Special Issue Article

Quality and Reliability Engineering International

(wileyonlinelibrary.com) DOI: 10.1002/qre.1641

Published online 16 June 2014 in Wiley Online Library

A Nonparametric Phase I Control Chart for Individual Observations Based on Empirical Likelihood Ratio

Wei Ning,[a] Arthur B. Yeh,[b*†] Xinqi Wu[c] and Boxiang Wang[d]

One common challenge in nonmanufacturing control chart applications is that many of the nonmanufacturing quality characteristics are not normally distributed. In these applications, normal transformation of the observations is certainly feasible; however, it will be done at the expense of the interpretability of the analysis that is particularly important to control chart users in nonmanufacturing industries.

Most of the existing nonparametric control charts are designed for Phase II monitoring. Little has been done in developing nonparametric Phase I control charts especially for individual observations that are prevalent in nonmanufacturing applications. In this work, we propose a new nonparametric Phase I control chart for monitoring the location parameter whose construction is essentially based on the empirical likelihood ratio test. The performance of the proposed chart, in terms of the signal probability, compares favorably with the recently developed charts for individual observations. A nonmanufacturing example is included in which the proposed chart and the other competing charts are applied and compared. Copyright © 2014 John Wiley & Sons, Ltd.

Keywords: change-point model; empirical likelihood ratio; nonnormal processes; signal probability; *X*-chart

1. Introduction

1.1. Background and literature eview

Of the techniques that form the core of statistical process control (SPC), the control chart is perhaps the most widely used technique in practical applications since it was first introduced by Shewhart[1]. In practice, the control chart applications are divided into two phrases. In the Phase I control, the control charts are used to assess the stability of the process, determine the process distribution, and estimate the population parameters that are needed for setting up the Phase II control charts. The primary goal of Phase II monitoring is to detect a process change as soon as it occurs.

For many years, the control chart applications can be found predominantly in manufacturing industries. Over the last two decades, they have begun to spread to nonmanufacturing industries such as health care, banking, and insurance, among others. As these nonmanufacturing applications continue to spread, new challenges will inevitably arise, which will require developing more effective control charts. One such challenge commonly encountered in nonmanufacturing applications is that many of the quality characteristics are not normally distributed. For example, in a health care application discussed in Jones-Farmer, Jordan, and Champ[2] (2009), the authors pointed out that the quality characteristic, the wait times of patients who underwent a colonoscopy procedure, does not follow a normal distribution. In order to continue to expand control chart applications into broader sectors, it is therefore important to develop control charts that are either more robust to or do not require normality assumption of the quality characteristic being monitored. Several authors, for example, Woodall and Montgomery[3] and Woodall[4], have also pointed out the need to develop nonparametric control charts. A comprehensive review of the development of nonparametric control charts up until 2013 can be found in Qiu[5].

In the last decade or so, numerous nonparametric control charts have appeared in statistical and quality engineering journals. Most existing nonparametric control charts utilize charting statistics that are based on ranking/ordering information of the observations across different time points. The charting statistics could be of Shewhart, cumulative sum, or exponentially weighted moving averages (EWMA) type. This work includes, for example, Albers and Kellenberg[6]; Bakir[7]; Chakraborti, van der Laan, and van de Wiel[8];

[a]*Department of Mathematics and Statistics, Bowling Green State University, Bowling Green, OH 43403, USA*
[b]*Department of Applied Statistics and Operations Research, Bowling Green State University, Bowling Green, OH 40403, USA*
[c]*Graduate University of Chinese Academy of Science, Beijing, 100049, People's Republic of China*
[d]*The School of Statistics, University of Minnesota, Minneapolis, MN 55455, USA*
**Correspondence to: Arthur B. Yeh, Department of Applied Statistics and Operations Research, Bowling Green State University, Bowling Green, OH 40403, USA.*
[†]*E-mail: byeh@bgsu.edu*

Bakir[9]; Chakraborti and Eryilmaz[10]; Albers and Kellenberg[11]; Chakraborti, Eryilmaz, and Human[12]; and Jones-Farmer, Jordan, and Champ[2]. Some recent work, for example, Zhou, Zou, Zhang, and Wang[13], Hawkins and Deng[14], and Zou and Tsung[15], examined the problem from a change-point model perspective and developed nonparametric control charts using rank or likelihood-ratio-based statistics. Most recently, Qiu and Li[16] developed a charting method based on first categorizing the original observations into categorical data and then using categorical data analysis techniques. They have shown that the method is often more effective than the ones based on ranking/ordering information. Also, the idea of finding a transformation based on an in-control (IC) data set and transforming the Phase II data so that a normality-based control chart can be used has been discussed recently in Qiu and Li[17] and Qiu and Zhang[18]. Nevertheless, these authors showed that the transformation-based approach is often ineffective in various cases. In addition, Liu, Zou, Zhang, and Wang[19] developed a nonparametric EWMA chart based on the standardized sequential rank of each incoming observation, and Ross and Adams[20] developed a nonparametric control chart based on measuring the difference between two empirical distribution functions by either the Cramer–von Mises (CM) or Kolmogorov–Smirnov (KS)-type distance. All of this aforementioned work focuses on univariate processes. As for the multivariate nonparametric control charts, refer to the work by, for example, Qiu and Hawkins[21,22], Qiu[23], Zou and Tsung[24], Zou, Wang, and Tsung[25], Sun and Zi [26], and Li, Dai, and Wang[27].

The vast majority of the existing nonparametric control charts focus on monitoring the location parameter. Some charts based on change-point model, such as those studied in Zou and Tsung[15], Qiu and Li[16], and Ross and Adams[20], are capable of detecting simply changes in the distribution function that could be a result of changes in the location parameter or scale parameter or both. Further, almost all of the existing charts are designed for Phase II monitoring. As such, they typically assume that the IC value of the parameter to be monitored is either known or can be reasonably estimated based on Phase I data that were collected when the process was IC. Some of these methodologies do not require that the IC parameter value be known; however, they assume the existence of an IC Phase I sample or that the process was IC. Refer to, for example, Zhou, Zou, Zhang, and Wang[13], Hawkins and Deng[14], Zou and Tsung[15], Qiu and Li[16], and Liu, Zou, Zhang, and Wang[19]. Because of the assumption of known IC process parameter values or the existence of an IC sample or an IC process, most of the aforementioned nonparametric Phase II control charts are not suitable for retrospective analysis in Phase I control.

Despite the importance of Phase I control in SPC and that Phase II monitoring is almost impossible without properly completing Phase I control, little has been done in developing nonparametric Phase I control charts. One such charting mechanism was the nonparametric Phase I control chart for a subgroup location studied in Jones-Farmer, Jordan, and Champ[2]. The authors proposed ranking each observation in each subgroup with respect to the entire sample and calculating the standardized average of the ranks in each subgroup. However, their proposed chart requires that the subgroup size is at least 3.

1.2. Problem formulation

There are many practical manufacturing as well as nonmanufacturing applications in which only individual observations can be collected for SPC implementation. Let X represent the quality characteristic to be monitored whose distribution, denoted by F, is unknown. Let $E(X)=\mu$ and $\mathrm{Var}(X)=\sigma^2$ denote the mean and variance of X, respectively. When the process is IC, we assume that $E(X)=\mu_0$ and $\mathrm{Var}(X)=\sigma_0^2$, where both μ_0 and σ_0^2 are unknown. In this work, our focus is on monitoring the process mean μ in a Phase I application; we therefore further assume that the process variance remains at σ_0^2 during the course of the Phase I application. Let X_1, X_2, …, X_n be n independent random observations from F, collected for the purpose of a Phase I application. That is, each X_j, $j=1,2,\ldots,n$, represents the only individual observation sampled from F at the jth sampling period. Under the framework just mentioned, the conventional Shewhart chart such as the X-chart plots each X_j against predetermined control limits, and the chart signals when any of the X_j's plots are outside of the control limits.

In order to introduce the proposed Phase I control chart in the later sections, we formulate the problem from a change-point model perspective. Let $E(X_j)=\mu(j)$, and $j=1,2,\ldots,n$. When the process is IC, we assume that $\mu(1)=\mu(2)=\cdots=\mu(n)=\mu_0$. When the process is out of control (OC), we assume that there exists one unknown sampling period k such that $\mu(1)=\mu(2)=\cdots=\mu(k)=\mu_0$, $\mu(k+1)=\mu(k+2)=\cdots=\mu(n)=\mu_1$, and $\mu_0\neq\mu_1$. Therefore, under the change-point model framework, the main problem to be tackled in this work is to develop a nonparamteric Phase I control chart that can effectively detect the occurrence of a change-point location as soon as the process becomes OC. Note that here we only consider one change-point location in the process for the simplicity of the discussion. For the multiple change-point locations scenario, it can be transferred to the single change-point location scenario by the binary segmentation procedure proposed by Vostrikova[28]. It should be noted that, in general, the Phase I control problem is essentially the same as the change-point detection problem in that in both cases, the sample size n is fixed, and one wishes to detect any changes in the underlying process distribution. Therefore, most existing change-point methods are relevant to the Phase I control problems. On the other hand, the Phase II monitoring is intrinsically different from the change-point detection problem in that the sample size increases n as monitoring progresses.

Recently, Hawkins and Deng[14] and Ross and Adams[20], while focusing on Phase II monitoring, briefly discussed nonparametric Phase I control charts based on a change-point model for individual observations. Hawkins and Deng[14] discussed using a two-sample Mann–Whitney (MW) test to detect a possible change in the location parameter. Ross and Adams[20] suggested comparing two empirical distribution functions using a CMor KS-type test statistic.

In parametric settings, control charts that are developed based on the log-likelihood ratio enjoy certain optimal properties (refer to, e.g., Moustakidas[29]). On the other hand, under nonparametric settings, the empirical likelihood ratio (Owen[30,31]) has been shown to provide efficient estimators and good testing power in many applications. This naturally leads to the thinking of whether

one can develop an effective nonparametric Phase I control chart based on the empirical likelihood ratio, as the existing charts were not developed based on the likelihood principle. Therefore, we are motivated in this work to develop a new nonparametric Phase I control chart for monitoring the location parameter based on the empirical likelihood ratio. The performance of the proposed chart, in terms of the signal probability, compares favorably with that of the aforementioned nonparametric Phase I charts, as well as the conventional *X*-chart, for individual observations.

The rest of the paper is organized as follows. We first briefly discuss in Section 2 the existing nonparametric Phase I control charts. In Section 3, we discuss the change point model based on empirical likelihood ratio and the proposed nonparametric Phase I control chart. Section 4 is devoted to comparing the performance of the proposed chart with that of the existing charts. A nonmanufacturing example is discussed in Section 5 in which the proposed chart and the existing charts are applied and compared. Section 6 gives a number of concluding remarks and possible future research along the same line.

2. The existing methodologies

Under the change-point model framework discussed in Section 1.2, Hawkins and Deng[14] suggested applying the two-sample MW test to test whether the location parameters are equal. Specifically, assume that k is the change-point location for $1 \le k < n$. To test H_0: $\mu_0 = \mu_1$ versus H_a: $\mu_0 \neq \mu_1$, the two-sample MW test statistic is equal to

$$MW_{k,n} = \sum_{i=1}^{k}\sum_{j=k+1}^{n} I(X_j < X_i) + \frac{1}{2}\sum_{i=1}^{k}\sum_{j=k+1}^{n} I(X_j = X_i), \tag{1}$$

where $I(A) = 1$ if A happens and 0 if otherwise. Also, define the standardized MW test statistic as

$$SMW_{k,n} = \frac{MW_{k,n} - E(MW_{k,n})}{\sqrt{\mathrm{Var}(MW_{k,n})}}, \tag{2}$$

where

$$E(MW_{k,n}) = \frac{k(n-k)}{2} \text{ and } \mathrm{Var}(MW_{k,n}) = \frac{k(n-k)(n+1)}{12}.$$

The change-point model formulation calls for evaluating

$$SMW_n = \max_{1 \le k \le n-1} |SMW_{k,n}|, \tag{3}$$

and that the MW-based nonparametric Phase I control chart signals when $SMW_n \ge h_{\alpha,SMW_n}$, where the threshold h_{α,SMW_n} depends on the desirable Phase I IC signal probability α and the sample size n. In this paper, we used Monte-Carlo simulations to approximate the distribution of SMW_n and consequently h_{α,SMW_n}.

The other two existing nonparametric Phase I control charts, both suggested by Ross and Adams[20], are based on comparing two empirical cumulative distribution functions (ECDFS), one obtained from the first k observations, $X_1, X_2, \ldots, X_k$, and the other from the last $n-k$ observations, $X_{k+1}, X_{k+2}, \ldots, X_n$. Specifically, define, for $1 < k < n$,

$$\hat{F}_{1,k}(x) = \frac{1}{k}\sum_{j=1}^{k} I(X_j < x) \text{ and } \hat{F}_{2,k}(x) = \frac{1}{n-k}\sum_{j=k+1}^{n} I(X_j < x).$$

Further, define the ECDF constructed based on the entire sample $X_1, X_2, \ldots, X_n$ as $\hat{F}_n(x) = 1/n\sum_{j=1}^{n} I(X_j < x)$. The KS-type distance between $\hat{F}_{1,k}(x)$ and $\hat{F}_{2,k}(x)$ is defined as

$$KS_{k,n} = \sup_x |\hat{F}_{1,k}(x) - \hat{F}_{2,k}(x)|, \tag{4}$$

whereas the CM-type distance is calculated as

$$CM_{k,n} = \frac{k(n-k)}{n} \times \int_{-\infty}^{\infty} |\hat{F}_{1,k}(x) - \hat{F}_{2,k}(x)|^2 d\hat{F}_n(x)$$

$$= \frac{k(n-k)}{n^2} \times \sum_{j=1}^{n} |\hat{F}_{1,k}(X_j) - \hat{F}_{2,k}(X_j)|^2. \tag{5}$$

For the nonparametric Phase I control chart based on the CM-type distance, the test statistic is equal to

$$CM_n = \max_{1<k<n} \frac{CM_{k,n} - E(CM_{k,n})}{\sqrt{\mathrm{Var}(CM_{k,n})}}, \tag{6}$$

where

$$E(CM_{k,n}) = \frac{n+1}{6n} \text{and} \mathrm{Var}(CM_{k,n}) = \frac{(n+1)\times[(1-3/4k)n^2 + (1-k)n - k]}{45n^2(n-k)}.$$

The CM-based chart signals when $CM_n > h_{\alpha,CM_n}$, where h_{α,CM_n} depends on the IC signal probability α and the sample size n. As for the chart based on the KS-type distance, Ross and Adams[20] suggested calculating

$$q_{k,n} = 1 - p_{k,n},$$

where $p_{k,n}$ is the probability of observing a more extreme value than $KS_{k,n}$. The test statistic of the KS-based nonparametric Phase I control chart is then defined as

$$KS_n = \max_{1<k<n} q_{k,n}, \tag{7}$$

and that the chart signals when $KS_n > h_{\alpha,KS_n}$. Again, the threshold h_{α,KS_n} depends on the sample size n and the IC signal probability α. Please refer to Ross and Adams[20] for details on how to calculate $q_{k,n}$. In this work, we used Monte-Carlo simulations to approximate h_{α,CM_n} as well as h_{α,KS_n}, for a given α and n.

3. The proposed methodology

3.1. *Change-point model based on empirical likelihood ratio*

The empirical likelihood is a data-driven method for statistical inference. It was first proposed by Thomas and Grunkemeier[32] and further developed by Owen[30,31]. It is a method that combines the advantages of parametric and nonparametric methods. On one hand, similar to the parametric likelihood method, it can be used to develop a likelihood-type function for finding efficient estimators and constructing tests with good power properties and confidence intervals (Owen[33]). On the other hand, it behaves similarly to a nonparametric method that makes no distributional assumption. Therefore, it enjoys the flexibility and robustness in many applications.

The pioneer work by Owen[30,31] extended the earlier work of Thomas and Grunkemeier[32] who applied a nonparametric likelihood ratio to constructing confidence intervals for the survival functions. Owen[30] showed that the likelihood-ratio-type statistic based on the empirical likelihood method, just as in the parametric settings, has an asymptotic χ^2 distribution. Since then, the empirical likelihood method has been employed in different statistical contexts such as regression models, change-point analysis, survival analysis, etc. Properties of the empirical likelihood ratio in the case of independent and identically distributed random variables were investigated in Owen [31], Hall[34], DiCiccio, Hall, and Romano[35], and Qin and Lawless[36]. Owen[33] extended the empirical likelihood method to the regression models. Kolaczyk[37] considered such a method in the class of generalized linear models. Kitamura[38,39] studied the empirical likelihood method for weakly dependent processes and its applications to the time series analysis. In survival analysis, Pan and Zhou[40] constructed an empirical likelihood ratio based on the cumulative hazard functions for censored data. Ren and Zhou[41] provided the inference for the Cox model using an empirical likelihood approach. In recent years, the empirical likelihood method has also been used to investigate the change-point problems. For instance, Zou, Liu, Qin, and Wang[42] employed this method for the general change-point problems and derived the asymptotic distribution of the empirical likelihood ratio statistic. Liu, Zou, and Zhang[43] developed the empirical likelihood method for the change-point problems in a linear regression model. Ning[44] and Ning, Pailden, and Gupta[45] studied different forms of the mean changes by using the empirical likelihood ratio test. For a more detailed account of the empirical likelihood method, readers are referred to Owen[46] and the references cited in the aforementioned papers. In the current work, we follow closely the theoretical development of Zou, Liu, Qin, and Wang[42] and apply it specifically to developing a nonparametric Phase I control chart for monitoring the location parameter.

The main idea of the empirical likelihood method is to place an unknown probability mass at each observation. Let $p_i = P(X = x_i)$ and the empirical likelihood function of F is defined as

$$L(F) = \prod_{i=1}^{n} p_i.$$

It is clear that $L(F)$, subject to the constraints

$$p_i \geq 0 \text{ and } \sum_i p_i = 1,$$

is maximized at $p_i = 1/n$, that is, the likelihood $L(F)$ attains its maximum n^{-n} under the full nonparametric model. When a population parameter μ, identified by $E(m(X,\mu)) = 0$, is of interest, the maximum of the empirical log-likelihood when μ has the true value μ_0 is obtained subject to an additional constraint $\sum p_i m(x_i, \mu_0) = 0$. The function $m(x;\mu)$ is a real-valued function through which the parameter of interest μ is specified by the estimating equation $E(m(X,\mu)) = 0$. In this paper, we take $m(x;\mu) = x - \mu$. More details of the estimating equations can be found in Owen[46]. The empirical log-likelihood ratio (ELR) statistic for testing $\mu = \mu_0$ is then given by

$$R(\mu_0) = \max_{p_1, p_2, \ldots, p_n} \left\{ \sum_i \log n p_i : p_i \geq 0, \sum_i p_i = 1, \sum_i p_i m(x_i, \mu_0) = 0 \right\}.$$

Similar to the likelihood ratio test statistic in a parametric model setting, Owen[30] showed that, under mild regularity conditions, $-2 \times R(\mu_0) \to \chi^2_r$ in distribution as $n \to \infty$ under the null hypothesis $\mu = \mu_0$, where r is the dimension of $m(x,\mu)$.

For the change-point model formulation described in Section 1, given a fixed k, the empirical log-likelihood function to test the hypothesis that $\mu_0 = \mu_1$ is equal to

$$l(\mu_0, \mu_1 | k) = i \sum \log u_i + j \sum \log v_j,$$

where $i = 1, \cdots, k; j = k+1, \cdots, n$, $u_i = P(X = x_i)$, and $v_j = P(X = x_j)$. With the constraints $\sum_i u_i = \sum_j v_j = 1$, the empirical log-likelihood function reaches the maximum value at $u_i = k^{-1}$ and $v_j = (n-k)^{-1}$ by a Lagrange multiplier method. Therefore, the ELR is equal to

$$\text{lr}(\mu_0, \mu_1 | k) = \sum_i \log(k u_i) + \sum_j \log\big((n-k) v_j\big).$$

Consequently, the profile ELR for a given μ_0 and μ_1 can be written as

$$R(\mu_0, \mu_1 | k) = \sup\{\text{lr}(\mu_0, \mu_1 | k) : \sum_i u_i = \sum_j v_j = 1, \sum_i u_i x_i = \mu_0, \sum_j v_j x_j = \mu_1\},$$

where $u_i \geq 0$ and $v_j \geq 0$. To test the hypothesis $\mu_0 = \mu_1$, the test statistic is defined as

$$Z_{n,k} = -2 \times R(\mu_0, \mu_0 | k) = -2 \times \mu_0 \sup \left\{ \sum_i \log(k u_i) + \sum_j \log\big((n-k) v_j\big) : \sum_i u_i = \sum_j v_j = 1, \sum_i u_i x_i = \sum_j v_j x_j = \mu_0 \right\}.$$

A more detailed discussion of how to calculate $Z_{n,k}$ for a given n, k, and μ_0 is given in the Appendix. Here, we used a Newton–Raphson algorithm to obtain numerical solutions to $Z_{n,k}$, which we will denote as $Z^*_{n,k}$. The R codes used to calculate the $Z^*_{n,k}$ are available from the first author. Note that in the simple univariate population mean case, the Newton–Raphson can always find unique solutions of the related objective functions. Refer to Owen[46] and Chen, Variyatha, and Abrahama[47] for more details about the computations of the empirical likelihood function. However, one needs to be careful in choosing the initial value so that the iterations will converge after several steps. In this paper, we adopt the algorithm proposed by Wang and Chen[48]. Further, in the change-point scenario, if the true change location is at the very beginning or at the very end of the sample, that is, there are only few observations for one of the two samples of observations, the solutions of the related objective functions may not exist. There, we recommend using a trimmed test statistic (Section 3.2).

3.2. Nonparametric Phase I control chart based on ELR

In the context of a change-point model, since k is unknown, it is natural to use the maximal empirical likelihood ratio statistic that is defined as

$$Z^*_n = \max_{1<k<n} \left\{ Z^*_{n,k} \right\},$$

and we will reject the null hypothesis with a significantly large value of Z^*_n. Note that if k or $n-k$ is too small, the empirical likelihood estimators of $Z^*_{n,k}$ may not exist; that is, our test may not detect the change point occurring at the very beginning or the very end of the sample. Therefore, we suggest the trimmed likelihood ratio statistic as

$$Z^*_n = \max_{k_0<k<n-k_1} \left\{ Z^*_{n,k} \right\}. \tag{8}$$

To obtain the asymptotic distribution of the test statistic in Equation (8), according to Csörgő and Horvath[49], we choose $k_0 = k_1 = 2[\log n]$, where $[x]$ denotes the largest integer not greater than x.

We shall call the proposed nonparametric Phase I control chart that is based on calculating the Z^*_n in Equation (8) the ELR chart. The chart signals when $Z^*_n > c_{\alpha, Z^*_n}$. In the current work, we used the asymptotic distribution of Z^*_n to approximate c_{α, Z^*_n}. The two main

asymptotic results, which are similar to those obtained by Csörgő and Horvath[43] under a parametric likelihood ratio method, were developed and discussed in Zou, Liu, Qin, and Wang[38]. When the process is IC, that, $\mu_0 = \mu_1$, we have

$$P\left(A(log(t(n)))\left(Z_n^*\right)^{1/2} \leq x + D_r(log(t(n))\right) - \rightarrow exp(-e^{-x}), \tag{9}$$

as $n \rightarrow \infty$ for all x, where

$$\begin{aligned} A(x) &= (2logx)^{1/2}, \\ D_r(x) &= 2logx + (r/2)log\ \ logx - log\Gamma(r/2), \\ t(n) &= \frac{n^2 + (2[logn])^2 - 2n[logn]}{(2[logn])^2}, \end{aligned}$$

and r is the dimensionality of the parameter space. Consequently, in our setting, the c_{α,Z_n^*} can be approximated by

$$c_{\alpha,Z_n^*} = (G_\alpha + D_r(log(t(n))))^2/(A(log(t(n))))^2, \tag{10}$$

where G_α is the $100(1-\alpha)\%$ percentile of a Gumbel distribution, $D_r(\cdot)$and $A(\cdot)$ are defined in Equation (9), n is the sample size, and $r = 1$ in the current work.

4. Performance evaluation and comparison

4.1. Simulation settings

In this section, we investigate the performance of the ELR chart and compare it with that of the conventional X-chart, as well as the MW, CM, and KS-based charts. The performance measure is defined as the signal probability that, for the proposed ELR chart, is the probability that $Z_n^* > c_{\alpha,Z_n^*}$. As for the X-chart, the signal probability is defined as the probability of having at least one observation plotted outside of the control limits. As for the MW, CM, and KS-based charts, the signal probability is defined in the same way as the ELR chart.

As in most Phase I control chart applications, the population parameters, and consequently the control limits, have to be estimated from the Phase I data. The control limits of the X-chart are estimated as

$$\text{LCL} = \overline{X} - L \times \frac{\overline{\text{MR}}}{d_2} \text{and UCL} = \overline{X} + L \times \frac{\overline{\text{MR}}}{d_2},$$

where $\overline{X} = \sum_{i=1}^{n} X_i/n$ is the sample average of the n Phase I observations, $\overline{\text{MR}} = \sum_{i=1}^{n-1} |X_{i+1} - X_i|/(n-1)$ is the average of moving ranges of length 2, $d_2 = 1.128$ is the adjusting constant for estimating the population standard deviation using the average of sample ranges of size 2, and L is the multiplier that depends on the IC performance of the X-chart as well as the sample size n. We set the IC signal probability to be 0.005 and consider two sample sizes $n = 50$ and 100. The values of L for the X-chart that produce approximately 0.005 IC signal probability are 3.945 and 4.093 for $n = 50$ and 100, respectively. For the proposed ELR chart, we obtained the UCLs (c_{pha,Z_n^*}) based on Equation (10). The UCLs for $n = 50$ and 100 are 21.4538 and 20.8743. For the MW-based chart, the UCLs (h_{α,SMW_n}) that produce approximately 0.005 IC signal probability for $n = 50$ and 100 are 3.431 and 3.586, respectively. For the same sample sizes and IC signal probability, the UCLs (h_{α,CM_n}) of the CM-based chart were determined to be 7.985 and 8.668 for $n = 50$ and 100, respectively. As for the KS-based chart, the UCLs (h_{α,KS_n}) are 0.99973 and 0.99980 for $n = 50$ and 100, respectively. The simulated IC signal probabilities of these control charts are summarized in Table I. For the X-chart, MW, CM, and KS-based charts, we used 300,000 simulations. As for the ELR chart, 10,000 simulations were used. Note that the simulated IC signal probabilities for the ELR chart under normal and t_3 distributions are closer to but generally more conservative than 0.005. On the other hand, when the underlying distribution is exponential, the simulated IC signal probabilities are slightly larger than the nominal level at 0.005.

For a given sample size, we compare the performance of the proposed ELR chart with that of the X-chart and the other competing nonparametric Phase I charts, assuming that the observations come from one of the three distributions, normal, exponential, and t_3, where t_3 denotes a t-distribution with 3 degrees of freedom. Here, we choose the exponential and t_3 distributions to represent the cases of continuous skewed and longer-tailed distributions. Similar to the existing literature in which Phase I control charts are evaluated and compared, we consider three possible OC scenarios, a step change in the process mean, the presence of outlying observations in the sample, and a gradual shift in the process mean. For each of the scenarios, we simulate and compare the OC signal probabilities of the competing charts for a range of process mean shifts, as characterized by $\mu_1 = \mu_0 + \delta\sigma_0$, where μ_0 and σ_0 are the IC mean and standard deviation, respectively, of a given distribution. The OC scenarios considered here are $\delta = 0.25, 0.50, 1.00, 1.50, 2.0,$

Table I. The simulated IC signal probabilities and the corresponding control limits for $\alpha = 0.005$

	X-chart	ELR	MW	CM	KS
Normal					
$n = 50$	.00502	.00454	.00496	.00480	.00480
	(3.945)	(21.4538)	(3.431)	(7.985)	(.9997)
$n = 100$	.00504	.00489	.00499	.00490	.00510
	(4.093)	(20.8743)	(3.586)	(8.668)	(.9999)
Exponential					
$n = 50$	.00505	.00892	.00510	.00498	.00498
	(7.999)	(21.4538)	(3.431)	(7.895)	(.9997)
$n = 100$	.00496	.00833	.00480	.00493	.00493
	(9.260)	(20.8743)	(3.586)	(8.668)	(.9998)
t_3					
$n = 50$	.00508	.00346	.00504	.00491	.00493
	(11.987)	(21.4538)	(3.431)	(7.895)	(.9997)
$n = 100$	.00505	.00387	.00501	.00503	.00511
	(17.413)	(20.8743)	(3.586)	(8.668)	(.9998)

The control limits are listed in the parentheses.

and 3.0. Note that, in the case of the *X*-chart, we standardize the observations before plotting them on the *X*-chart. More specifically, the OC scenarios considered include the following:

1. A step change in the process mean.

For $n = 50$, the change-point locations considered are $k = 10$, 25, and 40. As for $n = 100$, the change-point locations are set at $k = 20$, 50, and 80 of the set change point locations.

2. The presence of outlying observations.

We assume that two OC observations are present in the Phase I sample, and their locations are at 20 and 40 for $n = 50$, and at 40 and 80 for $n = 100$.

3. A gradual shift in the process mean.

In this scenario, we assume that the mean shift follows a linear trend. That is, for a given δ, the process mean $\mu(j)$, $j = 1, 2, \ldots, n$, at the j th observation is defined by

$$\mu(j) = \mu_0 + \frac{(j-1)}{n-1} \times \delta\sigma_0.$$

Note that $\mu(1) = \mu_0$ and $\mu(n) = \mu_1$.

4.2. Simulation results

Note that the value of L that determines the control limits used in the *X*-chart is specifically chosen so that the signal probability is approximately 0.005 when the IC process follows a standard normal distribution. However, when the IC process follows a nonnormal distribution, the actual IC signal probability is likely to be different if the same control limits are used. In fact, in our simulation settings, if the IC process is exponential, using the same control limits obtained under a normal process will produce IC signal probabilities of 0.4252 and 0.6557, respectively, for $n = 50$ and 100. These IC signal probabilities are much larger than the expected 0.005. Similarly, the actual IC signal probabilities when the IC process follows a t_3 distribution are 0.3930 and 0.6392 for $n = 50$ and 100, respectively. These larger than nominal IC signal probabilities indicate that the *X*-chart is not a suitable Phase I control charting mechanism when the IC process does not follow a normal distribution.

In order to compare the proposed ELR chart and the *X*-chart under nonnormal processes, similar to Jones-Farmer, Jordan, and Champ[2], we adjusted the values of L for the exponential and the t_3 distributions such that the IC signal probability is approximately 0.005. The L values for the exponential distribution are 7.999 and 9.260 for $n = 50$ and 100, respectively. As for the t_3 distribution, the corresponding Ls for $n = 50$ and 100 are 11.987 and 17.413, respectively. Refer to Table I for the corresponding simulated IC signal probabilities.

4.3. A step change in process mean

The simulated signal probabilities when the process mean incurs a step change are summarized in Table II ($n = 50$) and Table III ($n = 100$). In general, for any given control chart, except for the *X*-chart, the signal probability is higher if the change point occurs in the middle of the sample than early or later in the sample and if the sample size is larger. When the process is

Table II. The OC signal probabilities for the presence of outlying observations

Normal	$k=10$					$k=25$					$k=40$				
$\delta(\mu_0=0)$	*X*-chart	ELR	MW	CM	KS	*X*-chart	ELR	MW	CM	KS	*X*-chart	ELR	MW	CM	KS
0.25	.005	.007	.010	.010	.008	.006	.013	.016	.015	.015	.005	.004	.010	.009	.008
0.50	.007	.038	.031	.029	.022	.007	.025	.074	.066	.060	.007	.010	.031	.029	.022
1.00	.015	.243	.222	.210	.140	.017	.255	.516	.472	.410	.015	.035	.222	.208	.139
1.50	.040	.607	.653	.626	.465	.039	.773	.943	.924	.871	.040	.093	.654	.624	,463
2.00	.102	.896	.940	.929	.816	.080	.988	.999	.998	.994	.103	.203	.941	.928	.816
3.00	.414	.998	1.00	1.00	.998	.254	1.00	1.00	1.00	1.00	.413	.446	1.00	1.00	.998
Exponential	$k=10$					$k=25$					$k=40$				
$\delta(1/\lambda_0=1)$	*X*-chart	ELR	MW	CM	KS	*X*-chart	ELR	MW	CM	KS	*X*-chart	ELR	MW	CM	KS
0.25	.005	.035	.007	.005	.006	.007	.031	.011	.009	.009	.007	.024	.008	.006	.007
0.50	.007	.091	.013	.010	.011	.011	.055	.032	.023	.024	.013	.021	.018	.011	.015
1.00	.009	.186	.040	.028	.034	.025	.150	.127	.095	.104	.042	.047	.061	.035	.052
1.50	.012	.357	.089	.059	.079	.040	.335	.283	.218	.247	.091	.068	.136	.077	.114
2.00	.013	.535	.163	.109	.146	.057	.537	.466	.377	.421	.156	.078	.230	.138	.202
3.00	.015	.774	.334	.232	.307	.085	.786	.727	.646	.692	.291	.182	.402	.267	.368
t_3	$k=10$					$k=25$					$k=40$				
$\delta(\mu_0=0)$	*X*-chart	ELR	MW	CM	KS	*X*-chart	ELR	MW	CM	KS	*X*-chart	ELR	MW	CM	KS
0.25	.005	.006	.007	.007	.006	.005	.012	.010	.009	.009	.005	.004	.007	.007	.006
0.50	.005	.031	.015	.015	.012	.005	.031	.029	.028	.027	.005	.005	.015	.014	.012
1.00	.007	.262	.074	.072	.052	.007	.242	.187	.171	.156	.006	.017	.069	.067	.049
1.50	.009	.562	.246	.236	.165	.008	.621	.539	.509	.464	.007	.100	.229	.223	.158
2.00	.011	.796	.516	.499	.371	.011	.840	.854	.836	.787	.008	.172	.499	.489	.362
3.00	.016	.928	.893	.887	.789	.016	.979	.997	.997	.993	.013	.386	.923	.919	.817

$\alpha=0.005$

normal, it is surprising that the *X*-chart is even not as effective as the other nonparametric Phase I control charts. The signal probabilities of the *X*-chart under normal processes as shown in Tables II and III should give practitioners some caution as to whether the *X*-chart is an effective Phase I control chart even under normal processes, when the process mean and standard deviation have to be estimated using the Phase I data.

Of the three existing nonparametric Phase I charts, the MW-based chart is the best performing chart. This is consistent with the observations in Ross and Adams[19] in the context of Phase II monitoring as the MW test is specifically designed to test the difference of the location parameters coming from two samples of observations.

Comparing the proposed ELR chart with the MW-based chart, if the process is normal, the MW-based chart overall seems to perform better than does the proposed ELR chart. When the process is not normal, the proposed ELR chart has better performance when the change point occurs early or in the middle of the Phase I sample. On the other hand, if the change point occurs later in the Phase I sample, the MW-based chart tends to perform better.

In practice, the process distribution may not be easy to determine, especially if a large Phase I sample is not available. If the prevalent concern in Phase I is that the process mean may incur a sustained shift, the results in Tables I and II seem to indicate that the proposed ELR chart is a viable alternative to the MW-based chart under certain nonnormal processes.

4.4. The presence of outlying observations

Summarized in Table IV are the simulated OC signal probabilities with the presence of two outlying observations that occur on the 20th and 40th observations for $n = 50$ and on the 40th and 80th observations for $n = 100$. The results in Table IV indicate that none of the three existing nonparametric Phase I charts is effective in detecting the presence of outlying observations in the Phase I sample because all the OC signal probabilities are close to the IC signal probability of 0.005. The *X*-chart is also ineffective in detecting the presence of outlying observations, although it performs slightly better than the three existing nonparametric charts for normal and exponential distributions, especially when $\delta \geq 1.5$.

The proposed ELR chart, on the other hand, is much more effective, relative to the *X*-chart and the MW, CM, and KS-based charts, in detecting the presence of outlying observations. Intuitively, when there are only a few outlying observations present in the Phase I sample, the plotting statistics of the MW, CM, and KS-based charts will not change much, thus making these charts ineffective in detecting the presence of outlying observations. On the other hand, because of the constraints placed on the probability masses and the data-dependent weighted sample average, even if only a few outlying observations are present in the Phase I sample, the ELR test statistic $Z^*_{n,k}$, and thus Z^*_n in Equation (8), will likely incur more noticeable changes, thus making the ELR chart more effective than the three existing charts. It should be noted, however, that the OC signal probabilities of the ELR chart are not as high as those under step-change OC scenarios.

4.5. A gradual shift in process mean

When the process mean undergoes a gradual shift, the simulated OC signal probabilities are summarized in Table V for $n = 50$ and 100. When the process is normally distributed, the MW-based chart still has the best performance among the existing nonparametric Phase I charts. It also slightly outperforms the proposed ELR chart.

When the process follows a nonnormal distribution, such as the exponential and t_3 distributions considered here, the proposed ELR chart is more effective than the three existing charts, except when the process follows a t_3 distribution and that $\delta = 3$ under $n = 100$; in that case, both the MW and CM-based charts have slightly better performance. Note that the *X*-chart remains ineffective in detecting a gradual mean shift.

4.6. Summary

1. As compared with the proposed ELR chart and the three existing nonparametric charts, the conventional Phase I *X*-chart is not as effective in detecting OC processes not only when the process distribution is nonnormal but also when the process follows a normal distribution.
2. Of the three existing nonparametric Phase I control charts, the MW-based chart has the best performance in almost all of the OC scenarios considered regardless of the process distribution.
3. The proposed ELR chart, when compared with the MW-based chart, is a viable alternative for detecting step changes in the process location parameter. When a normal process undergoes a step change in its location parameter, the MW-based chart has better performance over all. On the other hand, if the process is nonnormal and the step change occurs early or in the middle of the Phase I sample, the proposed ELR chart performs better than the MW-based chart.
4. When outlying observations are present in the Phase I sample, the ELR chart is much more effective than the *X*-chart as well as the three existing nonparametric charts in all OC scenarios considered. In this case, caution should be exercised when using the existing nonparametric charts as their OC signal probabilities are all very close to the IC signal probability. If the cause of an OC process is because of a gradual shift in the process mean, the proposed ELR chart is more effective than the MW-based chart, especially if the process distribution is nonnormal. However, the MW-based chart has better performance under normal processes.

Table III. The OC signal probabilities for gradual mean shifts

Normal		$k=20$					$k=50$					$k=80$			
$\delta(\mu_0=0)$	*X*-chart	ELR	MW	CM	KS	*X*-chart	ELR	MW	CM	KS	*X*-chart	ELR	MW	CM	KS
0.25	.006	.012	.017	.016	.014	.006	.013	.030	.026	.024	.005	.008	.017	.016	.013
0.50	.007	.068	.088	.080	.058	.008	.051	.202	.178	.144	.007	.041	.088	.080	.058
1.00	.016	.446	.643	.604	.463	.019	.717	.918	.894	.812	.017	.406	.643	.606	.464
1.50	.048	.922	.984	.977	.926	.044	.994	1.00	1.00	.998	.048	.891	.984	.977	.927
2.00	.132	.998	1.00	1.00	.999	.097	1.00	1.00	1.00	1.00	.131	.996	1.00	1.00	.999
3.00	.561	1.00	1.00	1.00	1.00	.343	1.00	1.00	1.00	1.00	.561	1.00	1.00	1.00	1.00
Exponential		$k=20$					$k=50$					$k=80$			
$\delta(1/\lambda_0=1)$	*X*-chart	ELR	MW	CM	KS	*X*-chart	ELR	MW	CM	KS	*X*-chart	ELR	MW	CM	KS
0.25	.006	.048	.011	.008	.008	.007	.028	.019	.013	.014	.007	.022	.012	.009	.010
0.50	.007	.102	.030	.019	.023	.012	.086	.075	.050	.057	.015	.031	.042	.025	.031
1.00	.010	.351	.141	.090	.117	.028	.359	.360	.260	.308	.052	.105	.184	.111	.151
1.50	.012	.629	.347	.235	.303	.047	.698	.687	.558	.641	.116	.287	.392	.260	.345
2.00	.014	.811	.581	.435	.536	.068	.903	.887	.801	.862	.203	.528	.599	.442	.552
3.00	.017	.969	.870	.750	.847	.103	.997	.987	.969	.983	.378	.842	.831	.715	.805
t_3		$k=20$					$k=50$					$k=80$			
$\delta(\mu_0=0)$	*X*-chart	ELR	MW	CM	KS	*X*-chart	ELR	MW	CM	KS	*X*-chart	ELR	MW	CM	KS
0.25	.005	.012	.010	.010	.009	.005	.006	.015	.014	.014	.005	.008	.010	.010	.009
0.50	.005	.052	.033	.031	.026	.005	.052	.069	.063	.055	.005	.025	.032	.031	.026
1.00	.007	.412	.239	.226	.169	.006	.491	.497	.469	.397	.006	.293	.231	.219	.165
1.50	.009	.785	.669	.648	.528	.008	.899	.929	.917	.860	.006	.749	.670	.651	.528
2.00	.011	.922	.931	.924	.852	.010	.973	.998	.998	.992	.008	.944	.946	.938	.866
3.00	.016	.983	.999	.999	.997	.016	.997	1	1	1	.012	.993	1	1	1

$\alpha=0.005$

Table IV. The OC signal probabilities for the presence of outlying observations

			$n = 50$					$n = 100$		
Normal			(20, 40)					(40, 80)		
$\delta(\mu_0 = 0)$	*X*-chart	ELR	MW	CM	KS	*X*-chart	ELR	MW	CM	KS
0.25	.005	.326	.005	.005	.005	.005	.122	.005	.005	.005
0.50	.005	.374	.005	.004	.004	.005	.140	.005	.005	.005
1.00	.006	.428	.005	.004	.005	.006	.171	.005	.005	.005
1.50	.010	.454	.004	.004	.004	.010	.192	.005	.004	.005
2.00	.022	.488	.004	.004	.004	.025	.258	.005	.004	.005
3.00	.104	.540	.004	.004	.004	.147	.370	.004	.005	.005
Exponential			(20, 40)					(40, 80)		
$\delta(1/\lambda_0 = 1)$	*X*-chart	ELR	MW	CM	KS	*X*-chart	ELR	MW	CM	KS
0.25	.005	.378	.005	.004	.003	.005	.120	.005	.004	.004
0.50	.007	.368	.005	.004	.003	.006	.141	.005	.004	.004
1.00	.016	.401	.005	.004	.004	.015	.163	.005	.004	.004
1.50	.033	.412	.005	.004	.004	.032	.195	.005	.004	.004
2.00	.059	.386	.004	.004	.004	.060	.175	.004	.004	.004
3.00	.122	.458	.004	.004	.004	.128	.176	.005	.004	.003
t_3			(20,40)					(40, 80)		
$\delta(\mu_0 = 0)$	*X*-chart	ELR	MW	CM	KS	*X*-chart	ELR	MW	CM	KS
0.25	.005	.356	.005	.005	.005	.005	.138	.005	.005	.005
0.50	.005	.368	.005	.005	.005	.005	.144	.005	.005	.005
1.00	.005	.360	.005	.005	.005	.005	.164	.005	.005	.005
1.50	.005	.352	.005	.005	.005	.005	.144	.005	.005	.005
2.00	.005	.462	.004	.004	.004	.005	.156	.005	.005	.005
3.00	.006	.476	.004	.004	.004	.005	.248	.005	.005	.005

$\alpha = 0.005$.

Table V. The OC signal probabilities for gradual mean shifts

Normal			$n=50$					$n=100$		
$\delta(\mu_0=0)$	*X*-chart	ELR	MW	CM	KS	*X*-chart	ELR	MW	CM	KS
0.25	.005	.006	.009	.008	.008	.005	.009	.014	.013	.012
0.50	.006	.016	.025	.022	.020	.006	.025	.053	.048	.041
1.00	.010	.083	.135	.121	.100	.010	.162	.354	.323	.258
1.50	.017	.238	.403	.371	.305	.018	.567	.812	.778	.679
2.00	.032	.507	.729	.694	.598	.033	.931	.984	.977	.942
3.00	.102	.953	.990	.985	.959	.113	1.00	1.00	1.00	1.00
Exponential			$n=50$					$n=100$		
$\delta(1/\lambda_0=1)$	*X*-chart	ELR	MW	CM	KS	*X*-chart	ELR	MW	CM	KS
0.25	.006	.019	.007	.006	.005	.006	.027	.010	.008	.008
0.50	.008	.048	.015	.010	.012	.008	.065	.025	.019	.020
1.00	.014	.071	.040	.026	.030	.015	.126	.089	.058	.069
1.50	.024	.150	.075	.047	.060	.027	.266	.191	.127	.154
2.00	.037	.219	.121	.078	.098	.042	.400	.312	.212	.261
3.00	.068	.345	.210	.139	.177	.079	.662	.515	.376	.453
t_3			$n=50$					$n=100$		
$\delta(\mu_0=0)$	*X*-chart	ELR	MW	CM	KS	*X*-chart	ELR	MW	CM	KS
0.25	.005	.005	.007	.007	.007	.005	.009	.009	.009	.008
0.50	.005	.014	.016	.013	.012	.005	.026	.024	.023	.021
1.00	006	.062	.052	.048	.043	.006	.140	.127	.118	.099
1.50	.007	.224	.143	.132	.113	.007	.437	.372	.352	.295
2.00	.009	.414	.302	.282	.239	.009	.724	.681	.656	.573
3.00	.013	.783	.481	.656	.579	.012	.958	.973	.967	.936

$\alpha=0.005$.

5. In practical Phase I applications, the exact distribution of the process is usually not readily available. The results summarized in Tables 2–5 indicate that the proposed ELR chart is an effective nonparametric Phase I control chart in that it provides reasonably good detecting power overall under various OC scenarios and process distributions.

5. An example

In this section, we apply the proposed ELR chart to an example discussed in Jones-Farmer, Jordan, and Champ[2]. The data, collected from a regional medical center, consist of the wait times (in minutes) of 150 patients who underwent a colonoscopy procedure, taken over a period of 30 days with five patients per day. For a more detailed account of the example, please refer to Jones-Farmer, Jordan, and Champ[2].

In order to apply the rank-based Phase I control chart, Jones-Farmer, Jordan, and Champ[2] treated the data set as having 30 subgroups of 5 observations per subgroup. The standardized $\overline{X}$-chart is shown in Figure 1. The two sets of control limits, (LCL, UCL) = (−3.1485, 3.1485) and (−3.46, 3.46), give rise to IC signal probabilities of 0.05 and 0.005, respectively. As indicated in Figure 1, an OC signal was detected on day 15 on the $\overline{X}$-chart under 0.05 but not 0.005 IC signal probability. Because each subgroup of five patients was taken throughout the entire day, it is perhaps more appropriate to treat the wait time of each patient as an individual observation, with a total of 150 observations. The standardized *X*-chart of the patient wait times is shown in Figure 2. The two sets of control limits on Figure 2, (LCL, UCL) = (−3.59, 3.59) and (−4.18, 4.18), correspond to IC signal probabilities of 0.05 and 0.005, respectively.

When the IC signal probability is set at 0.05, observations 26, 71, 73, and 148 are OC on the *X*-chart. It is interesting to point out that for the same signal probability, the $\overline{X}$-chart also found that sample 15 was OC and that observations 71 and 73 fall on sample 15. Nevertheless, the *X*-chart already detected an OC signal at observation 71 that was the first observation of day 15, while the $\overline{X}$-chart did not signal until the end of day 15. Further, the *X*-chart also detected OC signals as early as observation 26 and as late as observation 148, while the $\overline{X}$-chart failed to detect OC signals at samples 6 and 30 to which observations 26 and 148 belong, respectively. When the IC signal probability is set at 0.005, the $\overline{X}$-chart no longer shows any OC sample, while the *X*-chart still detects an OC signal at observation 73.

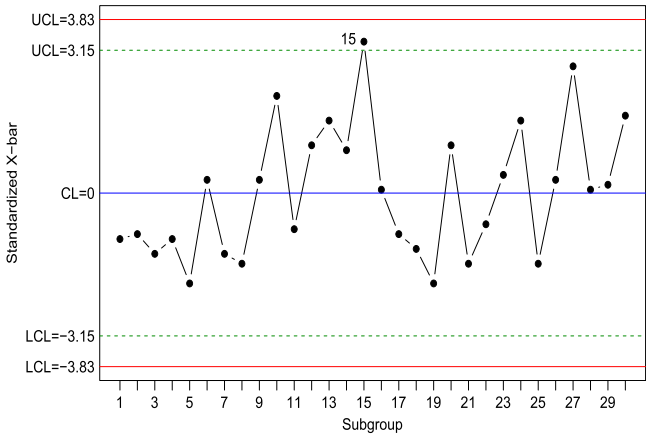

Figure 1. The $\overline{\mathbf{X}}$-chart of the colonoscopy data

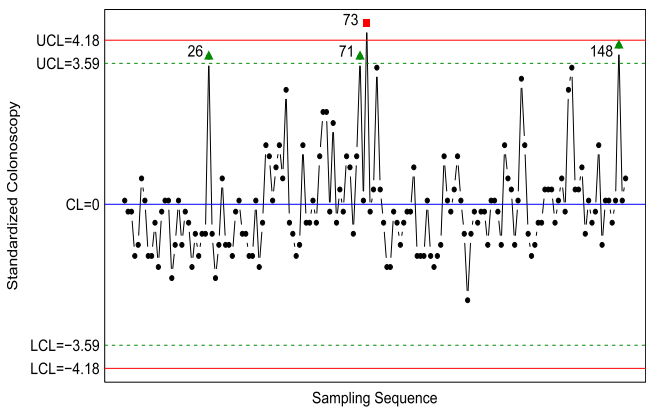

Figure 2. The *X*-chart of the colonoscopy data

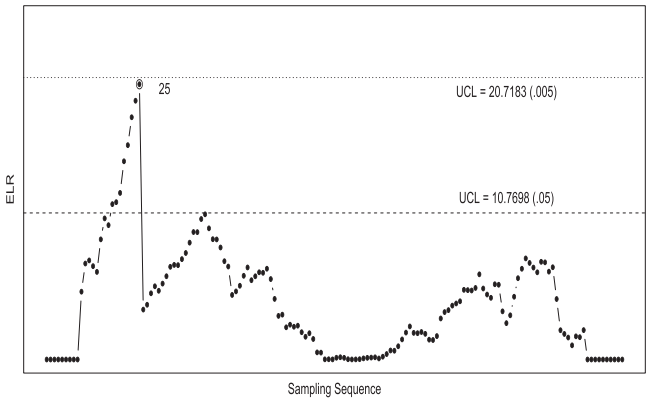

Figure 3. The ELR chart of the colonoscopy data

As suggested by Jones-Farmer, Jordan, and Champ[2], the histogram of the observations indicates that the distribution of the wait times is rightly skewed (Figure 1 of Jones-Farmer, Jordan, and Champ[2]). The violation of normality is likely to affect the performance of the *X*-chart in that the chart is probably more likely to signal than expected under a normal distribution. The proposed ELR chart, when applied to the 150 individual wait times, is shown in Figure 3. The two control limits 10.7698 and 20.7183 correspond to 0.05 and 0.005 IC signal probabilities, respectively.

As suggested in Figure 3, when the IC signal probability is set at 0.05, the ELR chart signals at sample 25, which indicates that the process mean shift starts at sample 26. This conclusion is consistent with what was observed on the *X*-chart in Figure 2.Note that the charting statistic of the ELR chart is the maxmimum of the calculated $Z^*_{n,k}$ (Equation (8)). In Figure 3, we simply plotted $Z^*_{n,k}$ versus the sampling sequence. In practice, once a change point is identified, the data set is split into two groups, the observations up to the change point and the observations after the change point. The ELR chart is further applied to both groups of observations in order to test whether additional change points exist in either of the two groups. In the colonoscopy example, we applied the same ELR calculation to the first 25 observations and the last 125 observations. For the first 25 observations, the test statistic was 3.3175, whereas the control limits for α=.05 and 0.005 were 9.5368 and 23.3197, respectively. As for the last 125 observations, the test statistic turned out to be 4.1573, which was not significant as judged against the control limits of 10.6656 and 20.7780 for α=.05 and 0.005, respectively. Unlike the *X*-chart, which indicated that samples 71, 73, and 148 were also OC, the ELR chart only identified sample 26 as the OC sample. This difference may be because of the increased IC signal probability of the *X*-chart when it is applied to nonnormal processes.

The MW-based Phase I chart of the colonoscopy data is shown in Figure 4. The two control limits for α=.05 and 0.005 are 3.0033 and 3.6508, respectively. The test statistic, obtained at sample 42, exceeded both control limits, indicating that the location parameter of the first 42 observations is different from that of the last 108 observations. Note that this conclusion is different from the *X*-chart and the proposed ELR chart. The data set was further split into two groups, the first 42 observations and the last 108 observations, and the MW-based Phase I chart was applied to both groups of observations. For the first 42 observations, the test statistic was 1.5592, while the control limits were 2.7797 and 3.3873 for α=.05 and 0.005, respectively. On the other hand, the test statistic for the last 108 observations was calculated to be 2.8929 whose control limits for α=.05 and 0.005 were determined to be 2.9599 and 3.6043, respectively. Therefore, similar to what was observed from the ELR chart, no additional change points were identified after the entire data set was split into two groups.

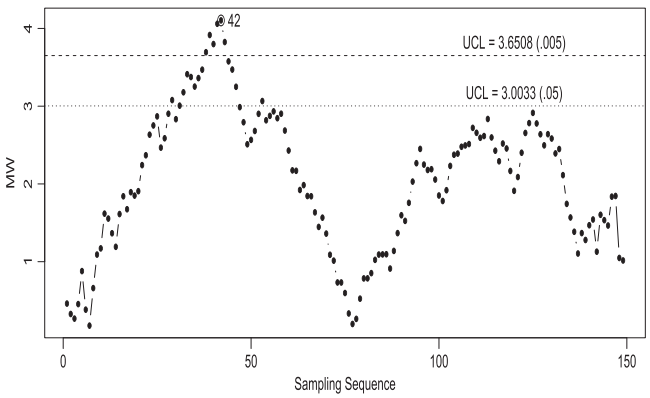

Figure 4. The MW-based chart of the colonoscopy data

Shown in Figure 5 is the CM-based Phase I chart of the colonoscopy data. The control limits are 5.9014 for α=.05 and 8.8669 for α=.005. The test statistic CM_n, which occurred in sample 42, was above both control limits, indicating that the distribution of the first 42 observations is different from that of the last 108 observations. Note that this conclusion is identical to what was observed from the MW-based chart. The entire sample of 150 observations was split into two groups, the first 42 observations and the last 108 observations, and the CM-based chart was applied to both groups of observations. For the first 42 observations, the test statistic was 1.7423, which was not significant, when judged against the control limits at α=.05 and 0.005, which were 5.3955 and 7.9569, respectively. As for the last 108 observations, the test statistic was determined to be 5.1218, which did not exceed the control limits of 5.8789 and 8.7082 at α=.05 and 0.005, respectively. Therefore, no additional change points were further identified in either of the two groups of observations, a conclusion similar to the results obtained from the ELR and MW-based charts.

The KS-based Phase I chart for the colonoscopy data is shown in Figure 6. Note that the scale was changed to $-\log(1-q_{k,n})$ for better visualization. The test statistic of the KS-based chart, which occurred at sample 41, exceeded the control limit at α=.05 but not at α=.005. The change point as identified at sample 41 is similar to that of the MW and CM-based charts. The entire sample was further split into two groups, and the KS-based chart was applied to both groups of observations. For the first 41 observations, the control limits were 0.99723 and 0.99967 for α=.05 and 0.005, respectively, and the test statistic 0.89910 was not significant. As for the last 109 observations, the test statistic was calculated to be 0.99421, which was not significant at α=.05 and 0.005, whose corresponding control limits were 0.99796 and 0.99983, respectively.

Summarizing the observations from Figures 2–6, the three existing nonparametric Phase I charts, the MW, CM, and KS-based charts, gave rise to very similar results. They produced an OC signal at almost the same time, sample 42 for the MW and CM-based charts and sample 41 for the KS-based chart. The *X*-chart, on the other hand, produced an OC signal at as early as sample 26. As indicated in Figure 2, the OC scenario for the colonoscopy data bears the closest resemblance to the case of having outlying observations in the Phase I sample. As pointed out earlier in Section 4, this is an OC scenario that the existing nonparametric charts are very ineffective in detecting. Similar to the *X*-chart, the proposed ELR chart also indicated a change point starting at sample 26.

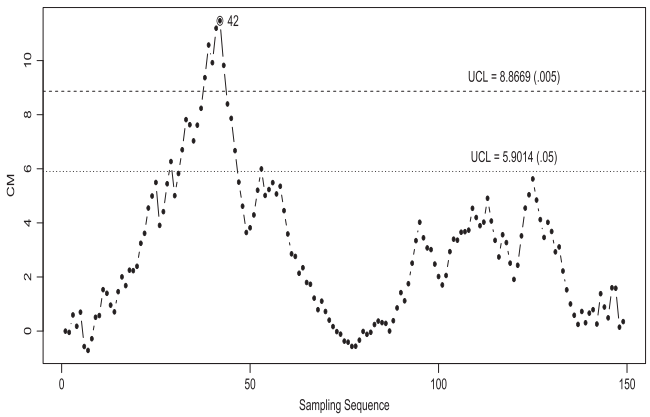

Figure 5. The CM-based chart of the colonoscopy data

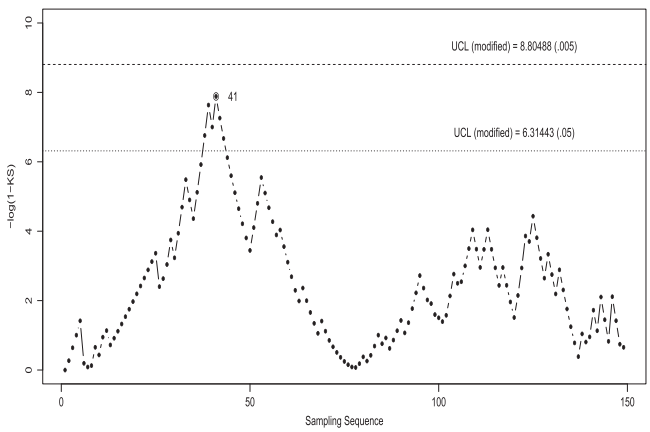

Figure 6. The KS-based chart of the colonoscopy data (the scale has been modified)

6. Concluding remarks

In this work, we proposed a nonparametric Phase I control chart for monitoring the process mean with individual observations. Based on a change-point model formulation, the proposed chart is based on the empirical likelihood ratio test of testing H_0: $\mu_0 = \mu_1$ versus H_a: $\mu_0 \neq \mu_1$. When compared with the conventional Shewhart *X*-chart when the process distribution is indeed normal, the simulation results indicated, rather surprisingly, that the proposed ELR chart and the other existing nonparametric Phase I charts outperform the *X*-chart. The simulation results also indicated that the performance of the proposed ELR chart is more robust against OC scenarios than are the existing charts. For example, the MW-based chart has the best overall performance when the process mean incurs a step change. However, the same chart is very ineffective when there are outlying observations in the sample. On the other hand, the proposed ELR chart is relatively effective in detecting either the presence of outlying observations or a gradual mean shift.

As mentioned earlier, the empirical likelihood ratio test approach is not limited to testing the process mean. It would be worthwhile to study how the same approach can be used to develop nonparametric Phase I control charts for monitoring the process variability, a topic that has received very little attention in the literature. In a recent paper by Yeh and Zerehsaz[50], the authors developed a Phase I control chart for monitoring simple linear profiles with individual observations under the conventional assumption that the error term follows a normal distribution. The empirical likelihood ratio test for linear models is well established in the literature. It naturally leads to the question as to whether it is possible to develop ELR-based nonparametric Phase I control charts for monitoring linear profiles with individual observations.

Several authors, for example, Qiu and Hawkins[21], Zou and Tsung[24], and Zou, Wang, and Tsung[25], have developed nonparametric multivariate control charts for Phase II monitoring. The flexibility of the ELR in dealing with multivariate distributions has been pointed out in the seminal work by Owen[31]. Recently, Sun and Zi[26] developed an ELR-based nonparametric multivariate EWMA chart for Phase II monitoring of the process mean vector. Li, Dai, and Wang[27] studied a nonparametric multivariate Phase I control chart based on the multivariate data depth concept. It would be worth it to develop ELR-based nonparametric multivariate Phase I control charts for monitoring the process mean vector or process variability and compare them with the data-depth-based charts.

Acknowledgements

We would like to thank the editor and two referees for their insights and numerous suggestions that help improve the presentation of the paper.

References

1. Shewhart WA. The application of statistics as an aid in maintaining the quality of manufactured product. *Journal of the American Statistical Association* 1925; **20**:546–548.
2. Jones-Farmer LA, Jordpn V, Champ CW. Distribution-free phase I control charts for subgroup location. *Journal of Quality Technology* 2009; **41**:304–316.
3. Woodall WH, Montgomery DC. Research issues and ideas in statistical process control. *Journal of Quality Technology* 1999; **31**:376–386.
4. Woodall WH. Controversies and contradictions in statistical process control (with discussion). *Journal of Quality Technology* 2000; **28**:341–378
5. Qiu P. Introduction to Statistical Process Control. Chapman& Hall/CRC: Boca Raton, FL, 2014.
6. Albers W, Kallenberg WCM. Empirical nonparametric control charts: Estimation effects and corrections. *Journal of Applied Statistics* 2005: **31**:345–360.
7. Bakir ST. A distribution-free Shewhart quality control chart based on singed-ranks. *Quality Engineering* 2004; **16**:611–623.
8. Chakraborti S, van der Laan P, van de Wiel MA. A class of distribution-free control charts. *Journal of the Royal Statistical Society, Series C* 2004; **53**:443–462.
9. Bakir ST. Distribution-free quality control charts based on signed-rank-like statistics. *Communications in Statistics: Theory and Methods* 2006; **35**:743–757.
10. Chakraborti S, Eryilmaz S. A nonparametric Shewhart-type signed-rank control chart based on runs. *Communications in Statistics: Simulations and Computations* 2007; **36**:335–356.
11. Albers W, Kallenberg WCM. CUMIN charts. *Metrika* 2009; **70**:111–130.
12. Chakraborti S, Eryilmaz S, Human SW. A phase II nonparametric control chart based on precedence statistics with runs-type signaling rules. *Computational Statistics and Data Analysis* 2009; **53**:1054–1065.
13. Zhou C, Zou C, Zhang Y, Wang Z. Nonparametric control chart based on change-point model. *Statistics Papers* 2009; **50**:13–28.
14. Hawkins DM, Deng, Q. A nonparametric change-point control chart. *Journal of Quality Technology* 2010; **42**:165–173.
15. Zou C, Tsung F. Likelihood ratio-based distribution-free EWMA control charts. *Journal of Quality Technology* 2010; **42**:174–196.
16. Qiu, P, Li Z. On nonparametric statistical process control of univariate processes. *Technometrics* 2011; **53**:390–405.
17. Qiu P, Li Z. Distribution-free monitoring of univariate processes. *Statistics and Probability Letters* 2011; **81**:1833–1840.
18. Qiu, P, Zhang J. On Phase II SPC in cases when normality is invalid. *Quality and Reliability Engineering International* 2014; to appear.
19. Liu L, Zou C, Zhang J, Wang Z. A sequential rank-based nonparametric adaptive EWMA control chart. *Communications in Statistics: Simulation and Computation* 2013; **42**:841–859. DOI: 10.1080/03610818.2012.655829.
20. Ross GJ, Adams NM. Two nonparametric control charts for detecting arbitrary distribution changes. *Journal of Quality Technology* 2012; **44**:102–116.
21. Qiu P, Hawkins DM. A rank based multivariate CUSUM procedures. *Technometrics* 2001; **43**:120–132.
22. Qiu P, Hawkins DM. A nonparametric cumulative sum procedure for detecting shifts in all directions. *Journal of the Royal Statistical Society, Series D* 2003; **52**:151–164.
23. Qiu P. Distribution-free multivariate process control based on log-linear modeling. *IIE Transactions on Quality and Reliability Engineering* 2008; **40**:664–677.
24. Zou C, Tsung F. A multivariate sign EWMA control chart. *Technometrics* 2011; **53**:84–97.
25. Zou C, Wang Z, Tsung F. A spatial rank-based multivariate EWMA control chart. *Naval Research Logistics* 2012; **59**:91–110.

26. Sun G, Zi X. An empirical likelihood based multivariate EWMA control scheme. *Communications in Statistics-Theory and Methods* 2013; **42**:429–446.
27. Li Z, Dai Y, Wang Z. Multivariate change point control chart based on data depth for phase I analysis. *Communications in Statistics: Simulation and Computation* 2013; **42**: DOI: 10.1080/03610918.2012.735319.
28. Vostrikova LJ. Detecting 'disorder' in multidimensional random process. *Soviet Mathematics Doklady* 1981; **24**:55–59.
29. Moustakides GV. Optimal stopping times for detecting changes in distributions. *The Annals of Statistics* 1986; **14**:1379–1387.
30. Owen AB. Empirical likelihood ratio confidence interval for a single functional. *Biometrika* 1988; **75**:237–249.
31. Owen AB. Empirical likelihood confidence regions. *The Annals of Statistics* 1990; **18**:90–120.
32. Thomas DR, Grunkemeier GL. Confidence interval estimation of survival probability for censored data. *Journal of the American Statistical Association* 1975; **70**:865–871.
33. Owen AB. Empirical likelihood for linear models. *The Annals of Statistics* 1991; **19**:1725–1747.
34. Hall P. Pseudo-likelihood theory for empirical likelihood. *The Annals of Statistics* 1990; **18**:121–140.
35. DiCiccio T, Hall P, Romano J. Empirical likelihood is Bartlett-correctable. *The Annals of Statistics* 1991; **19**:1053–1061.
36. Qin J, Lawless J. Empirical likelihood and general estimating equations. *The Annals of Statistics* 1994; **22**:300–325.
37. Kolaczyk E. Empirical likelihood for generalized linear models. *Statistica Sinica* 1994; **4**:199–218.
38. Kitamura Y. Empirical likelihood methods with weakly dependent processes. *The Annals of Statistics* 1997; **25**:2084–2102.
39. Kitamura Y. Asymptotic optimality of empirical likelihood for testing moment restrictions. *Econometrica* 2001; **69**:1661–1672.
40. Pan XR, Zhou, M. Empirical likelihood ratio in terms of cumulative hazard function for xensored data. *Journal of Multivariate Analysis* 2002; **80**:166–188.
41. Ren J, Zhou, M. Full likelihood inferences in the cox model: An empirical likelihood approach. *Annals of the Institute of Statistical Mathematics* 2001; **63**:1005–1018.
42. Zou C, Liu Y, Qin P, Wang Z. Empirical likelihood ratio test for the change-point problem. *Statistics and Probability Letters* 2007; **77**:374–382.
43. Liu Y, Zou C, Zhang R. Empirical likelihood ratio test for a change-point in linear regression model. *Communications in Statistics: Theory and Methods* 2008; **37**:2551–2563.
44. Ning W. Empirical likelihood ratio test for a mean change point model with a linear trend followed by an abrupt change. *Journal of Applied Statistics* 2012; **39**:947–961. DOI:10.1080/02664763.2011.628647.
45. Ning W, Pailden J, Gupta AK. The empirical likelihood ratio test for the epidemic change point model. *Journal of Data Science* 2012; **10**:107–127.
46. Owen AB. Empirical Likelihood. Chapman and Hall/CRC: London, 2001.
47. Chen J, Variyatha A, Abrahama B. Adjusted empirical likelihood and its properties. *Journal of Computational and Graphical Statistics* 2008; **17**:426–443.
48. Wang XL, Chen H. Empirical likelihood methods for detecting unknown change points in climate data. Unpublished manuscript.
49. Csörgő M, Horvath L. Limit Theorems in Change-Point Analysis. Wiley: NY, 1997.
50. Yeh AB, Zerehsaz, Y. Phase I control of simple linear profiles with individual observations. *Quality and Reliability Engineering International* 2013; **29**:829–840. DOI: 10.1002/qre.1439.

Appendix

Calculating $\mathbf{Z}^*_{\mathbf{n,k}}$

Using the Lagrange multiplier method, we define

$$G(\mu_0,\lambda_1,\eta_1,\lambda_2,\eta_2,u_i,v_j)=\sum_i log(ku_i)-n\lambda_1\left(\sum_i u_i x_i-\mu_0\right)+\eta_1\left(\sum_i u_i-1\right)$$
$$+\sum_j log\left((n-k).v_j\right)-n\lambda_2(\sum_j v_j x_j-\mu_0)+\eta_2(\sum_j v_j-1),$$

where $i=1,\cdots,k; j=k+1,\cdots,n$. By taking the first derivative of G with respect to u_i, we obtain

$$\frac{\partial G}{\partial u_i}=\frac{1}{u_i}-\lambda_1(x_i-\mu_0)+\eta_1=0$$
$$\Rightarrow u_i=\frac{1}{n\lambda_1(x_i-\mu_0)-\eta_1}.$$

From the previously mentioned equation, we have the following results immediately:

$$\sum u_i\frac{\partial G}{\partial u_i}=\sum u_i\left\{\frac{1}{u_i}-\lambda_1(x_i-\mu_0)+\eta_1\right\}=0$$
$$\Rightarrow\eta_1=-k$$
$$\Rightarrow u_i=\frac{1}{k+n\lambda_1(x_i-\mu_0)}.$$

Similarly, one can also obtain

$$v_j=\frac{1}{(n-k)+n\lambda_2(x_j-\mu_0)}.$$

Finally, taking the first derivative of G with respect to μ_0, it leads to

$$\frac{\partial G}{\partial \mu_0} = \lambda_1 + \lambda_2 = 0 \Rightarrow \lambda_2 = -\lambda_1.$$

Therefore,

$$v_j = \frac{1}{(n-k) - n\lambda_1(x_j - \mu_0)}.$$

For convenience, we denote $\theta_k = \frac{k}{n}$ and $\lambda_1 = \lambda$. It follows that

$$u_i = \frac{1}{n\theta_k + n\lambda(x_i - \mu_0)} = \frac{1}{n\theta_k} \times \frac{1}{1 + \theta_k^{-1}\lambda(x_i - \mu_0)},$$
$$v_j = \frac{1}{n(1-\theta_k) - n\lambda(x_j - \mu_0)} = \frac{1}{n(1-\theta_k)} \times \frac{1}{1 - (1-\theta_k)^{-1}\lambda(x_i - \mu_0)}.$$

Hence,

$$Z_{n,k} = Z(\theta_k, \lambda, \mu_0) = 2\left\{\sum_i log\left(1 + \theta_k^{-1}\lambda(x_i - \mu_0)\right) + \sum_j log\left(1 - (1-\theta_k)^{-1}\lambda(x_j - \mu_0)\right)\right\}.$$

Define the score functions

$$\phi_1(\lambda, \mu_0) = \frac{\partial Z(\theta_k, \lambda, \mu_0)}{2\partial\lambda} = \sum_i \frac{\theta_k^{-1}(x_i - \mu_0)}{1 + \theta_k^{-1}\lambda(x_i - \mu_0)} - \sum_j \frac{(1-\theta_k)^{-1}(x_j - \mu_0)}{1 - (1-\theta_k)^{-1}\lambda(x_j - \mu_0)}$$

and

$$\phi_2(\lambda, \mu_0) = \frac{\partial Z(\theta_k, \lambda, \mu_0)}{-2\lambda\partial\mu_0} = \sum_i \frac{\theta_k^{-1}}{1 + \theta_k^{-1}\lambda(x_i - \mu_0)} - \sum_j \frac{(1-\theta_k)^{-1}}{1 - (1-\theta_k)^{-1}\lambda(x_j - \mu_0)}.$$

Then, $\left(\hat{\lambda}(\theta_k), \hat{\mu}_0(\theta_k)\right)$ are determined by

$$\phi_1\left(\hat{\lambda}(\theta_k), \hat{\mu}_0(\theta_k)\right) = 0,$$
$$\phi_2\left(\hat{\lambda}(\theta_k), \hat{\mu}_0(\theta_k)\right) = 0.$$

Note that the previously mentioned two equations $\phi_1(\lambda, \mu_0)$ and $\phi_2(\lambda, \mu_0)$ do not have closed form solutions. Here, we used a Newton–Raphson algorithm to obtain numerical solutions to these equations. Therefore, we have

$$Z^*_{n,k} = Z\left(\theta_k, \hat{\lambda}(\theta_k), \hat{\mu}_0(\theta_k)\right).$$

Authors' biographies

Dr. Wei Ning received his PhD in statistics from Syracuse University in 2006. Currently, he is an associate professor of statistics in the Department of Mathematics and Statistics at Bowling Green State University. His research interests include change point analysis, empirical likelihood method, sequential analysis, and time series analysis.

Dr. Arthur Yeh is a professor of statistics and chair of the Department of Applied Statistics and Operations Research at Bowling Green State University. Over the years, Dr. Yeh has conducted and published research in several areas of industrial statistics, including, among others, optimal experimental designs, computer experiments, univariate and multivariate control charts, multivariate process capability indices, univariate and multivariate run-by-run process control, and statistical profile monitoring. He currently serves as an associate editor for *The Statistical Papers*. He was an associate editor for *The American Statistician* in the past. He has also served in the past as the president of the Northwest Ohio Chapter of the American Statistical Association and the chair of the Toledo Section of the American Society for Quality.

Dr. Xinqu Wu received his PhD in statistics from the Graduate University of Chinese Academy of Science in Beijing in 2014. Currently, he is a research associate in Alibaba Group in Hang Zhou, China.

Boxiang Wang received his BS degree in statistics from Nankai University, China, in 2010. He graduated from the Department of Applied Statistics and Operations Research at Bowling Green State University with an MS degree in 2012. He is currently a PhD student in the School of Statistics at the University of Minnesota. His main research interests include statistical quality control, experimental design, and machine learning, with a special focus on classification.