Statistics and Its Interface Volume 11 (2018) 281–293

Nonparametric multivariate Polya tree EWMA control chart for process changepoint detection*

Yuhui Chen[†,‡], Mingwei Sun[§], and Timothy Hanson[¶]

In this article, we propose a nonparametric multivariate control scheme for simultaneously monitoring several related characteristics of a process in time. Through the use of a novel weighted multivariate Polya tree, the proposed method can quickly detect small mean and/or variance shifts in various types of longitudinal processes, Gaussian or non-Gaussian. Briefly, we center a weighted multivariate Polya tree at an initial parametric model on the monitored process, such as multivariate Gaussian; then by adding more details via data, departures from the parametric model will be captured and used for adjusting the initial model to obtain robust estimation. By weighting the Polya tree in the test statistic, the proposed chart thus can heighten the sensitivity of detecting one or more out of control characteristics. Examples show that our chart performs good for monitoring a process where the normality assumption is violated. Particularly, the proposed chart is more sensitive to variance shifts compared to the multivariate EWMA and multivariate CUSUM charts.

Keywords and phrases: Changepoints, Control Charts, Exponentially Weighted Predictive Densities, MCUSUM, MEWMA, Multivariate Polya Trees, Nonparametric Modelings, Statistical Process Control.

1. INTRODUCTION

Recently, the rapid growth of data acquisition technology has led to an interest in simultaneously monitoring several quality characteristics. Consequently, the multivariate control chart has received increasingly more attention due to its ability to monitor several correlated characteristics. To this point, many multivariate control schemes and their extensions, see Crosier (1988), Lowry et al. (1992), Prabhu and Runger (1997), and Testik and Borror (2004), grouped by the multivariate Shewhart charts (Hotelling, 1947), the multivariate CUSUM charts (Woodall and Ncube, 1985), and the multivariate EWMA charts (Lowry et al., 1992) were proposed over the last few decades. But many of them focused on detecting only the mean shift of a multivariate process. To further address the effects on charts when changes in the covariance matrix occurred, Yeh et al. (2004) developed a likelihood-ratio-based EWMA chart for monitoring the variability shift for multivariate Gaussian processes. Along this line, see also Huwang et al. (2007) and Hawkins and Maboudou-Tchao (2008). To further improve the charts for detecting changes in the mean and covariance matrix simultaneously, Zhang et al. (2010) proposed a chart which integrates the EWMA procedure with the generalized likelihood ratio test to jointly monitor the multivariate process mean and variability under the normality assumption, see also Cheng and Thaga (2005), Zhang and Chang (2008), Reynolds and Cho (2011), and Wang et al. (2014).

However, it's well known that parametric control charts may signal incorrectly when the underlying normality assumption is violated. Many nonparametric charts (Willemain and Runger, 1996; Albers and Kallenberg, 2004; Chakraborti and Eryilmaz, 2007; and Chen 2015) were thus developed for robustness; however none of them can be used for multivariate processes. To this point, Qiu and Hawkins (2012) proposed a multivariate CUSUM procedure to detect mean shift upon the cross-sectional antiranks of the measurements (ARCUSUM); Zou and Tsung (2011) developed a method to monitor a location parameter by adapting a multivariate sign test to online sequential monitoring; and Li et al. (2013) proposed two nonparametric charts for monitoring location and scale changes separately. Other nonparametric methods proposed for detecting mean shifts include Liu et al. (2013), Sun and Zi (2013), and Holland and Hawkins (2014). However, there are only few nonparametric charts proposed for monitoring the mean and variability shifts simultaneously for multivariate processes, see Li et al. (2014). To this end, we propose a nonparametric multivariate control scheme based on the weighted Polya tree (PT) predictive density for monitoring the mean, variability, and overall distributional shape changes simultaneously. Briefly, we center a multivariate Polya tree at an initial parametric guess on the monitored process, such as multivariate Gaussian; then by adding more details via data, any departure from this guess will be captured and used for adjusting the initial to robustly detect longitudinal changes in the process

*This research was supported in part by 2015 Research Grants Committee, University of Alabama.
[†]Corresponding author. E-mail: ychen164@ua.edu.
[‡]Assistant Professor, Department of Mathematics, The University of Alabama.
[§]Student, Department of Mathematics, The University of Alabama.
[¶]Professor, Department of Statistics, The University of South Carolina.

when the underlying normality assumption is not satisfied. To allow the density estimate to favor more recent observations and heighten sensitivity to detect out of controls, we weight recent observations more to obtain a novel local predictive PT density. Consequently, the proposed chart inherits both merits of robustness of the nonparametric models and sensitivity of the EWMA charts.

The rest of this paper is organized as follows. In Section 2, we introduce the exponentially weighted multivariate Polya tree priors. In Section 3, we construct a nonparametric multivariate chart based upon the proposed weighted Polya tree, and also develop important aspects of the chart, including obtaining the simulated control limits based on the in-control parameters and diagnosing distributional changes in the process. Simulation studies are then conducted in Section 4 and we then illustrate the use of the chart based on the pseudo real data simulated for a healthy condition process; finally real data on a chemical quality process are analyzed in Section 5. Conclusions are given in Section 6.

2. EXPONENTIALLY WEIGHTED MULTIVARIATE POLYA TREE PRIORS

The Polya trees (PT) have been used for providing flexible and robust inferences in several statistical domains including regression models (Hanson and Johnson, 2002), survival analysis (Walker and Mallick, 1999; Hanson, 2006), time series (Denison and Mallick, 2007), analysis for truncated data (Chen and Hanson, 2014a), and hypothesis testing (Chen and Hanson, 2014b; Cipolli et al., 2016). Most applications have dealt with univariate data. Recent developments on multivariate PT by Paddock (1999), Hanson (2006), and Chen and Hanson (2014b) provide a way for multivariate nonparametric modeling.

We start by defining a multivariate finite PT prior for a distribution G with dimension $d > 1$. Analogous to univariate PT priors, and like the most popular charts which have a Gaussian underlying assumption, the centering distribution for multivariate PT can be assumed to be a multivariate Gaussian denoted as $\Phi_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu}$ is the $d \times 1$ location vector and $\boldsymbol{\Sigma}$ is the $d \times d$ covariance matrix. A multivariate PT prior for G with finite J levels is characterized by a collection of increasingly refined partitions on domain $\mathbb{R}^d$ and a collection of conditional independent probabilities that link a parent set in a given level and its 2^d offspring sets in the subsequent level. Let $\Pi_1, \ldots, \Pi_J$ denote the series of nested partitions such that Π_j is a refined partition of Π_{j-1} in that each set in Π_{j-1} is the union of 2^d offspring sets in Π_j, where $1 < j \leq J$, and Π_0 represents the entire domain $\mathbb{R}^d$.

Consider the base sets $B_0(j; \mathbf{k})$ for the jth level starting with Cartesian products of intervals obtained as quantiles from the standard normal distribution, i.e.,

$$B_0(j; \mathbf{k}) = \overset{d}{\underset{l=1}{\times}} \left(\Phi^{-1}((k_l - 1)/2^j), \Phi^{-1}(k_l/2^j) \right)$$

where $\mathbf{k} = (k_1, \ldots, k_d) \in \{1, \ldots, 2^j\}^d$ with each component $k_l \in \{1, \ldots, 2^j\}$ for $l = 1, 2, \ldots, d$. The collection of all of the sets $\{B_0(j; \mathbf{k}) : \mathbf{k} \in \{1, \ldots, 2^j\}^d\}$ thus partitions the domain $\mathbb{R}^d$ at the level j. Consequently, partitions at the same level j induced by the location-scale parameters $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ in the centering distribution $\Phi_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ are obtained by

$$B(j; \mathbf{k}) = \{\boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\mathbf{z} : \mathbf{z} \in B_0(j; \mathbf{k})\}$$

with $\Pi_j = \{B(j; \mathbf{k}) : \mathbf{k} \in \{1, \ldots, 2^j\}^d\}$.

In addition to the nested partitions $\{\Pi_j\}_{j=1}^J$, the finite multivariate PT prior for G is also defined through independent vectors of conditional probabilities denoted as $\mathcal{X} = \cup_{j=1}^J \{\mathcal{X}(j; \mathbf{k}) : \mathbf{k} \in \{1, \ldots, 2^j\}^d\}$. Specifically, let's consider an observation $\mathbf{y} = (y_1, \ldots, y_d)$ with $\mathbf{y} \mid G \sim G$. Given $\mathbf{y}$ is contained in set $B(j; \mathbf{k})$ at the jth level, then $\mathbf{y}$ must belong to one of the 2^d offspring sets in Π_{j+1}. Denote those offspring sets by $\{B(j+1; 2k_1 - s_1, \ldots, 2k_d - s_d) : (s_0, \ldots, s_d) \in \{0, 1\}^d\}$ and let the vector $\mathcal{X}(j; \mathbf{k})$ grouped by the corresponding conditional probabilities of $\mathbf{y}$ being in those 2^d sets given $\mathbf{y}$ in their parent set $B(j; \mathbf{k})$. For all $\mathbf{k} \in \{1, \ldots, 2^j\}^d$, we model $\mathcal{X}(j; \mathbf{k}) \in \mathcal{X}$, for $j = 1, \ldots, J$, with $\mathcal{X}(j; \mathbf{k}) \sim \text{Dirichlet}(\mathbf{1}_{2^d} c\rho(j))$, where $\mathbf{1}_{2^d}$ is a vector of length 2^d containing all ones, $c > 0$ and $\rho(j) > 0$. By specifying that G follows the centering $\Phi_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ on the sets at level J, then a random distribution G on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ is said to have a finite multivariate PT prior with parameters $(c, \rho(j), \Phi_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}))$, written as $G \sim PT_d^J(c, \rho(j), \Phi_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}))$. Various functions for $\rho(j)$ have been considered, see Lavine (1992), Berger and Guglielmi (2001), and Branscum and Hanson (2008), and throughout this paper, we set $\rho(j) = j^2$ (Hanson, 2006; Chen and Hanson, 2014b). The parameter c is used for controlling how close G is to $\Phi_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$; as $c \to \infty$ the centering $\Phi_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ will be obtained. For more details about univariate and multivariate Polya trees, see Hanson (2006).

Let $\mathbf{y}_{1:i} = (\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_i)'$ be the first i observations that follow the nonparametric finite multivariate PT prior centered at $\Phi_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, written as

$$\mathbf{y}_1, \ldots, \mathbf{y}_i | G \overset{i.i.d}{\sim} G, \quad G | c, \boldsymbol{\mu}, \boldsymbol{\Sigma} \sim PT_d^J(c, \Phi_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})). \tag{1}$$

Note here, we consider $\rho(j) = j^2$ and thus neglect it from the notation for $PT_d^J(\cdot)$. The predictive density for a new observation $\mathbf{y}$ given $\mathbf{y}_{1:i}$ (Hanson, 2006; Chen and Hanson, 2014b) is then written as

(2)

$$p(\mathbf{y} | \mathbf{y}_{1:i}, c, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \phi_d(\mathbf{y} | \boldsymbol{\mu}, \boldsymbol{\Sigma}) \times \prod_{j=1}^{J} \frac{cj^2 + \sum_{k=1}^{i} I\{\boldsymbol{\epsilon}_j(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \boldsymbol{\epsilon}_j(\mathbf{y}_k; \boldsymbol{\mu}, \boldsymbol{\Sigma})\}}{cj^2 + 2^{-d} \sum_{k=1}^{i} I\{\boldsymbol{\epsilon}_{j-1}(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \boldsymbol{\epsilon}_{j-1}(\mathbf{y}_k; \boldsymbol{\mu}, \boldsymbol{\Sigma})\}},$$

where $\phi_d(\cdot | \boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the d-dimensional Gaussian density associated with $\Phi_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $I(\cdot)$ is an indicator function with

$I(A) = 1$ if A holds and 0 otherwise, and

$$\begin{aligned} \boldsymbol{\epsilon}_j(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) &\stackrel{def}{=} (\lceil 2^j \Phi(z_1) \rceil, \ldots, \lceil 2^j \Phi(z_d) \rceil)', \\ \mathbf{z} &= (z_1, \ldots, z_d)' = \boldsymbol{\Sigma}^{-\frac{1}{2}}(\mathbf{y} - \boldsymbol{\mu}) \end{aligned}$$

is a $d \times 1$ vector indicating which partition set $\mathbf{y}$ resides in at the level j, here $\lceil x \rceil$ is the usual ceiling function, giving the largest integer less than or equal to x. We should mention here that there are a continuum of matrix square roots of $\boldsymbol{\Sigma}$. Consider the usual spectral decomposition $\boldsymbol{\Sigma} = \mathbf{M}\boldsymbol{\Lambda}\mathbf{M}'$. Two obvious square roots are $\boldsymbol{\Sigma}^{-\frac{1}{2}} = \mathbf{M}\boldsymbol{\Lambda}^{-\frac{1}{2}}$ (asymmetric) or $\boldsymbol{\Sigma}^{-\frac{1}{2}} = \mathbf{M}\boldsymbol{\Lambda}^{-\frac{1}{2}}\mathbf{M}'$ (symmetric), see Chen and Hanson (2014b). Others can be considered, e.g. $\boldsymbol{\Sigma}^{-\frac{1}{2}} = \mathbf{M}\boldsymbol{\Lambda}^{-\frac{1}{2}}\mathbf{O}$ where $\mathbf{O}$ is a matrix with orthonormal columns and rows (Hanson, Monteiro, and Jara, 2011).

In the usual predictive density estimate (2), the elements in $\mathbf{y}_{1:i}$ contribute the same weight to the predictive density of $\mathbf{y}$, regardless of how recently they occurred. To allow the density estimate to favor more recent observations and heighten sensitivity to a process that is beginning to fall out of control, the density should weight more recent observations more highly. Following Zou and Tsung (2010), we exponentially weight the observations, obtaining the local predictive process written as

(3)
$$p_\lambda(\mathbf{y}|\mathbf{y}_{1:i}, c, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \phi_d(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) \times \prod_{j=1}^{J} \frac{cj^2 + \sum_{k=1}^{i} I\{\boldsymbol{\epsilon}_j(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \boldsymbol{\epsilon}_j(\mathbf{y}_k; \boldsymbol{\mu}, \boldsymbol{\Sigma})\}(1-\lambda)^{i-k}}{cj^2 + 2^{-d}\sum_{k=1}^{i} I\{\boldsymbol{\epsilon}_{j-1}(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \boldsymbol{\epsilon}_{j-1}(\mathbf{y}_k; \boldsymbol{\mu}, \boldsymbol{\Sigma})\}(1-\lambda)^{i-k}},$$

where λ is a smoothing parameter that satisfies $0 \leq \lambda < 1$. With $\lambda = 0$, an unweighted predictive density is obtained. This density estimate eventually "forgets" observations that occurred in the distant past, thus allowing focus to shift onto the recent observations that we wish to classify as "in-control" or "out-of-control".

3. MULTIVARIATE POLYA TREE EWMA CONTROL CHART

3.1 Test statistic

Assume we observe a sequence of observations $\mathbf{y}_1, \ldots, \mathbf{y}_m, \mathbf{y}_{m+1}, \ldots$, where $\mathbf{y}_1, \ldots, \mathbf{y}_m$ are known in-control observations collected in Phase I and $\mathbf{y}_{m+1}, \ldots$, are obtained as unknowns in Phase II. The t observations are collected over time and follows the change point model characterized by a finite multivariate PT as

$$\begin{cases} \text{In-control:} & \mathbf{y}_t|G_0 \overset{i.i.d}{\sim} G_0, \quad G_0|c_0, \boldsymbol{\mu}^0, \\ & \boldsymbol{\Sigma}^0 \sim PT_d^J(c_0, \Phi_d(\boldsymbol{\mu}^0, \boldsymbol{\Sigma}^0)), \\ & \text{for } t = 1, \ldots, m, m+1, \ldots, \eta, \\ \text{Out-of-control:} & \mathbf{y}_t|G_1 \overset{i.i.d}{\sim} G_1, \quad G_1|c_1, \boldsymbol{\mu}^1, \\ & \boldsymbol{\Sigma}^1 \sim PT_d^J(c_1, \Phi_d(\boldsymbol{\mu}^1, \boldsymbol{\Sigma}^1)), \\ & \text{for } t = \eta + 1, \ldots, \end{cases}$$

where η is the unknown change point, and G_0 and G_1 are the distribution functions for in-control and out-of-control processes, respectively, with $G_0 \neq G_1$. To test if t is the change point, we thus wish to test the null hypothesis $H_0 : \mathbf{y}_t|G_0 \sim G_0$ versus the alternative hypothesis $H_1 : \mathbf{y}_t|G_1 \sim G_1$. An intuitive way to conduct this testing is to compare the predictive density of $\mathbf{y}_t$ given the observations from the Phase II to the one given the observations from both Phases. In addition, to further allow the test to favor more recent observations, we thus consider the test statistic as

(4)
$$T_\lambda(t) = \sum_{i=m+1}^{t} (1-\lambda)^{t-i} \left| \log \frac{p_\lambda\left(\mathbf{y}_i|\mathbf{y}_{m+1:i}, c_{1i}, \boldsymbol{\mu}_i^\lambda, \boldsymbol{\Sigma}_i^\lambda\right)}{p_0\left(\mathbf{y}_i|\mathbf{y}_{1:i-1}, c_{0i}, \boldsymbol{\mu}_i^0, \boldsymbol{\Sigma}_i^0\right)} \right|,$$

see Zhang (2002) and Zou and Tsung (2010), here an unweighted density estimate $p_0(\cdot)$ is defined in Eq. (2) and a weighted density $p_\lambda(\cdot)$ is given in Eq. (3); c_{1i} and c_{0i} are chosen such that $p_\lambda(\cdot)$ and $p_0(\cdot)$ are maximized over the grid $\left\{\exp\left(\frac{14}{19}(j-1) - 7\right)\right\}_{j=1}^{20}$, respectively, see Chen and Hanson (2014b). The estimated mean and covariance for the centering distribution in $p_0(\cdot)$ can be obtained as the sample mean and sample covariance matrix for a multivariate Gaussian given as

$$\hat{\boldsymbol{\mu}}_i^0 = \frac{1}{i-1}\sum_{k=1}^{i-1}\mathbf{y}_k \text{ and } \hat{\boldsymbol{\Sigma}}_i^0 = \frac{1}{i-1}\sum_{k=1}^{i-1}\left(\mathbf{y}_k - \hat{\boldsymbol{\mu}}_i^0\right)\left(\mathbf{y}_k - \hat{\boldsymbol{\mu}}_i^0\right)',$$

respectively. Similarly, the mean and covariance for $p_\lambda(\cdot)$ are estimated from the weighted sample mean and sample covariance matrix given by

$$\hat{\boldsymbol{\mu}}_i^\lambda = \frac{\lambda}{1-(1-\lambda)^{i-m}}\sum_{k=m+1}^{i}(1-\lambda)^{i-k}\mathbf{y}_k,$$

$$\hat{\boldsymbol{\Sigma}}_i^\lambda = \frac{\lambda}{1-(1-\lambda)^{i-m}}\sum_{k=m+1}^{i}(1-\lambda)^{i-k}\left(\mathbf{y}_k - \hat{\boldsymbol{\mu}}_i^\lambda\right)\left(\mathbf{y}_k - \hat{\boldsymbol{\mu}}_i^\lambda\right)',$$

respectively, with the fact that $\sum_{k=m+1}^{i}(1-\lambda)^{i-k} = \left[1-(1-\lambda)^{i-m}\right]/\lambda$.

The test statistic constructed upon the predictive densities in Eq. (4) performs in the similar way to the likelihood ratio test but with higher weights for more recent observations. It thus provides an evidence that a new observation $\mathbf{y}_t$, for $t = m+1, \ldots$, is out of control if its statistic $T_\lambda(t) > U_t$, where U_t is chosen based on an arbitrarily preselected in-control ARL and changes with t. Note that: as $c \to \infty$, a multivariate Gaussian EWMA control chart obtains, but where the density is used to determine process control rather than the mean. The computational efficiency is achieved by its recursive format, see Proposition 1 given as follows.

Proposition 1. *The statistic $T_\lambda(t)$ defined in Eq. (4) has a recursive format*

$$T_\lambda(t) = R_\lambda(t) + (1-\lambda)T_\lambda(t-1), \tag{5}$$

here $R_\lambda(t) = \left| \log \frac{p_\lambda(\mathbf{y}_t|\mathbf{y}_{m+1:t}, c_{1t}, \boldsymbol{\mu}_t^\lambda, \boldsymbol{\Sigma}_t^\lambda)}{p_0(\mathbf{y}_t|\mathbf{y}_{1:t-1}, c_{0t}, \boldsymbol{\mu}_t^0, \boldsymbol{\Sigma}_t^0)} \right|$, *for* $t = m+1, \ldots,$ *with the initial* $T_\lambda(m) = 0$.

The proof is trivial and thus we ignore it here, but one can refer to the article by Lowry et al. (1992) for details. By the following theorem, we found that as $m \to \infty$ and $t \to \infty$, $R_\lambda(t)$ computed upon the observations collected from any continuous underlying process will converge in probability to $R_\lambda(t)$ obtained based on the observations from a Gaussian process. It thus suggests an efficient way to obtain the simulated control limits, i.e., for a large t, we could use the value of $R_\lambda(t)$ just from the Gaussian underlying to help obtain the test statistics for any other continuous underlying processes.

Theorem 1. *Let*

$$R_\lambda(i) = \left| \log \frac{p_\lambda(\mathbf{y}_i \mid \mathbf{y}_{m+1:i}, c_{1i}, \boldsymbol{\mu}_i^\lambda, \boldsymbol{\Sigma}_i^\lambda)}{p_0(\mathbf{y}_i \mid \mathbf{y}_{1:i-1}, c_{0i}, \boldsymbol{\mu}_i^0, \boldsymbol{\Sigma}_i^0)} \right|$$

obtained upon observations $\mathbf{y}_1, \ldots, \mathbf{y}_m, \ldots, \mathbf{y}_t, \ldots$ *from a d-dimensional Gaussian distribution* N_d. *Similarly, let* $R_\lambda^*(i)$ *be obtained upon observations* $\mathbf{y}_1^*, \ldots, \mathbf{y}_m^*, \ldots, \mathbf{y}_t^*, \ldots$ *from a d-dimensional distribution* F_d *which is different from* N_d. *Given* N_d *and* F_d *are known, as* $i \to \infty$, $R_\lambda^*(i)$ *converges in probability to* $R_\lambda(i)$.

The proof is given in Appendix.

3.2 Implementation issues

Control limits

Given a false alarm rate α according to the in-control ARL_0, i.e., $\text{ARL}_0 = 1/\alpha$, the control limit U_t can be approximated (Hawkins et al., 2003; Zhou et al., 2009; and Zou and Tsung, 2010) by

$$\alpha = \begin{cases} Pr(T_\lambda(t) > U_t | T_\lambda(m) = 0); & t = m+1 \\ Pr(T_\lambda(t) > U_t | \bigcap_{j=m+1}^{t-1} T_\lambda(j) \le U_j); & t > m+1. \end{cases} \tag{6}$$

Note, usually in order to calculate the values of U_t for $m+1 \le t \le m+d$, one could take the last d observations from the historical data as the pseudo-future observations, see Zou and Tsung (2010). Then those d historical observations together with the observations in Phase II are sufficiently used to obtain the control limits, U_t, for $t = m+1, \ldots,$. In our simulation studies, we instead simulate in-control samples of size $m+d$ with the extra d data points combined with the observations in Phase II for calculating the control limits.

In Table 1, we provide the simulated control limits for a multivariate Gaussian process, $\mathcal{N}_d(\boldsymbol{\mu} = \mathbf{0}_{d\times 1}, \boldsymbol{\Sigma} = \mathbf{I}_{d\times d})$, and a multivariate Student-t process, $\mathcal{T}_d(\boldsymbol{\mu} = \mathbf{0}_{d\times 1}, \boldsymbol{\Sigma} = \mathbf{I}_{d\times d}, \text{d.f.}=10)$, based on the in-control $\text{ARL}_0 = 200$ and the different combinations of d, λ, and m, i.e., $(d = 3, \lambda = 0.05, m = 100)$, $(d = 3, \lambda = 0.2, m = 300)$, and $(d = 5, \lambda = 0.1, m = 100)$. The simulated control limits are based on the total of 10,000 in-control samples. We observed the control limit increases as t goes up till it will be fixed at a certain t. Following the discussion in Zou and Tsung (2010), we therefore suggest to compute the first 200 control limits and the last one in this sequence is then used for the remains. In addition, the control limits do depend on the in-control mean and covariance, and thus one may need to simulate the control limits for each particular setup to guarantee to achieve the best performance of the chart. The control limits for other in-control setups are available from the authors upon request.

Diagnostic

In addition to detecting the change point of a process, in practice, it is also critical to diagnose a change in the multivariate process of interest. Reynolds and Cho (2006) pointed out that in today's environment of multivariate monitoring, control charts are almost always plotted by computer, so additional control charts or other plots can be called up when needed for diagnostic. But it is not that easy since a number of characteristics are involved at one time and also correlations exist among them. The identification of out-of-control characteristics after warning signals has been an interesting topic for many researchers, see Alt (1985), Jackson (1991), Hayter and Tsui (1994), Sepulveda and Nachlas (1997), Niaki and Abbasi (2005), and Zou et al. (2013). Among them, Zou et al. (2013) presented an appealing lasso-based diagnostic framework for multivariate statistical process control, which is also adaptive to our proposed chart for diagnosing the changes. Briefly, let $\mathbf{Z}_1$ and $\mathbf{Z}_2$ be two sets of independent observations before and after a parameter change in the monitored process with the mean vector of the in-control $\boldsymbol{\mu}_1$ and out-of-control $\boldsymbol{\mu}_2$, respectively. Further let $\boldsymbol{\mu}_2 = \boldsymbol{\mu}_1 + \boldsymbol{\delta}$. Zou et al. (2013) pointed out that in many applications, it is rare that all parameters shift at the same time, and thus some components of $\boldsymbol{\delta}$ are zero. They thus presented a method for determining which components of $\boldsymbol{\delta}$ are not zero by using a lasso-based method to shrink all nuisance components to be exactly equal to zero. Consequently, the non-zero components give a hint on which parameters are shifted. For each component of $\boldsymbol{\delta}$, if one considers it either zero or not, the full model space contains total $2^d - 1$ candidate models, where d is the multivariate dimension. Thus they further provided a model comparison method within the Bayesian framework based on posterior probability derived from their proposed pseudo-likelihood, which eventually leads to minimize an extended family of BIC (EBIC), see Chen and Chen (2008). For details, one can refer to Zou et al. (2013), but this is beyond our scope for this article.

Table 1. The Simulated Control Limits, U_t, for $t = m + k$ and $k = 1, 2, \ldots, 300$

	$d = 3, \lambda = 0.05, m = 100$		$d = 3, \lambda = 0.2, m = 300$		$d = 5, \lambda = 0.1, m = 100$	
k	$\mathcal{N}_d$	$\mathcal{T}_d$	$\mathcal{N}_d$	$\mathcal{T}_d$	$\mathcal{N}_d$	$\mathcal{T}_d$
1	18.928	23.600	19.377	23.471	28.821	35.622
2	30.973	34.220	29.309	33.501	48.009	55.404
3	41.428	45.369	36.920	41.059	63.816	69.345
4	50.607	54.753	42.627	47.088	77.687	84.582
5	59.097	62.710	46.847	50.905	89.600	96.635
6	66.694	70.643	50.031	53.916	99.943	106.817
7	73.845	78.234	52.253	57.255	109.599	116.395
8	80.723	85.505	54.153	59.231	117.464	125.593
9	87.003	91.534	55.440	60.866	124.911	132.432
10	92.916	96.817	56.773	61.980	131.255	138.222
11	98.415	102.692	57.921	62.329	136.760	144.741
12	103.290	107.859	58.078	62.620	142.064	149.956
13	108.237	113.365	59.025	63.497	146.797	154.078
14	112.692	117.550	59.359	63.652	150.601	160.077
15	117.250	121.356	59.578	63.706	154.577	162.530
16	121.111	125.671	59.892	63.817	157.661	164.423
17	124.825	129.177	60.099	63.966	160.590	167.949
18	128.267	133.057	60.260	64.176	162.688	170.538
19	131.608	136.586	60.471	64.305	164.982	173.441
20	134.773	138.928	60.521	64.453	167.232	174.789
22	140.466	144.487	60.627	64.539	170.409	179.613
25	147.731	151.957	60.782	64.653	174.401	181.041
28	153.928	158.291	60.831	64.718	177.141	185.644
30	157.383	161.185	60.933	64.810	178.172	186.108
35	164.630	168.417	60.967	64.898	180.059	187.310
40	169.714	173.884	61.016	65.007	181.557	188.124
50	176.646	181.072	61.098	65.106	182.005	188.984
60	180.491	184.050	61.189	65.198	182.341	189.792
70	182.579	186.257	61.263	65.287	182.687	190.266
80	183.449	187.047	61.313	65.366	183.019	190.447
90	184.077	187.586	61.352	65.442	183.210	190.627
100	184.385	188.412	61.382	65.695	183.357	190.754
120	184.535	188.366	61.427	65.732	183.455	190.887
140	184.656	188.172	61.458	65.774	183.525	190.993
160	184.775	188.885	61.492	65.801	183.603	191.175
180	184.878	188.923	61.538	65.828	183.664	191.207
200	184.965	188.954	61.546	65.838	183.680	191.225
230	184.987	188.972	61.546	65.839	183.687	191.283
260	184.987	188.973	61.546	65.839	183.687	191.283
300	184.987	188.973	61.546	65.839	183.687	191.283

4. SIMULATION STUDIES

To investigate the performance of the proposed multivariate Polya tree EWMA (MPTEWMA) chart, we conduct simulation studies under several shifting scenarios for multivariate Gaussian $\mathcal{N}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and multivariate Student-t $\mathcal{T}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \text{d.f.} = 10)$ processes. Let the in-control mean vector and covariance matrix be $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$ and the out-of-control mean vector and covariance matrix be $\boldsymbol{\mu}_1$ and $\boldsymbol{\Sigma}_1$. The different types of shifts considered here are: (a) mean shift only; (b) variance shift only; and (c) mean and variance shift.

Under the scenario (a), we consider two different situations: (i) all variables have changes in the location, i.e., $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_0 + \delta \cdot (\sigma_{01}, \sigma_{02}, \ldots, \sigma_{0d})'$; and (ii) we arbitrarily choose to shift the location of the second variable X_2, i.e., $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_0 + \delta \cdot (0, \sigma_{02}, 0, \ldots)'$. We set the in-control ARL as $\text{ARL}_{\text{ic}} = 200$ so that the false alarm rate $\alpha = 0.005$. For comparison purposes, we also include the simulation results from two popular multivariate charts: the multivariate EWMA (MEWMA) chart and the multivariate CUSUM (MCUSUM) chart. Specifically, the MEWMA and MCUSUM charts could be obtained by using a built-in R function called *mult.chart()* in the R package **MSQC** (Santos-Fernandez, 2016). For choosing the design parameters of an MCUSUM chart, one can refer to the articles by Healy (1987) and Pignatiello and Runger (1990). The sim-

ulated ARL results based on 2,000 datasets are reported in Table 2 with different in-control $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$ for three combinations of d, λ, and m; i.e., $(d = 3, \lambda = 0.05, m = 100)$, $(d = 3, \lambda = 0.2, m = 300)$, and $(d = 5, \lambda = 0.1, m = 100)$. δ is considered from the set of $\{0, 0.5, 1, 1.5\}$ with $\delta = 0$ for in-control purpose. The simulated control limits do depend on the in-control mean and covariance, but to investigate the effects of using the same control limits for different in-control setups, in Table 2 [C], we report two results under a Gaussian process for different combinations of d, λ, and m. The column labeled as $\mathcal{N}_d$ gives the ARLs based on the control limits simulated upon its own in-control parameters, i.e., $\boldsymbol{\mu}_0 = (1, 4, 8)'$ and $\boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}_0^*$, where

$$\boldsymbol{\Sigma}_0^* = \begin{pmatrix} 1 & 1.8 & -0.6 \\ 1.8 & 4 & -2.4 \\ -0.6 & -2.4 & 9 \end{pmatrix},$$

and the column labeled as $\mathcal{N}_d^*$ provides the ARLs obtained upon the control limits from the in-control $\boldsymbol{\mu}_0 = \mathbf{0}_{d\times 1}$ and $\boldsymbol{\Sigma}_0 = \mathbf{I}_{d\times d}$. The ARL results shown in both columns ($\mathcal{N}_d$ and $\mathcal{N}_d^*$) for the MPTEWMA chart are very close to each other when the underlying process is a Gaussian. However we found it will not be always true for a non-Gaussian process. Thus, we suggest, in order to better monitor a process, one should simulate the control limits for that particular process, especially when the process is not Gaussian distributed. Otherwise, false alarms will be triggered often. Apparently, the MEWMA and MCUSUM charts work better than the MPTEWMA chart when the monitored process follows a Gaussian distribution, see the scenario (a) [A], [B], and [C] for a Gaussian underlying. However, when processes are non-Gaussian distributed, see the scenario (a) [A] for a Student-t, the MPTEWMA performs better. But it performs not as good as the MEWMA and MCUSUM charts in the scenario (a) [B] when a mean shift occurs only on one variable, it is possibly due to a small sample size of Phase I since our proposed nonparametric method needs more data to precisely estimate the chart parameters. We also found the chart parameter λ works fine when it takes a value from the set of $\{0.05, 0.1, 0.2\}$. Those values are also recommended in other articles, see Lowry et al. (1992).

Under the scenario (b), we also consider two different situations: (i) all variables have changes in their variances, i.e., $\sigma_{1s} = \sqrt{\delta}\sigma_{0s}$ for $s = 1, 2, \ldots, d$; and (ii) we arbitrarily choose that a variance change occurs at the third variable X_3, i.e., $\sigma_{13} = \sqrt{\delta}\sigma_{03}$. We report the simulated ARL results in Table 3 again with different in-control $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$ for three combinations of d, λ, and m. δ is considered from the set of $\{1, 2, 4, 6\}$ with $\delta = 1$ for in-control purpose. Similar to what we did for the mean shift only scenario, in Table 3 [C], we also report two results for a Gaussian process with different combinations of d, λ, and m. Again, we can observe the same phenomenon that the ARLs of the MPTEWMA chart are quite close to each other even if the control limits are obtained from different in-control setups. As before, this conclusion does not usually hold for a non-Gaussian process. From Table 3, obviously, we could tell that our proposed method works the best in detecting variance shifts for either Gaussian processes or non-Gaussian processes. This again shows our nonparametric method can relax certain restrictive assumptions on the underlying processes and thus it is more robust for various types of processes.

Under the scenario (c), we consider that the underlying process is a multivariate Student-t, $\mathcal{T}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \text{d.f.}=3)$, with $d = 5$. The in-control mean vector and covariance matrix are setup as

$$\boldsymbol{\mu}_0 = (2, 5, 0, 10, 4)' \quad \text{and}$$

$$\boldsymbol{\Sigma}_0^{**} = \begin{pmatrix} 4 & 4 & 0 & -2 & 6.4 \\ 4 & 16 & 2.4 & 4 & 4.8 \\ 0 & 2.4 & 1 & 3.5 & 0 \\ -2 & 4 & 3.5 & 25 & 2 \\ 6.4 & 4.8 & 0 & 2 & 16 \end{pmatrix},$$

respectively. We arbitrarily choose to change the mean of X_2 and the variance of X_5, i.e., $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_0 + \delta \cdot (0, \sigma_{02}, 0, 0, 0)'$, $\sigma_{15} = \sqrt{\delta^*}\sigma_{05}$, and others leave the same as the in-control setups. We consider five different combinations of δ and δ^* including $\delta = 0$ and $\delta^* = 1$ for the in-control status. The simulated ARL results are reported in Table 4. Obviously, the in-control ARLs of the MEWMA and MCUSUM charts, i.e., 88.38 and 128.68, respectively, are much smaller than 200, which suggests warning signals will be falsely triggered often when no shift occurs. In contrast, the proposed method works perfectly for the in-control status.

We should note that for a non-Gaussian distributed process, the proposed chart usually takes longer to send a warning signal. For example, in Table 2 [B] for $(d = 3, \lambda = 0.05, m = 100)$, when the mean of X_2 shifts to the right by 0.5, the ARL is 99.62 under a Gaussian process compared to the value of 144.27 for a Student-t. The key reason is that the proposed chart considered the Polya tree centered at a Gaussian, it thus can more efficiently detect the out-of-control status for a Gaussian process. Consequently, a pilot study on dataset is preferred in order to find a better initial guess (the centering parametric distribution) on the process to achieve a better monitoring.

To better illustrate our proposed method also suitable to moderate dimensions, we give one more simulation with the dimension $d = 20$ and the underlying process follows a Student's t-distribution with the degree of freedom equal to 10. We choose $\lambda = 0.1$ and $m = 100$. The covariance matrix is set up such as the diagonal entries equal to 1 and the off-diagonal entries is obtained by $\text{cov}(i, j) = 0.5^{|j-i|}$, where $\text{cov}(i, j)$ denotes the entry of the covariance matrix at the row i and the column j. We arbitrarily choose to shift the variance of X_2 by $\sigma_2^* = \sqrt{\delta}\sigma_2$ with $\delta = \{1, 2, 3, 4\}$. The ARL results are given in Table 5. Again, the proposed method performs the best. In other words, compared to the MEWMA and MCUSUM charts, the proposed method can

Table 2. Scenario (a): Mean Shift Only

[A]: Mean Shift On All Variables: $\boldsymbol{\mu}_0 = \mathbf{0}_{d\times 1}$ and $\boldsymbol{\Sigma}_0 = \mathbf{I}_{d\times d}$

$d = 3, \lambda = 0.05, m = 100$

	MPTEWMA		MEWMA		MCUSUM	
δ	$\mathcal{N}_d$	$\mathcal{T}_d$	$\mathcal{N}_d$	$\mathcal{T}_d$	$\mathcal{N}_d$	$\mathcal{T}_d$
0.0	200.16	201.12	201.64	202.87	198.99	201.64
0.5	27.28	60.34	13.00	83.53	12.55	74.68
1.0	7.18	13.38	4.01	24.81	4.23	24.89
1.5	3.66	6.87	2.25	7.67	2.31	7.71

$d = 3, \lambda = 0.2, m = 300$

	MPTEWMA		MEWMA		MCUSUM	
δ	$\mathcal{N}_d$	$\mathcal{T}_d$	$\mathcal{N}_d$	$\mathcal{T}_d$	$\mathcal{N}_d$	$\mathcal{T}_d$
0.0	199.45	200.28	198.63	203.39	198.34	202.32
0.5	20.54	61.13	15.16	80.05	11.72	73.39
1.0	6.14	15.58	4.21	25.06	4.01	24.41
1.5	3.08	6.23	2.35	7.73	2.13	7.61

$d = 5, \lambda = 0.1, m = 100$

	MPTEWMA		MEWMA		MCUSUM	
δ	$\mathcal{N}_d$	$\mathcal{T}_d$	$\mathcal{N}_d$	$\mathcal{T}_d$	$\mathcal{N}_d$	$\mathcal{T}_d$
0.0	200.48	200.49	199.66	201.95	203.02	202.78
0.5	23.56	99.58	10.99	101.54	10.24	101.11
1.0	6.68	22.40	3.41	24.04	3.38	24.11
1.5	3.70	9.50	1.80	10.28	1.83	10.32

[B]: Mean Shift Only On X_2: $\boldsymbol{\mu}_0 = \mathbf{0}_{d\times 1}$ and $\boldsymbol{\Sigma}_0 = \mathbf{I}_{d\times d}$

$d = 3, \lambda = 0.05, m = 100$

	MPTEWMA		MEWMA		MCUSUM	
δ	$\mathcal{N}_d$	$\mathcal{T}_d$	$\mathcal{N}_d$	$\mathcal{T}_d$	$\mathcal{N}_d$	$\mathcal{T}_d$
0.0	200.13	200.17	202.44	199.44	199.27	203.87
0.5	99.62	144.27	31.20	94.57	36.54	80.34
1.0	22.30	126.85	9.83	31.12	10.40	31.63
1.5	9.73	46.27	5.14	8.85	5.21	9.06

$d = 3, \lambda = 0.2, m = 300$

	MPTEWMA		MEWMA		MCUSUM	
δ	$\mathcal{N}_d$	$\mathcal{T}_d$	$\mathcal{N}_d$	$\mathcal{T}_d$	$\mathcal{N}_d$	$\mathcal{T}_d$
0.0	199.36	199.41	203.51	199.45	198.73	202.87
0.5	97.57	179.01	41.976	85.67	33.21	84.14
1.0	27.24	135.04	11.494	33.92	9.29	30.46
1.5	9.81	68.55	5.44	9.36	5.24	8.81

$d = 5, \lambda = 0.1, m = 100$

	MPTEWMA		MEWMA		MCUSUM	
δ	$\mathcal{N}_d$	$\mathcal{T}_d$	$\mathcal{N}_d$	$\mathcal{T}_d$	$\mathcal{N}_d$	$\mathcal{T}_d$
0.0	199.82	200.57	201.53	198.01	202.67	197.44
0.5	144.25	178.02	51.92	107.76	42.52	105.73
1.0	61.30	154.94	12.70	34.55	12.86	33.24
1.5	19.05	110.98	6.19	10.30	6.17	10.20

[C]: Mean Shift Only On X_2: $\boldsymbol{\mu}_0 = (1, 4, 8)'$ and $\boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}_0^*$

	$d = 3, \lambda = 0.05, m = 100$				$d = 3, \lambda = 0.2, m = 300$			
	MPTEWMA		MEWMA	MCUSUM	MPTEWMA		MEWMA	MCUSUM
δ	$\mathcal{N}_d$	$\mathcal{N}_d^*$	$\mathcal{N}_d$	$\mathcal{N}_d$	$\mathcal{N}_d$	$\mathcal{N}_d^*$	$\mathcal{N}_d$	$\mathcal{N}_d$
0.0	199.94	197.31	201.89	198.79	200.24	205.23	200.01	202.68
0.5	12.30	12.26	6.25	6.51	12.31	12.33	6.66	6.43
1.0	3.64	3.63	2.14	2.37	2.87	2.88	2.05	2.25
1.5	1.97	1.97	1.24	1.33	1.54	1.55	1.22	1.34

Table 3. Scenario (b): Variance Shift Only

[A]: Variance Shift On All Variables: $\boldsymbol{\mu}_0 = \mathbf{0}_{d\times 1}$ and $\boldsymbol{\Sigma}_0 = \mathbf{I}_{d\times d}$

$d = 3, \lambda = 0.05, m = 100$

	MPTEWMA		MEWMA		MCUSUM	
$\sqrt{\delta}$	$\mathcal{N}_d$	$\mathcal{T}_d$	$\mathcal{N}_d$	$\mathcal{T}_d$	$\mathcal{N}_d$	$\mathcal{T}_d$
$\sqrt{1}$	200.05	200.35	202.39	202.85	199.42	203.88
$\sqrt{2}$	20.04	30.79	28.27	39.57	22.04	31.05
$\sqrt{4}$	5.18	8.07	5.72	8.77	5.54	8.19
$\sqrt{6}$	3.24	4.69	3.53	4.73	3.27	4.70

$d = 3, \lambda = 0.2, m = 300$

	MPTEWMA		MEWMA		MCUSUM	
$\sqrt{\delta}$	$\mathcal{N}_d$	$\mathcal{T}_d$	$\mathcal{N}_d$	$\mathcal{T}_d$	$\mathcal{N}_d$	$\mathcal{T}_d$
$\sqrt{1}$	199.76	200.48	201.15	195.55	198.36	201.64
$\sqrt{2}$	14.65	25.74	17.42	25.87	19.12	25.02
$\sqrt{4}$	4.36	7.23	4.86	7.99	4.56	7.91
$\sqrt{6}$	2.78	4.31	3.20	4.55	3.29	4.51

$d = 5, \lambda = 0.1, m = 100$

	MPTEWMA		MEWMA		MCUSUM	
$\sqrt{\delta}$	$\mathcal{N}_d$	$\mathcal{T}_d$	$\mathcal{N}_d$	$\mathcal{T}_d$	$\mathcal{N}_d$	$\mathcal{T}_d$
$\sqrt{1}$	200.11	200.25	203.66	203.68	201.37	204.73
$\sqrt{2}$	14.17	32.10	16.17	32.81	20.91	33.77
$\sqrt{4}$	3.89	7.46	4.20	8.11	4.15	7.55
$\sqrt{6}$	2.45	4.38	2.80	4.78	2.66	4.64

[B]: Variance Shift Only On X_3: $\boldsymbol{\mu}_0 = \mathbf{0}_{d\times 1}$ and $\boldsymbol{\Sigma}_0 = \mathbf{I}_{d\times d}$

$d = 3, \lambda = 0.05, m = 100$

	MPTEWMA		MEWMA		MCUSUM	
$\sqrt{\delta}$	$\mathcal{N}_d$	$\mathcal{T}_d$	$\mathcal{N}_d$	$\mathcal{T}_d$	$\mathcal{N}_d$	$\mathcal{T}_d$
$\sqrt{1}$	199.64	200.33	202.97	202.64	202.77	203.60
$\sqrt{2}$	73.86	97.79	94.81	102.87	80.16	105.72
$\sqrt{4}$	15.93	34.55	25.72	36.26	22.23	36.68
$\sqrt{6}$	8.97	18.15	14.62	18.74	10.91	19.17

$d = 3, \lambda = 0.2, m = 300$

	MPTEWMA		MEWMA		MCUSUM	
$\sqrt{\delta}$	$\mathcal{N}_d$	$\mathcal{T}_d$	$\mathcal{N}_d$	$\mathcal{T}_d$	$\mathcal{N}_d$	$\mathcal{T}_d$
$\sqrt{1}$	200.20	200.72	200.18	197.10	202.53	200.82
$\sqrt{2}$	53.02	87.67	61.56	99.27	58.78	96.96
$\sqrt{4}$	12.18	32.98	14.3	41.61	18.75	41.12
$\sqrt{6}$	7.12	15.62	8.47	17.42	10.22	17.03

$d = 5, \lambda = 0.1, m = 100$

	MPTEWMA		MEWMA		MCUSUM	
$\sqrt{\delta}$	$\mathcal{N}_d$	$\mathcal{T}_d$	$\mathcal{N}_d$	$\mathcal{T}_d$	$\mathcal{N}_d$	$\mathcal{T}_d$
$\sqrt{1}$	199.94	200.76	201.36	201.54	203.86	199.53
$\sqrt{2}$	92.14	103.73	94.51	119.22	95.76	108.74
$\sqrt{4}$	28.42	69.31	33.75	76.49	36.02	75.12
$\sqrt{6}$	13.23	43.30	17.36	50.95	17.64	44.48

[C]: Variance Shift Only On X_3: $\boldsymbol{\mu}_0 = (1, 4, 8)'$ and $\boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}_0^*$

	$d = 3, \lambda = 0.05, m = 100$				$d = 3, \lambda = 0.2, m = 300$			
	MPTEWMA		MEWMA	MCUSUM	MPTEWMA		MEWMA	MCUSUM
$\sqrt{\delta}$	$\mathcal{N}_d$	$\mathcal{N}_d^*$	$\mathcal{N}_d$	$\mathcal{N}_d$	$\mathcal{N}_d$	$\mathcal{N}_d^*$	$\mathcal{N}_d$	$\mathcal{N}_d$
$\sqrt{1}$	199.70	198.69	201.65	202.17	200.21	204.55	196.78	203.02
$\sqrt{2}$	66.19	65.57	68.63	69.98	40.99	42.65	45.04	47.54
$\sqrt{4}$	13.73	13.49	18.42	15.75	10.00	10.02	10.14	14.15
$\sqrt{6}$	7.58	7.57	10.72	8.93	5.33	5.32	6.72	6.67

Table 4. Scenario (c): Mean and Variance Shift

[A]: Mean Shift On X_2 and Variance Shift On X_5: $\boldsymbol{\mu}_0 = (2,5,0,10,4)'$ and $\boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}_0^{**}$					
$d = 5, \lambda = 0.1, m = 100$					
$(\delta, \sqrt{\delta^*})$	$(0, \sqrt{1})$	$(0.5, \sqrt{2})$	$(0.5, \sqrt{6})$	$(1, \sqrt{4})$	$(1.5, \sqrt{2})$
MPTEWMA-ARL	199.67	45.09	29.96	24.78	11.37
MEWMA-ARL	88.38	9.16	3.82	2.85	2.11
MCUSUM-ARL	128.68	13.43	7.74	5.66	4.46

Table 5. A Moderate Dimension

$d = 20, \lambda = 0.1, m = 100$			
$\sqrt{\delta}$	MEWMA	MCUSUM	MPTEWMA
$\sqrt{1}$	202.31	201.56	203.91
$\sqrt{2}$	121.43	135.44	117.14
$\sqrt{3}$	67.59	84.70	57.58
$\sqrt{4}$	43.29	60.38	35.05

quickly trigger a warning signal when an out-of-control occurs.

5. EXAMPLES

5.1 A simulated example

In this section, we will illustrate the proposed chart for monitoring the health condition of human beings. We assume that we are interested in monitoring four characteristics: RBC (Red Blood Corpuscle) Count, MCHC (Mean Corpuscular Hemoglobin Concentration), VA (Visual Acuity), and BUN (Blood Urea Nitrogen). Briefly, RBC count measures the number of red blood cells, and the normal RBC range is $4.2 - 6.1 \times 10^{10}$ $cells/cL$ (cells per centiliter). Their logarithm values are thus in the range of $24.4 - 24.8$. MCHC measures the concentration of hemoglobin in a given volume of packed red blood cells, and its normal range is $3.0 - 3.6 \times 10^4$ mg/dL (milligrams per deciliter). We also consider the values of their logarithm in our model, i.e., $10.3 - 10.5$. VA commonly refers to the clarity of vision and usually is measured as a fraction from the eye chart. BUN provides the important information for one's kidney and liver function, and the normal range for the amount of urea nitrogen in blood is $9 - 20$ mg/dL.

According to the illustration on the above four characteristics, we simulate 300 data points from a multivariate Student-t underlying, $\mathcal{T}_4(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \text{d.f.}=12)$, for Phase I using the in-control mean vector and covariance matrix as

$$\boldsymbol{\mu}_0 = (24.6, 10.4, 2, 16), \boldsymbol{\Sigma}_0 = \begin{pmatrix} 25 & 18 & 0 & -6 \\ 18 & 16 & 0 & -4 \\ 0 & 0 & 1 & 0.2 \\ -6 & -4 & 0.2 & 4 \end{pmatrix},$$

respectively. Based on the facts, we assume a positive relationship between the pairs of RBC count and MCHC, and VA and BUN, a negative relationship between the pairs of RBC count and BUN, MCHC and BUN, and no relationship between the pairs of RBC count and VA, MCHC and VA in the covariance matrix. The testing sample of size 100 in Phase II is generated the same way as the first 80 data points from the in-control process, and starting from the 81st data point, changes in the location of RBC count occur, which means an unhealthy condition is observed. We further assume RBC count increases by $0.2\sigma = 0.2\sqrt{25} = 1$ on the logarithm scale. As usual, the in-control ARL are set as 200 and let $\lambda = 0.1$. The control limits of the proposed chart are obtained based on 10,000 in-control samples, and the control limits of MEWMA and MSEWMA are 12.723 and 11.896, respectively. To investigate the performance of the proposed chart (MPTEWMA), we compare it to the multivariate EWMA (MEWMA) chart by Lowry et al. (1992) and the multivariate sign EWMA (MSEWMA) chart by Zou and Tsung (2011) and report the results in Figure 1, Figure 2, and Figure 3.

Obviously, compared to the MEWMA and MSEWMA charts, the proposed chart performs the best in the sense that it can quickly detect the changes and thus send signal warnings when an operation problem occurs in the process. Specifically, from Figure 3, the proposed chart starts to correctly send warnings from the 84th observation. However, in contrast, the MEWMA chart can not detect any shift and thus no signal is triggered, see Figure 1. Instead of warning the out-of-control status, the MSEWMA chart actually has two false nuisance alarms for the 27th and the 76th observations, see Figure 2.

5.2 Real data application

In this section, we use the proposed chart to monitor a real chemical process (Montgomery, 2009). We are interested in simultaneously monitoring four characteristics. 20 data points are collected in Phase I, see Table 6, and the in-control mean vector and correlation matrix are

$$\begin{aligned} \boldsymbol{\mu}_0 &= (9.955, 20, 14.68, 15.765) \\ \boldsymbol{\Sigma}_0 &= \begin{pmatrix} 1.0000 & 0.9302 & 0.2060 & 0.3595 \\ 0.9302 & 1.0000 & 0.1669 & 0.4502 \\ 0.2060 & 0.1669 & 1.0000 & 0.3439 \\ 0.3595 & 0.4502 & 0.3439 & 1.0000 \end{pmatrix}, \end{aligned}$$

respectively. Following the discussion in Montgomery (2009), the underlying process is assumed to be a multivari-

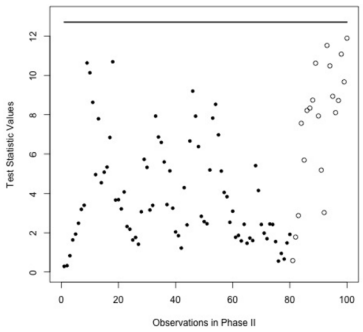

Figure 1. MEWMA.

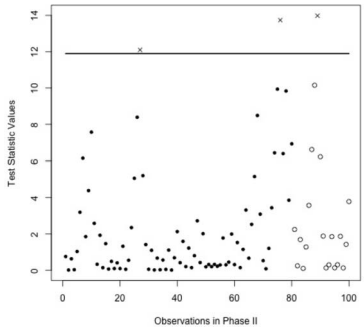

Figure 2. MSEWMA.

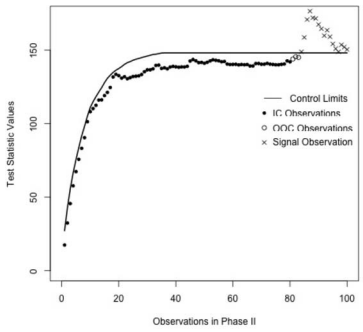

Figure 3. MPTEWMA.

Table 6. Chemical Process in Phase I

Observation	x_1	x_2	x_3	x_4	Observation	x_1	x_2	x_3	x_4
1	10	20.7	13.6	15.5	11	10.5	20.3	17	16.5
2	10.5	19.9	18.1	14.8	12	9.2	19	11.5	16.3
3	9.7	20	16.1	16.5	13	11.3	21.6	14	18.7
4	9.8	20.2	19.1	17.1	14	10	19.8	14	15.9
5	11.7	21.5	19.8	18.3	15	8.5	19.2	17.4	15.8
6	11	20.9	10.3	13.8	16	9.7	20.1	10	16.6
7	8.7	18.8	16.9	16.8	17	8.3	18.4	12.5	14.2
8	9.5	19.6	13.6	14.5	18	11.9	21.8	14.1	16.2
9	10.1	19.4	16.2	15.8	19	10.3	20.5	15.6	15.1
10	9.5	19.6	13.6	14.5	20	8.9	19	8.5	14.7

Table 7. The Fitted Control Charts

					MPTEWMA		MEWMA	MSEWMA
t	x_1	x_2	x_3	x_4	U_t	$T_\lambda(t)$	T_t^2	Q_t
1	9.9	20	15.4	15.9	27.877	13.865	0.033	0.760
2	8.7	19	9.9	16.8	43.188	30.355	2.162	0.010
3	11.5	21.8	19.3	12.1	55.994	53.262	4.668	0.624
4	15.9	24.6	14.7	15.3	67.599	82.978	21.473	2.058
5	12.6	23.9	17.1	14.2	76.412	386.325	245.867	4.009
6	14.9	25	16.3	16.6	84.863	365.050	161.950	7.046
7	9.9	23.7	11.9	18.1	92.138	344.028	34.664	2.890
8	12.8	26.3	13.5	13.7	98.733	327.653	80.557	2.284
9	13.1	26.1	10.9	16.8	103.933	310.868	170.454	2.328
10	9.8	25.8	14.8	15	109.197	296.950	424.396	4.438

* The control limits for the MEWMA and MSEWMA charts are 12.723 and 11.896, respectively.

ate Gaussian; we thus use the Henze-Zirkler's approach to test the normality assumption and obtain the p-value equal to 0.974, which again supports Montgomery's discussion. We further compare the proposed chart to the MEWMA and MSEWMA charts with the smoothing parameter $\lambda = 0.1$ and the in-control $\text{ARL}_0 = 200$. The fitted results along with the 10 observations in Phase II are reported in Table 7.

Obviously, the proposed chart starts to trigger warnings from the 4th observation, coincides with the fact that the process went out-of-control starting from the 4th one. The same situation is also observed from the MEWMA chart with its control limit 12.723. It turns out that for a multivariate Gaussian process, the proposed nonparametric chart can perform as well as the parametric MEWMA chart. However, in contrast, the nonparametric MSEWMA chart fails to detect changes with all test statistics smaller than its control limit 11.896. This shows that the proposed nonparametric chart is more stable compared to the MSEWMA chart when the in-control sample has a small size.

6. CONCLUSIONS

We proposed a multivariate nonparametric control chart for detecting changes in the process using a novel weighted

Polya tree. Simulation results and examples showed that the proposed chart performs well to detect various types of shifting for Gaussian or non-Gaussian processes. However, one disadvantage of the proposed chart is that the chart itself is nonparametric instead of distribution-free. Thus, we suggest: to better screen a process, one should simulate the control limits for each monitored process, especially when the process is not Gaussian distributed.

In addition, there are several issues that are not thoroughly addressed in this paper, which could be appealing in practice. For example: (1) although the proposed chart can quickly detect changes for mean and variance shifts in the monitored characteristics, it does not clearly address which one or more characteristics contribute to signal warnings; (2) we assume that the observations in a process is stochastically independent in our simulation studies, but we are not sure the use of the chart for monitoring the autocorrelated process since it is well known that the traditional charts designed for stochastically independent observations may signal incorrectly and thus weaken the effectiveness of detecting shifts when the underlying process is autocorrelated (Chen, 2017; Chen and Hanson, 2017); and (3) our proposed chart is constructed for single change-point cases, however, we don't clearly know if this chart is also suitable to multiple change-point cases which are quite common in these days. We need do more research to investigate this situation.

APPENDIX

Proof. Given N_d and F_d are known, it is equivalent to m being sufficiently large. Let $\epsilon > 0$, we then have

$$\begin{aligned}
&\lim_{\substack{m\to\infty\\ i\to\infty\\ i\gg m}} Pr\Big(\big|R^*_\lambda(i) - R_\lambda(i)\big| \ge \epsilon\Big) \\
&\quad= \lim_{\substack{m\to\infty\\ i\to\infty\\ i\gg m}} Pr\left(\left|\log \frac{p^*_\lambda\left(\mathbf{y}^*_i \mid \mathbf{y}^*_{m+1:i}, c^*_{1i}, \boldsymbol{\mu}^{\lambda*}_i, \boldsymbol{\Sigma}^{\lambda*}_i\right)}{p^*_0\left(\mathbf{y}^*_i \mid \mathbf{y}^*_{1:i-1}, c^*_{0i}, \boldsymbol{\mu}^{0*}_i, \boldsymbol{\Sigma}^{0*}_i\right)}\right|\right. \\
&\qquad\left. - \left|\log \frac{p_\lambda\left(\mathbf{y}_i \mid \mathbf{y}_{m+1:i}, c_{1i}, \boldsymbol{\mu}^{\lambda}_i, \boldsymbol{\Sigma}^{\lambda}_i\right)}{p_0\left(\mathbf{y}_i \mid \mathbf{y}_{1:i-1}, c_{0i}, \boldsymbol{\mu}^{0}_i, \boldsymbol{\Sigma}^{0}_i\right)}\right| \ge \epsilon\right).
\end{aligned}$$

By the triangle inequality,

$$\begin{aligned}
&\left|\log \frac{p^*_\lambda\left(\mathbf{y}^*_i \mid \mathbf{y}^*_{m+1:i}, c^*_{1i}, \boldsymbol{\mu}^{\lambda*}_i, \boldsymbol{\Sigma}^{\lambda*}_i\right)}{p^*_0\left(\mathbf{y}^*_i \mid \mathbf{y}^*_{1:i-1}, c^*_{0i}, \boldsymbol{\mu}^{0*}_i, \boldsymbol{\Sigma}^{0*}_i\right)}\right| \\
&\qquad - \left|\log \frac{p_\lambda\left(\mathbf{y}_i \mid \mathbf{y}_{m+1:i}, c_{1i}, \boldsymbol{\mu}^{\lambda}_i, \boldsymbol{\Sigma}^{\lambda}_i\right)}{p_0\left(\mathbf{y}_i \mid \mathbf{y}_{1:i-1}, c_{0i}, \boldsymbol{\mu}^{0}_i, \boldsymbol{\Sigma}^{0}_i\right)}\right| \\
&\quad\le \left|\log \frac{p^*_\lambda\left(\mathbf{y}^*_i \mid \mathbf{y}^*_{m+1:i}, c^*_{1i}, \boldsymbol{\mu}^{\lambda*}_i, \boldsymbol{\Sigma}^{\lambda*}_i\right)}{p^*_0\left(\mathbf{y}^*_i \mid \mathbf{y}^*_{1:i-1}, c^*_{0i}, \boldsymbol{\mu}^{0*}_i, \boldsymbol{\Sigma}^{0*}_i\right)}\right. \\
&\qquad\left. - \log \frac{p_\lambda\left(\mathbf{y}_i \mid \mathbf{y}_{m+1:i}, c_{1i}, \boldsymbol{\mu}^{\lambda}_i, \boldsymbol{\Sigma}^{\lambda}_i\right)}{p_0\left(\mathbf{y}_i \mid \mathbf{y}_{1:i-1}, c_{0i}, \boldsymbol{\mu}^{0}_i, \boldsymbol{\Sigma}^{0}_i\right)}\right|.
\end{aligned}$$

Thus,

$$\begin{aligned}
&\lim_{\substack{m\to\infty\\ i\to\infty\\ i\gg m}} Pr\Big(\big|R^*_\lambda(i) - R_\lambda(i)\big| \ge \epsilon\Big) \\
&\quad\le \lim_{\substack{m\to\infty\\ i\to\infty\\ i\gg m}} Pr\left(\left|\log \frac{p^*_\lambda\left(\mathbf{y}^*_i \mid \mathbf{y}^*_{m+1:i}, c^*_{1i}, \boldsymbol{\mu}^{\lambda*}_i, \boldsymbol{\Sigma}^{\lambda*}_i\right)}{p^*_0\left(\mathbf{y}^*_i \mid \mathbf{y}^*_{1:i-1}, c^*_{0i}, \boldsymbol{\mu}^{0*}_i, \boldsymbol{\Sigma}^{0*}_i\right)}\right.\right. \\
&\qquad\left.\left. - \log \frac{p_\lambda\left(\mathbf{y}_i \mid \mathbf{y}_{m+1:i}, c_{1i}, \boldsymbol{\mu}^{\lambda}_i, \boldsymbol{\Sigma}^{\lambda}_i\right)}{p_0\left(\mathbf{y}_i \mid \mathbf{y}_{1:i-1}, c_{0i}, \boldsymbol{\mu}^{0}_i, \boldsymbol{\Sigma}^{0}_i\right)}\right| \ge \epsilon\right).
\end{aligned}$$

Let

$$\begin{aligned}
L_i &= \log \frac{p^*_\lambda\left(\mathbf{y}^*_i \mid \mathbf{y}^*_{m+1:i}, c^*_{1i}, \boldsymbol{\mu}^{\lambda*}_i, \boldsymbol{\Sigma}^{\lambda*}_i\right)}{p^*_0\left(\mathbf{y}^*_i \mid \mathbf{y}^*_{1:i-1}, c^*_{0i}, \boldsymbol{\mu}^{0*}_i, \boldsymbol{\Sigma}^{0*}_i\right)} - \log \frac{p_\lambda\left(\mathbf{y}_i \mid \mathbf{y}_{m+1:i}, c_{1i}, \boldsymbol{\mu}^{\lambda}_i, \boldsymbol{\Sigma}^{\lambda}_i\right)}{p_0\left(\mathbf{y}_i \mid \mathbf{y}_{1:i-1}, c_{0i}, \boldsymbol{\mu}^{0}_i, \boldsymbol{\Sigma}^{0}_i\right)} \\
&= \log p^*_\lambda\left(\mathbf{y}^*_i \mid \mathbf{y}^*_{m+1:i}, c^*_{1i}, \boldsymbol{\mu}^{\lambda*}_i, \boldsymbol{\Sigma}^{\lambda*}_i\right) - \log p^*_0\left(\mathbf{y}^*_i \mid \mathbf{y}^*_{1:i-1}, c^*_{0i}, \boldsymbol{\mu}^{0*}_i, \boldsymbol{\Sigma}^{0*}_i\right) \\
&\quad - \big[\log p_\lambda(\mathbf{y}_i \mid \mathbf{y}_{m+1:i}, c_{1i}, \boldsymbol{\mu}^{\lambda}_i, \boldsymbol{\Sigma}^{\lambda}_i) - \log p_0(\mathbf{y}_i \mid \mathbf{y}_{1:i-1}, c_{0i}, \boldsymbol{\mu}^{0}_i, \boldsymbol{\Sigma}^{0}_i)\big] \\
&= \log p^*_\lambda\left(\mathbf{y}^*_i \mid \mathbf{y}^*_{m+1:i}, c^*_{1i}, \boldsymbol{\mu}^{\lambda*}_i, \boldsymbol{\Sigma}^{\lambda*}_i\right) - \log p_\lambda\left(\mathbf{y}_i \mid \mathbf{y}_{m+1:i}, c_{1i}, \boldsymbol{\mu}^{\lambda}_i, \boldsymbol{\Sigma}^{\lambda}_i\right) \\
&\quad - \big[\log p^*_0(\mathbf{y}^*_i \mid \mathbf{y}^*_{1:i-1}, c^*_{0i}, \boldsymbol{\mu}^{0*}_i, \boldsymbol{\Sigma}^{0*}_i) - \log p_0(\mathbf{y}_i \mid \mathbf{y}_{1:i-1}, c_{0i}, \boldsymbol{\mu}^{0}_i, \boldsymbol{\Sigma}^{0}_i)\big] \\
&= \log \frac{p^*_\lambda\left(\mathbf{y}^*_i \mid \mathbf{y}^*_{m+1:i}, c^*_{1i}, \boldsymbol{\mu}^{\lambda*}_i, \boldsymbol{\Sigma}^{\lambda*}_i\right)}{p_\lambda\left(\mathbf{y}_i \mid \mathbf{y}_{m+1:i}, c_{1i}, \boldsymbol{\mu}^{\lambda}_i, \boldsymbol{\Sigma}^{\lambda}_i\right)} - \log \frac{p^*_0\left(\mathbf{y}^*_i \mid \mathbf{y}^*_{1:i-1}, c^*_{0i}, \boldsymbol{\mu}^{0*}_i, \boldsymbol{\Sigma}^{0*}_i\right)}{p_0\left(\mathbf{y}_i \mid \mathbf{y}_{1:i-1}, c_{0i}, \boldsymbol{\mu}^{0}_i, \boldsymbol{\Sigma}^{0}_i\right)}.
\end{aligned}$$

Let $\phi_d(\cdot)$ and $f_d(\cdot)$ be the true density functions of N_d and F_d, respectively. Based on the Polya tree properties, when $m \to \infty$ and $i \gg m$,

$$\begin{aligned}
p_\lambda\left(\mathbf{y}_i \mid \mathbf{y}_{m+1:i}, c_{1i}, \boldsymbol{\mu}^{\lambda}_i, \boldsymbol{\Sigma}^{\lambda}_i\right) &= \phi_d(\mathbf{y}_i), \\
p_0\left(\mathbf{y}_i \mid \mathbf{y}_{1:i-1}, c_{0i}, \boldsymbol{\mu}^{0}_i, \boldsymbol{\Sigma}^{0}_i\right) &= \phi_d(\mathbf{y}_i).
\end{aligned}$$

Similarly, when $m \to \infty$ and $i \gg m$,

$$\begin{aligned}
p^*_\lambda\left(\mathbf{y}^*_i \mid \mathbf{y}^*_{m+1:i}, c^*_{1i}, \boldsymbol{\mu}^{\lambda*}_i, \boldsymbol{\Sigma}^{\lambda*}_i\right) &= f_d(\mathbf{y}^*_i), \\
p^*_0\left(\mathbf{y}^*_i \mid \mathbf{y}^*_{1:i-1}, c^*_{0i}, \boldsymbol{\mu}^{0*}_i, \boldsymbol{\Sigma}^{0*}_i\right) &= f_d(\mathbf{y}^*_i).
\end{aligned}$$

Hence, as $i \to \infty$,

$$\begin{aligned}
L_i &= \log \frac{p^*_\lambda\left(\mathbf{y}^*_i \mid \mathbf{y}^*_{m+1:i}, c^*_{1i}, \boldsymbol{\mu}^{\lambda*}_i, \boldsymbol{\Sigma}^{\lambda*}_i\right)}{p_\lambda\left(\mathbf{y}_i \mid \mathbf{y}_{m+1:i}, c_{1i}, \boldsymbol{\mu}^{\lambda}_i, \boldsymbol{\Sigma}^{\lambda}_i\right)} - \log \frac{p^*_0\left(\mathbf{y}^*_i \mid \mathbf{y}^*_{1:i-1}, c^*_{0i}, \boldsymbol{\mu}^{0*}_i, \boldsymbol{\Sigma}^{0*}_i\right)}{p_0\left(\mathbf{y}_i \mid \mathbf{y}_{1:i-1}, c_{0i}, \boldsymbol{\mu}^{0}_i, \boldsymbol{\Sigma}^{0}_i\right)} \\
&= \log \frac{f_d(\mathbf{y}^*_i)}{\phi_d(\mathbf{y}_i)} - \log \frac{f_d(\mathbf{y}^*_i)}{\phi_d(\mathbf{y}_i)} = 0.
\end{aligned}$$

Therefore,

$$\lim_{\substack{m\to\infty\\ i\to\infty\\ i\gg m}} Pr\Bigg(\Big|\log\frac{p^*_\lambda\left(\mathbf{y}^*_i \mid \mathbf{y}^*_{m+1:i}, c^*_{1i}, \boldsymbol{\mu}^{\lambda*}_i, \boldsymbol{\Sigma}^{\lambda*}_i\right)}{p^*_0\left(\mathbf{y}^*_i \mid \mathbf{y}^*_{1:i-1}, c^*_{0i}, \boldsymbol{\mu}^{0*}_i, \boldsymbol{\Sigma}^{0*}_i\right)} - \log\frac{p_\lambda\left(\mathbf{y}_i \mid \mathbf{y}_{m+1:i}, c_{1i}, \boldsymbol{\mu}^{\lambda}_i, \boldsymbol{\Sigma}^{\lambda}_i\right)}{p_0\left(\mathbf{y}_i \mid \mathbf{y}_{1:i-1}, c_{0i}, \boldsymbol{\mu}^{0}_i, \boldsymbol{\Sigma}^{0}_i\right)}\Big| \geq \epsilon\Bigg)$$
$$= \lim_{\substack{m\to\infty\\ i\to\infty\\ i\gg m}} Pr\left(\Big|L_i\Big| \geq \epsilon\right) = 0.$$

Thus, $\lim_{\substack{m\to\infty\\ i\to\infty\\ i\gg m}} Pr\Big(\big|R^*_\lambda(i) - R_\lambda(i)\big| \geq \epsilon\Big) = 0$. We proved the theorem. □

ACKNOWLEDGEMENTS

Chen's research was supported in part by 2015 Research Grants Committee, the University of Alabama.

Received 14 October 2016

REFERENCES

Albers, W. and Kallenberg, W. (2004). Empirical non-parametric control charts: estimation effects and corrections. *Journal of Applied Statistics*, **31**, 345–360. MR2061387

Alt, F. (1985). Multivariate quality control. In *Encyclopedia of statistical science*. Edited by Kotz, S., Nelson, N., New York: Wiley.

Berger, J. and Guglielmi, A. (2001). Bayesian testing of a parametric model versus nonparametric alternatives. *Journal of the American Statistical Association*, **96**, 174–184. MR1952730

Branscum, A. and Hanson, T. (2008). Bayesian nonparametric meta-analysis using Polya tree mixture models. *Biometrics*, **64**, 825–833. MR2526633

Chakraborti, S. and Eryilmaz, S. (2007). A nonparametric Shewhart-type signed-rank control charts based on runs. *Communications in Statistics-Simulation and Computation*, **36**, 335–356. MR2370905

Chen, J. and Chen, Z. (2008). Extended Bayesian information criteria for model selection with large model spaces. *Biometrics*, **95**, 759–771. MR2443189

Chen, Y. (2015). A new type of Bayesian nonparametric control charts for individual measurements. *Journal of Statistical Theory and Practice*, **10**, 226–238. MR3453038

Chen, Y. (2017). EWMA control charts for multivariate autocorrelated processes. *Statistics and Its Interface*, **10**, 575–584. MR3662773

Chen, Y. and Hanson, T. (2014a). Bayesian nonparametric density estimation for doubly-truncated data. *Statistics and Its Interface*, **7**, 455–463. MR3302374

Chen, Y. and Hanson, T. (2014b). Bayesian nonparametric k-sample tests for censored and uncensored data. *Computational Statistics and Data Analysis*, **71**, 335–346. MR3131974

Chen, Y. and Hanson, T. (2017). Semiparametric regression control charts. *Journal of Statistical Theory and Practice*, **11**, 126–144. MR3606924

Cheng, S. and Thaga, K. (2005). Multivariate max-CUSUM chart. *Quality Technology Quantitative Management*, **2**, 221–235. MR2222007

Cipolli, W., Hanson, T., and McLain, A. (2016). Bayesian nonparametric multiple testing. *Computational Statistics & Data Analysis*, **101**, 64–79. MR3504836

Crosier, R. (1988). Multivariate generalizations of cumulative sum quality control schemes. *Technometrics*, **30**, 291–303. MR0959530

Denison, D. and Mallick, B. (2007). Analyzing financial data using Polya trees. In *Bayesian statistics and its application*. Edited by Upadhyay, S., Singh, U. and Dey, D., New Delhi: Anamaya Publishers.

Hanson, T. (2006). Inference for mixtures of finite Polya tree models. *Journal of the American Statistical Association*, **101**, 1548–1565. MR2279479

Hanson, T. and Johnson, W. (2002). Modeling regression error with a mixture of Polya trees. *Journal of the American Statistical Association*, **97**, 1020–1033. MR1951256

Hanson, T., Monteiro, J., and Jara, A. (2011). The Polya tree sampler: towards efficient and automatic independent Metropolis-Hastings proposals. *Journal of Computational and Graphical Statistics*, **20**, 41–62. MR2816537

Hawkins, D. and Maboudou-Tchao, E. (2008). Multivariate exponentially weighted moving covariance matrix. *Technometrics*, **50**, 155–166. MR2439876

Hawkins, D., Qiu, P., and Kang, C. (2003). The changepoint model for statistical process control. *Journal of Quality Technology*, **35**, 355–366.

Hayter, A. and Tsui, K. (1994). Identification and qualification in multivariate quality control problems. *Journal of Quality Technology*, **26**, 197–208.

Healy, J. (1987). A note on multivariate CUSUM procedures. *Technometrics*, **29**, 409–412.

Holland, M. and Hawkins, D. (2014). A control chart based on a nonparametric multivariate change-point model. *Journal of Quality Technology*, **46**, 63–77.

Hotelling, H. (1947). Multivariate quality control-illustrated by the air testing of sample bombsights. In *Techniques of Statistical Analysis*. Edited by Eisenhart, C., Hastay, M. and Wallis, W., New York: McGraw Hill.

Huwang, L., Yeh, A., and Wu, C. (2007). Monitoring multivariate process variability for individual observations. *Journal of Quality Technology*, **39**, 258–278.

Jackson, J. (1991). *A user guide to principal components*. New York: Wiley.

Lavine, M. (1992). Some aspects of Polya tree distributions for statistical modeling. *The Annals of Statistics*, **20**, 1222–1235. MR1186248

Li, J., Zhang, X., and Jeske, D. (2013). Nonparametric multivariate CUSUM control charts for location and scale changes. *Journal of Nonparametric Statistics*, **25**, 1–20. MR3039967

Li, Z., Dai, Y., and Wang, Z. (2014). Multivariate change point control chart based on data depth for phase I analysis. *Communications in Statistics-Simulation and Computation*, **43**, 1490–1507. MR3215788

Liu L., Zi, X., Zhang, J., and Wang, Z. (2013). A sequential rank-based nonparametric adaptive EWMA control chart. *Communications in Statistics-Simulation and Computation*, **42**, 841–859. MR3039617

Lowry, C., Woodall, W., Champ, C., and Rigdon, S. (1992). A multivariate EWMA control chart. *Technometrics*, **34**, 46–53.

Montgomery, D. (2009). *Introduction to statistical quality control*. 6th ed., Hoboken: John Wiley & Sons Inc.

Niaki, S. and Abbasi, B. (2005). Fault diagnosis in multivariate control charts using artificial neural networks. *Quality and Reliability Engineering International*, **21**, 825–840.

Paddock, S. (1999). Randomized Polya trees: Bayesian nonparametrics for multivariate data analysis. Doctoral dissertation, Duke University. MR2699876

Pignatiello, J. and Runger, G. (1990). Comparisons of Multivariate CUSUM Charts. *Journal of Quality Technology*, **22**, 173–186.

rabhu, S. and Runger, G. (1997). Designing a multivariate EWMA control chart. *Journal of Quality Technology*, **29**, 8–15.

Qiu, P. and Hawkins, D. (2012). A rank-based multivariate CUSUM procedure. *Technometrics*, **43**, 120–132. MR1954134

Reynolds Jr, M. and Cho, G. (2006). Multivariate control charts for monitoring the mean vector and covariance matrix. *Journal of Quality Technology*, **38**, 230–253.

Reynolds Jr, M. and Cho, G. (2011). Multivariate control charts for monitoring the mean vector and covariance matrix with variable sampling intervals. *Sequential Analysis*, **30**, 1–40. MR2770703

Santos-Fernandez, E. (2016). R package **MSQC**.

Sepulveda A. and Nachlas, J. (1997). A simulation approach to multivariate control. *Computers and Industrial Engineering*, **33**, 113–116.

Sun, G. and Zi, X. (2013). An empirical-likelihood-based multivariate EWMA control scheme. *Communications in Statistics-Theory and Methods*, **42**, 429–446. MR3005788

Testik, M. and Borror, C. (2004). Design strategies for the multivariate exponentially weighted moving average control chart. *Quality and Reliability Engineering International*, **20**, 571–577.

Walker, S. and Mallick, B. (1999). Semiparametric accelerated life time model. *Biometrics*, **55**, 477–483. MR1705102

Wang, K., Yeh, A., and Li, B. (2014). Simultaneous monitoring of process mean vector and covariance matrix via penalized likelihood estimation. *Computational Statistics and Data Analysis*, **78**, 206–217. MR3212167

Willemain, T. and Runger, G. (1996). Designing control charts using an empirical reference distribution. *Journal of Quality Technology*, **28**, 31–38.

Woodall, W. and Ncube, M. (1985). Multivariate CUSUM quality control procedures. *Technometrics*, **27**, 285–292. MR0797567

Yeh, A., Huwang, L., and Wu, Y. (2004). A likelihood-ratio-based EWMA control chart for monitoring variability of multivariate normal processes. *IIE Transactions*, **36**, 865–879.

Zhang, G. and Chang, S. (2008). Multivariate EWMA control charts using individual observations for process mean and variance monitoring and diagnosis. *International Journal of Production Research*, **46**, 6855–6881.

Zhang, J. (2002). Powerful goodness-of-fit tests based on the likelihood ratio. *Journal of the Royal Statistical Society: Series B*, **64**, 281–294. MR1904705

Zhang, J., Li, Z., and Wang, Z. (2010). A multivariate control chart for simultaneously monitoring process mean and variability. *Computational Statistics and Data Analysis*, **54**, 2244–2252. MR2720485

Zhou, C., Zou, C., Zhang, Y., and Wang, Z. (2009). Nonparametric-control chart based on change-point model. *Statistical Papers*, **50**, 13–28. MR2476166

Zou, C., Jiang, W., and Tsung, F. (2013). A LASSO-based diagnostic framework for multivariate statistical process control. *Quality control and applied statistics*, **58**, 33–36. MR2867503

Zou, C. and Tsung, F. (2010). Likelihood ratio-based distribution-free EWMA control charts. *Journal of Quality Technology*, **42**, 174–196.

Zou, C. and Tsung, F. (2011). A multivariate sign EWMA control chart. *Technometrics*, **53**, 84–97. MR2791949

Yuhui Chen
Department of Mathematics
The University of Alabama
Tuscaloosa, AL
USA
E-mail address: ychen164@ua.edu

Mingwei Sun
Department of Mathematics
The University of Alabama
Tuscaloosa, AL
USA
E-mail address: msun@crimson.ua.edu

Timothy Hanson
Department of Statistics
The University of South Carolina
Columbia, SC
USA
E-mail address: hansont@stat.sc.edu