL = 0.3 $\gamma_{\mathrm{RUCL}} = 0.819\,39$	$n_t = 100$, $\alpha_i^* = 0.1$ RLCL = 0.01 $\gamma_{\mathrm{RLCL}} = 0.160\,28$ RUCL = 0.26 $\gamma_{\mathrm{RUCL}} = 0.905\,46$	$n_t = 200$, $\alpha_i^* = 0.1$ RLCL = 0.02 $\gamma_{\mathrm{RLCL}} = 0.063\,544$ RUCL = 0.235 $\gamma_{\mathrm{RUCL}} = 0.603\,45$
0.0001	10.616	1.0062	1.0011
0.001	11.012	1.0635	1.0120
0.02	24.031	3.4929	1.8423
0.04	55.372	12.673	5.7924
0.06	128.66	47.547	24.531
0.08	280.23	176.81	122.76
0.10	370.40	370.40	370.40
0.12	200.78	175.92	155.48
0.14	82.917	60.301	46.735
0.16	37.087	24.140	17.389
0.18	18.720	11.417	7.9250
0.20	10.540	6.2304	4.2913
0.22	6.5197	3.8402	2.6883

Table II. ARL_0 and ARL_1 values are listed for various $\tilde{\alpha}_i^*$ and n_t with $(\alpha_i^*, \alpha_s) = (0.05, 100)$ and $\tilde{\alpha}_s = 100$. ARL_0 is given when $\tilde{\alpha}_i^* = 0.05$

$\tilde{\alpha}_i^*$	$n_t = 50$, $\alpha_i^* = 0.05$ RLCL = 0 $\gamma_{\mathrm{RLCL}} = 0.010\,793$ RUCL = 0.2 $\gamma_{\mathrm{RUCL}} = 0.146\,26$	$n_t = 100$, $\alpha_i^* = 0.05$ RLCL = 0 $\gamma_{\mathrm{RLCL}} = 0.046\,660$ RUCL = 0.18 $\gamma_{\mathrm{RUCL}} = 0.938\,62$	$n_t = 200$, $\alpha_i^* = 0.05$ RLCL = 0 $\gamma_{\mathrm{RLCL}} = 0.36344$ RUCL = 0.16 $\gamma_{\mathrm{RUCL}} = 0.877\,27$
0.0001	93.027	21.581	2.7820
0.001	96.500	22.976	3.0721
0.01	139.37	43.076	8.3100
0.02	209.19	86.894	25.261
0.03	304.71	173.44	76.863
0.04	390.75	312.22	218.39
0.05	370.40	370.40	370.40
0.06	249.85	239.48	232.96
0.07	143.50	120.61	104.49
0.08	81.487	61.507	49.295
0.09	48.302	33.738	25.669
0.10	30.219	19.962	14.655
0.11	19.914	12.644	4.0703

the three control charts in monitoring the second defect type when the control limits are constructed from the estimate $\widehat{\boldsymbol{\alpha}}$ of α. It is seen that ARLs do not vary that much among these three cases. This indicates that the ARL is somewhat robust to the estimation error of $\widehat{\boldsymbol{\alpha}}$. Figures 4–6 show the marginal distribution of x_{2t} under (a) $(\widehat{\alpha}_2^*, \widehat{\alpha}_{s,\mathrm{PML}}) = (0.1004, 88.260)$, (b) $(\widehat{\alpha}_2^*, \widehat{\alpha}_{s,\mathrm{PML}}) = (0.098, 79.766)$, and (c) $(\widehat{\alpha}_2^*, \widehat{\alpha}_{s,\mathrm{PML}}) = (0.104\,87, 88.262)$, which correspond to the 10 000th, 50 000th, and 90 000th ordered samples in the above ARL study, respectively.

6. A NUMERICAL EXAMPLE

Consider a process that has four different defect types. Assume that the yield/defect probability vector $\mathbf{p}_t$ follows Dirichlet($\boldsymbol{\alpha}$) with $\boldsymbol{\alpha} = (60, 15, 10, 10, 5)'$. Let $n_t = 100$ and $T = 200$. For $t = 1, \ldots, T$, we first generate $\mathbf{p}_t$

Table III. ARL_0 and ARL_1 values are listed for various $\tilde{\alpha}_i^*$ and n_t with $(\alpha_i^*, \alpha_s) = (0.15, 100)$ and $\tilde{\alpha}_s = 100$. ARL_0 is given when $\tilde{\alpha}_i^* = 0.15$

$\tilde{\alpha}_i^*$	$n_t = 50$, $\alpha_i^* = 0.15$ RLCL = 0 $\gamma_{\text{RLCL}} = 0.916\,20$ RUCL = 0.36 $\gamma_{\text{RUCL}} = 0.126\,06$	$n_t = 100$, $\alpha_i^* = 0.15$ RLCL = 0.03 $\gamma_{\text{RLCL}} = 0.375\,24$ RUCL = 0.33 $\gamma_{\text{RUCL}} = 0.932\,66$	$n_t = 200$, $\alpha_i^* = 0.15$ RLCL = 0.045 $\gamma_{\text{RLCL}} = 0.300\,48$ RUCL = 0.30 $\gamma_{\text{RUCL}} = 0.170\,69$
0.0001	1.0959	1.0005	1.0001
0.001	1.1368	1.0058	1.0007
0.03	3.7596	1.8025	1.2785
0.06	13.369	5.6619	3.1059
0.09	49.113	23.977	13.186
0.12	180.05	121.41	83.672
0.15	370.40	370.40	370.40
0.18	161.90	129.46	107.59
0.21	52.270	34.402	24.935
0.24	20.221	12.035	8.2167
0.27	9.3997	5.3903	3.6652
0.30	5.1074	2.9683	2.0986
0.33	3.1666	1.9392	1.4669

Table IV. ARL_0 and ARL_1 values are listed for various $\tilde{\alpha}_i^*$ and n_t with $(\alpha_i^*, \alpha_s) = (0.5, 100)$ and $\tilde{\alpha}_s = 100$. ARL_0 is given when $\tilde{\alpha}_i^* = 0.5$

$\tilde{\alpha}_i^*$	$n_t = 50$, $\alpha_i^* = 0.5$ RLCL = 0.24 $\gamma_{\text{RLCL}} = 0.845\,45$ RUCL = 0.76 $\gamma_{\text{RUCL}} = 0.84\,545$	$n_t = 100$, $\alpha_i^* = 0.5$ RLCL = 0.29 $\gamma_{\text{RLCL}} = 0.738\,76$ RUCL = 0.71 $\gamma_{\text{RUCL}} = 0.738\,76$	$n_t = 200$, $\alpha_i^* = 0.5$ RLCL = 0.32 $\gamma_{\text{RLCL}} = 0.493\,74$ RUCL = 0.68 $\gamma_{\text{RUCL}} = 0.493\,74$
0.20	1.3166	1.0626	1.0119
0.25	1.9942	1.3131	1.1080
0.30	3.8600	2.1479	1.5427
0.35	9.6535	5.0309	3.2297
0.40	31.305	17.495	11.245
0.45	129.52	91.041	67.939
0.50	370.40	370.40	370.40
0.55	129.52	91.041	67.939
0.60	31.305	17.495	11.245
0.65	9.6535	5.0309	3.2297
0.70	3.8600	2.1479	1.5427
0.75	1.9942	1.3131	1.1080
0.80	1.3166	1.0626	1.0119

from the Dirichlet($\boldsymbol{\alpha}$) distribution and then generate count data $\mathbf{x}_t$ from the multinomial(n_t; $\mathbf{p}_t$) distribution given $\mathbf{p}_t$. For this data set, we get $\widehat{\boldsymbol{\alpha}}^* = (0.6009, 0.1538, 0.1033, 0.095, 0.047)$ and $\widehat{\alpha}_{s,\text{PML}} = 105.31$.

The RUCL, median (the center line), and RLCL can be obtained for the yield and each defect type as follows:

	x_{0t}	x_{1t}	x_{2t}	x_{3t}	x_{4t}
RUCL	0.82	0.34	0.27	0.26	0.17
γ_{RUCL}	0.243 80	0.320 32	0.584 96	0.897 07	0.032 113
median	0.63	0.16	0.1	0.1	0.04
RLCL	0.42	0.04	0.01	0.01	0
γ_{RLCL}	0.714 98	0.082 920	0.446 30	0.161 43	0.044 996

Table V. Effect of estimation error. ARL_0 and ARL_1 values are listed for various $\tilde{\alpha}_2^*$ with respect to the 10 000th, 50 000th, and 90 000th ordered samples out of 100 000 samples, where $n_t = 50$, $T = 300$. The control charts are constructed for $(\alpha_2^*, \alpha_s) = (\widehat{\alpha}_2^*, \widehat{\alpha}_{s,\mathrm{PML}})$ and $\tilde{\alpha}_s = \widehat{\alpha}_{s,\mathrm{PML}}$ instead of the true $(\alpha_2^*, \alpha_s) = (0.1, 100)$ and $\tilde{\alpha}_s = 100$

$\tilde{\alpha}_2^*$	$\widehat{\alpha}_2^* = 0.1004$ $\widehat{\alpha}_s = 88.260$ RLCL $= 0$ $\gamma_{\mathrm{RLCL}} = 0.087\,143$ RUCL $= 0.30$ $\gamma_{\mathrm{RUCL}} = 0.454\,95$	$\widehat{\alpha}_2^* = 0.098$ $\widehat{\alpha}_s = 79.766$ RLCL $= 0$ $\gamma_{\mathrm{RLCL}} = 0.072\,280$ RUCL $= 0.30$ $\gamma_{\mathrm{RUCL}} = 0.392\,21$	$\widehat{\alpha}_2^* = 0.104\,87$ $\widehat{\alpha}_s = 88.262$ RLCL $= 0$ $\gamma_{\mathrm{RLCL}} = 0.105\,81$ RUCL $= 0.30$ $\gamma_{\mathrm{RUCL}} = 0.143\,67$
0.0001	11.521	13.889	9.4890
0.001	11.942	14.386	9.8352
0.02	25.602	30.377	21.086
0.04	57.864	67.520	47.665
0.06	131.70	150.55	108.81
0.08	280.52	305.68	239.75
0.10	371.79	361.43	376.10
0.12	210.86	193.71	254.25
0.14	89.876	83.731	111.31
0.16	40.855	38.926	49.738
0.18	20.780	20.173	24.730
0.20	11.725	11.548	13.639
0.22	7.2385	7.2057	8.2540

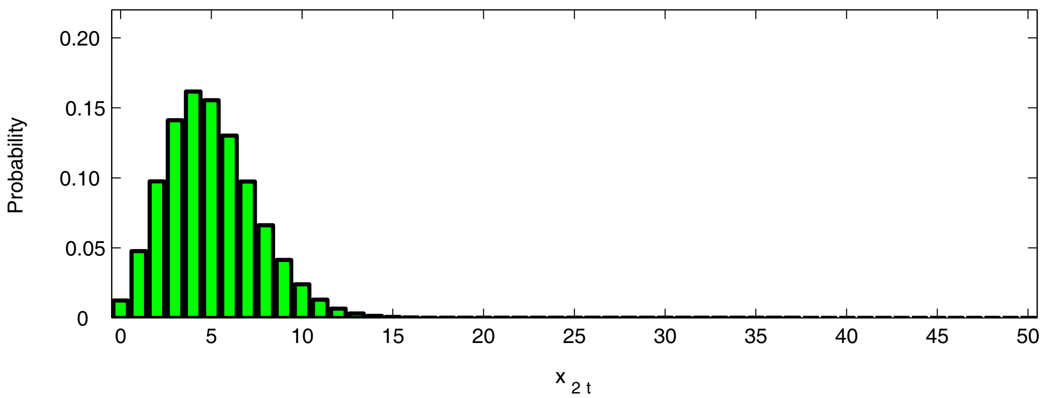

Figure 4. The marginal p.m.f. of x_{2t} with $n_t = 50$ and $(\alpha_2^*, \alpha_s) = (\widehat{\alpha}_2^*, \widehat{\alpha}_{s,\mathrm{PML}}) = (0.1004, 88.260)$

Now we can monitor the process for polytomous data with the control scheme constructed above. Assume that the process has shifted such that the yield/defect probability vector $\mathbf{p}_t$ is now distributed as Dirichlet($\tilde{\boldsymbol{\alpha}}$), where $\tilde{\boldsymbol{\alpha}} = (55, 15, 10, 10, 10)'$. That is to say, the chance of the fourth defect type occurring has increased while the chances of other defect types remain the same. Generate 100 $\mathbf{x}_t$ for this situation. Figures 7–11 give the control charts for each component of $(\mathbf{x}_t/n_t)$. It is noted that the control charts for both x_{0t}/n_t and x_{4t}/n_t show that the process is out of control while the other control charts show that the process remains in control. This demonstrates the effectiveness of the monitoring scheme for this example since only $\alpha_0 = 60$ is shifted to $\tilde{\alpha}_0 = 55$ and $\alpha_4 = 5$ to $\tilde{\alpha}_4 = 10$.

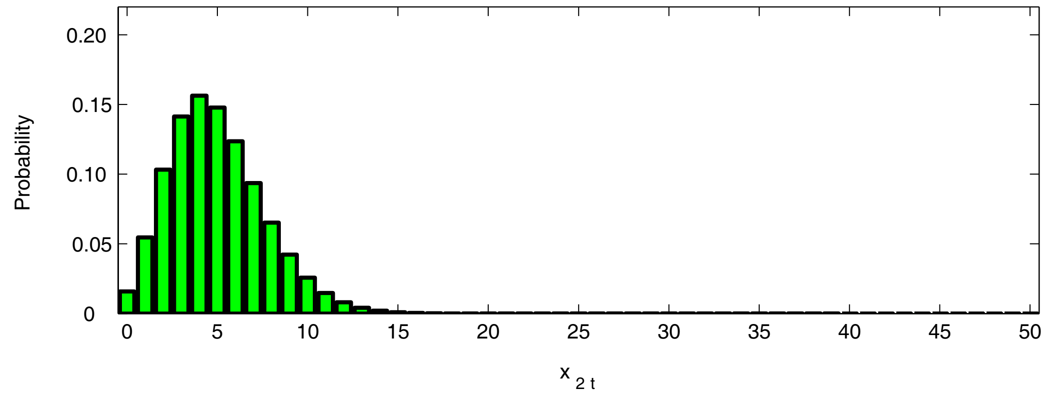

Figure 5. The marginal p.m.f. of x_{2t} with $n_t = 50$ and $(\alpha_2^*, \alpha_s) = (\widehat{\alpha}_2^*, \widehat{\alpha}_{s,\mathrm{PML}}) = (0.098, 79.766)$

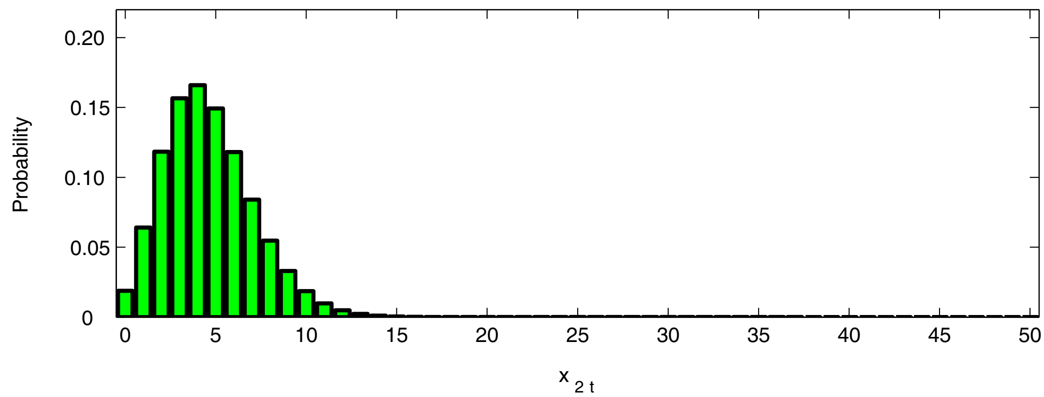

Figure 6. The marginal p.m.f. of x_{2t} with $n_t = 50$ and $(\alpha_2^*, \alpha_s) = (\widehat{\alpha}_2^*, \widehat{\alpha}_{s,\mathrm{PML}}) = (0.104\,87, 88.262)$

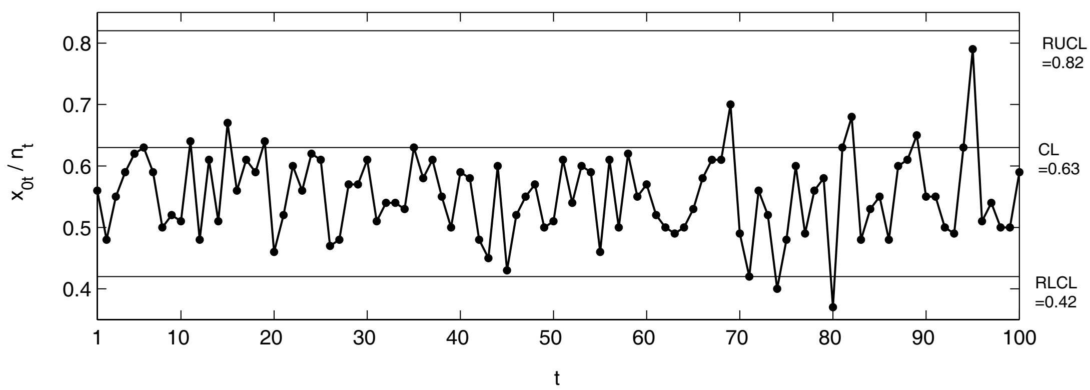

Figure 7. The control chart of (x_{0t}/n_t) for $t = 1, \ldots, 100$ with $n_t = 100$, $\boldsymbol{\alpha} = (55, 15, 10, 10, 10)'$, RLCL = 0.42, CL (center line) ≡ median = 0.63, RUCL = 0.82, $\gamma_{\mathrm{RLCL}} = 0.71\,498$, and $\gamma_{\mathrm{RUCL}} = 0.243\,80$

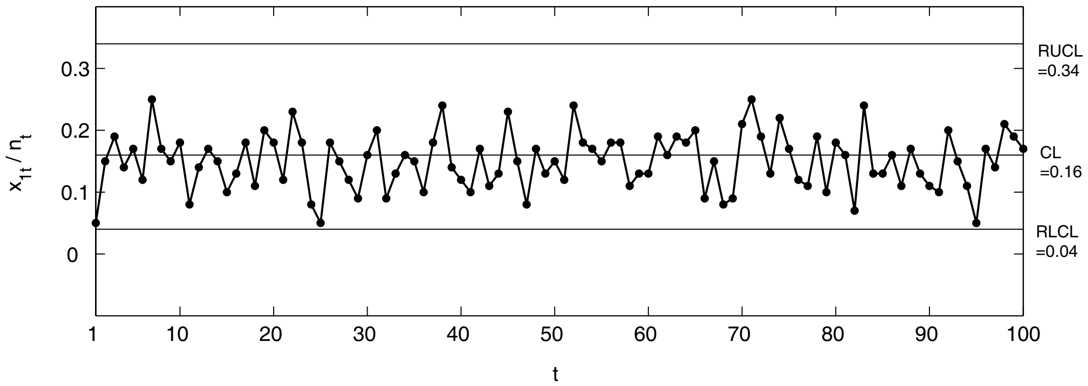

Figure 8. The control chart of (x_{1t}/n_t) for $t = 1, \ldots, 100$ with $n_t = 100$, $\boldsymbol{\alpha} = (55, 15, 10, 10, 10)'$, RLCL $= 0.04$, CL (center line) $\equiv$ median $= 0.16$, RUCL $= 0.34$, $\gamma_{\text{RLCL}} = 0.082\,920$, and $\gamma_{\text{RUCL}} = 0.320\,32$

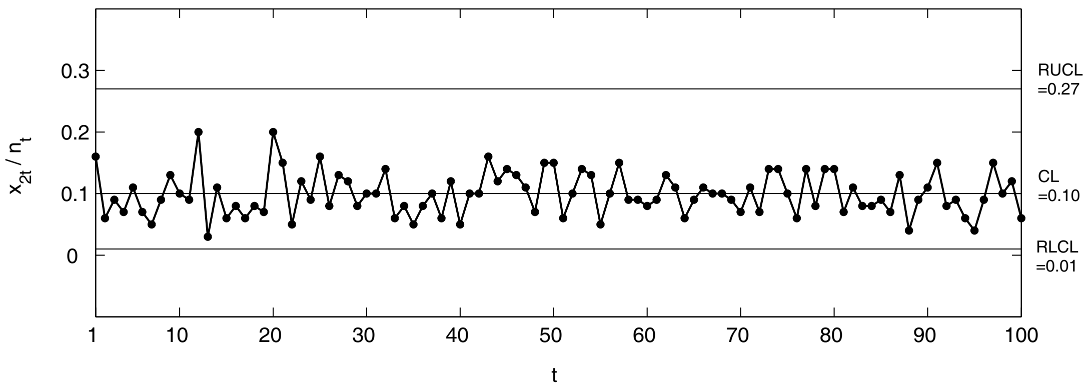

Figure 9. The control chart of (x_{2t}/n_t) for $t = 1, \ldots, 100$ with $n_t = 100$, $\boldsymbol{\alpha} = (55, 15, 10, 10, 10)'$, RLCL $= 0.01$, CL (center line) $\equiv$ median $= 0.1$, RUCL $= 0.27$, $\gamma_{\text{RLCL}} = 0.446\,30$, and $\gamma_{\text{RUCL}} = 0.584\,96$

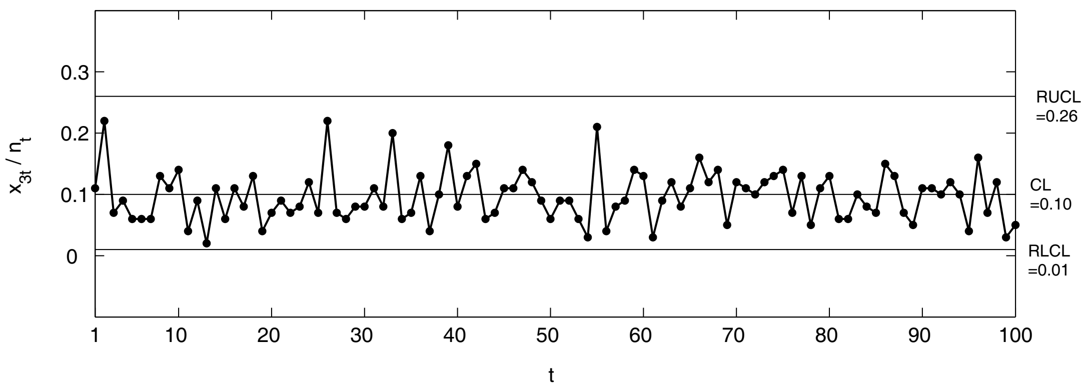

Figure 10. The control chart of (x_{3t}/n_t) for $t = 1, \ldots, 100$ with $n_t = 100$, $\boldsymbol{\alpha} = (55, 15, 10, 10, 10)'$, RLCL $= 0.01$, CL (center line) $\equiv$ median $= 0.1$, RUCL $= 0.26$, $\gamma_{\text{RLCL}} = 0.161\,43$, and $\gamma_{\text{RUCL}} = 0.897\,07$

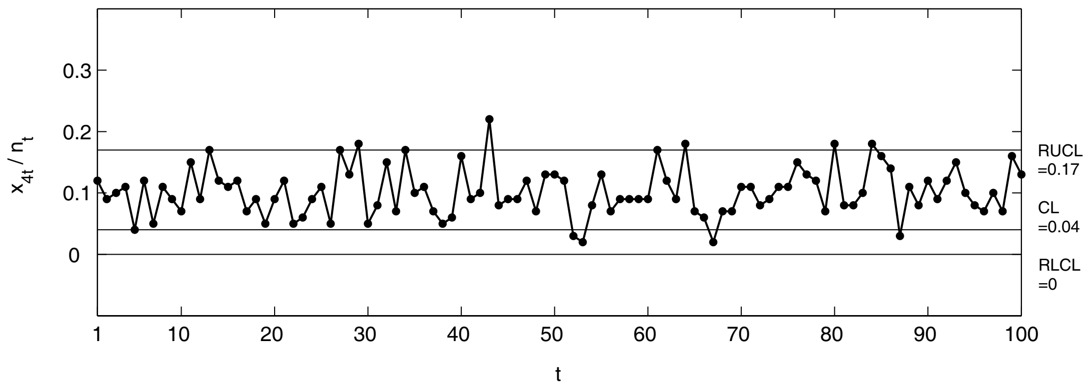

Figure 11. The control chart of (x_{4t}/n_t) for $t = 1, \ldots, 100$ with $n_t = 100$, $\boldsymbol{\alpha} = (55, 15, 10, 10, 10)'$, RLCL = 0, CL (center line) ≡ median = 0.04, RUCL = 0.17, $\gamma_{\mathrm{RLCL}} = 0.044\,996$, and $\gamma_{\mathrm{RUCL}} = 0.032\,113$

7. CONCLUSION

In this paper, we develop an empirical Bayes process monitoring scheme for polytomous data. The yield/defect count data vector $\mathbf{x}_t$ is modeled by the multinomial$(n_t; \mathbf{p}_t)$ distribution given $\mathbf{p}_t$. The yield/defect probability vector $\mathbf{p}_t$ is allowed to vary, and is modeled by the Dirichlet$(\boldsymbol{\alpha})$ distribution with unknown hyperparameter vector $\boldsymbol{\alpha}$. Thus the variation in the yield/defect count data is decomposed into two sources of variation: (i) the sampling variation (i.e. the expectation of the conditional variance–covariance matrix of $\mathbf{x}_t/n_t$ given $\mathbf{p}_t$), and (ii) the process variation (i.e. the variance–covariance matrix of $\mathbf{p}_t$).

To estimate $\boldsymbol{\alpha}$, two estimation methods, the method of moments and the pseudo maximum likelihood method, are proposed and studied. It is found by simulation that the PMLE performs slightly better than the MME.

Using the above empirical Bayes methodology control charts are developed, based on the marginal distributions of the components of $\mathbf{x}_t/n_t$. To achieve the usual false alarm rate when the process is in control, we develop a randomized-control-limits scheme. The performance of the control scheme is studied via the average run length. It is found that if the marginal distribution of the defect/yield proportion is not too skewed, then the detecting power of the proposed control chart is fairly good. It is also found that the ARLs are somewhat robust to the estimation error of $\widehat{\boldsymbol{\alpha}}$. Finally, an illustrative example demonstrates the potential usefulness of the proposed monitoring scheme.

Acknowledgements

The authors would like to thank Ching-Chou Chien and Hsiao-Hui Liao for providing the computing work and carrying out the simulations. The authors would also like to thank the referee for his/her valuable comments and suggestions. Jyh-Jen Shiau's work was partially supported by the National Science Council of the Republic of China Grant No. NSC91-2118-M-009-007.

REFERENCES

1. McCullagh P, Nelder JA. *Generalized Linear Models* (2nd edn). Chapman and Hall: London, 1989.
2. Montgomery DC. *Introduction to Statistical Quality Control* (4th edn). Wiley: New York, 2001.
3. Voss B, Dunert J, Dahms S, Weiss H. A multinomial model for the quality control of colony counting procedures. *Biometrical Journal* 2000; **42**:263–278.
4. Lindley DV. The Bayesian analysis of contingency tables. *Annals of Mathematics and Statistics* 1964; **35**:1622–1643.
5. Walley P. Inferences from multinomial data: Learning about a bag of marbles. *Journal of Royal Statistical Society* B 1996; **58**:3–57.

6. Nair VN, Tang B, Xu L. Bayesian inference for some mixture problems in quality and reliability. *Journal of Quality Technology* 2001; **33**:16–28.
7. Sturm GW, Feltz CJ, Yousry MA. An empirical Bayes strategy for analysing manufacturing data in real time. *Quality and Reliability Engineering International* 1991; **7**:159–167.
8. Yousry MA, Sturm GW, Feltz CJ, Noorossana R. Process monitoring in real time: Empirical Bayes approach—discrete case. *Quality and Reliability Engineering International* 1991; **7**:123–132.
9. Gong G, Samaniego FJ. Pseudo maximum likelihood estimation: Theory and applications. *Annals of Statistics* 1981; **9**:861–869.
10. Carlin BP, Louis TA. *Bayes and Empirical Bayes Methods for Data Analysis* (2nd edn). Chapman and Hall/CRC: Boca Raton, FL, 2000.
11. Gelman A, Carlin JB, Stern HS, Rubin DB. *Bayesian Data Analysis*. Chapman and Hall: London, 1995.
12. Johnson NL, Kotz S, Balakrishnan N. *Discrete Multivariate Distributions*. Wiley: New York, 1997.

Authors' biographies

Jyh-Jen Horng Shiau received her BS in mathematics from the National Taiwan University, her MS in applied mathematics from the University of Maryland Baltimore, her SM in computer science and PhD in statistics from the University of Wisconsin-Madison. She has taught at the Southern Methodist University, the University of Missouri at Columbia and National Tsing Hua University. She worked for the Engineering Research Center of AT&T Bell Labs before she moved to Hsinchu, Taiwan where she is a professor at the Institute of Statistics, National Chiao Tung University. Her interests include industrial statistics and non-parametric function estimation. She is a member of the Institute of Mathematical Statistics.

Chih-Rung Chen is an assistant professor at the Institute of Statistics, Chiao Tung University. He received his BS and MS in mathematics from the National Taiwan University, and his MS and PhD in statistics from the University of Chicago. His research interests include likelihood inference, estimating functions, transformations, and industrial statistics.

Carol J. Feltz is an associate professor at the Division of Statistics, Northern Illinois University. After receiving her PhD in statistics from the University of Missouri-Columbia, she worked at AT&T's Engineering Research Center in Princeton, New Jersey. While there, she developed methods of improving quality at AT&T's manufacturing facilities. In 1989, she moved to her current position. Her research interests include developing statistical methods for quality improvement, life-testing and goodness-of-fit tests.

