Research Article

A Nonparametric Shewhart-Type Quality Control Chart for Monitoring Broad Changes in a Process Distribution

Saad T. Bakir

College of Business Administration, Alabama State University, P.O. Box 271, Montgomery, AL 36101, USA

Correspondence should be addressed to Saad T. Bakir, bakir00@yahoo.com

Received 7 May 2012; Revised 17 July 2012; Accepted 22 July 2012

Academic Editor: Xiaohu Li

Copyright © 2012 Saad T. Bakir. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper develops a distribution-free (or nonparametric) Shewhart-type statistical quality control chart for detecting a broad change in the probability distribution of a process. The proposed chart is designed for grouped observations, and it requires the availability of a reference (or training) sample of observations taken when the process was operating in-control. The charting statistic is a modified version of the two-sample Kolmogorov-Smirnov test statistic that allows the exact calculation of the conditional average run length using the binomial distribution. Unlike the traditional distribution-based control charts (such as the Shewhart X-Bar), the proposed chart maintains the same control limits and the in-control average run length over the class of all (symmetric or asymmetric) continuous probability distributions. The proposed chart aims at monitoring a broad, rather than a one-parameter, change in a process distribution. Simulation studies show that the chart is more robust against increased skewness and/or outliers in the process output. Further, the proposed chart is shown to be more efficient than the Shewhart X-Bar chart when the underlying process distribution has tails heavier than those of the normal distribution.

1. Introduction

Most traditional statistical quality control charts assume that the monitored process has a prespecified known probability distribution (usually normal for continuous measurements). Consequently, the chart properties (control limits, false alarm rate, and the in-control average run length) would be in error if the process distribution were misspecified. To remedy this, a number of distribution-free (or nonparametric) schemes that maintain the same chart properties over a class of distributions have been proposed in the literature. For an overview of nonparametric control charts, see Chakraborti et al. [1, 2].

Another problem is that traditional control charts aim at monitoring a change in one parameter (usually a location or scale) of a process distribution. Realistically, however, when a special cause influences a process, it may cause a shift in more than one parameter (location, scale, skewness, etc.) of the process distribution. To remedy this, we need control charts designed to monitor a broad rather than a one-parameter change in a process distribution. To our knowledge, Bakir [3] was first to suggest such charts based on the two-sample

Kolmogorov-Smirnov and the Cramer-von Mises statistics. Zou and Tsung [4] proposed a nonparametric likelihood ratio chart for monitoring broad changes in a process distribution. Ross and Adams [5] developed nonparametric charts based on the two-sample Kolmogorov-Smirnov and the Cramer-von Mises statistics. Their charts, however, are designed for individual observations whereas the chart proposed in this paper is designed for grouped observations.

In this paper, we propose a nonparametric Shewhart-type control chart for monitoring a broad change in a process probability distribution. To develop the chart, we assume the availability of a training (or reference) sample taken when the process was operating in statistical control. The idea of assuming a training sample was first used by Park and Reynolds [6] to develop distribution-free charts based on the Urban and Wolfe [7] placement statistic. Later, Hackl and Ledolter [8], Willemain and Runger [9], and Chakraborti et al. [10] proposed nonparametric charts assuming the availability of a reference sample. Our proposed chart works by taking a random sample (test sample) from the process output at each monitoring stage. The charting statistic is a modified version of the two-sample Kolmogorov-Smirnov

test statistic where the difference of the reference and test empirical distribution functions is maximized only over the training sample values. Such modification allows exact calculation of the conditional average run length of the proposed chart using the binomial distribution. Unlike the traditional distribution-based Shewhart X-Bar (Shew-XB) chart, the proposed chart maintains the same control limits and the same in-control average run length (ARL_0) over the class of all (symmetric or asymmetric) continuous distributions. The Shew-XB and its average run length (ARL) will be discussed in Section 5. Given the training sample, the exact conditional ARL of the proposed nonparametric chart is computed using the binomial probability distribution. The unconditional ARL can then be computed approximately using simulations. A preliminary simulation study shows that the proposed nonparametric chart is more efficient (has smaller out-of-control ARL) than the Shew-XB chart under distributions with tails heavier than those of the normal distribution. If the process distribution is actually normal, then the Shew-XB chart is more efficient, as expected. The simulation study also indicates that the proposed chart is more robust against increased skewness and/or outliers in the process output.

The rest of the paper is organized as follows: Section 2 presents notational preliminaries. Section 3 develops the proposed nonparametric control chart, and Section 4 develops its ARL. Section 5 discusses the Shew-XB chart and its ARL. Section 6 investigates the effects of skewness and outliers on the two charts and presents efficiency comparisons.

2. Preliminaries

We assume the availability of a training random sample, $\vec{X}_0 = (X_{01}, X_{02}, \dots, X_{0m})$ of size $m > 1$ observations taken when the process was operating in-control. The in-control process distribution is assumed to have a continuous cumulative distribution function (CDF), F_0 . Let $S_0(z)$ denote the empirical distribution function (EDF) of the training sample, as defined by

$$S_0(z) = \begin{cases} 0 & \text{if } z < X_{0(1)}, \\ \frac{i}{m} & \text{if } X_{0(j)} \leq z < X_{0(j+1)} \text{ for } j = 1, 2, \dots, m-1, \\ 1 & \text{if } z \geq X_{0(m)}. \end{cases} \quad (1)$$

Here, $X_{0(j)}$ is the j th order statistic of the training sample, $\vec{X}_0$. Then at each sampling instance t , $t = 1, 2, \dots$, we obtain one test sample $\vec{Y}_t = (Y_{t1}, Y_{t2}, \dots, Y_{tn})$ of size $n > 1$ from the process output, which is assumed to have a continuous CDF, F_y . Let $S_t(z)$ be the empirical distribution function of the test sample, $\vec{Y}_t$, given by

$$S_t(z) = \begin{cases} 0 & \text{if } z < Y_{t(1)}, \\ \frac{i}{n} & \text{if } Y_{t(i)} \leq z < Y_{t(i+1)} \text{ for } i = 1, 2, \dots, n-1, \\ 1 & \text{if } z \geq Y_{t(n)}. \end{cases} \quad (2)$$

Here, $Y_{t(i)}$ is the i th order statistic of the test sample, $\vec{Y}_t$.

FIGURE 1: $H_a^- : F_y(z) > F_0(z)$.

In practice, we may need to detect one of the following three situations.

Situation 1. Detect whether or not the process tends to produce stochastically smaller observations than the observations of the in-control state. In the terminology of statistical hypothesis testing, we are interested in testing the following null and alternative hypotheses:

$$\begin{aligned} H_0^- : F_y(z) &\leq F_0(z) \quad \text{for all } -\infty < z < \infty, \\ H_a^- : F_y(z) &\geq F_0(z) \quad \text{for all } z, \\ F_y(z) &> F_0(z) \quad \text{for at least one } z. \end{aligned} \quad (3)$$

Figure 1 depicts Situation 1 graphically and it shows that the process CDF, F_y , has shifted to the left of the in-control CDF F_0 .

Situation 2. Detect whether or not the process tends to produce stochastically larger observations than the observations of the in-control state. That is, we are testing the following null and alternative hypotheses:

$$\begin{aligned} H_0^+ : F_y(z) &\geq F_0(z) \quad \text{for all } -\infty < z < \infty, \\ H_a^+ : F_y(z) &\leq F_0(z) \quad \text{for all } z, \\ F_y(z) &< F_0(z) \quad \text{for at least one } z. \end{aligned} \quad (4)$$

Figure 2 depicts Situation 2 graphically and it shows that the process CDF F_y has shifted to the right of the in-control CDF F_0 .

Situation 3. Detect whether or not the process tends to produce smaller and/or larger observations than the in-control state. That is, we are testing the following null and alternative hypotheses:

$$\begin{aligned} H_0 : F_y(z) &= F_0(z) \quad \text{for all } -\infty < z < \infty, \\ H_a : F_y(z) &\neq F_0(z) \quad \text{for at least one } z. \end{aligned} \quad (5)$$

Figure 3 depicts Situation 3 graphically.

FIGURE 2: $H_a^+ : F_y(z) < F_0(z)$. FIGURE 3: $H_a : F_y(z) \neq F_0(z)$.

3. The Proposed Nonparametric Control Chart

In this section, we develop the steps for constructing a distribution-free control chart of the Shewhart type, that is, based on a modified version of the two-sample Kolmogorov-Smirnov statistic. The proposed chart, hereafter, is abbreviated to Shew-KS chart.

Step 1: control characteristic. The characteristic to be controlled (monitored) is the process theoretical probability distribution represented by the CDF, F_y . The purpose is to detect whether or not F_y has shifted away from the process in-control CDF, F_0 .

Step 2: sampling plan. Obtain a training sample $\vec{X}_0 = (X_{01}, X_{02}, \dots, X_{0m})$ of size $m > 1$ when the process was operating in-control. Then obtain a test sample $\vec{Y}_t = (Y_{t1}, Y_{t2}, \dots, Y_{tn})$ of size $n > 1$ from the process output at each sampling instance t , $t = 1, 2, \dots$

Step 3: assumptions. Observations on the process output are independent. The test samples are drawn from unknown continuous distribution with CDF, F_y . The process in-control underlying distribution is assumed continuous with unknown CDF, F_0 .

Step 4: pivot statistics. Calculate $S_0(z)$, the EDF of the training sample $\vec{X}_0$. Then at each sampling instance t , $t = 1, 2, \dots$, calculate the EDF, $S_t(z)$, of the test sample

$\vec{Y}_t = (Y_{t1}, Y_{t2}, \dots, Y_{tn})$. The pivot statistic for Situation 1 (the lower-sided Shew-KS chart) is

$$\psi_t^- = \min_{z=x_{0j}} [S_0(z) - S_t(z)]. \quad (6)$$

Note that ψ_t^- tends to be negative when the process produces observations smaller than the in-control state, see Figure 1.

The pivot statistic for Situation 2 (the upper-sided Shew-KS chart) is

$$\psi_t^+ = \max_{z=x_{0j}} [S_0(z) - S_t(z)]. \quad (7)$$

Note that ψ_t^+ tends to be positive when the process produces larger observations, see Figure 2.

The pivot statistic for Situation 3 (the two-sided Shew-KS chart) is

$$\psi_t = \max_{z=x_{0j}} |S_0(z) - S_t(z)|. \quad (8)$$

Note 1. The pivot statistics in (6), (7), and (8) will assume only integer values if each is multiplied by the constant mn .

Note 2. The pivot statistics are modified versions of the traditional two-sample Kolmogorov-Smirnov statistic ([11], pp 456–462) where maximization is taken only over the training sample observations, $\vec{X}_0 = (X_{01}, X_{02}, \dots, X_{0m})$.

Step 5: control sequence (or charting statistics). The control sequences for the lower-sided, the upper-sided, and the two-sided Shew-KS charts, respectively, are

$$\begin{aligned} \{\psi_t^-, t = 1, 2, \dots\}, \quad \{\psi_t^+, t = 1, 2, \dots\}, \\ \{\psi_t, t = 1, 2, \dots\}. \end{aligned} \quad (9)$$

Step 6: control limits. For simplification, we consider one upper-sided control limit, L , and let the lower-sided control limit be $-L$. Because the Shew-KS chart is distribution free, the control limit, L , is a constant (design parameter) that depends only on m , n , and the desired in-control ARL_0 of the chart. This control limit, however, does not depend on the functional form of the in-control process distribution.

Step 7: signaling rules. The two-sided Shew-KS signals if $\psi_t \geq L$. The lower-sided and upper-sided control charts signal, respectively, if $\psi_t^- \leq -L$, and $\psi_t^+ \geq L$.

Illustration. Let $N(\theta, \sigma^2)$ denote a normal probability distribution with mean θ and variance σ^2 . As an illustration of the proposed Shew-KS chart, we generated 20 observations from the standard normal distribution $N(0,1)$ to represent the in-control reference X -sample. Four test Y -samples, each of size 10, were generated. The first two samples, $Y1$ and $Y2$, have a $N(0,1)$ distribution. The third and fourth samples, $Y3$ and $Y4$, have an $N(2,1)$ and an $N(3,4)$ distributions, respectively. Table 1 depicts the generated samples and the required calculations for the two-sided charting statistic. The resulting Shew-KS chart, shown in Figure 4, gives an out-of-control signal at the third sample when the process mean shifted from zero to two.

FIGURE 4: Shew-KS chart.

4. Calculating the ARL of the Shew-KS Chart

Values of the ARL are needed for the implementation and the performance evaluation of control charts. The implementation of a control chart requires values of the control limits that lead to some desired values of the in-control ARL_0 . When the successive charting statistics of a Shewhart type control chart are independent, the run length distribution is geometric and the $ARL = 1/\Pr(\text{signal})$. Unfortunately, this property of independence does not hold for the proposed Shew-KS chart because the successive charting statistics, ψ_t , $t = 1, 2, \dots$, all depend on the same training sample $\vec{X}_0$. In this section, we develop a method for calculating the ARL of the chart by first conditioning on the training sample $\vec{X}_0$, a method used by Chakraborti [12, 13] and by Vermaat et al. [14].

Recall that the proposed two-sided Shew-KS chart signals at the first sampling instance t , $t = 1, 2, \dots$ for which $\max_{z=x_{0j}} |S_0(z) - S_t(z)| \geq L$, where $L > 0$. Suppose that the maximum occurs at a value, say, $z = z_{\max}$. Thus, a signal occurs if

$$|S_0(z_{\max}) - S_t(z_{\max})| \geq L. \quad (10)$$

Equivalently, a signal occurs if

$$S_0(z_{\max}) - S_t(z_{\max}) \leq -L, \quad (11)$$

or

$$S_0(z_{\max}) - S_t(z_{\max}) \geq L. \quad (12)$$

It is seen that (11) represents the branch of the chart that detects if the process output, $\vec{Y}_t$, is stochastically smaller than the in-control output, $\vec{X}_0$, see Figure 1. Similarly, (12) detects if the process output is stochastically larger than the in-control output, see Figure 2.

We will work first on the lower branch, (11), of the chart. After rearranging terms and multiplying the inequality by the test sample size, n , (11) becomes

$$nS_t(z_{\max}) \geq n[L + S_0(z_{\max})]. \quad (13)$$

Note that because maximization in (6)–(8) is defined over the X values only, $z_{\max}$ becomes fixed when we condition on $\vec{X}_0$. Consequently, given the training sample, $\vec{X}_0$, $nS_t(z_{\max})$ becomes a binomial random variable, $B_{n,\pi}$, with number of trials $= n$ and probability of success $\pi = F_y(z_{\max})$. Given $\vec{X}_0$, the exact conditional probability of a signal and the exact conditional ARL of the lower branch of the chart, respectively, are

$$\begin{aligned} \text{cond}P_{\text{Shew-KS}}^- &= \Pr(B_{n,\pi} \geq n[S_0(z_{\max}) + L] \mid \vec{X}_0), \\ \text{cond}ARL_{\text{Shew-KS}}^- &= \frac{1}{\text{cond}P_{\text{Shew-KS}}^-}. \end{aligned} \quad (14)$$

Upon taking expectations over the training sample $\vec{X}_0$, the unconditional probability of a signal and the unconditional ARL for the lower branch of the Shew-KS chart are

$$\begin{aligned} P_{\text{Shew-KS}}^- &= E(\text{cond}P_{\text{Shew-KS}}^-), \\ ARL_{\text{Shew-KS}}^- &= E(\text{cond}ARL_{\text{Shew-KS}}^-). \end{aligned} \quad (15)$$

Similarly, (12) can be transformed to show that the signal conditional probability and the conditional ARL of the upper branch of the chart, respectively, are

$$\begin{aligned} \text{cond}P_{\text{Shew-KS}}^+ &= \Pr(B_{n,\pi} \leq n[S_0(z_{\max}) - L] \mid \vec{X}_0), \\ \text{cond}ARL_{\text{Shew-KS}}^+ &= \frac{1}{\text{cond}P_{\text{Shew-KS}}^+}. \end{aligned} \quad (16)$$

The two-sided Shew-KS chart signals if either one of the lower or the upper branch signals. Therefore, given the training sample $\vec{X}_0$, the conditional probability of a signal and the conditional ARL of the two-sided chart, respectively, are

$$\begin{aligned} \text{cond}P_{\text{Shew-KS}} &= \text{cond}P_{\text{Shew-KS}}^- + \text{cond}P_{\text{Shew-KS}}^+ \\ \text{cond}ARL_{\text{Shew-KS}} &= \frac{1}{\text{cond}P_{\text{Shew-KS}}}. \end{aligned} \quad (17)$$

Theoretically, the unconditional probability of a signal and the unconditional ARL of the one-sided and two-sided charts are the expectations, over the training sample, of the respective conditional expressions.

Unfortunately, the required unconditional expectations over the training sample, $\vec{X}_0$, cannot be expressed directly into a closed form. In this paper we use a large number of simulations, $M = 1$ million runs, to estimate these unconditional expectations. At each simulation run, a training sample $\vec{X}_0$ and a test sample $\vec{Y}_t$ are generated where the conditional probability of a signal and the conditional ARL are calculated according to their exact formulas. The International Mathematical and Statistical Library (IMSL) is used to generate pseudo random variables (assuming a certain probability distribution) for the training and test samples, calculate the empirical CDFs, identify $z = z_{\max}$, calculate the exact binomial probabilities, and finally calculate the exact conditional probabilities of a signal and the ARLs. Then we

average out these conditional values over the number of simulations to get estimates of the required unconditional expectations. For example, the estimated values for the unconditional probability of a signal and the unconditional ARL for the two-sided Shew-KS chart, respectively, are

$$\begin{aligned}\hat{P} &= \frac{1}{M} \sum_{r=1}^M \text{cond}P_r, \\ \hat{\text{ARL}} &= \frac{1}{M} \sum_{r=1}^M \text{condARL}_r,\end{aligned}\quad (18)$$

where r is the simulation run number. Similar calculations are applied to estimate the unconditional expectations of the one-sided charts. The above methods for calculating the signal probability and the ARL can play an important role in the design and implementation of the proposed Shew-KS chart because they allow for calculating control limits that correspond to certain desired values of the in-control ARL for various values of m and n .

5. Calculating the ARL of the Traditional Shew-XB Chart

In this section, we outline an efficient method for evaluating the ARL of the traditional Shew-XB chart in order to compare it to the proposed Shew-KS chart.

The traditional Shew-XB control chart is based on charting the sequence of means $\bar{Y}_t = (1/n) \sum_{j=1}^n Y_{tj}$ of the test samples $\vec{Y}_t = (Y_{t1}, Y_{t2}, \dots, Y_{tn})$, $t = 1, 2, \dots$. The control limits are calculated using the sample mean $\bar{X}_0 = (1/m) \sum_{j=1}^m X_{0j}$ and the sample standard deviation $S_0 = \sqrt{\sum_{j=1}^m (X_{0j} - \bar{X}_0)^2 / (m-1)}$ of the in-control training sample $\vec{X}_0 = (X_{01}, X_{02}, \dots, X_{0m})$. The two-sided Shew-XB chart gives an out-of-control signal at the first sampling instance, t , for which $\bar{Y}_t \leq \bar{X}_0 - kS_0/\sqrt{n}$ or $\bar{Y}_t \geq \bar{X}_0 + kS_0/\sqrt{n}$, where k is a constant chosen (usually equals 3) to achieve a desired in-control ARL. One-sided Shew-XB charts can be obtained by employing one of the signaling rules.

Because the successive signaling events (e.g., $\bar{Y}_t \leq \bar{X}_0 - kS_0/\sqrt{n}$) of the Shew-XB control chart all use the same control limits as estimated from the same training sample, they are no longer independent. Therefore, we cannot use the geometric distribution argument that the $\text{ARL} = 1/\text{Pr}(\text{signal})$. Jensen et al. [15] presented a literature review on the effects of parameter estimation on control charts properties. Chakraborti [12, 13] used conditional expectation arguments to derive exact formulas for the run length distribution and the ARL of the Shew-XB chart when the in-control mean and/or the variance are estimated. However, almost all studies regarding the effects of parameter estimation on control charts properties assume that the underlying process distribution is normal. This dogmatic restriction to the normal distribution is not appropriate to the distribution-free world of nonparametric statistics where we need to compare the performance of the competing charts under distributions other than the normal.

In this section, we use a conditional expectation argument and simulations to obtain reasonable estimates for the values of the unconditional ARL of the Shew-XB chart under several underlying process distributions.

Given the training sample, $\vec{X}_0$, the exact conditional probabilities of signals for the lower and the upper branches of the Shew-XB chart, respectively, are

$$\begin{aligned}\text{cond}P_{\text{Shew-XB}}^- &= F_{\bar{Y}}\left(\bar{X}_0 - \frac{kS_0}{\sqrt{n}} \mid \vec{X}_0\right), \\ \text{cond}P_{\text{Shew-XB}}^+ &= 1 - F_{\bar{Y}}\left(\bar{X}_0 + \frac{kS_0}{\sqrt{n}} \mid \vec{X}_0\right),\end{aligned}\quad (19)$$

where $F_{\bar{Y}}$ is the theoretical CDF of the sample mean of the test sample, $\bar{Y}_t$. For example, if the test sample has a normal distribution with mean μ_y and variance σ_y^2 , then the conditional probability of a signal of the lower and the upper branches of the Shew-XB charts, respectively, are

$$\begin{aligned}\text{cond}P_{\text{Shew-XB}}^- &= \Phi\left(\frac{\bar{X}_0 - \mu_y - k(S_0/\sqrt{n})}{\sigma_y/\sqrt{n}} \mid \vec{X}_0\right), \\ \text{cond}P_{\text{Shew-XB}}^+ &= 1 - \Phi\left(\frac{\bar{X}_0 - \mu_y + k(S_0/\sqrt{n})}{\sigma_y/\sqrt{n}} \mid \vec{X}_0\right),\end{aligned}\quad (20)$$

where Φ is the CDF of the standard normal distribution. The conditional probability of a signal for the two-sided chart is

$$\text{cond}P_{\text{Shew-XB}} = \text{cond}P_{\text{Shew-XB}}^- + \text{cond}P_{\text{Shew-XB}}^+. \quad (21)$$

The exact CDFs of the sample mean are known for many populations beside the normal. We state some results concerning the CDF of the mean of a sample of size n drawn from gamma, Cauchy and Laplace distributions.

- (1) For a 3-parameter gamma (shape = α , scale = β , location = θ) distribution with probability density function (PDF)

$$f_y(z) = \frac{(z - \theta)^{\alpha-1} [e^{-(z-\theta)/\beta}]}{\beta^\alpha \Gamma(\alpha)}. \quad (22)$$

The mean of a random sample of size n also has a gamma (shape = $n\alpha$, scale = β/n , location = θ) distribution.

- (2) For a Cauchy distribution (scale = β , location = θ), the mean of a random sample of size n has the same Cauchy distribution with PDF

$$f_y(z) = f_{\bar{y}}(z) = \frac{\beta}{\pi [\beta^2 + (z - \theta)^2]}, \quad (23)$$

$$\text{for } -\infty < z < \infty, -\infty < \theta < \infty, \beta > 0.$$

- (3) For a Laplace distribution (scale = β , location = θ) with PDF

$$f_y(z) = \frac{1}{\beta} e^{-|z-\theta|/\beta}, \quad (24)$$

$$\text{for } -\infty < z < \infty, -\infty < \theta < \infty, \beta > 0,$$

the distribution of the sample mean is a little bit complicated. Using basic results in Johnson et al. ([16], pp 167), we can express the CDF of the sample mean (when $z > 0$ and $\theta = 0$) as

$$F_{\bar{y}}(z) = 1 - \sum_{j=0}^{n-1} 2^{j+1-2n} \binom{2n-j-2}{n-1} \Pr(G_{j+1} > z), \quad (25)$$

where G_{j+1} is a 2-parameter gamma (shape = $j + 1$, scale = β/n) random variable.

The conditional ARLs for the lower, upper, and two-sided Shew-XB chart, respectively, are $\text{condARL}_{\text{Shew-XB}}^- = 1/\text{cond}P_{\text{Shew-XB}}^-$, $\text{condARL}_{\text{Shew-XB}}^+ = 1/\text{cond}P_{\text{Shew-XB}}^+$, and $\text{condARL}_{\text{Shew-XB}} = 1/\text{cond}P_{\text{Shew-XB}}$.

The unconditional probabilities of signals and the unconditional ARLs for the Shew-XB chart are obtained by taking expectations of their respective conditional values over the training sample $\vec{X}_0$. In practice, we use large number of simulations to estimate these unconditional quantities as described in Section 4, (18).

6. Effects of Skewness, Outliers, and Efficiency Comparisons

In this section, we conduct simulation studies to investigate the sensitivity against skewness, outliers, and the efficiency of both the traditional Shew-XB and the proposed Shew-KS control charts.

6.1. Effects of Skewness. We now examine the effect of skewness on the in-control ARLs of the Shew-XB and the Shew-KS charts. The control limits for the two charts are adjusted so that both charts have the same in-control ARL of 170 under the standard normal distribution. The sample sizes of the training and test sample, respectively, are $m = 39$ and $n = 10$. We used IMSL to generate pseudo random numbers from the three-parameter gamma distribution in (22). We varied the shape parameter α to obtain extremely skewed to almost symmetric distributions. To have a gamma distribution with mean = 0 and variance = 1, the scale and location parameters are chosen as $\beta = \alpha^{-1/2}$ and $\theta = -\sqrt{\alpha}$. The ARLs of both charts are calculated by first getting the exact conditional ARL and then using one million simulation runs to get the unconditional ARL by averaging the conditional ARL. Table 2 shows that ARL_0 of the Shew-KS is chart not affected at all by the skewness of the distribution. The Shew-XB chart, however, changes dramatically as we move from extremely skewed to symmetric distributions. The ARL_0 of the Shew-XB chart becomes close to the normal theory value of 170 only when the shape parameter of the gamma distribution is at least 16. Table 2 depicts two anomalies in the ARL of the Shew-XB chart when the shape parameter $\alpha = 2$ or 3. For explanation of these anomalies, refer to Vermaat et al. ([14], pp. 343).

6.2. Effects of Outliers. There are situations in which the in-control process output is contaminated by few outliers; for

TABLE 2: In-control ARLs (in groups of size 10) of the Two-sided Shew-XB and the Shew-KS charts for a process with a gamma (shape α , scale β , location θ) Distribution. Training Sample Size $m = 39$ and Test Sample Size $n = 10$.

Shape α	Scale $\beta = \alpha^{-1/2}$	Location $\theta = -\sqrt{\alpha}$	Shew-XB $k = 2.75$	Shew-KS $CL = 0.541$
1	1	-1	1099.2	165
2	0.7071	-1.4142	2125.4	167.3
3	0.5774	-1.7321	1764.9	164.3
4	0.5000	-2.0000	320.5	176.7
5	0.4472	-2.2361	289.1	173.1
9	0.3333	-3.0000	227	166.6
10	0.3162	-3.1623	181.6	181.0
12	0.2887	-3.4641	235.7	171.6
16	0.2500	-4.0000	189.9	160.5
20	0.2236	-4.4721	168.5	179.5
Normal	1.0	0.0	170	170

Note: for the Shew-XB chart, 2.75 is the “ k ” value in the control limits $\bar{X}_0 \pm kS_0/\sqrt{n}$.

example, a process involving complex analytical measurements. A single extreme outlying observation may trigger an out-of-control signal while in fact the process is in-control, thus increasing the false alarm rate and decreasing the in-control ARL of the control chart. A good model for generating normally distributed processes with occasional outliers is the contaminated normal distribution, the CDF of which is

$$N_P(\theta, \sigma^2) = (1 - p)N(\theta, 1) + pN(\theta, \sigma^2), \quad (26)$$

where $0 \leq p \leq 1$. We will refer to p and σ^2 as the *percentage of contamination* and the *extremity of contamination*, respectively. When $\theta = 0$, the process is in-control though producing occasional outliers. When $p = 0.0$ and $\theta = 0$, (26) becomes the standard normal CDF. In each simulation run, we generated 500 reference samples, of size $m = 39$ each, from the standard normal distribution. For each reference sample thus generated, we generated 500 test samples, of size $n = 10$ each, from the contaminated normal distribution all with $\theta = 0$ and all the possible combinations of (σ^2, p) where $\sigma^2 = 4, 9, 16$ and $p = 0.01, 0.01, 0.20$. Table 3 shows the simulated values of the two-sided in-control ARLs (in groups of size $n = 10$) of the Shew-XB and the Shew-KS charts for various levels of contamination.

Table 3 shows that the effect of outliers depends on the contamination severity (p, σ^2) , and the effect is more pronounced on the Shew-XB than on the Shew-KS chart. Keeping in mind that the in-control ARL of the traditional Shew-XB chart for a process operating with no outliers is 163 (in groups of size $n = 10$), we make the following observations on the results in Table 3.

- (i) Under very light percentage $p = 1\%$ and light extremity $\sigma^2 = 4$ of contamination, outliers have no

TABLE 3: Simulated Values of the ARL_0 (in groups of size $n = 10$) of the Two-sided Shew-XB and the Shew-KS Charts for an in-control Processes ($\theta = 0$) with outliers.

%	Extremity $\sigma^2 = 4$		$\sigma^2 = 9$		$\sigma^2 = 16$	
	Shew-XB $k = 2.80$	KS $L = 0.541$	Shew-XB $k = 2.80$	KS $L = 0.541$	Shew-XB $k = 2.80$	KS $L = 0.541$
$P = 0.0\%$	163	171	163	171	163	171
$P = 1\%$	159	171	115	171	70	171
$P = 5\%$	90	147	39	148	21	143
$P = 10\%$	61	144	22	127	12	118
$P = 15\%$	41	120	15	120	8	107
$P = 20\%$	33	112	11	94	6	88

Note: the case when $P = 0.0\%$ represents a process operating under a standard normal distribution with no outliers.

effect on both charts as the in-control ARLs of the two charts do not change.

- (ii) Under very light percentage $p = 1\%$ but moderate extremity $\sigma^2 = 9$ of contamination, outliers have a noticeable effect on the traditional Shew-XB chart as its ARL_0 drops to 115, which entails about $163/115 = 1.4$ times as many false alarms as the expected ARL_0 of 163. When the extremity of contamination grows to $\sigma^2 = 16$, outliers have substantial effect on the Shew-XB chart as its ARL_0 drops to 70, which entails about 2.3 times as many false alarms. In contrast, the light percentage of contamination $p = 1\%$ has no effect on the Shew-KS neither when $\sigma^2 = 9$ nor when $\sigma^2 = 16$.
- (iii) Under a moderate percentage $p = 10\%$ and light extremity $\sigma^2 = 4$ of contamination, outliers have noticeable effect on the Shew-XB as its ARL_0 drops to 61, which entails about 2.6 times as many false alarms as the expected ARL_0 of 163. In contrast, the Shew-KS triggers $171/144 = 1.2$ times as many false alarms. When the extremity of contamination grows to $\sigma^2 = 9$, outliers have a greater effect on the Shew-XB chart as its ARL_0 drops to 22, entailing about 7.4 times as many false alarms. In contrast, the Shew-KS triggers 1.4 times as many false alarms. With severe extremity of contamination, $\sigma^2 = 16$, the ARL_0 of the Shew-XB chart drops to 12, entailing 13.8 times as many false alarms. In contrast, the Shew-KS triggers 1.5 times as many false alarms.
- (iv) Outliers can have a more dramatic effect on the traditional Shew-XB chart when the percentage of contamination is as large as $p = 20\%$. With light extremity of contamination, $\sigma^2 = 4$, the ARL_0 of the Shew-XB chart drops to 33, entailing about 4.9 times as many false alarms as the expected ARL_0 of 163. In contrast, the Shew-KS triggers 1.5 times as many false alarms. With moderate extremity of contamination, $\sigma^2 = 9$, the ARL_0 of the Shew-XB chart drops to 11, entailing 14.8 times as many false alarms. In contrast, the Shew-KS triggers 1.8 times as many false alarms. With severe extremity of contamination, $\sigma^2 = 16$, the ARL_0 of the Shew-XB chart drops to just 6, entailing

about 27 times as many false alarms. In contrast, the Shew-KS triggers 1.9 times as many false alarms.

To sum up the results of Table 3, we conclude that for monitoring processes contaminated by outliers, one should not use the traditional Shew-XB, unless the percentage and the severity of contamination are both very light, around $(\sigma^2, p) = (4, 0.01)$. Otherwise, the traditional Shew-XB would trigger many folds of false alarm signals as those for an uncontaminated process. Outliers have some effect on the Shew-KS chart when the percentage of contamination is as high as $(\sigma^2, p) = (16, 0.20)$.

6.3. A Simulation Study for Efficiency. To compare two control charts, we adjust their control limits so that their in-control ARLs become approximately equal and then compare their out-of-control ARLs at various levels of change in the monitored quality characteristic. The chart with the smaller out-of-control ARL is considered to be more efficient.

In this section, we perform a simulation study to compare the efficiencies of the Shew-XB and the Shew-KS charts. The competing charts are compared for processes operating under a normal distribution with a standard deviation of 1.0, a Cauchy distribution and a Laplace distribution. Equation (23) gives the PDF of the Cauchy distribution with center θ (=median, mean does not exist) and scale β . Equation (24) gives the PDF of the Laplace distribution with center θ (=mean = median) and scale β . In (23), the scale β is set to equal 0.2605 so that the Cauchy distribution with center 0 has a probability of 0.05 to the right of 1.645, the same as that of the standard normal distribution. Since the Laplace distribution has variance $=2\beta^2$, the scale β in (24) is set to be $\beta = 1/\sqrt{2}$ so that the Laplace distribution has a standard deviation of 1.0. Efficiency comparisons are made when the median θ of the process is shifted from the in-control value of 0.0 to 1.0 in increments of 0.2. We used a training sample size $m = 39$ and a test sample size $n = 10$ in all comparisons. As mentioned in Sections 4 and 5, the ARLs of both charts are calculated by first getting the exact conditional ARLs and then using one million simulation runs to get the unconditional ARL by averaging the conditional ARLs. Tables 4, 5, and 6 show the simulated values of the two-sided ARLs (in groups of size $n = 10$). The in-control ARLs (when $\theta = 0.0$) of the competing control

TABLE 4: ARLs (in groups of size 10) of the Two-sided Shew-XB and the Shew-KS Charts under a Normal Distribution. Training Sample Size $m = 39$ and Test Sample Size $n = 10$.

	Shew XB $k = 2.75$	Shew-KS $CL = 0.541$
Median = 0.0	168.0	171.0
0.2	109.1	125.9
0.4	35.6	57.3
0.6	9.	15.4
0.8	3.4	6.2
1.0	1.8	3.0

Note: for the Shew-XB chart, 2.75 is the “ k ” value in the control limits $\bar{X}_0 \pm kS_0/\sqrt{n}$.

TABLE 5: ARLs (in groups of size 10) of the Two-sided Shew-XB and the Shew-KS Charts under a Laplace distribution. Training Sample Size $m = 39$ and Test Sample Size $n = 10$.

	Shew XB $k = 2.80$	Shew-KS $CL = 0.541$
Median = 0.0	170.0	171.0
0.2	126.0	100.1
0.4	58.0	25.0
0.6	20.0	5.3
0.8	9.3	2.3
1.0	5.0	1.6

Note: for the Shew-XB chart, 2.80 is the “ k ” value in the control limits $\bar{X}_0 \pm kS_0/\sqrt{n}$.

TABLE 6: ARLs (in groups of size 10) of the Two-sided Shew-XB and the Shew-KS Charts under a Cauchy distribution. Training Sample Size $m = 39$ and Test Sample Size $n = 10$.

	Shew-XB $k = 6.70$	Shew-KS $CL = 0.541$
Median = 0.0	171.0	171.0
0.2	165.1	38.7
0.4	165.1	3.2
0.6	161.2	1.4
0.8	160.5	1.1
1.0	159.4	1.0

Note: for the Shew-XB chart, 6.70 is the “ k ” value in the control limits $\bar{X}_0 \pm kS_0/\sqrt{n}$.

charts are made approximately equal by adjusting the control limits under each distribution.

Examinations of Tables 4, 5, and 6 lead to the following findings.

- (i) Table 4: For monitoring processes operating under a normal distribution, the Shew-KS is less efficient (has larger out-of-control ARLs) than the traditional Shew-XB chart.
- (ii) Table 5: For monitoring processes operating under a Laplace distribution, the proposed Shew-KS is more efficient (has smaller out-of-control ARLs) than the

traditional Shew-XB chart at all shifts in the process center.

- (iii) Table 6: For monitoring processes operating under a Cauchy distribution, the Shew-KS becomes dramatically more efficient than the traditional Shew-XB chart at all shifts in the process center. For example, the Shew-KS chart is quicker than the traditional Shew-XB chart by about 4-times, 52-times, 115-times, 146-times, and 159-times to signal at respective shifts of $\theta = 0.2, 0.4, 0.6, 0.8$, and 1.0 in the process center.

To sum up, the results in Tables 4, 5, and 6 lead to the following recommendations.

To monitor processes operating under moderate or heavy-tailed underlying distributions (heavier than those of the normal), the proposed Shew-KS is more efficient than the traditional Shew-XB chart. This is in addition to the advantage that the Shew-KS chart maintains same control limits over the class of (symmetric or asymmetric) continuous distributions. If one is sure that the process underlying distribution is normal, then the traditional Shew-XB chart is recommended over the Shew-KS.

7. Summary and Suggestions for Further Research

In this paper, a distribution-free (or nonparametric) Shewhart-type statistical quality control chart is developed for detecting broad changes in the underlying probability distribution of a process. We assume the availability of a random sample, called training sample, taken when the process was operating in-control. At each sampling instance, we take a random sample from the process output and calculate a modified version of the two-sample Kolmogorov-Smirnov test statistic, which will serve as the charting statistic. A signal is given if the charting statistic falls outside the control limits. Unlike the traditional distribution-based control charts (such as the Shew-XB), the proposed chart maintains the same in-control ARL_0 value over the class of all (symmetric or asymmetric) continuous distributions. Consequently, the control limits of the proposed chart need not be adjusted according to an assumed underlying process distribution. Given the training sample, the conditional ARL of the proposed chart is computed exactly using the binomial probability distribution. The unconditional ARL can then be estimated by simulations. A preliminary simulation study shows that the proposed Shew-KS chart is more efficient than the Shew-XB chart if the process underlying distribution has tails heavier than those of the normal. If the underlying process distribution can be assumed normal, then the Shew-XB chart is more efficient, as expected. The simulation study also indicates that the proposed chart is more robust against increased skewness and/or outliers in the process output.

Further simulation studies are needed to expand the efficiency comparisons of the proposed Shew-KS chart with charts other than Shew-XB. Tabulated values of the control limits are needed for the implementation of the proposed chart. It is worthwhile to investigate how the Kolmogorov-Smirnov statistic can be used with other charting schemes,

such as the exponentially weighted moving average (EWMA) and the cumulative sum (CUSUM.)

References

- [1] S. Chakraborti, P. van der Laan, and S. T. Bakir, "Nonparametric control charts: an overview and some results," *Journal of Quality Technology*, vol. 33, no. 3, pp. 304–315, 2001.
- [2] S. Chakraborti and M. A. Graham, "Nonparametric control charts," in *Encyclopedia of Statistics in Quality and Reliability*, Wiley, New York, NY, USA, 2007.
- [3] S. T. Bakir, "Quality control charts for detecting a general change in a process," in *Proceedings of the Section on Quality and Productivity (ASA '97)*, pp. 53–56, American Statistical Association, 1997.
- [4] C. Zou and F. Tsung, "Likelihood ratio-based distribution-free EWMA control charts," *Journal of Quality Technology*, vol. 42, no. 2, pp. 174–196, 2010.
- [5] G. J. Ross and N. M. Adams, "Two nonparametric control charts for detecting arbitrary distribution changes," *Journal of Quality Technology*, vol. 44, no. 2, pp. 102–116, 2012.
- [6] C. Park and M. R. Reynolds Jr., "Nonparametric procedures for monitoring a location parametric based on linear placement statistics," *Sequential Analysis*, vol. 6, no. 4, pp. 303–323, 1987.
- [7] J. Orban and D. A. Wolfe, "A class of distribution-free two-sample tests based on placements," *Journal of the American Statistical Association*, vol. 77, pp. 666–670, 1982.
- [8] P. Hackl and J. Ledolter, "A control chart based on ranks," *Journal of Quality Technology*, vol. 23, pp. 117–124, 1991.
- [9] T. R. Willemain and G. C. Runger, "Designing control charts using an empirical reference distribution," *Journal of Quality Technology*, vol. 28, no. 1, pp. 31–38, 1996.
- [10] S. Chakraborti, P. van der Laan, and M. A. van de Wiel, "A class of distribution-free control charts," *Journal of the Royal Statistical Society C*, vol. 53, no. 3, pp. 443–462, 2004.
- [11] W. J. Conover, *Practical Nonparametric Statistics*, Wiley, New York, NY, USA, 3rd edition, 1999.
- [12] S. Chakraborti, "Run length, average run length and false alarm rate of shewhart X-bar chart: exact derivations by conditioning," *Communications in Statistics Part B*, vol. 29, no. 1, pp. 61–81, 2000.
- [13] S. Chakraborti, "Parameter estimation and design considerations in prospective applications of the $\bar{X}$ chart," *Journal of Applied Statistics*, vol. 33, no. 4, pp. 439–459, 2006.
- [14] M. B. Vermaat, R. A. Ion, R. J. M. M. Does, and C. A. J. Klaassen, "A comparison of Shewhart individuals control charts based on normal, non-parametric, and extreme-value theory," *Quality and Reliability Engineering International*, vol. 19, no. 4, pp. 337–353, 2003.
- [15] W. A. Jensen, L. A. Jones-Farmer, C. W. Champ, and W. H. Woodall, "Effects of parameter estimation on control chart properties: a literature review," *Journal of Quality Technology*, vol. 38, no. 4, pp. 349–364, 2006.
- [16] N. L. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions*, vol. 2, Wiley, New York, NY, USA, 1995.