Computational Statistics and Data Analysis ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

A phase II nonparametric control chart based on precedence statistics with runs-type signaling rules

S. Chakraborti [a,*], S. Eryilmaz [b], S.W. Human [c]

[a] *Department of Information Systems, Statistics and Management Science, University of Alabama, Tuscaloosa, Box 870226, AL 35487, USA*
[b] *Department of Mathematics, Izmir University of Economics, Balcova, Izmir 35330, Turkey*
[c] *Department of Statistics, University of Pretoria, Pretoria 0002, South Africa*

ARTICLE INFO

Article history:
Received 3 June 2008
Received in revised form 10 September 2008
Accepted 16 September 2008
Available online xxxx

ABSTRACT

Nonparametric control charts do not require knowledge about the shape of the underlying distribution and can thus be attractive in certain situations. Two new Shewhart-type nonparametric control charts are proposed for monitoring the unknown location parameter of a continuous population in Phase II (prospective) applications. The charts are based on control limits given by two specified order statistics from a reference sample, obtained from a Phase I (retrospective) analysis, and using some runs-type signaling rules. The plotting statistic can be any order statistic in a Phase II sample; the median is used here for simplicity and robustness. Exact run length distributions of the proposed charts are derived using conditioning and some results from the theory of runs. Tables are provided for practical implementation of the charts for a given in-control average run length (ARL_0) between 300 and 500. Comparisons of the average run length ARL, the standard deviation of run length (SDRL) and some run length percentiles show that the charts have robust in-control performance and are more efficient when the underlying distribution is t (symmetric with heavier tails than the normal) or gamma (1, 1) (right-skewed). Even for the normal distribution, the new charts are quite competitive. An illustrative numerical example is given. An added advantage of these charts is that they can be applied before all the data are collected which might lead to savings in time and resources in certain applications.

1. Introduction

Statistical process control (SPC) techniques strive to distinguish between two sources of variation in a process: those that are random or natural and therefore cannot be identified and removed under reasonable economic constraints (so called chance causes) and those that can be (so called assignable causes). When a process operates only under chance causes, at or around an acceptable target, it is said to be in statistical control (hereafter in-control, denoted IC). Control charts help users identify assignable causes so that corrective actions, if necessary, can be taken as soon as possible. In the process control environment with variables data the output is typically assumed to follow a parametric distribution such as the normal. It is well-known that if the underlying process is not as assumed, the performance of these parametric charts can be significantly degraded. In this context, one key problem is the lack of IC robustness of some of the well-known parametric charts (see, e.g. Chakraborti et al. (2001, 2004)). This, for example, means that there could be too many false alarms than what is nominally expected and obviously, this could mean considerable loss of time and resources. Thus, it is desirable, from a practical point

* Corresponding author.
E-mail addresses: schakrab@cba.ua.edu (S. Chakraborti), serkan.eryilmaz@ieu.edu.tr (S. Eryilmaz), schalk.human@up.ac.za (S.W. Human).

0167-9473/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2008.09.025

of view, to develop and apply a set of control charts that are not designed under the assumption of normality (or any other parametric distribution). Distribution-free or nonparametric control charts can serve this purpose.

Nonparametric control charts have much to offer in SPC and have received considerable attention during the recent years. If the IC run length distribution of a control chart is the same for every continuous probability distribution the chart is called distribution-free or nonparametric. A recent review of nonparametric control charts can be found in Chakraborti and Graham (2007). The main advantage of these charts is the flexibility that their applications do not require the assumption of any specific probability distribution for the underlying process. In addition, nonparametric control charts are likely to share the well-known robustness properties of nonparametric tests and confidence intervals and are therefore are expected to be less impacted by outliers.

Chakraborti et al. (2004) considered a class of Phase II Shewhart-type nonparametric charts based on two order statistics from the reference sample. This is called the (basic) 1-*of*-1 precedence chart or simply the precedence chart. It was shown that compared to the classical Shewhart $\overline{X}$ chart the precedence chart was IC robust and was equally or more efficient in detecting location shifts. However, the construction of a nonparametric control chart with better shift detection capabilities remains an interesting possibility. Chakraborti and Eryilmaz (2007) adapted the signed rank (SR) charts of Bakir (2004) using runs-type signaling rules. These charts can be used for the specified median of a continuous symmetric distribution. They showed that (i) the new charts are nonparametric and, (ii) they have smaller and more desirable false alarm rates (and longer IC average run lengths) and better out-of-control (OOC) performance (shorter OOC average run lengths) for some heavy-tailed distributions. These observations provide motivation for the present work. Our goal is to consider Phase II Shewhart-type nonparametric precedence control charts with runs-type signaling rules to monitor the unknown IC process median. It will be seen that the addition of these signaling rules can significantly enhance the performance of the precedence charts and thus provide a new class of more powerful nonparametric charts in practice.

Note that in this paper we consider the case when the process measurement is univariate. There are some distribution-free SPC procedures for the multivariate case; see, e.g., Liu (1995), Qiu and Hawkins (2001, 2003) and Qiu (2008).

2. Distribution-free control charts

2.1. Basic (1-of-1) precedence control charts

Suppose that a reference sample of size m is available from an IC process with an unknown continuous distribution function F. A reference sample is typically obtained after a suitable Phase I analysis. Let $X_{1:m} < X_{2:m} < \cdots < X_{m:m}$ denote the order statistics of the reference sample. The control limits for the 1-*of*-1 precedence chart are given by $LCL = X_{a:m}$ and $UCL = X_{b:m}$, $1 \le a < b \le m$. In Phase II, the so-called monitoring phase, test samples, each of size n, are drawn sequentially and independently of one another as well as of the reference sample, are monitored. The plotting statistic for the hth test sample, in general, can be any order statistic $Y_{j:n}^h$, however the median is a popular choice in practice, since it is easily interpretable and is known to be robust. Thus, for example, when the subgroup size n is equal to 5, which is fairly common in SPC applications, the plotting statistic is the 3rd order statistic of the test sample. The process is declared OOC when, for the first time, a test sample median falls on or outside of either of the control limits. Note that the precedence charts can be applied as soon as the necessary order statistics are available and this can be a practical advantage in some applications.

Define the indicator random variables for the hth test sample:

$$Z_h = \begin{cases} 1, & Y_{j:n}^h \notin (LCL, UCL) \\ 0, & Y_{j:n}^h \in (LCL, UCL) \end{cases}, \qquad h = 1, 2, 3 \dots. \tag{1}$$

Thus, the Z's are signaling indicators; if $Z_h = 1$, a signal is indicated and the process is declared to be OOC on the hth test sample, whereas the opposite is true if $Z_h = 0$. The event when $Z_h = 1$, that is when $Y_{j:n}^h \notin (LCL, UCL)$, is called a signaling event and its probability, $P(Z_h = 1) = p$, say, is called the (unconditional) signaling probability. It is assumed that the test samples all come from a continuous distribution with c.d.f. G and hence the superscript h on $Y_{j:n}^h$ is suppressed hereafter until it's necessary to avoid any confusion.

Because the control limits are order statistics from the same reference sample, the signaling events and therefore the signaling indicators $Z_1, Z_2, \dots$ are *dependent* binary random variables. So the implementation, analysis and interpretation of the control charts must take account of this dependence. Note that given (or conditionally on) $X_{a:m} = x_1$ and $X_{b:m} = x_2$, the probability of a no-signal is

$$P(x_1 < Y_{j:n} < x_2 | X_{a:m} = x_1, X_{b:m} = x_2) = G_j(x_2) - G_j(x_1) = 1 - p(x_1, x_2, j, G) \tag{2}$$

where G_j denotes the c.d.f. of the jth order statistic in a sample of size n from a distribution with c.d.f. G. Since $G_j(x) = I_{G(x)}(j, n-j+1)$, where $I_a(b, c)$ denotes the incomplete beta function, the probability in (2) can be expressed as $I_{G(x_2)}(j, n-j+1) - I_{G(x_1)}(j, n-j+1)$. The unconditional probability of a no-signal, denoted $1-p$, can be found by averaging this over the joint distribution of $X_{a:m}$ and $X_{b:m}$. Thus, transforming results to (0,1), we get

$$1 - p = P(Z_h = 0) = \int_0^1 \int_0^y [I_{GF^{-1}(y)}(j, n-j+1) - I_{GF^{-1}(x)}(j, n-j+1)] f_{a,b}(x, y) \mathrm{d}x\mathrm{d}y, \tag{3}$$

Please cite this article in press as: Chakraborti, S., et al., A phase II nonparametric control chart based on precedence statistics with runs-type signaling rules. Computational Statistics and Data Analysis (2008), doi:10.1016/j.csda.2008.09.025

where $f_{a,b}(x, y)$ denotes the well-known joint p.d.f. of the ath and the bth order statistics in a sample of size m from the uniform (0, 1) distribution (see, e.g. Gibbons and Chakraborti (2003)). The unconditional probability of a signal, p, can be obtained from (3).

Note that since the observations are assumed to be continuous, the probability of a plotting statistic falling on the control limit equals 0, theoretically, yet this can happen in practice and thus as a convention we include this possibility in the definition of a signal without altering the probability.

2.2. Precedence control charts with signaling rules

We consider generalizing the 1-*of*-1 precedence charts by incorporating signaling rules involving runs of the plotting statistic above and/or below the control limits. Similar extensions have been considered in the literature for the parametric Shewhart charts, see, e.g., Nelson (1984) and Klein (2000). Thus, the charts proposed here can be viewed as "runs rule enhanced" Phase II nonparametric control charts. Note that only two-sided control charts are studied here, one-sided charts can be considered along similar lines.

Three signaling rules: (a), (b), and (c) are considered that lead to control charts. According to these, a process is declared OOC when

(a) a single point (plotting statistic: median) falls on or outside the control limits (the 1-*of*-1*chart or the precedence chart*)
(b) two consecutive points (plotting statistics: medians of consecutive test samples) (i) both fall on or above the *UCL* or, (ii) both fall on or below the *LCL* or, (iii) one falls on or above the *UCL* and the next one falls on or below the *LCL* or, (iv) one falls on or below the *LCL* and the next falls on or above the *UCL* (Derman and Ross, 1997; hereafter the 2-*of*-2 *DR chart*)
(c) two consecutive points (medians of consecutive test samples) both fall on or above the UCL or both fall on or below the LCL (Klein, 2000; hereafter 2-*of*-2*KL chart)*.

Rule (a) is the simplest and most frequently used in the control charting literature. In the present setting, this corresponds to the 1-*of*-1 precedence chart studied in Chakraborti et al. (2004), mentioned here for reference and comparison purposes. It is clear that rule (a) is a special case of rules (b) and (c).

Performance of Phase II control charts is evaluated on the basis of its run length distribution. Note that the run length also be viewed as the waiting time until the first signal. These waiting times, which are positive integer valued random variables, are defined as follows.
Case (a): The waiting time for chart (a) is

$$T_1 = \min\{t : Z_t = 1\}. \tag{4}$$

Case (b): For chart (b) the waiting time is defined as

$$T_2 = \min\{t : Z_{t-1} = 1, Z_t = 1\} \tag{5}$$

where the indicator random variables, Z's, are given by (1) and,
Case (c): For chart (c) the waiting time is defined as

$$T_2' = \min\left\{T_2^{(1)}, T_2^{(2)}\right\} \tag{6}$$

where

$$T_2^{(1)} = \min\left\{t : Z'_{t-1} = Z'_t = 1\right\}, \qquad T_2^{(2)} = \min\left\{t : Z'_{t-1} = Z'_t = 2\right\}$$

and

$$Z'_h = \begin{cases} 0 & \text{if } Y_{j:n} \in (X_{a:m}, X_{b:m}) \\ 1 & \text{if } Y_{j:n} \geq X_{b:m} \\ 2 & \text{if } Y_{j:n} \leq X_{a:m}. \end{cases}$$

The three charts are illustrated in Fig. 1. It is seen that here the 1-*of*-1, the 2-*of*-2 DR and the 2-*of*-2 KL charts signal (for the first time) at sample numbers 3, 6, and 7, respectively. Accordingly, the run lengths (or, the waiting times) corresponding to these charts are 3, 6, and 7, respectively.

It may be noted that if the 2-*of*-2 KL chart signals so does the 2-*of*-2 DR chart but the reverse is not always true. In addition, the 2-*of*-2 KL chart seems more suitable when detecting a shift in the process location, either up or down is of interest, whereas the 2-*of*-2 DR rule can detect a possible swing.

The run length distributions for the proposed charts are studied in the following section.

Please cite this article in press as: Chakraborti, S., et al., A phase II nonparametric control chart based on precedence statistics with runs-type signaling rules. Computational Statistics and Data Analysis (2008), doi:10.1016/j.csda.2008.09.025

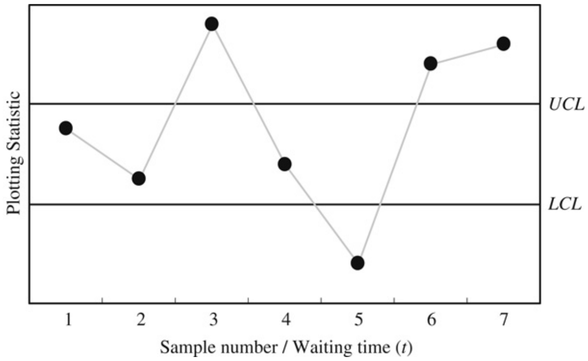

Fig. 1. An illustration of the 1-*of*-1, the 2-*of*-2 DR and the 2-*of*-2 KL charts.

3. Waiting time (run length) distributions

3.1. Distribution of T_2: Run Length distribution for the 2-of-2 DR chart

Chakraborti et al. (2004) derived the distribution of T_1 by conditioning on the order statistics $X_{a:m}$ and $X_{b:m}$. However, application of the same technique here leads to an expression for the conditional distribution of T_2 that is too complex, particularly for $x \geq 3$, to attempt a direct derivation of a closed form expression for the unconditional distribution. Instead, we find the distribution of T_2 by first conditioning on the total number of successes $S_n = \sum_{i=1}^{n} Z_i$ in the sequence of random variables $Z_1, Z_2, \ldots$. To this end, first note that $Z_1, Z_2, \ldots$ is a sequence of dependent binary random variables, and in fact it can be seen that they are exchangeable or symmetrically dependent (i.e., any permutation of any subset of these random variables has the same distribution). Using this result, an exact expression for the probability distribution of T_2 is now derived.

George and Bowman (1995) derived the distribution of the total number of successes in a sequence of n exchangeable binary trials. According to their result

$$P(S_n = s) = \binom{n}{s} \sum_{i=0}^{n-s} (-1)^i \binom{n-s}{i} \lambda_{s+i} \tag{7}$$

where $\lambda_t = P(Z_1 = 1, \ldots, Z_t = 1)$ for $t = 1, 2, \ldots, n$. We use this result in deriving the unconditional run length distribution of the 2-*of*-2 DR chart, given in Theorem 1. Note that since conditionally λ_t equals $[p(x_1, x_2, j, G)]^t$, unconditionally

$$\lambda_t = P(Z_1 = 1, \ldots, Z_t = 1) = \int_0^1 \int_0^y [1 - \{I_{GF^{-1}(y)}(j, n-j+1) - I_{GF^{-1}(x)}(j, n-j+1)\}]^t f_{a,b}(x, y)\mathrm{d}x\mathrm{d}y. \tag{8}$$

The unconditional distribution of T_2 is given by

Theorem 1.

$$P(T_2 = x) = \begin{cases} 0 & \text{if } 0 \leq x < 2 \\ \lambda_2 & \text{if } x = 2 \end{cases} \tag{9}$$

and for $x \geq 3$.

$$P(T_2 = x) = \sum_{y=1}^{x-2} \sum_{j=0}^{\min\left(y, \left[\frac{x-y-2}{2}\right]\right)} \sum_{i=0}^{y} (-1)^j (-1)^i \binom{y}{j} \binom{y}{i} \binom{x - 2(j+1) - 1}{y-1} \lambda_{x-y+i}.$$

Proof. Given in the Appendix. □

Corollary 1. *The IC* ($F = G$) *run length distribution of the 2-of-2 DR chart is given by Theorem 1 where* $\lambda_t = \int_0^1 \int_0^y [1 - (I_y(j, n-j+1) - I_x(j, n-j+1))]^t f_{a,b}(x, y)\mathrm{d}x\mathrm{d}y$.

Thus, the IC run length distribution is free from either F or G and the chart is distribution-free.

Please cite this article in press as: Chakraborti, S., et al., A phase II nonparametric control chart based on precedence statistics with runs-type signaling rules. Computational Statistics and Data Analysis (2008), doi:10.1016/j.csda.2008.09.025

3.2. Distribution of T_2': Run Length distribution for the 2-of-2 KL chart

Again, observe that conditionally on $X_{a:m}$ and $X_{b:m}$, the random variables $Z_1', Z_2', \ldots$ are i.i.d. with

$$p_L = P\left\{Y_{j:n} \le X_{a:m} | X_{a:m} = x_1\right\} = I_{G(x_1)}(j, n-j+1) \tag{10}$$

and

$$p_U = P\left\{Y_{j:n} \ge X_{b:m} | X_{b:m} = x_2\right\} = 1 - I_{G(x_2)}(j, n-j+1). \tag{11}$$

We obtain the distribution of T_2' by applying some results in Fu and Lou (2003). Given $X_{a:m}$ and $X_{b:m}$, let T_2^* denote the waiting time for two consecutive 1's or two consecutive 2's in the sequence of i.i.d. trials $Z_1', Z_2', \ldots$. This is called a compound pattern $\Lambda = \Lambda_1 \cup \Lambda_2$, where $\Lambda_1 = \{11\}$ and $\Lambda_2 = \{22\}$. Now applying Theorem 5.2 of Fu and Lou (2003, page 68), the distribution of T_2^* is obtained

$$P\left\{T_2^* = x | X_{a:m}, X_{b:m}\right\} = \underline{\xi} N^{x-1} (I - N)\, \underline{1}', \quad x \ge 2, \tag{12}$$

where

$$N = \begin{bmatrix} 0 & 1-p_L-p_U & p_U & p_L \\ 0 & 1-p_L-p_U & p_U & p_L \\ 0 & 1-p_L-p_U & 0 & p_L \\ 0 & 1-p_L-p_U & p_U & 0 \end{bmatrix}, \qquad \underline{\xi} = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix}, \qquad \underline{1} = \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix}.$$

It may be noted that this result follows from more general results on the distribution of waiting time for the first occurrence of a compound pattern in a sequence of i.i.d. or homogeneous Markov dependent m-state trials derived in Fu and Lou (2003, Chapter 5). In our case (Case (c)), m equals 3 and the necessary imbedded Markov chain is defined on the state space $\{\phi, 0, 1, 2, \alpha_1, \alpha_2\}$ where $\alpha_1 = \{11\}$ and $\alpha_2 = \{22\}$ are the two absorbing states (when the process is declared to be OOC). The associated transition probability matrix $M_{6\times 6}$ can be written as $M = \begin{bmatrix} N & C \\ O & I \end{bmatrix}$ where $N_{4\times 4}$ is defined above,

$$C_{4\times 2} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ p_U & 0 \\ 0 & p_L \end{bmatrix}, \qquad O_{2\times 4} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \qquad I_{2\times 2} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

and the probabilities p_L and p_U are defined in (10) and (11), respectively.

The unconditional distribution of T_2' is then obtained from the conditional distribution of T_2^* by averaging over the joint distribution of $X_{a:m}$ and $X_{b:m}$. This is given below.

Theorem 2.

$$P\left\{T_2' = x\right\} = \int_{-\infty}^{\infty} \int_{-\infty}^{x_2} P\left\{T_2^* = x | X_{a:m} = x_1, X_{b:m} = x_2\right\} h_{a,b}(x_1, x_2) \mathrm{d}x_1 \mathrm{d}x_2 \quad \text{for } x \ge 2, \tag{13}$$

where $P\left\{T_2^* = x | X_{a:m} = x_1, X_{b:m} = x_2\right\}$ is given by (12) and $h_{a,b}(x_1, x_2)$ is the joint p.d.f. of $X_{a:m}$ and $X_{b:m}$ from a continuous c.d.f. F.

Corollary 2. *The IC ($F = G$) run length distribution of the 2-of-2 KL chart is given by*

$$P(T_2' = x) = \int_0^1 \int_0^v \underline{\xi} N_0^{x-1} (I - N_0)\, \underline{1}' f_{a,b}(u, v) \mathrm{d}u \mathrm{d}v$$

where

$$N_0 = \begin{bmatrix} 0 & I_v(j,n-j+1) - I_u(j,n-j+1) & 1 - I_v(j,n-j+1) & I_u(j,n-j+1) \\ 0 & I_v(j,n-j+1) - I_u(j,n-j+1) & 1 - I_v(j,n-j+1) & I_u(j,n-j+1) \\ 0 & I_v(j,n-j+1) - I_u(j,n-j+1) & 0 & I_u(j,n-j+1) \\ 0 & I_v(j,n-j+1) - I_u(j,n-j+1) & 1 - I_v(j,n-j+1) & 0 \end{bmatrix}. \tag{14}$$

Once again the IC run length distribution is seen to be free from either F or G, so that the chart is distribution-free.

4. *ARL*, *VAR* and *FAR* calculations

In order to study the performance of a Phase II control chart, it is common to examine the average and the standard deviation (or the variance) of the run length distribution. The false alarm rate (*FAR*) of the chart is also of interest. Expressions for these can be obtained exactly and are shown below.

Please cite this article in press as: Chakraborti, S., et al., A phase II nonparametric control chart based on precedence statistics with runs-type signaling rules. Computational Statistics and Data Analysis (2008), doi:10.1016/j.csda.2008.09.025

4.1. ARL, VAR and FAR of 2-of-2 DR chart

The required quantities are most conveniently derived by conditioning, noting that given $X_{a:m}$ and $X_{b:m}$, the random variable T_2 has a geometric distribution of order 2 (see Balakrishnan and Koutras (2002)). The conditional expected value and the variance of T_2 are known to be

$$E\left(T_2\left|X_{a:m}=x_1, X_{b:m}=x_2\right.\right)=\frac{1+p}{p^2},$$
$$\text{and}\quad \operatorname{VAR}\left(T_2\left|X_{a:m}=x_1, X_{b:m}=x_2\right.\right)=\frac{1-5(1-p)p^2-p^5}{(1-p)^2p^4}, \tag{15}$$

where $p=p(x_1,x_2,j,G)=1-[G_j(x_2)-G_j(x_1)]=1-[I_{G(x_2)}(j,n-j+1)-I_{G(x_1)}(j,n-j+1)]$.

So the unconditional ARL for the 2-*of*-2 DR chart is given by

$$\begin{aligned} ARL_{DR} &= E_{X_{a:m},X_{b:m}}E\left(T_2\left|X_{a:m},X_{b:m}\right.\right)\\ &=\int_0^1\int_0^y\left[\frac{2-\{G_j(F^{-1}(y))-G_j(F^{-1}(x))\}}{[1-\{G_j(F^{-1}(y))-G_j(F^{-1}(x))\}]^2}\right]f_{a,b}(x,y)\mathrm{d}x\mathrm{d}y\\ &=\int_0^1\int_0^y\left[\frac{2-\{I_{GF^{-1}(y)}(j,n-j+1)-I_{GF^{-1}(x)}(j,n-j+1)\}}{[1-\{I_{GF^{-1}(y)}(j,n-j+1)-I_{GF^{-1}(x)}(j,n-j+1)\}]^2}\right]f_{a,b}(x,y)\mathrm{d}x\mathrm{d}y. \end{aligned}\tag{16}$$

The IC average run length (ARL_0) is obtained by substituting $F=G$ in (16).

The unconditional variance of the run length distribution for the 2-*of*-2 DR chart can be calculated by using the formula

$$VARL_{DR}=E_{X_{a:m},X_{b:m}}VAR(T_2|X_{a:m},X_{b:m})+VAR_{X_{a:m},X_{b:m}}(E(T_2|X_{a:m},X_{b:m}))$$

where the conditional expectation and the variance formulas are given in (15). The formula for the IC unconditional variance can be obtained by substituting $F=G$ in the resulting expression. The final expressions are much to complicated to be presented here. We calculate and discuss these later.

The FAR of the 2-*of*-2 DR chart may be calculated as

$$\begin{aligned} FAR_{DR} &= P\left\{Y_{j:n}^{h-1}\le X_{a:m}, Y_{j:n}^{h}\le X_{a:m}\left|F=G\right.\right\}+P\left\{Y_{j:n}^{h-1}\ge X_{b:m}, Y_{j:n}^{h}\ge X_{b:m}\left|F=G\right.\right\}\\ &\quad+P\left\{Y_{j:n}^{h-1}\ge X_{b:m}, Y_{j:n}^{h}\le X_{a:m}\left|F=G\right.\right\}+P\left\{Y_{j:n}^{h-1}\le X_{a:m}, Y_{j:n}^{h}\ge X_{b:m}\left|F=G\right.\right\}\\ &=\int_0^1\left[1-I_y(j,n-j+1)\right]^2f_b(y)\mathrm{d}y+\int_0^1[I_x(j,n-j+1)]^2f_a(x)\mathrm{d}x\\ &\quad+2\int_0^1\int_0^y\left(1-I_y(j,n-j+1)\right)I_x(j,n-j+1)f_{a,b}(x,y)\mathrm{d}x\mathrm{d}y \end{aligned}\tag{17}$$

where $f_a(x)$ and $f_b(x)$ are the p.d.f.'s of the beta $(a,m-a+1)$ and beta $(b,m-b+1)$ distribution, respectively.

4.2. ARL, VAR and FAR of 2-of-2 KL chart

Again, given $X_{a:m}$ and $X_{b:m}$, the conditional expected value of T_2' is known to be (see Klein (2000))

$$E\left(T_2'\left|X_{a:m},X_{b:m}\right.\right)=\frac{1}{p-\frac{p_U}{1+p_U}-\frac{p_L}{1+p_L}}=\frac{1}{\frac{p_U^2}{1+p_U}+\frac{p_L^2}{1+p_L}}\tag{18}$$

where p_L and p_U are defined in (10) and (11), respectively, and $p=p_L+p_U$ (note the slightly different notation in Klein; he defines p to be the probability of no-signal whereas we define p to be the probability of a signal).

Now, averaging over the distribution of $X_{a:m}$ and $X_{b:m}$, the ARL of the 2-*of*-2 KL chart is obtained

$$ARL_{\mathrm{KL}}=\int_0^1\int_0^y\left[\frac{\left(1+(1-G_j(F^{-1}(y)))\right)\left(1+G_j(F^{-1}(x))\right)}{\left(G_j(F^{-1}(x))\right)^2\left(1+(1-G_j(F^{-1}(y)))\right)+\left(1-G_j(F^{-1}(y))\right)^2\left(1+G_j(F^{-1}(x))\right)}\right]f_{a,b}(x,y)\mathrm{d}x\mathrm{d}y.\tag{19}$$

Noting that $G_j(F^{-1}(x))=I_{GF^{-1}(x)}(j,n-j+1)$, the IC average run length can be obtained by substituting $F=G$ in (19).

In order to calculate the variance of the run length distribution note that if the Z's are i.i.d., the second raw moment of T_2' is given by (Fu et al., 2002) $\underline{\xi}(I+N)(I-N)^{-2}\underline{1}'$. Using this result conditionally on $X_{a:m}$ and $X_{b:m}$ and the formula for the unconditional variance from the conditional expectation and variance formulas shown earlier, we can calculate the unconditional variance of the run length distribution of the 2-*of*-2 KL chart. From this, the unconditional variance of the IC run length distribution can be found by substituting $F=G$. Like in the case of the DR chart, the expression for the unconditional variance is complicated and is not presented here; we calculate and discuss these later.

Please cite this article in press as: Chakraborti, S., et al., A phase II nonparametric control chart based on precedence statistics with runs-type signaling rules. Computational Statistics and Data Analysis (2008), doi:10.1016/j.csda.2008.09.025

Table 1
In-control average run length (ARL_0), false alarm rate (FAR) and constants (a, b) (The three rows of each cell shows the achieved ARL_0, the FAR, and the chart constants (a, b), respectively.) for the 2-of-2 DR nonparametric chart for $m = 50, 100, 200, 500$ and $(j, n) = (3, 5), (4, 7), (5, 9)$.

$j = 3, n = 5$				$j = 4, n = 7$				$j = 5, n = 9$			
$m = 50$	100	200	500	$m = 50$	100	200	500	$m = 50$	100	200	500
605.44 0.0072 (8, 43)	548.99 0.0040 (15, 86)	537.62 0.0029 (29, 172)	536.72 0.0023 (71, 430)	597.80 0.0090 (10, 41)	509.54 0.0048 (19, 82)	597.72 0.0027 (36, 165)	526.08 0.0024 (90, 411)	976.53 0.0084 (11, 40)	739.47 0.0040 (21, 80)	558.51 0.0031 (42, 159)	528.95 0.0024 (104, 397)
275.30 0.0121 (9, 42)	373.31 0.0055 (16, 85)	443.56 0.0034 (30, 171)	496.90 0.0025 (72, 429)	264.91 0.0150 (11, 40)	345.93 0.0065 (20, 81)	490.44 0.0033 (37, 164)	487.01 0.0026 (91, 410)	383.92 0.0144 (12, 39)	481.18 0.0056 (22, 79)	456.18 0.0037 (43, 158)	488.41 0.0026 (105, 396)
	261.69 0.0074 (17, 84)	368.80 0.0040 (31, 170)	460.60 0.0026 (73, 428)		241.21 0.0088 (21, 80)	405.20 0.0039 (38, 163)	451.33 0.0028 (92, 409)	172.47 0.0236 (13, 38)	322.26 0.0077 (23, 78)	375.04 0.0044 (44, 157)	451.43 0.0028 (106, 395)
		308.82 0.0047 (32, 169)	427.48 0.0028 (74, 427)			336.97 0.0046 (39, 162)	418.70 0.0030 (93, 408)		221.57 0.0104 (24, 77)	310.28 0.0053 (45, 156)	417.68 0.0030 (107, 394)
		260.37 0.0056 (33, 168)	397.20 0.0031 (75, 426)			281.98 0.0054 (40, 161)	388.83 0.0032 (94, 407)			258.24 0.0062 (46, 155)	386.83 0.0033 (108, 393)
			369.50 0.0033 (76, 425)				361.45 0.0034 (95, 406)				358.60 0.0035 (109, 392)
			344.12 0.0035 (77, 424)				336.33 0.0037 (96, 405)				332.75 0.0038 (110, 391)
			320.83 0.0037 (78, 423)				313.25 0.0039 (97, 404)				309.06 0.0041 (111, 390)

The FAR for the KL chart is given by

$$FAR_{KL} = P\left\{Y_{j:n}^{h-1} \geq X_{b:m}, Y_{j:n}^{h} \geq X_{b:m} \,|F = G\right\} + P\left\{Y_{j:n}^{h-1} \leq X_{a:m}, Y_{j:n}^{h} \leq X_{a:m} \,|F = G\right\},$$

$$= \int_0^1 \left[1 - I_y(j, n-j+1)\right]^2 f_b(y)\mathrm{d}y + \int_0^1 \left[I_x(j, n-j+1)\right]^2 f_a(x)\mathrm{d}x. \tag{20}$$

5. Implementation of charts

For practical implementation of the proposed charts we need the control limits, which means we need to find the indices a and b $(1 \leq a < b \leq n)$, respectively, of the reference sample order statistics.

5.1. Determination of control limits

Typically, in Phase II applications, control limits are determined so that a specified IC average run length (say ARL_0^*) is obtained. Since 370 is a standard ARL_0 value used in the industry, we need to solve, for example, $ARL_{0,DR} = 370$ for the 2-*of*-2 DR chart (and $ARL_{0,KL} = 370$ for the 2-*of*-2 KL chart) for a and b. The ARL_0 expressions are obtained from (16) and (19), respectively, substituting $F = G$. Since the Phase II (test) sample median is used as the charting statistic and the sample size is assumed odd, it seems reasonable to use symmetric limits (see, e.g., Chakraborti et al. (2004)) and we take $b = m - a + 1$, so that only the constant a needs to be determined. The solutions to the equations are obtained using the software package Mathcad. In Tables 1 and 2 we display the various choices of the constants a and b for the 2-*of*-2 DR and the 2-*of*-2 KL chart, for a given ARL_0^* in the neighborhood of 300 and 500, when reference samples of size $m = 50$, 100, 200 and 500 are used to estimate the control limits in Phase I and these limits are used in Phase II monitoring using medians of (test) samples of size $n = 5$, 7, or 9, respectively. Thus, j equals 3, 4, and 5, respectively, in the tables. For each combination of values of n, j, and m, the tables display (in each cell) a combination of the attained ARL_0, the attained FAR and the (a, b) values, in that order, where the ARL_0 values are close to the specified values 300 and 500. Note that these values (in the neighborhood of 300 and 500, which can be considered reasonably large), are provided since it is rare to be able to achieve an ARL_0 (or FAR) exactly as specified with the nonparametric charts as the IC distribution of the charting statistic is discrete. For example, from Table 1, for $m = 500$, $n = 5$ and $j = 3$, one set of constants for the DR chart are given by $a = 72$ and $b = 500 - 72 + 1 = 429$ so that $LCL = X_{72:500}$ and $UCL = X_{429:500}$. In this case the achieved ARL_0 and the FAR are 496.90 and 0.0025, respectively. We emphasize these are the exact values and the control limits can be used for all continuous distributions. For a more moderate reference sample size, say $m = 50$, Table 1 shows that ARL_0 values such as 275.30 or 605.44, which may be

Table 2
In-control average run length (ARL_0), false alarm rate (FAR) and chart constants (a, b) (The three rows of each cell shows the achieved ARL_0, the FAR, and the chart constants (a, b), respectively.) for the 2-of-2 KL nonparametric chart for $m = 50,\ 100, 200, 500$ and $(j, n) = (3, 5), (4, 7), (5, 9)$.

$j = 3, n = 5$				$j = 4, n = 7$				$j = 5, n = 9$			
$m = 50$	100	200	500	$m = 50$	100	200	500	$m = 50$	100	200	500
1010.37 0.0048 (8, 43)	650.75 0.0033 (16, 85)	559.01 0.0026 (32, 169)	524.39 0.0023 (80, 421)	985.39 0.0063 (10, 41)	594.56 0.0041 (20, 81)	504.01 0.0031 (40, 161)	506.61 0.0024 (99, 402)	1591.68 0.0062 (11, 40)	547.12 0.0049 (23, 78)	548.41 0.0031 (45, 156)	530.19 0.0023 (112, 389)
460.89 0.0079 (9, 42)	456.52 0.0044 (17, 84)	471.18 0.0031 (33, 168)	490.21 0.0024 (81, 420)	437.32 0.0102 (11, 40)	414.67 0.0054 (21, 80)	424.10 0.0036 (41, 160)	472.95 0.0026 (100, 401)	626.67 0.0103 (12, 39)	376.11 0.0066 (24, 77)	456.29 0.0036 (46, 155)	493.12 0.0025 (113, 388)
237.00 0.0123 (10, 41)	328.69 0.0057 (18, 83)	399.60 0.0036 (34, 167)	458.70 0.0026 (82, 419)	217.33 0.0160 (12, 39)	296.08 0.0070 (22, 79)	358.81 0.0042 (42, 159)	441.90 0.0027 (101, 400)	281.29 0.0165 (13, 38)	264.69 0.0086 (25, 76)	381.78 0.0042 (47, 154)	459.05 0.0027 (114, 387)
	242.15 0.0074 (19, 82)	340.87 0.0041 (35, 166)	429.62 0.0027 (83, 418)			305.16 0.0048 (43, 158)	413.24 0.0029 (102, 399)			321.15 0.0049 (48, 153)	427.69 0.0029 (115, 386)
		292.37 0.0047 (36, 165)	402.76 0.0029 (84, 417)			260.82 0.0056 (44, 157)	386.77 0.0031 (103, 398)			271.54 0.0057 (49, 152)	398.81 0.0031 (116, 385)
			377.91 0.0031 (85, 416)				362.28 0.0033 (104, 397)				372.18 0.0033 (117, 384)
			354.91 0.0033 (86, 415)				339.62 0.0035 (105, 396)				347.61 0.0035 (118, 383)
			333.60 0.0035 (87, 414)				318.62 0.0037 (106, 395)				324.92 0.0038 (119, 382)
			313.83 0.0037 (88, 413)				299.16 0.0040 (107, 394)				303.95 0.0040 (120, 381)

deemed reasonably large in practice, are achievable. Obviously as m and/or n increase, the available choices for the ARL_0 values also increase.

Similar behavior is observed in the case of the 2-*of*-2 KL chart as can be seen from the entries in Table 2. For instance, when $m = 500$, $n = 5$ and $j = 3$, and one uses $LCL = X_{80:500}$ and $UCL = X_{421:500}$, the ARL_0 of the 2-*of*-2 KL chart is 524.39, whereas the FAR is 0.0023. However, if instead one chooses to use $LCL = X_{81:500}$ and $UCL = X_{420:500}$, the ARL_0 decreases to 490.21, whereas the FAR slightly increases to 0.0024.

It may be noted that when some order statistic other than the median is used as the Phase II charting statistic, determination of the constants a and b is a more involved problem, since in this case setting $b = m - a + 1$ becomes questionable due to a lack of symmetry of the IC distribution. We don't discuss this any further in this paper.

5.2. Example

We illustrate the nonparametric charts using the data given in Tables 5.1 and 5.2 of Montgomery (2001). The goal of this study was to establish statistical control of the inside diameter of the piston rings for an automotive engine manufactured in a forging process. Twenty-five retrospective or Phase I samples, each of size five, were collected when the process was thought to be IC. As shown in Montgomery (2001), the traditional Shewhart $\bar{X}$ and R charts provide no indication of an OOC condition, so these data are considered to be Phase I reference data and these "trial" limits were adopted for use in on-line process control.

In order to obtain the control limits the constants a and b are needed. Possible symmetric control limits $(a, b = m - a + 1)$ for the three charts are shown in Table 3, for $m = 25 \times 5 = 125$, $n = 5$ and $j = 3$, along with the corresponding FAR and ARL_0 values.

Using Table 3, we take $a = 7$ so that $b = 119$, and the control limits for the 1-*of*-1 precedence chart are the 7th and the 119th ordered values of the reference sample: $LCL = X_{7:125} = 73.984$ and $UCL = X_{119:125} = 74.017$, which yield an exact ARL_0 of 413.80 and a FAR of 0.0044. A plot of the medians for the 1-*of*-1 chart is shown in Fig. 2 for all forty samples, the first twenty five of which are from Phase I and the remaining are from Phase II. It is seen that while the Phase I samples are in control, the 1-*of*-1 precedence chart signals on the 12th sample in the prospective phase (Phase II).

For the 2-*of*-2 DR chart, again using Table 3, we take $a = 19$ so that the resulting limits, $LCL = X_{19:125} = 73.992$ and $UCL = X_{107:125} = 74.012$, yield an ARL_0 and FAR of 464.38 and 0.0040, respectively. The 2-*of*-2 DR chart is shown in Fig. 3.

Finally, for the 2-*of*-2 KL chart we take $a = 21$ so that $b = 125 - 21 + 1 = 105$ and thus $LCL = X_{21:125} = 73.992$ and $UCL = X_{105:125} = 74.011$. This yields an ARL_0 of 460.54 and a FAR of 0.0038, respectively. This 2-*of*-2 KL chart is almost

Please cite this article in press as: Chakraborti, S., et al., A phase II nonparametric control chart based on precedence statistics with runs-type signaling rules. Computational Statistics and Data Analysis (2008), doi:10.1016/j.csda.2008.09.025

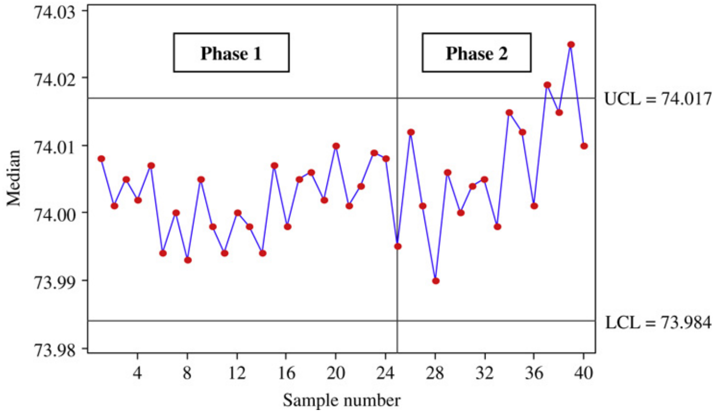

Fig. 2. 1-*of*-1 phase II precedence chart for the Montgomery (2001) piston-ring data.

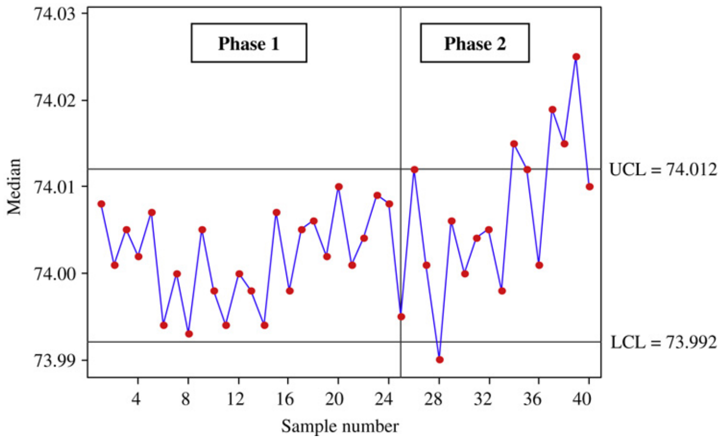

Fig. 3. 2-*of*-2 phase II DR chart for the Montgomery (2001) piston-ring data.

Table 3
In-control average run length (ARL_0), false alarm rate (FAR) and chart constants (a, b) for the 1-*of*-1, 2-*of*-2 DR and 2-*of*-2 KL precedence charts when $m = 125$, $n = 5$ and $j = 3$.

1-*of*-1				2-*of*-2 DR				2-*of*-2 KL			
a	b	ARL_0	FAR	a	b	ARL_0	FAR	a	b	ARL_0	FAR
5	121	1315.98	0.0019	19	107	464.38	0.0040	19	107	819.47	0.0024
6	120	695.09	0.0029	20	106	344.73	0.0052	20	106	608.81	0.0030
7	119	413.80	0.0044	21	105	260.69	0.0066	21	105	460.54	0.0038
8	118	267.40	0.0062	22	104	200.46	0.0084	22	104	354.09	0.0048

identical to the chart shown in Fig. 3 and is thus omitted. Note that both the 2-*of*-2 DR and KL charts signal on the 10th sample in the prospective phase, slightly earlier than the 1-*of*-1 chart.

6. Performance comparisons

The performance of Phase II charts is typically compared by first designing each chart to (roughly) have the same ARL_0 and then examining their ARL_1 (OOC average run length) values at some OOC values of the parameter of interest. The control chart with the shorter (or smaller) OOC average run length is usually preferred. Since the proposed Phase II charts are nonparametric Shewhart-type charts, their main competitor is the basic 1-*of*-1 precedence control chart of Chakraborti et al. (2004). The normal, the t and the gamma distribution were used in the performance comparisons. The t-distribution was used to study the effects of heavier tails and the gamma distribution was used to study the effect of skewness. In order for the results to be comparable, the distributions were scaled so that each had a mean of 0 and a variance of 1; thus, the normal (0,1), the $t(4)$ and the gamma (1, 1) distributions were used in the sequel. The parametric Shewhart $\bar{X}$ chart was included in the comparison for the normal distribution but not for the t and the gamma distribution since the $\bar{X}$ chart is well-known to be nonrobust under nonnormality; see, e.g., Chakraborti et al. (2004).

Please cite this article in press as: Chakraborti, S., et al., A phase II nonparametric control chart based on precedence statistics with runs-type signaling rules. Computational Statistics and Data Analysis (2008), doi:10.1016/j.csda.2008.09.025

Table 4
ARL and SDRL values for the 2-*of*-2 DR, 2-*of*-2 KL, and the basic (1-*of*-1) precedence chart and the Shewhart chart for the normal distribution when $m = 500, j = 3, n = 5$.

Shift	2-*of*-2DR				2-*of*-2 KL				Basic precedence (1-*of*-1)				Shewhart X-bar	
	ARL	*SDRL*	*ARL*	*SDRL*	*ARL*	*SDRL*	*ARL*	*SDRL*	*ARL*	*SDRL*	*ARL*	*SDRL*	*ARL*	*SDRL*
0.00	496.90	573.05	536.72	621.20	490.21	554.18	524.39	594.55	460.22	538.61	520.27	613.67	500.00	571.14
0.25	233.82	278.56	250.23	299.59	170.07	203.00	180.06	215.88	233.27	290.26	261.60	329.17	184.12	216.66
0.50	58.22	66.10	61.33	69.99	39.37	43.17	41.11	45.28	70.42	85.43	77.73	95.38	43.38	48.51
0.75	17.55	17.85	18.23	18.64	12.99	12.60	13.39	13.06	23.74	27.01	25.79	29.64	13.12	13.71
1.00	7.36	6.41	7.56	6.63	5.99	4.90	6.12	5.04	9.58	10.11	10.26	10.93	5.19	4.93
1.50	2.88	1.52	2.91	1.55	2.67	1.26	2.69	1.29	2.66	2.21	2.76	2.34	1.67	1.08
2.00	2.13	0.49	2.14	0.50	2.10	0.41	2.10	0.42	1.36	0.72	1.39	0.75	1.09	0.32
2.50	2.01	0.14	2.01	0.15	2.01	0.12	2.01	0.12	1.06	0.26	1.07	0.27	1.01	0.08
3.00	2.00	0.03	2.00	0.03	2.00	0.02	2.00	0.02	1.01	0.08	1.01	0.09	1.00	0.01
Chart constants	($a = 72$, $b = 429$)		($a = 71$, $b = 430$)		($a = 81$, $b = 420$)		($a = 80$, $b = 421$)		($a = 25$, $b = 476$)		($a = 24$, $b = 477$)		3.0845	

Table 5
ARL and SDRL values for the 2-*of*-2 DR, 2-*of*-2 KL and the basic (1-*of*-1) precedence chart for the $t(4)$ distribution when $m = 500, j = 3, n = 5$.

Shift	2-*of*-2 DR				2-*of*-2 KL				Basic precedence (1-*of*-1)			
	ARL	*SDRL*	*ARL*	*SDRL*	*ARL*	*SDRL*	*ARL*	*SDRL*	*ARL*	*SDRL*	*ARL*	*SDRL*
0.00	496.90	573.05	536.72	621.20	490.21	554.18	524.39	594.55	460.22	538.61	520.27	613.67
0.25	200.92	248.00	215.95	268.18	138.19	170.25	146.90	182.03	288.43	370.47	328.18	426.13
0.50	38.68	45.31	40.98	48.41	25.09	27.66	26.28	29.19	102.82	143.88	117.63	167.75
0.75	10.01	9.77	10.41	10.26	7.43	6.66	7.65	6.92	32.84	45.71	37.43	53.44
1.00	4.26	3.11	4.35	3.22	3.61	2.36	3.67	2.43	11.19	14.51	12.58	16.84
1.50	2.23	0.67	2.24	0.68	2.17	0.55	2.17	0.56	2.25	1.97	2.40	2.20
2.00	2.02	0.19	2.02	0.20	2.02	0.16	2.02	0.16	1.16	0.46	1.18	0.49
2.50	2.00	0.06	2.00	0.06	2.00	0.05	2.00	0.05	1.02	0.13	1.02	0.14
3.00	2.00	0.02	2.00	0.02	2.00	0.02	2.00	0.02	1.00	0.04	1.00	0.04
Chart constants	($a = 72, b = 429$)		($a = 71, b = 430$)		($a = 81, b = 420$)		($a = 80, b = 421$)		($a = 25, b = 476$)		($a = 24, b = 477$)	

Tables 4–6 display the results when a reference sample of size $m = 500$ is used to estimate the control limits and to monitor the location with Phase II samples of size $n = 5$, using $Y_{3:5}$, their median, as the plotting statistic. The charts were designed so that an ARL_0 close to 500 was achieved. Two combinations of chart constants are used for each nonparametric chart for which the ARL_0 was slightly below and slightly above the target value, 500. The tables show both the average (*ARL*) and the standard deviation of run length (SDRL); the shift refers to a shift in the mean. From Table 4 it is seen that even under the normal distribution, the nonparametric charts can be quite efficient. The 2-*of*-2 KL chart is almost as efficient as the $\bar{X}$ chart, with a shorter *ARL* but a slightly higher SDRL, especially for small shifts. When the distribution is $t(4)$, that is symmetric yet with heavier tails than the normal, Table 5 shows that both the 2-*of*-2 DR and the 2-*of*-2 KL charts perform better than the basic precedence 1-*of*-1 chart in detecting small shifts, with the 2-*of*-2 KL chart being the best. Thus, the two new nonparametric Shewhart-type charts with signaling rules provide better alternatives, especially for smaller shifts. The same observation applies in the case of a right-skewed distribution such as the gamma (1, 1) as shown in Table 6.

Hence, the runs-type signaling rules enhance the nonparametric chart's sensitivity to a location shift. The gain in efficiency can be substantial; for example, for the $t(4)$ distribution for a shift of 0.5, the *ARL* values of the 1-*of*-1, 2-*of*-2 DR and the 2-*of*-2 KL charts are 117.63, 40.98 and 26.28, respectively, when the corresponding ARL_0 values are very comparable, 520.27, 536.72 and 524.39, respectively. Note that in Table 6 for the gamma (1, 1) distribution the 2-*of*-2 DR and the basic precedence chart display somewhat of a strange behavior in that both the *ARL* and SDRL values first increase from their corresponding values for the IC case for a shift of 0.25; thereafter the *ARL* and *SDRL* values decrease for increasing shifts as it might be expected. We have not been able to fully explain this phenomenon. A repeat of the simulations produced similar results.

Because the run length distribution is highly right-skewed, percentiles, in addition to the *ARL* and the SDRL values, should be examined (as recommended in the literature; see, e.g., Chakraborti (2007)). To this end, a number of percentiles of the unconditional run length distribution were calculated, including the three quartiles for each of the three distributions, respectively, both for the IC and the OOC cases. The results (not shown here) lead to the same general observation that the newly proposed nonparametric charts are more efficient than the basic precedence chart, with the 2-*of*-2 KL chart having a slight edge. For example, in the OOC case for a shift of 0.5 in the mean of a normal distribution, the quartiles for the 2-*of*-2 KL chart (with $a = 81$ and $b = 420$) are all shorter: 7, 16 and 33, respectively, compared to those of both the 2-of-2 DR chart (with $a = 72$ and $b = 429$): 11, 24, and 50 and the 1-*of*-1 precedence chart (with $a = 25$ and $b = 476$): 23, 57, and 127. This shows that the 2-*of*-2 KL chart is far superior, since shorter percentiles in the OOC case means quicker signals, to the 1-*of*-1 chart and is somewhat superior to the 2-*of*-2 DR chart.

Please cite this article in press as: Chakraborti, S., et al., A phase II nonparametric control chart based on precedence statistics with runs-type signaling rules. Computational Statistics and Data Analysis (2008), doi:10.1016/j.csda.2008.09.025

Table 6
ARL and SDRL values for the 2-*of*-2 DR, 2-*of*-2 KL and the basic (1-*of*-1) precedence chart for the gamma (1, 1) distribution when $m = 500$, $j = 3$, $n = 5$.

Shift	2-*of*-2 DR				2-*of*-2 KL				Basic precedence (1-*of*-1)			
	ARL	*SDRL*	*ARL*	*SDRL*	*ARL*	*SDRL*	*ARL*	*SDRL*	*ARL*	*SDRL*	*ARL*	*SDRL*
0.00	496.90	573.05	536.72	621.20	490.21	554.18	524.39	594.55	460.22	538.61	520.27	613.67
0.25	233.82	815.92	250.23	887.98	310.12	405.49	331.64	436.02	527.27	730.48	600.16	844.59
0.50	58.22	216.59	61.33	234.84	88.52	111.41	94.24	119.33	255.49	351.96	290.46	406.50
0.75	17.55	61.43	18.23	66.27	28.03	33.05	29.65	35.23	124.76	170.53	141.61	196.68
1.00	7.36	18.96	7.56	20.33	10.26	10.74	10.75	11.39	61.56	83.20	69.72	95.80
1.50	2.88	2.30	2.91	2.45	2.61	1.34	2.66	1.42	15.70	20.35	17.67	23.33
2.00	2.13	0.13	2.14	0.15	2.00	0.03	2.00	0.04	4.47	5.19	4.96	5.92
2.50	2.01	0.00	2.01	0.00	2.00	0.00	2.00	0.00	1.63	1.32	1.75	1.52
3.00	2.00	0.00	2.00	0.00	2.00	0.00	2.00	0.00	1.03	0.23	1.05	0.29
Chart constants	($a = 72, b = 429$)		($a = 71, b = 430$)		($a = 81, b = 420$)		($a = 80, b = 421$)		($a = 25, b = 476$)		($a = 24, b = 477$)	

7. Summary and recommendations

The key advantage of a nonparametric chart is its IC robustness, that is, the fact that the IC run length distribution and hence all the associated properties (mean, standard deviation, percentiles, etc.) remain the same for all continuous distributions. This allows the practitioners to have greater confidence in their charts. The new nonparametric Shewhart-type control charts enhance the performance of the 1-*of*-1 precedence control chart. Because these charts (the control limits as well as the charting statistic) are based on order statistics, they can be applied as soon as the required order statistics are observed and this can be an advantage in certain applications. Charting constants are provided for the control limits along with the achieved ARL_0 and *FAR* values; these would help in practical implementation. The performance of the proposed charts is seen to be either on per or better than the Shewhart $\bar{X}$ chart and the 1-*of*-1 precedence chart. The new charts can thus be useful for the quality practitioner and are recommended in practice. A choice between the two new charts would depend on the kind of process change one expects to detect. The 2-*of*-2 KL chart is more suitable when the process shifts either up or down (shift in one direction) while the 2-*of*-2 DR chart is appropriate if the process swings from one time point to the other. Thus, in general, the chart that best matches the hypothesized or expected shift pattern should be chosen in practice.

Finally, one can further consider enhancements of precedence charts using, say *k-of-k* rules, $k \geq 3$. The run length random variable associated with *k-of-k* DR chart is defined as $T_k = \min\{t : Z_{t-k+1} = 1, \ldots, Z_t = 1\}$. Similarly, *k-of-m*, $k \leq m$, rules are also possible. Although, such rules may improve the sensitivity of the charts, the 2-*of*-2 charts studied here seem more attractive from a practical point of view.

Acknowledgements

The authors would like to thank the referees for their constructive comments and suggestions. Part of this work was done during the first author's visit to the Department of Mathematics, Izmir University of Economics, Turkey. Support from the Scientific and Technological Research Council of Turkey (TUBITAK) and from the Department of Information Systems, Statistics and Management Science, University of Alabama, is appreciated.

Appendix

Proof of Theorem 1. The proof is straightforward for $x \leq 2$, note that $P(T_2 = 2) = P(Z_1 = 1, Z_2 = 1) = \lambda_2$, as defined in (8). For $x \geq 3$, we write for the unconditional distribution of T_2

$$P(T_2 = x) = \sum_{y=1}^{x-2} P(T_2 = x | S_x = x - y) P(S_x = x - y). \tag{A.1}$$

First, consider the conditional probability $P\{T_2 = x | S_x = x - y\}$. By de Finetti's theorem, a sequence of exchangeable random variables is conditionally i.i.d. Hence, the conditional distribution of T_2 given the number of successes for *exchangeable* binary random variables is the same as that for a sequence of i.i.d. binary variables. This latter distribution has been worked out in the literature (see, e.g., Balakrishnan and Koutras (2002) page 56; note a typo) and is given by

$$P(T_2 = x | S_x = x - y) = \frac{\sum_{j=0}^{[\frac{x-y-2}{2}]} (-1)^j \binom{y}{j} \binom{x - 2(j+1) - 1}{y - 1}}{\binom{x}{y}}. \tag{A.2}$$

Now, using (7)

$$P(S_x = x - y) = \binom{x}{y} \sum_{i=0}^{y} (-1)^i \binom{y}{i} \lambda_{x-y+i}. \tag{A.3}$$

The proof is completed by substituting (A.2) and (A.3) in (A.1). □

References

Bakir, S.T., 2004. A distribution-free Shewhart quality control chart based on signed-ranks. Quality Engineering 16, 613–623.

Balakrishnan, N., Koutras, M.V., 2002. Runs and Scans with Applications. John Wiley, New York.

Chakraborti, S., Van der Laan, P., Bakir, S.T., 2001. Nonparametric control charts: An overview and some results. Journal of Quality Technology 33, 304–315.

Chakraborti, S., Van der Laan, P., Van de Wiel, M.A., 2004. A class of distribution-free control charts. Journal of the Royal Statistical Society, Series C 53, 443–462.

Chakraborti, S., 2007. Run length distribution and percentiles: The Shewhart X-bar chart with unknown parameters. Quality Engineering 19, 119–127.

Chakraborti, S., Eryilmaz, S., 2007. A nonparametric Shewhart-type signed-rank control chart based on runs. Communications in Statistics: Simulation and Computation 36, 335–356.

Chakraborti, S., Graham, M.A., 2007. Nonparametric control charts. In: Encyclopedia of Quality and Reliability. John Wiley, New York.

Derman, C., Ross, S.M., 1997. Statistical Aspects of Quality Control. Academic Press, San Diego.

Fu, J.C., Spiring, F.A., Xie, H., 2002. On the average run lengths of quality control schemes using a Markov chain approach. Statistics and Probability Letters 56, 369–380.

Fu, J.C., Lou, W.Y.W., 2003. Distribution Theory of Runs and Patterns and Its Applications: A Finite Markov Chain Imbedding Approach. World Scientific Publishing, Singapore.

George, E.O., Bowman, D., 1995. A full likelihood procedure for analyzing exchangeable binary data. Biometrics 51, 512–523.

Gibbons, J.D., Chakraborti, S., 2003. Nonparametric Statistical Inference, fourth ed. Marcel Dekker, New York.

Klein, M., 2000. Two alternatives to the Shewhart $\bar{X}$ control chart. Journal of Quality Technology 32, 427–431.

Liu, R., 1995. Control charts for multivariate processes. Journal of the American Statistical Association 90, 1380–1387.

Montgomery, D.C., 2001. Introduction to Statistical Quality Control, fourth ed. John Wiley, New York.

Nelson, L.S., 1984. The Shewhart control chart-tests for special causes. Journal of Quality Technology 16, 237–239.

Qiu, P., Hawkins, D., 2001. A rank-based multivariate cumulative sum procedure. Technometrics 43, 120–132.

Qiu, P., Hawkins, D., 2003. A nonparametric multivariate CUSUM procedure for detecting shifts in all directions. Journal of the Royal Statistical Society, Series D 52, 151–164.

Qiu, P., 2008. Distribution-free multivariate process control based on log-linear modeling. IIE Transactions 40, 664–677.

Please cite this article in press as: Chakraborti, S., et al., A phase II nonparametric control chart based on precedence statistics with runs-type signaling rules. Computational Statistics and Data Analysis (2008), doi:10.1016/j.csda.2008.09.025