Department of Applied Mathematics
Faculty of EEMCS

University of Twente
The Netherlands

P.O. Box 217
7500 AE Enschede
The Netherlands

Phone: +31-53-4893400
Fax: +31-53-4893114

Email: memo@math.utwente.nl
www.math.utwente.nl/publications

Memorandum No. 1764

Minimum control charts

W. Albers and W.C.M. Kallenberg

June, 2005

ISSN 0169-2690

Minimum control charts

Willem Albers and Wilbert C. M. Kallenberg

Department of Applied Mathematics
University of Twente
P.O. Box 217, 7500 AE Enschede
The Netherlands

Abstract Standard control charts are very sensitive to estimation effects and/or deviations from normality. Hence a program has been carried out to remedy these problems. This is quite adequate in most circumstances, but not in all. In the present paper, the remaining complication is attacked: what to do if a nonparametric approach is indicated, but too few Phase I observations are available? It is shown that grouping the observations during Phase II works well. Surprisingly, rather than using the group averages, it is definitely preferable to compare the minimum for each group to a suitably chosen upper control limit. (And in the two-sided case, also the maximum to an analogous lower control limit.) This 'minimum control chart' is demonstrated to be quite attractive: it is easy to explain and to implement. Moreover, while it is truly nonparametric, its power of detection is comparable to that of the customary, normality assuming, charts based on averages.

Keywords and phrases: Statistical Process Control, Phase II control limits, order statistics

2000 Mathematics Subject Classification: 62P30, 62C05

1 Introduction and motivation

Consider the traditional Shewhart $\overline{X}$-chart for monitoring the mean of a production process. An upper limit UL and a lower limit LL are selected and an Out-of-Control (OoC) signal is produced as soon as a newly arriving measurement falls outside these limits. During the period in which the process is In-Control (IC), the false alarm rate (FAR) should be p, which is usually very small, e.g. $p = 0.001$. Typically, the distribution from which the measurements stem is not known. Hence before starting the actual control, a number, say n, of Phase I observations need to be used to estimate the extreme quantiles UL and LL.

The customary approach narrows the problem down by assuming the underlying distribution to be normal. After this, it merely remains to estimate its mean and variance.

But even then, a large n is required to reduce the effects of the corresponding estimation errors to bearable proportions: see e.g. Ghosh et al. (1981), Quesenberry (1993), Roes (1995), Chen (1997), Woodall and Montgomery (1999) (p. 379) and Chakraborti (2000). Therefore, Albers and Kallenberg (2003, 2004a, b) (to be denoted for short as AK (2003, 2004a, b) in the sequel) have demonstrated how this can be solved by using relatively simple corrections.

But assuming normality is a step which often is questionable, see e.g. Chan et al. (1988), Pappanastos and Adams (1996) and Albers, Kallenberg and Nurdiati (2004a) (AKN (2004a) for short). Clearly, without this assumption, estimation becomes essentially more difficult. As a first improvement, AKN (2002, 2004a) have developed parametric charts. These generalize the usual normal ones by estimating a shape parameter, in addition to the mean and the variance. As a second alternative possibility, nonparametric charts are considered in AK (2004c) (for some related work, see e.g. Willemain and Runger (1996), Ion et al. (2000) and Chakraborti et al. (2001).)

In going from normal charts via parametric ones to the nonparametric case, two opposing effects can be observed. The normal chart has only two parameters which need to be estimated, so the stochastic error (SE) involved is small, at least in comparison to the alternative proposals. But if normality fails to hold, one is actually estimating the wrong quantity, and thus a non-vanishing model error (ME) turns up. Adding the third parameter in the parametric case will increase the SE. But on the other hand, the increased flexibility of the parametric family helps to reduce the ME. Often this will be sufficient to produce an adequate solution, but not always. In the latter case one has to resort to a nonparametric solution. Here the ME is nicely 0, but the SE can be huge. As an example, just imagine how to estimate the upper 0.001-quantile of a distribution in a nonparametric way, using an order statistic based on a customary sample of say $n = 100$ observations from Phase I.

In practice one needs to select one of the three choices offered, either on the basis of prior belief about the type of the underlying distribution, or by using a data driven approach (see AKN (2004b)). In many instances, the result will be satisfactory. But, as is clear from the above, a problem occurs if one winds up with the nonparametric chart and moreover only has a sample size n of a magnitude which is common in practice (and hence too small). For this reason, in AKN (2004b) a modified version of this chart is proposed, but even there n should be at least 500. If this is not the case, two simple options exist. The first is obvious: increase n to the desired level after all. The second is pretty straightforward too: increase p to a less extreme level like 0.01, which will also help to mollify the estimation effects. But if neither of these two alternatives is possible or acceptable, the question remains: "How to get suitable nonparametric control charts for less than 500 Phase I observations?" Providing a satisfactory and simple answer to this question is in fact the purpose of the present paper.

A possible solution originates from combining the incoming observations during Phase II into groups of size m, with $m > 1$. Incidentally, in practice quite often one already automatically is faced with such grouped observations, with as customary values $m = 3$, 4 or 5. The potential advantage is that a few observations together may tell us more than each of these observations considered separately. Postponing a decision till e.g. 2 - 5 observations have come in thus might be profitable. Notice, however, that the step from individual observations (i.e. $m = 1$) to grouped ones presents several complications. First of all, the question arises how the step towards $m > 1$ affects the behavior of the chart once the process gets OoC. If the estimation problems are solved by going to the grouped case at the cost of a substantial loss of detection power during OoC, the solution is of little value.

An additional complication is that actually it is not just 'the' step from $m = 1$ to $m > 1$. For an individual measurement X matters are simple: a signal occurs if either $X > UL$ or $X < LL$. But for larger m, we first have to figure out which statistic based on $X_1, \ldots, X_m$ we want to use and thus there are various ways to proceed from $m = 1$. Clearly, under normality the sample mean (AVE) is optimal and easy to work with. Actually, the grouped case is reduced to the individual one in a few simple steps (cf. AK (2003)). But in the nonparametric case, AVE turns out to be neither optimal nor easy to work with. (Simple central limit theory does not help: m is far too small, especially as we are dealing with tail probabilities.) Hence it is really necessary to investigate more possibilities.

The combination of having to take the OoC-behavior into account as m varies, together with the fact that each value of m itself offers different possibilities, strongly suggests to first study the resulting picture separately, before taking the estimation effects into account. In other words, even though our whole exercise is motivated by the estimation problems described above, it pays to first look at matters for the case of known underlying distributions. Fortunately, this task has already been executed in AK (2004d). The results obtained in that paper can be summarized for our present purpose as follows. To begin with, also when normality fails to hold, AVE remains at first sight an obvious candidate for the statistic to use. Moreover, even under normality, for AVE the value of m which is optimal with respect to detection power, will depend on the magnitude of the shift d which the process is supposed to experience during OoC. Roughly speaking, the larger d, the smaller the best value of m, a result which is intuitively clear in a qualitative sense. (For quantitative details, consult AK (2004d).) But d is typically unknown and by definition only manifests itself during the OoC situation. Hence, unlike the shape of the underlying distribution, d cannot be estimated during Phase I using data driven techniques. Consequently, a specific choice of m has to depend solely on some prior belief about d. However, a global type of conclusion can be drawn: for a wide range of d-values of practical interest, AVE-based charts for values of m ranging from 2 to 5 perform better than the individual chart (see again AK (2004d)). In other words, increasing m somewhat for estimation purposes, as we intend to do in this paper, definitely does not destroy the performance of the chart.

After studying AVE, the attention in AK (2004d) is directed towards finding suitable competitors. The main conclusion is that a very nice possibility is offered by charts based on MIN, the minimum of the m observations in the group, for dealing with the upper control limit. At first sight MAX, the maximum rather than the minimum, might seem a more likely choice for this upper case, while MIN would be suitable for comparison to the lower control limit. In fact, previous proposals using order statistics all went in this direction (see AK (2004d) for further discussion and references). However, in our opinion this should really be done the other way around: MIN at UL and MAX at LL. We call such a chart a minimum control chart or a MIN chart for short. To avoid switching from MIN to MAX all the time, from now on we tacitly concentrate on the upper limit, unless stated otherwise.

Studying the OoC performance immediately reveals that MAX at UL is a poor choice indeed, whereas MIN performs quite well. An intuitive explanation is that once a shift occurs, it affects all observations in the group. Hence this translates itself in all these observations being pretty large, but not necessarily extremely large. So MIN really uses the group, while MAX in connection with the upper limit is nothing but a repeated individual chart. Of course MIN loses some detection power in comparison to AVE when normality holds after all, but the MIN-chart for $m = 2 - 5$, just like AVE, outperforms the individual chart for the interesting d-values. Moreover, outside the normal model, examples are easily found (cf. AK (2004d) again) where MIN also outperforms AVE. Consequently, from the important point of view of OoC behavior, the MIN-charts are definitely admissible.

The excursion to the case of known underlying distributions in AK (2004d) thus has produced AVE and MIN as well-performing candidates. Next we return to the unknown case and face the estimation problems concerned. First we consider AVE, for which we need to estimate the upper $(mp)^{th}$-quantile. In this way, the comparison is fair, as the average time to signal (ATS) will then be $m/FAR = 1/p$, which thus agrees with ATS of the straightforward individual chart based on the upper p^{th}-quantile. As demonstrated in AK (2004d), relying on approximate normality in general still goes wrong for such extreme values: the resulting ME is often far too large to be acceptable. Hence we have to proceed nonparametrically by generalizing the approach for $m = 1$. Instead of simply using a suitable extreme upper order statistic of the n individual observations, we now first consider all groups of size m taken from these Phase I observations and compute the corresponding values of AVE. From among these 'n over m' pseudo-observations we select a suitable order statistic to estimate the $(mp)^{th}$-quantile. The idea is of course that this will produce a more accurate estimate. However, verifying this conjecture presents us with new and nontrivial theoretical problems. To be more precise, it involves a rather delicate analysis of the tail behavior of empirical distribution functions for convolutions (see AK (2004e)). Unfortunately, it turns out that the improvement is very superficial from a practical point of view: the estimation step still requires uncomfortably large values of n.

Hence by now the field has been narrowed down to MIN as the sole candidate for a satisfactory answer to the aforementioned motivating question: "How to get suitable nonparametric control charts for less than 500 Phase I observations?" Fortunately, it is immediately clear that matters here are both more simple and more promising than with AVE: to obtain ATS equal to $1/p$, in case of MIN we should use the upper $\{(mp)^{(1/m)}\}^{th}$-quantile. But this is much less extreme than either the p^{th}- or the $(mp)^{th}$-quantile, and thus can be estimated quite well for ordinary sample sizes. For example, for $m = 3$ and $p = 0.001$, we obtain that $(mp)^{(1/m)} = 0.144$. Consequently, this introduction indeed has demonstrated that it is worthwhile to study such 'minimum control charts'.

In section 2 we will introduce these charts in a systematic manner. Section 3 is devoted to studying the bias and deriving corrections to remove it. Likewise, section 4 is concerned with exceedance probabilities and ways to control these. The impact of the corrections from the last two sections on the OoC behavior, is the subject of section 5. A further explanation of why going to $m > 1$ does indeed help, as well as a brief comparison to the normal chart, are presented in section 6. The exposition in section 2-6 is illustrated with quite a few (partly continuing) examples. In addition, an application of the MIN chart to real data is presented as well.

2 Definition and basic properties of MIN

Let X be a random variable (rv) with a continuous distribution function (df) F. For ease of presentation, we shall concentrate on the one-sided case; the two-sided case can be treated in a similar fashion and will lead to completely analogous results. (Just remember to switch from MIN to MAX for the lower control limit!) First consider the individual case ($m = 1$) for a moment. Hence for given p, we need UL such that simply $P(X > UL) = p$. For any df H we write $\overline{H} = 1-H$ and H^{-1} and $\overline{H}^{-1}$ for the respective inverse functions, and thus $UL = F^{-1}(1-p) = \overline{F}^{-1}(p)$. Usually F is (at least partly) unknown and an estimated version $\widehat{UL}$ is required, using a sample $X_1, \ldots, X_n$ from F (the Phase I observations). In the simple normal case, $F(x) = \Phi((x-\mu)/\sigma)$, in which Φ stands for the standard normal df, and thus $UL = \mu + \sigma u_p$, with $u_p = \overline{\Phi}^{-1}(p)$ (e.g. $u_{0.001} = 3.09$). The corresponding $\widehat{UL}$ then is given by $\widehat{UL} = \hat{\mu} + \hat{\sigma} u_p$, with $\hat{\mu}$ and $\hat{\sigma}$ e.g. the sample mean and sample standard deviation, respectively.

But if F is completely unknown, a nonparametric approach is required. Let $F_n(x) = n^{-1}\#\{X_i \leq x\}$ be the empirical df and F_n^{-1} the corresponding quantile function, i.e. $F_n^{-1}(t) = \inf\{x | F_n(x) \geq t\}$. Then it follows that $F_n^{-1}(t)$ equals $X_{(i)}$ for $(i-1)/n < t \leq i/n$, where $X_{(1)} < \ldots < X_{(n)}$ are the order statistics corresponding to $X_1, \ldots, X_n$. Hence, letting $\overline{F}_n^{-1}(t) = F_n^{-1}(1-t)$, we get for the individual case: a signal occurs if for a new observation Y we have

$$Y > \widehat{UL}, \text{ with } \widehat{UL} = \overline{F}_n^{-1}(p) = X_{(n-r)}, \tag{2.1}$$

where moreover $r = [np]$, with $[y]$ the largest integer $\leq y$. Note that for ordinary p and n, like $p = 0.001$ and $n = 100$, we will have $r = 0$, and thus $\widehat{UL} = X_{(n)}$. This already shows that such a chart cannot be satisfactory. For $n = 100$ we get the same $\widehat{UL}$ for all $0 < p < 0.01$ and then, obviously, FAR cannot be close to $p = 0.001$ and $p = 0.02$ at the same time. In fact, for $p = 0.001$, r will remain 0 until n is at least 1000. The behavior of this chart, as well as that of suitably corrected versions, is amply studied in AK (2004c).

Next we move on to the grouped case, where $m > 1$ (and typically $2 \leq m \leq 5$). The situation where the process is IC will be our starting point. Hence after Phase I, let

$Y_1, \dots, Y_m$ be a new group of observations from F, and consider

$$T = T(m) = \min(Y_1, \dots, Y_m) \tag{2.2}$$

as our control statistic. (Here and in what follows we add '(m)' to the quantities we define when needed to avoid confusion, but often we use the abbreviated notation.) As in this case $P(T > UL) = \overline{F}(UL)^m$, it follows that a fair comparison to the individual chart is obtained by choosing $UL = UL(m) = \overline{F}^{-1}((mp)^{1/m})$. In analogy to the above, the estimation step then leads to our basic proposal for the nonparametric minimum based control chart

$$T > \widehat{UL}, \text{ with } \widehat{UL} = \widehat{UL}(m) = \overline{F}_n^{-1}((mp)^{1/m}) = X_{(n-r(m))}, \tag{2.3}$$

and where now

$$r = r(m) = [n(mp)^{1/m}]. \tag{2.4}$$

Taking once more $p = 0.001$ and $n = 100$, raising the value of m to e.g. $m = 3$ then produces $r = r(3) = 14$: instead of the maximal value $X_{(100)}$ obtained for $m = 1$, it is now okay to use the much less extreme $X_{(86)}$. To illustrate matters, we have the following:

Application 2.1. We consider a real life example concerning the production of electric shavers by Philips. In an electrochemical process razor heads are formed. The measurements concern the thickness of these razor heads on a particular spot on the head. Available are two samples of 835 measurements each. In AKN (2004b) this large data set has already been utilized to illustrate the data driven approach. The first sample is used there as Phase I observations and the control chart thus obtained is applied to the second sample. A histogram of the first sample (see Figure 1 in AKN (2004b)) already suggests that the right tail might be normal, but that the left tail is too thick. Indeed, the data driven chart simply selects the normal chart for the upper limit, but for the lower limit it rejects the normal choice, as well as the parametric one, and winds up with the nonparametric choice there. Consequently, in this application it would not have been appropriate to apply a straightforward, normality based, Shewhart chart.

But we should realize that this approach was feasible because we had $n = 835$, which is quite large. At this point once again the motivating question from the Introduction springs to mind: "How to get suitable nonparametric control charts for less than 500 Phase I observations?" Using the results above, we can now indeed provide a first step towards a satisfactory answer. Instead of using the full first sample, only take the first 100 observations, i.e. let $n = 100$. For this Phase I sample, the average $\overline{x} = 42.50$, while the smallest and largest observations are 32.27 and 51.04, respectively. As we saw above, taking $p = 0.001$ and $m = 3$ leads to $r(3) = r = 14$, and hence to $X_{(86)}$ as UL. In the present example, $x_{(86)} = 45.45$. Hence for the remaining sample, a signal will occur whenever the minimum of one of the 523 triples contained in these 1570 observations exceeds 45.45. If we for a moment switch to the lower control limit, it is also immediate how to proceed there. Just select $X_{(15)}$ as LL, which here produces $x_{(15)} = 39.14$ and produce a signal when the maximum of one of the 523 triples falls below this value.

In the next sections we shall continue with this application. □

Using a stochastic limit as in (2.3) in the control chart means that the fixed p from the case of known F is in fact replaced by

$$P_n = P_n(m) = \frac{P(T > \widehat{UL}|(X_1, \ldots, X_n))}{m} = \frac{\{\overline{F}(X_{(n-r)}\}^m}{m}. \tag{2.5}$$

(In what follows we will usually write $P_n = P(T > \widehat{UL})/m$ without explicitly stating that we work conditionally on $(X_1, \ldots, X_n)$). Let $U_{(1)} < \ldots < U_{(n)}$ denote order statistics for a sample of size n from the uniform df on (0,1), then it is immediate from (2.5) that the basic proposal from (2.3) and (2.4) leads to

$$P_n \cong \frac{\{U_{(r+1)}\}^m}{m}, \tag{2.6}$$

with '$\cong$' denoting 'distributed as'.

The performance of this estimated chart can be judged in terms of FAR, by comparing P_n directly to p. Alternatively, one can use ATS as a starting point, which means comparing $1/P_n$ to $1/p$. We shall consider both FAR and ATS. Thus, denoting the corresponding relative errors by $W = W_i$, $i = 1, 2$, with

$$W_1 = \frac{P_n}{p} - 1 \text{ and } W_2 = \frac{p}{P_n} - 1, \tag{2.7}$$

respectively, the problem boils down to studying the variability of these W_i. For what values of p, n and m are quantities like EW and $P(W > \varepsilon)$, with ε small, sufficiently close to 0? What kind of modifications to $\widehat{UL}$ given by (2.3) and (2.4) can help to make them closer to 0?

3 Bias criterion

First consider EW, which defines a bias criterion: we focus on controlling the average behavior of the chart during a long series of separate applications. To begin with we look at EW_1 in (2.7), i.e. we simply compare EP_n to p. From (2.6) it is immediate that the basic proposal gives

$$EP_n = \frac{\prod_{j=1}^{m} \frac{r+j}{n+j}}{m} = \frac{\binom{r+m}{m}}{m\binom{n+m}{m}}. \tag{3.1}$$

Now let $\delta = n(mp)^{1/m} - r$, then it follows from (2.4) that $0 \le \delta < 1$ and $p = \{(r+\delta)/n\}^m/m$. In case $m = 1$, we simply have that $EP_n = (r+1)/(n+1) > (r+\delta)/n = p$, unless the rare situation occurs where $\delta \ge 1 - p$. For $m \ge 2$ it is straightforward to verify that $EP_n > p$ will always hold for the range of values of r which are of interest. (In fact, $r < (m-1)n/(m+1) - 2(m^2+2)/\{3(m+1)\}$ and $r < n - m$ suffices.)

From the above it is clear that EW_1 attains its maximum with respect to δ for $\delta = 0$. Hence

$$EW_1 \le \frac{\prod_{j=1}^{m} \frac{r+j}{n+j}}{(r/n)^m} - 1 \le \exp\left(\frac{m(m+1)}{2r}\right) - 1. \tag{3.2}$$

As an example, for the customary value $p = 0.001$ and for $r = r(m)$ as given by (2.4), we obtain for each of the values $m = 3$, 4 or 5 that $m(m+1)/(2r)$ in (3.2) equals about $40/n$. Notice to begin with that this indeed provides an enormous improvement over the case where $m = 1$, as in that case the maximal EW_1 almost equals $1/r = 1000/n$. Hence grouping indeed helps to reduce the bias! On the other hand, a relative error like $40/n$ is still not really negligible and below we will suggest a suitably corrected version of the basic chart.

But let us first also briefly consider ATS, rather than just FAR. Hence look at EW_2 now. From (2.6) it is immediate that

$$E\frac{1}{P_n} = m \prod_{j=0}^{m-1} \frac{n-j}{r-j}. \tag{3.3}$$

For the standard Shewhart chart, it is a well known phenomenon that both P_n and $1/P_n$ have a positive bias. Indeed this is the case here as well, as $E(1/P_n) > m(n/r)^m \ge m(n/(r+\delta))^m = 1/p$. The maximal bias thus occurs here for $\delta = 1$ and some straightforward calculation shows that for $m/(r+1) < 0.68$

$$EW_2 \le \exp\left(\frac{m(m+1)}{2(r+1)} + \frac{m(m+1)(2m+1)}{6(r+1)^2}\right) - 1. \tag{3.4}$$

Comparison of (3.4) to (3.2) shows that the maximal relative bias of $1/P_n$ behaves pretty much like that of P_n itself.

Next we shall introduce corrected versions in order to remove the bias. This can be done in a very simple way by randomizing between consecutive, somewhat shifted, order statistics. Let V be independent of $(X_1, \ldots, X_n, Y_1, \ldots, Y_m)$, with $P(V = 1) = 1 - P(V = 0) = \lambda$ and let k be some integer. Then replace the basic $\widehat{UL}$ from (2.3) by

$$\widehat{UL}(k, \lambda) = (1 - V)X_{(n+k+1-r)} + V\,X_{(n+k-r)}, \tag{3.5}$$

with r once more as in (2.4) and using the convention $X_{(n^*)} = \infty$ for $n^* \ge n+1$. Then we obtain:

Lemma 3.1. *$EW_1 = 0$, and thus $EP_n = p$, will result by taking k and λ in* (3.5) *such that*

$$\binom{r-k-1+m}{m} \le mp\binom{n+m}{m} < \binom{r-k+m}{m},\ \lambda = \frac{mp\binom{n+m}{m} - \binom{r-k-1+m}{m}}{\binom{r-k+m}{m} - \binom{r-k-1+m}{m}}. \quad (3.6)$$

Proof. From the definition of the new $\widehat{UL}$ in (3.5), together with (2.5) and (2.6), it is immediate that $EP_n = \{(1-\lambda)EU^m_{(r-k)} + \lambda EU^m_{(r-k+1)}\}/m$. In view of (3.1), this in its turn shows that $EP_n = \{(1-\lambda)\binom{r-k-1+m}{m} + \lambda\binom{r-k+m}{m}\}/\{m\binom{n+m}{m}\}$, and the interpolation result in (3.6) follows. □

Remark 3.1. The idea in (3.6) obviously is to lower the initial $r = r(m)$ 'a bit' in order to compensate for the positive bias in P_n. Actually, as $\Pi_{j=1}^m (r+j-k) \approx r^m\{1 + m((m+1)/2-k))/r\}$, while $(r+\delta)^m \approx r^m(1+m\delta/r)$, equality of these two expressions will entail $k \approx m/2 + 1/2 - \delta$. As $0 \le \delta < 1$, a fair first guess for k in (3.6) is $[m/2]$, i.e. $k=1$ for $m = 2$ or 3 and $k = 2$ for $m = 4$ or 5. □

Example 3.1. Once more use the example mentioned following (2.4), with $p = 0.001$, $n = 100$ and $m = 3$, which led to $r = r(3) = 14$. We now get in addition $mp\binom{n+m}{m} = 530.6$, $\binom{13+m}{m} = 560$ and $\binom{12+m}{m} = 455$, showing that $k = 1$ is indeed right here and moreover that $\lambda = 0.72$. Hence the basic choice $\widehat{UL} = X_{(86)}$ is now modified into $\widehat{UL}(1, 0.72) = (1-V)X_{(88)} + V\,X_{(87)}$, where $P(V = 1) = 1 - P(V = 0) = 0.72$. A signal occurs as soon as the minimum T of a triple Y_1, Y_2 and Y_3 of new observations exceeds this bound. □

Remark 3.2. The exact solution from Lemma 3.1 can be simplified in various ways into approximately unbiased ones. First, replacing V by its expectation λ will not make too much of a difference. Hence an alternative to (3.5) is the deterministic mixture $\widehat{UL}^*(k,\lambda) = (1-\lambda)X_{(n+k+1-r)} + \lambda\,X_{(n+k-r)}$. A second type of simplification is to aim solely at the average value $\delta = \frac{1}{2}$. In view of Remark 3.1, this suggests to use $X_{(n+m/2-r)}$ for m even and $\{X_{(n+(m+1)/2-r)} + X_{(n+(m-1)/2-r)}\}/2\}$ for m odd. In Example 3.1 this would mean replacing $X_{(86)}$ by $\{X_{(88)} + X_{(87)}\}/2$. □

Application 3.1. We continue with Application 2.1: in addition to $x_{(86)} = 45.45$, we now also use $x_{(87)} = 45.47$ and $x_{(88)} = 45.84$. Hence e.g. the deterministic $\widehat{UL}^*(1, 0.72)$ from Remark 3.2 results in widening UL from 45.45 to 45.57. In the same fashion, on the left hand side we have $x_{(14)} = 38.98$ and $x_{(13)} = 38.59$ leading to LL which is lowered from 39.14 to 38.87. Let us now also carry out the actual inspection of the 523 triples. It turns out that the three largest minima are 45.76, 45.63 and 45.24, respectively. Hence our changing UL with respect to bias removal leaves the number of signals unchanged at 2; at LL the result is also the same for the uncorrected and the bias corrected chart: no signals occur in either case, as the smallest maximum among the 523 turns out to equal 39.75. □

The above correction exercise can be repeated for ATS, but we shall not go into details here. One reason is that it is quite straightforward: the results are completely similar, albeit in precisely the opposite direction. For example, the approximate version from Remark 3.2 would now lead to $X_{(n-m/2-r)}$ for m even and $\{X_{(n-(m+1)/2-r)} + X_{(n-(m-1)/2-r)}\}/2$ for m odd, and hence in Example 3.1 to replacing $X_{(86)}$ by $\{X_{(84)} + X_{(85)}\}/2$. More important, however, is the fact that this type of correction does not seem to be terribly attractive for use in practice. As $1/P_n$ does have a positive bias, to correct this it is indeed necessary to move the bound 'a bit' inward, rather than outward, as was the case with P_n itself. But this type of solution may look a bit awkward: the rare occurrence of very small outcomes of P_n inflates $E(1/P_n)$ and in this way causes the positive bias. This suggests that $E(1/P_n)$ might not be the most suitable criterion to judge the behavior of the run length. Hence the correction in itself provides the right answer, but it answers a question which will not often be asked.

4 Exceedance criterion

In the previous section we have analyzed the behavior of the basic proposal from (2.3) and (2.4) with respect to the bias. Through Lemma 3.1 we have supplied a corrected version which is exactly unbiased, followed by some simplified proposals which provide approximate unbiasedness. In itself this is quite satisfactory, but we should keep in mind that so far we have only been dealing with the average behavior of the chart, i.e. over a long series of separate applications. It remains to be seen how bad things can get in one given application of the chart. Note that this is a serious issue: in the Introduction it has been explained how nonparametric charts nicely avoid the ME that spoils parametric charts, but that the cost for this might be a (too) large SE. Actually, for $m = 1$ this is known to be the case, unless n is really large. In fact, this prompted our consideration of the grouped case. In the previous section we already observed that grouping indeed helped: the actual bias was reduced substantially by going from $m = 1$ to $m > 1$ (cf. the example where $1000/n$ for $m = 1$ was replaced by $40/n$ for $m = 3$, 4 or 5). Hence it remains to figure out to what extent this is true where variability rather than bias is concerned.

To analyze how likely 'bad' values of P_n are, it makes sense to look at exceedance probabilities like $P(W_1 > \varepsilon) = P(P_n > p(1 + \varepsilon))$. Incidentally, note that $P(W_2 < -\varepsilon^*)$ boils down to $P(P_n > p/(1-\varepsilon^*))$. Hence the ATS case immediately follows from the FAR case by choosing $\varepsilon = \varepsilon^*/(1 - \varepsilon^*)$ and we can thus restrict attention to W_1. Starting again with the standard proposal from (2.3) and (2.4), we arrive through (2.6) immediately at

$$P(W_1 > \varepsilon) = P(U_{(r+1)} > \{mp(1 + \varepsilon)\}^{1/m}) = B(n, \{mp(1 + \varepsilon)\}^{1/m}, r), \qquad (4.1)$$

where $B(n, p^*, j)$ stands for the cumulative binomial probability $P(Z \le j)$, with Z bin(n, p^*). Note that for any given configuration (n, p, m, ε) the result in (4.1) already gives the exact answer for the desired exceedance probability. Nevertheless, it makes sense to devote some additional attention to (4.1), in order to figure out how $P(W_1 > \varepsilon)$ behaves as a function of these underlying parameters.

To begin with, observe that $n\{mp(1+\varepsilon)\}^{1/m} - r = (r+\delta)(1+\varepsilon)^{1/m} - r > r\varepsilon/m$, which is a positive multiple of n. This means that the binomial probability in (4.1) will tend to 0 as $n \to \infty$, and hence so will $P(W_1 > \varepsilon)$. Consequently, for large sample sizes 'bad' values of P_n, in the sense that p is exceeded by more than a fraction ε, will occur with small probability only. However, especially for small m, the convergence will be slow. This is most clearly visible in the boundary case $m = 1$. Then $r = r(1) = [np]$ will be really small and we have $P(W_1 > \varepsilon) \approx Po(np(1+\varepsilon), r)$, where $Po(\lambda^*, j)$ stands for the cumulative Poisson probability $P(Z^* \le j)$, with Z^* Poisson distributed with parameter λ^*. For illustrative purposes we shall now first present an explicit example.

Example 4.1. As usual, let $p = 0.001$. In addition, let us choose $\varepsilon = 0.2$: a relative deviation of more than 20% will be deemed 'bad'. Starting with the boundary case $m = 1$, we then have that even for n as large as 10000 the exceedance probability approximately equals $Po(12, 10) = 0.35$ (cf. AK (2004c)). Still more than a third of the applications of the chart will wind up with using a 'bad' outcome of P_n! Next we move on to $m = 2$. The improvement is tremendous: here n can be reduced from 10000 to about 500 in order to get an exceedance probability of 0.35 again. For $m = 3, 4, 5$, the required n are about 300, 225 and 215, respectively. □

The desired transparent approximation to $P(W_1 > \varepsilon)$ can be obtained as follows. Let

$$q_\varepsilon = q_\varepsilon(p, m) = (mp)^{1/m}(1+\varepsilon)^{1/m}, \tag{4.2}$$

and in particular $q = q_0 = (mp)^{1/m}$. Then

Lemma 4.1. *For given p, m and ε we have that for n large*

$$P(W_1 > \varepsilon) \approx \Phi\left(\frac{r + 1/2 - nq_\varepsilon}{\{nq_\varepsilon(1-q_\varepsilon)\}^{1/2}}\right) \approx \Phi\left(-\frac{\varepsilon n^{1/2} q^{1/2}}{(1-q)^{1/2} m}\right), \tag{4.3}$$

where the last step holds for ε small.

Proof. As in view of (4.1) and (4.2) we see that $P(W_1 > \varepsilon) = B(n, q_\varepsilon, r)$, the standard approximation for the binomial distribution with continuity correction readily provides the first result in (4.3). By noting that $r = nq - \delta$, it follows that $r + 1/2$ can be replaced by nq in this result. As moreover for ε small we observe that $q - q_\varepsilon = q\{1 - (1+\varepsilon)^{1/m}\} \approx -\varepsilon q/m$, while $q_\varepsilon(1-q_\varepsilon) \approx q(1-q)$, the second step in (4.3) also easily follows. □

Note that the second result in (4.3) nicely separates the effect on $P(W_1 > \varepsilon)$ of the underlying parameters. Especially the dependence on n and ε is straightforward. As concerns p and m, observe that for the customary value $p = 0.001$ the function

$$h(p, m) = \frac{q^{1/2}}{(1-q)^{1/2} m} \tag{4.4}$$

attains the values 0.032, 0.108, 0.137, 0.145 and 0.146, for $m = 1, 2, 3, 4$ and 5, respectively. This illustrates that going to $m > 1$ does indeed help, especially in the first steps (cf. Example 4.1). Moreover, the approximations work quite well as soon as $m > 1$, as is illustrated in the next example.

Example 4.2. Continue with $p = 0.001$ and $\varepsilon = 0.2$ and first let $m = 2$ and $n = 500$, then the exact value of the exceedance probability equals 0.349 (cf. Example 4.1), while the first approximation in (4.3) produces 0.340 and the second 0.314. For $m = 4$ and $n = 225$, these three numbers are 0.344 (cf. Example 4.1 again), 0.340 and 0.332, respectively. □

Hence the good news so far is that working with groups not only helps to reduce the bias but also the variability, and moreover that through (4.1) and (4.3) we both have exact, as well as transparent approximate, results to assess the remaining damage. However, it is also clear that we are not done yet. Just reconsider Example 4.1 for a moment. It is nice that the huge sample size 10000 has been brought down to a few hundreds, but the exceedance probability of 0.35 involved is still quite large. Consequently, just as in the previous section, corrected versions are called for. In fact, the basic idea proposed there, can be used here as well. Just replace the standard $\widehat{UL}$ from (2.3) once again by a $\widehat{UL}(k, \lambda)$ as in (3.5). This time the shifting towards a slightly more extreme (pair of) order statistic(s) is not aimed at the reduction of $|EW_1|$ but at that of $P(W_1 > \varepsilon)$. Only the aim now is not simply 0 (as is the case with unbiasedness), but rather some acceptable level for the occurrence rate of 'bad' values. Hence in addition to ε we need a second small quantity α, in order to specify as our criterion that

$$P(W_1 > \varepsilon) \leq \alpha. \tag{4.5}$$

In complete analogy to Lemma 3.1 we thus obtain

Lemma 4.2. *Equality in* (4.5) *will result by selecting k and λ in* (3.5) *such that*

$$B(n, q_\varepsilon, r-k-1) \leq \alpha < B(n, q_\varepsilon, r-k),\ \lambda = \frac{\alpha - B(n, q_\varepsilon, r-k-1)}{b(n, q_\varepsilon, r-k)}, \tag{4.6}$$

where $b(n, p^*, j)$ stands for the binomial probability $P(Z = j)$, with Z bin(n, p^*).

Proof. In view of (3.5), in combination with (4.1) and (4.2), it is immediate that $P(W_1 > \varepsilon) = (1-\lambda)P(U_{(r-k)} > q_\varepsilon) + \lambda\, P(U_{(r-k+1)} > q_\varepsilon) = (1-\lambda)B(n, q_\varepsilon, r-k-1) + \lambda\, B(n, q_\varepsilon, r-k) = B(n, q_\varepsilon, r-k-1) + \lambda\, b(n, q_\varepsilon, r-k)$, from which (4.5) with equality follows. □

Remark 4.1. For $m = 1$ we need to stick to the Poisson approximation, but as soon as $m > 1$, the results from (4.3) can be used to find approximate values for k and λ. Actually, arguing as in Lemma 4.1 we have that $B(n, q_\varepsilon, r-k) \approx \Phi((r-k+1/2-nq_\varepsilon)/\{nq_\varepsilon(1-q_\varepsilon)\}^{1/2})$. Equating this to the desired boundary value $\Phi(-u_\alpha) = \alpha$ shows that to first order $k = [k_1]$ (and thus $1 - \lambda = k_1 - [k_1]$), where

$$k_1 = u_\alpha\{nq_\varepsilon(1-q_\varepsilon)\}^{1/2} + \{r + 1/2 - nq_\varepsilon\} \approx k_2 = u_\alpha\{nq(1-q)\}^{1/2} - \frac{\varepsilon nq}{m}, \tag{4.7}$$

with k_2 as a further approximation for ε small. Also note that $k_2 \approx u_\alpha\{r(1-r/n)\}^{1/2} - \varepsilon r/m$. □

Example 4.3. Continuing with $p = 0.001$ and $\varepsilon = 0.2$, we first quote an example from AK (2004c) for the boundary case $m = 1$. Let $n = 5000$, and thus $r = 5$, and observe that $Po(6, z)$ for $z = 5$, 4 and 3, produces the values 0.45, 0.29 and 0.15, respectively. Consequently, suppose we let $\alpha = 0.2$ as well, then it follows that $k = 1$ and $\lambda = 0.36$. Hence the uncorrected $\widehat{UL} = X_{(4995)}$ is replaced through using the Poisson approximation in (4.5) by $X_{(4996)}$ with probability 0.36 and by $X_{(4997)}$ otherwise. This example shows that we can indeed remedy the problems from the uncorrected case (cf. Example 4.1) for $m = 1$ as well. But on the other hand, a very large n remains necessary and the bounds obtained are still very close to the boundary. □

Example 4.4. Next we reconsider the situation of Example 3.1, with $p = 0.001$, $n = 100$ and $m = 3$, leading to $r = 14$. Using once more $\varepsilon = 0.2$, we obtain for $B(100, 0.1533, 14-j)$ the outcomes 0.421, 0.315, 0.220 and 0.143 for $j = 0$, 1, 2 and 3, respectively. Hence if $\alpha = 0.2$, we arrive at $k = 2$ and $\lambda = 0.74$. In the exceedance case the basic choice $\widehat{UL} = X_{(86)}$ is thus modified into $\widehat{UL}(2, 0.74) = (1-V)X_{(89)} + V\,X_{(88)}$, where $P(V = 1) = 1 - P(V = 0) = 0.74$. The approximations from (4.7) produce $k_1 = 2.21$ and $k_2 = 2.00$, respectively. This leads to $k = 2$ in either case and to $\lambda = 0.79$ and $\lambda = 1.00$, respectively, which are quite close to the exact values $k = 2$ and $\lambda = 0.74$. Finally, if we compare the present result to the outcome $\widehat{UL}(1, 0.72)$ obtained in the bias case from Example 3.1, we observe that the outward shift is increased by about one step in going to the exceedance case. This illustrates that protection against exceedance effects requires a larger correction than against bias. □

Application 4.1. Continuing with Applications 2.1 and 3.1, it follows from Example 4.4 that $x_{(89)} = 45.92$ and $x_{(12)} = 38.42$ are also brought into the action. This leads to using e.g. $\widehat{UL}^*(2, 0.74)$ (cf. Remark 3.2) and thus to a further widening of the limits to $UL = 45.86$ and $LL = 38.55$, respectively. Hence in the upper case the two largest minimum values 45.76 and 45.63 now fall within the limit as well and the number of signals is reduced from 2 to 0. Clearly, at the lower end, the number stays 0. Figures 1 and 2 below illustrate the behavior of the chart at the upper and lower limit, respectively. □

The approximations from (4.7) are again quite suitable to make the behavior of the correction transparent. The first term in both k_1 and k_2 is positive and typically dominates the second one, which is negative. The approximation k_1 still uses r itself and as such nicely follows the slightly oscillatory behavior of the exact k as a function of n. Consequently, k_1 is very accurate. The approximation k_2 ignores this effect and thus is slightly less accurate, but still amply adequate. As it has the very simple form $k_2 = An^{1/2} - Bn$, it trivially follows that for approximately $n = n_{max} = \{A/(2B)\}^2$ the approximately maximal correction $k_{max} = Bn_{max}$ occurs. Likewise, the correction becomes about 0, and thus superfluous, for $n \approx 4n_{max}$. Plugging the actual values of A and B from (4.7) into

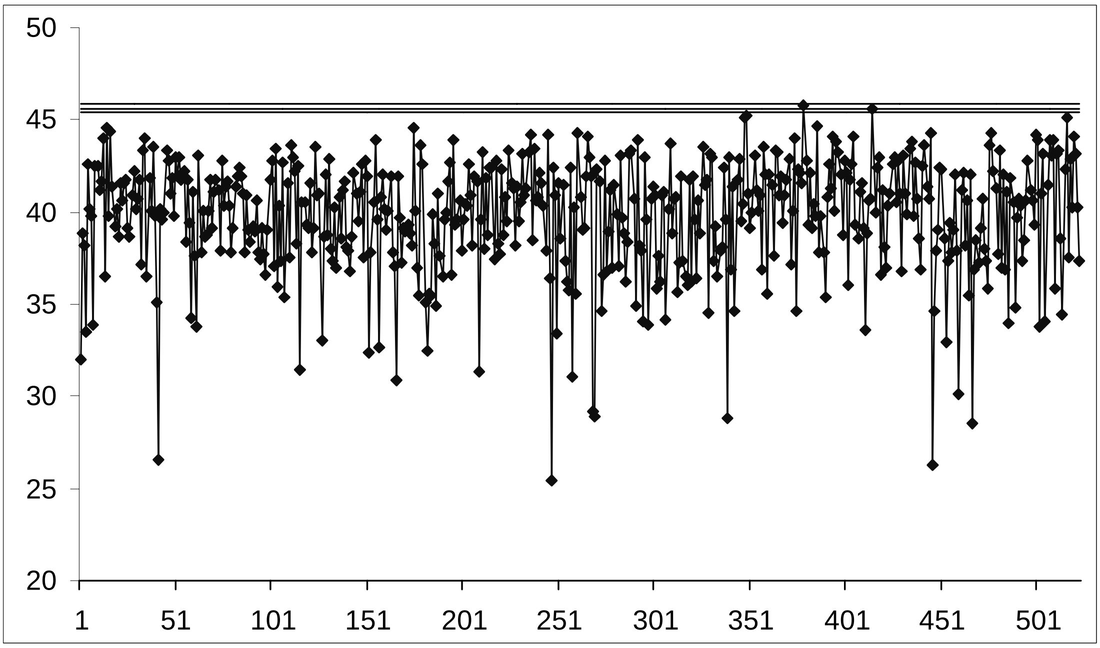

Figure 1: *MIN-chart with* $\widehat{UL} = 45.45$ *(uncorrected),* $\widehat{UL} = 45.57$ *(bias corrected) and* $\widehat{UL} = 45.86$ *(exceedance corrected).*

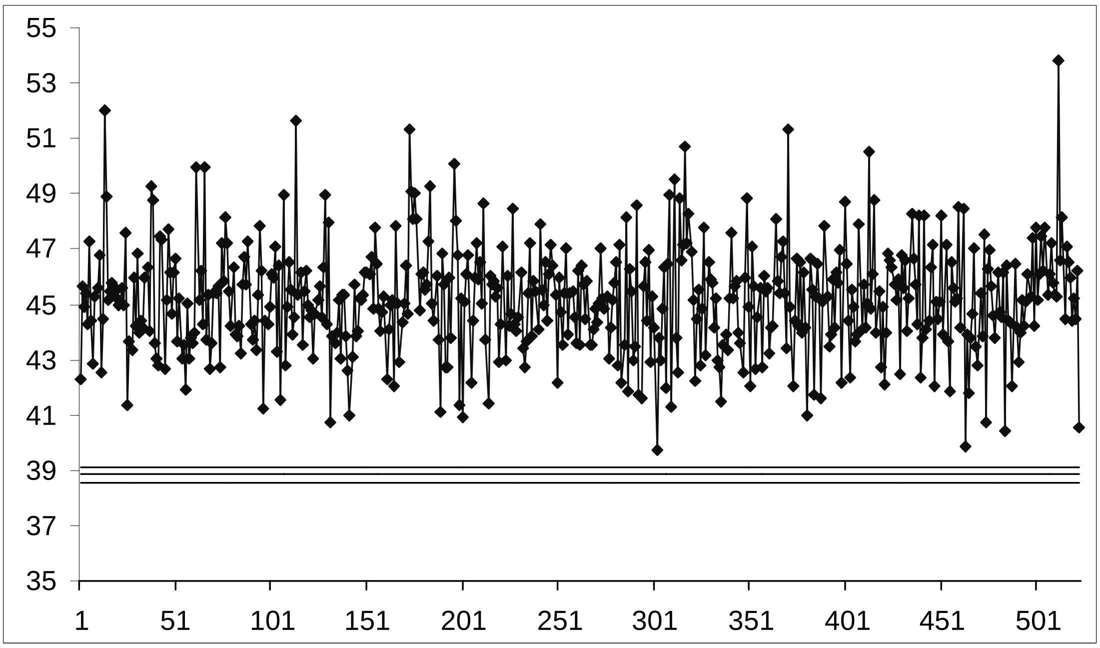

Figure 2: *MIN-chart with* $\widehat{LL} = 39.14$ *(uncorrected),* $\widehat{LL} = 38.87$ *(bias corrected) and* $\widehat{LL} = 38.55$ *(exceedance corrected).*

these results, we obtain using (4.2) and (4.4) that

$$n_{max} = \left(\frac{mu_\alpha}{2\varepsilon}\right)^2 \frac{1-q}{q} = \left(\frac{u_\alpha}{2\varepsilon h(p,m)}\right)^2, \text{ while } k_{max} = \frac{mu_\alpha^2(1-q)}{4\varepsilon}. \tag{4.8}$$

Example 4.5. Continuing once more with the values $p = 0.001$, $\varepsilon = 0.2$ and $\alpha = 0.2$, we obtain for $m = 2$, 3, 4 and 5 the following outcomes for (n_{max}, k_{max}): (378, 1.69), (236, 2.27), (211, 2.65) and (209, 2.89), respectively. Hence, for the present ε and α, a shift of about 2 to 3 order statistics is as bad as it gets. The effect of changing the parameters ε and α is immediate from (4.8): e.g. in k_{max} just recalculate u_α^2/ε. □

Remark 4.2. Also note that the correction approximately replaces r by $r(1 + \varepsilon/m) - u_\alpha\{r(1-r/n)\}^{1/2}$. Clearly, this is a more prominent change than in the bias case, where r is replaced by approximately $r - m/2$. This is in line with the fact that in the exceedance case we aim at controlling the behavior of the chart for each individual application, rather than just during a long series of separate applications. □

5 Impact on the Out-of-control behavior

For both the bias and the exceedance case we now have corrected versions which meet the specifications given. The remaining question is to what extent these corrections affect the OoC behavior: the price for controlling the IC behavior of the chart should be acceptable in terms of loss of detection power during OoC. Of course, we should realize that the notion 'price' is somewhat diffuse here and thus should be viewed from the right perspective. In fact, the corrected charts can be compared to two alternatives. The first is the chart for known F, the second the uncorrected proposal from (2.3).

As concerns the comparison to the case of a known underlying distribution, the situation is in principle quite straightforward. No real choice seems to be involved: if one knows F, of course one should use the chart for this fortunate situation. But in general one does not, and hence then there is no way to avoid paying the estimation price. However, as should be clear from the ample discussion in the Introduction, there is a bit more to it than just this blunt observation. Actually, there is a two-step procedure involved. The first step consists of selecting alternative proposals to the standard Shewhart chart which have competitive detection power during OoC under known F. From among these, in the second step a further selection is made by looking at the price to be paid for the estimation part, when going to the more typical situation of an unknown underlying distribution. For example, in the Introduction is was mentioned that the at first sight more obvious choice MAX was already eliminated during the first step, because of its poor power properties for known F. Surviving contenders to the individual Shewhart chart after the first round were AVE and MIN. The situation of AVE in the nonparametric setup was analyzed in AK (2004e) and found to be rather unrewarding: complicated and apparently no substantial reduction of estimation price as m increases. Hence only MIN remains and by now we have demonstrated in sections 3 and 4 that there estimation aspects can be dealt with in a relatively simple and exact manner. But indeed it remains interesting to look at the price involved, to make the appraisal of this alternative complete.

The second comparison concerns that of the corrected and the uncorrected version. Here both charts operate under the same condition of an unknown underlying distribution. Once again one could argue that no real choice is involved: the uncorrected chart simply ignores the impact of the estimation step during IC. If one wants to control these effects, at this point as well there is no way to avoid paying the estimation price. But once again, it is of some interest to figure out how much this protection actually costs. For example, if the price is acceptable, it becomes even less reasonable to keep ignoring the estimation effects by merely plugging in the estimated values and thus using the uncorrected chart.

Having put the idea of 'price' into the proper perspective, we now turn to its actual evaluation. As concerns the OoC situation, we shall focus on the case where $Y_1, \ldots, Y_m$ come from a shifted df $F(x-\Delta)$, where Δ is such that $p_\Delta = \{\overline{F}(\overline{F}^{-1}(q)-\Delta))\}^m/m$, with q as in (4.2), may be small, but not extremely so, like p. Just as in section 2, let $P_{n,\Delta}$ be the stochastic counterpart of p_Δ when using the uncorrected $\widehat{UL}$ from (2.3). Likewise, write $P_{n,\Delta}(k,\lambda)$ in case we use $\widehat{UL}(k,\lambda)$ from (3.5) for certain k and λ (e.g. from Lemma 3.1 or Lemma 4.2). The effect applying a correction can now be expressed in terms of the relative change RC, where

$$RC = \frac{EP_{n,\Delta} - EP_{n,\Delta}(k,\lambda)}{EP_{n,\Delta}}. \tag{5.1}$$

We have the following result.

Lemma 5.1. *Let $g = g_F = f/\overline{F}$, then RC from* (5.1) *approximately equals*

$$\frac{m(k+1-\lambda)}{r} \frac{g(\overline{F}^{-1}(q)-\Delta)}{g(\overline{F}^{-1}(q))}. \tag{5.2}$$

Proof. Because $p_\Delta = \{\overline{F}(\overline{F}^{-1}(q)-\Delta)\}^m/m$, unlike p itself, is not very small, in this case the relative error in $EP_{n,\Delta}$ is small and we can approximate this latter quantity by p_Δ. (Hence in this approximation step we essentially replace $\overline{F}_n^{-1}$ simply by $\overline{F}^{-1}$.) From e.g. the proof of Lemma 4.2 it is clear that replacing $\widehat{UL}$ from (2.3) by $\widehat{UL}(k,1)$ or $\widehat{UL}(k,0)$ (cf. (3.5)) implies replacing r by $r-k$ or $r-k-1$, respectively. As q is $(r+\delta)/n$, this means that e.g. $EP_{n,\Delta}(k,1) \approx \{\overline{F}(\overline{F}^{-1}(q-k/n)-\Delta)\}^m/m$. Application of a standard Taylor type argument to (5.1) then shows in a second approximation step that $RC \approx \eta m f(\overline{F}^{-1}(q)-\Delta))/\{f(\overline{F}^{-1}(q))\overline{F}(\overline{F}^{-1}(q)-\Delta)\}$, where $\eta = (k+1-\lambda)/n$. By noting that $\eta = \{(k+1-\lambda)/r\}(r/n) \approx \{(k+1-\lambda)/r\}(\overline{F}(\overline{F}^{-1}(q))$ and using the notation $g = f/\overline{F}$, the result follows. □

Remark 5.1. Due to the discrete steps in r as function of n, the approximations remain somewhat 'jumpy'. Actually, a completely similar argument as in the proof of Lemma 5.1

shows that (5.2) remains just as valid with r replaced by $r+1$. (Just observe that the fact that $P_{n,\Delta} \cong \{\overline{F}(\overline{F}^{-1}(U_{(r+1)}) - \Delta)\}^m/m$, while $EU_{(r+1)} = q^* = (r+1)/(n+1)$, suggests to use $p^*_\Delta = \{\overline{F}(\overline{F}^{-1}(q^*) - \Delta)\}^m/m$ rather than p_Δ as a starting point.) Especially in the case where $m = 1$, we deal with small values of $r = [n(mp)^{1/m}]$, even if n is very large, and using either r or $r+1$ makes a considerable difference. Of course, this merely illustrates the coarseness of the approximations for the individual case (cf. AK (2004c)). □

Remark 5.2. To analyze the behavior of RC, it is useful to observe that the expression in (5.2) consists of two factors. The first one equals $m(k+1-\lambda)/r$ and is independent of Δ. Hence this factor could be viewed as RC for the IC case, obtained by letting $\Delta = 0$ in (5.2). In the bias case, the average value for $k+1-\lambda$ is $m/2$ (cf. e.g. Remark 4.2) and thus this $RC \approx m^2/(2r)$. On the other hand, from (5.1) it is clear that for $\Delta = 0$ we actually have that $RC \approx EW_1$. Indeed the value $m^2/(2r)$ is nicely in line with the upper bound for EW_1 from (3.2). In Remark 4.2 it was moreover observed that for the exceedance case k will typically be somewhat larger, as this correction is more stringent. Note that this implies that the thus corrected chart will have a slight negative bias, required to bring the exceedance probability down to level α. Using k_2 from (4.7), it is immediate that then

$$RC \approx m u_\alpha \left\{\frac{1}{r} - \frac{1}{n}\right\}^{1/2} - \varepsilon. \tag{5.3}$$

The interpretation of (5.3) is straightforward. In the limit, $P(P_n > p) = 1/2$ and indeed $RC = 0$ for $\alpha = 1/2$ and $\varepsilon = 0$. Likewise, to obtain the even more liberal $P(P_n > p(1+\varepsilon)) \geq 1/2$, the negative value $RC = -\varepsilon$ occurs. But typically $\alpha < 1/2$ and the u_α-term in (5.3) will dominate, leading to a positive RC. □

For $\Delta > 0$, we of course need to consider the ratio of the hazard rates g in (5.2) as well. Fortunately this is a rather well studied type of function. For example, in the normal case, $1/g = \overline{\Phi}/\varphi$, which is known as Mills' ratio. Typically, g will be increasing and thus the ratio of g's in (5.2) will be smaller than 1 for $\Delta > 0$. Hence this ratio can be viewed as the reduction factor which appears when going from IC to OoC. Moreover, it is decreasing in Δ, and thus in p_Δ: indeed the estimation effects become smaller as the signal probability becomes less extreme. It is also clear that this reduction factor will depend on the actual underlying distribution F. Just us in nonparametric testing, the notion 'nonparametric' is connected to the IC behavior (cf. the null hypothesis case). The performance during OoC (cf. the power under the alternative hypothesis) does depend on F again. To illustrate these general observations we now consider an example of a specific F.

Example 5.1. In the normal case g is indeed increasing. Actually, it is well known that $g(x)$ behaves like x for large x. A slight adaptation is already obtained (cf. AK (2004c)) by using $4(1+x)/5$, which works well for $0 \leq x \leq 3.1$. However, in view of the fact that now m will typically be larger than 1, we have to reckon with negative x as well in the present application. To be more specific, let us for example assume that p_Δ attains values between

0.05 and 0.1, corresponding to ATS between 10 and 200. As $p_\Delta = \{\overline{\Phi}(u_q - \Delta)\}^m/m$, we have that $u_q - \Delta = u_v$, where $v = (mp_\Delta)^{1/m}$. Note that for the extreme case $m = 5$ and $p = 0.1$ we obtain $v = 0.5^{1/5} = 0.87$, with $u_v = -1.13$. It turns out that a quadratic approximation like $(10 + 7x + x^2)/12$ is adequate for $-2 < x < 3.1$, which amply covers our region of interest. Hence, in this way (5.2) for $F = \Phi$ produces the further, even more simple, approximation

$$RC \approx \frac{m(k+1-\lambda)}{r} \frac{10 + 7u_v + u_v^2}{10 + 7u_q + u_q^2}. \tag{5.4}$$

The first of these two factors thus equals $m^2/(2r)$ for the bias case and $mu_\alpha\{1/r - 1/n\}^{1/2} - \varepsilon$ for the exceedance case (see (5.3)), whereas the second factor represents the reduction of RC due to going to the OoC situation, i.e. letting Δ be positive. □

Clearly, the above is an example in the sense that a specific choice for F is singled out. The next step is to consider some specific values as a further illustration.

Example 5.2. We continue with Example 5.1, in combination with Example 4.4. Hence once more we let $p = 0.001$, $n = 100$, $m = 3$, $\varepsilon = 0.2$ and $\alpha = 0.2$. According to Remark 5.2, for $\Delta = 0$ the bias case $RC \approx m^2/(2r) = 0.32$, while according to (5.3) the exceedance case $RC \approx mu_\alpha\{1/r - 1/n\}^{1/2} - \varepsilon = 0.43$. Next we go to the OoC situation, where $\Delta > 0$. Choose $\Delta = u_q - u_v$ such that $p_\Delta = 0.05$, then the ratio of g's from (5.2) for $F = \Phi$ equals 0.48. Hence this reduction factor brings the bias case RC down to 0.15 and the exceedance case RC to 0.20. In view of (5.1), $EP_{n,\Delta}(k, \lambda) = (1 - RC)EP_{n,\Delta}$, which means that $EP_{n,\Delta}$ is lowered by applying the two types of corrections from about $p_\Delta = 0.05$ to 0.042 and 0.040, respectively. In other words, ATS goes from 20 to 23.6 and 25.1, respectively, which looks quite acceptable. This conclusion also seems to hold for other group sizes: for $m = 2$, 3, 4 and 5, the reduction factor equals 0.53, 0.48, 0.43 and 0.39, respectively. (If the quadratic approximation from (5.4) is used instead, these values become 0.55, 0.51, 0.47 and 0.43, respectively.

6 Additional explanation

In the last sections we have repeatedly noted that going from $m = 1$ to $m > 1$ indeed provides a very substantial improvement (see e.g. Example 4.1). Apparently this step greatly reduces the variability in W from (2.7) and brings the SE down to acceptable proportions. Here we shall attempt to shed some more light on this phenomenon. Typically, the corrections we have introduced consist of widening the control limits somewhat, by shifting to slightly more extreme order statistics. We can also try to interpret this as replacing the original p by a slightly smaller value, say $p(1 - \kappa)$ for some small κ. As $r = [n(mp)^{1/m}]$, this then approximately boils down to replacing r by the reduced value $r(1 - \kappa/m)$. Denote the corresponding relative error W in this situation by $W(\kappa) = P_n(\kappa)/p - 1$, where $P_n(\kappa) = \{\overline{F}(X_{(n-r(1-\kappa/m))})\}^m/m$. In section 4 it was remarked that, when dealing with exceedance probabilities, we can immediately reduce the situation for W_2 to that for W_1. Therefore we shall again simply restrict attention to the latter case.

It follows from (2.6) and (2.7) that

$$W_1(\kappa) \cong \left\{ \frac{nU_{(r(1-\kappa/m)+1)}}{r+\delta} \right\}^m - 1. \tag{6.1}$$

Hence to first order $EW_1(\kappa) = -\kappa$, which means that the expectation of W_1 is reduced by about κ by changing r into $r(1-\kappa/m)$. (Actually this is rather obvious: replacing p by $p(1-\kappa)$ should indeed lower the average of $P_n/p - 1$ by this same κ.) Consequently, for $m > 1$ use that $P(W_1(\kappa) > \varepsilon) = P(W_1(\kappa) + \kappa > \varepsilon + \kappa) \approx \overline{\Phi}((\varepsilon + \kappa)/\sigma_{W_1})$, which will equal α for

$$\kappa \approx u_\alpha \sigma_{W_1} - \varepsilon. \tag{6.2}$$

To see the connection with the results from sections 4 and 5, just note that var $(W_1) \approx m^2\{1/r - 1/n\}$, and thus $\kappa \approx u_\alpha m\{1/r - 1/n\}^{1/2} - \varepsilon$, which indeed agrees with $r\kappa/m = k_2 \approx u_\alpha\{r(1-r/n)\}^{1/2} - \varepsilon r/m$ (cf. (4.7)) and moreover equals RC from (5.3).

In view of (6.2), the correction κ is governed by

$$\sigma_{W_1} \approx m r^{-1/2} \approx m(mp)^{-1/(2m)} n^{-1/2}. \tag{6.3}$$

Incidentally, observe that for $m = 1$ the normal approximation leading to (6.2) may only be of use for n huge, but that nevertheless the exact $\sigma_{W_1} = \{\text{var } (U_{(r+1)}/p)\}^{1/2}$ is readily available here for comparison to the situation where $m > 1$ (cf. (14) in AK (2004c)). In fact, for $m = 1$,

$$\sigma_{W_1} = \frac{\{(n-r)(r+1)\}^{1/2}}{(n+1)(n+2)^{1/2}p} \approx \frac{(r+1)^{1/2}}{np}, \tag{6.4}$$

which for huge n indeed also behaves like $r^{-1/2}$ or $(np)^{-1/2}$, but for $r = 0$ like $(np)^{-1}$. By way of illustration, we present an example.

Example 6.1. Let $p = 0.001$ and $n = 100$. For $m = 1$, 2, 3, 4 and 5, we have $r = 0$, 4, 14, 25 and 34, respectively, leading to $\sigma_{W_1} = 9.80 \approx 10$ (from (6.4)), 1, 0.80, 0.80 and 0.86 (from (6.3)), respectively. Indeed, the drop in going from $m = 1$ to $m > 1$ is spectacular.□

The above is not only useful for showing the relation between the standard deviation σ_{W_1} from (6.3) and the correction κ from (6.2), but also for making the connection to e.g. the normal chart. This latter chart is based on the sample mean (AVE) and works under the assumption of normality, i.e. $F(x) = \Phi((x-\mu)/\sigma)$. It gives a signal if $\overline{Y} = m^{-1}\Sigma_{i=1}^m Y_i > \widehat{UL} = \hat{\mu} + m^{-1/2}\hat{\sigma}u_{mp}$, where $\hat{\mu}$ and $\hat{\sigma}$ are e.g. the sample mean and sample standard deviation, respectively (cf. (2.3)). For this situation the stochastic version (cf. (2.6)) of p is $P_n^* = \overline{\Phi}(u_{mp} + V^*)/m$, with $V^* = m^{1/2}(\hat{\mu} - \mu)/\sigma + u_{mp}\{(\hat{\sigma}/\sigma) - 1\}$. Let W^* be the corresponding relative error (cf. (2.7)), and in analogy to the above, let $W^*(\kappa^*)$ be this error for the case where mp is replaced by $mp(1-\kappa^*)$ (and hence also in V^*). Hence (6.1) now becomes

$$W_1^*(\kappa^*) = \frac{\overline{\Phi}(u_{mp(1-\kappa^*)} + V^*)}{mp} - 1. \tag{6.5}$$

As $u_{mp(1-\kappa^*)} \approx u_{mp} + c$ with $c = \kappa^* mp/\varphi(u_{mp})$, it is again straightforward to verify that also $EW_1^*(\kappa^*) \approx -\kappa^*$ and thus that $\kappa^* = u_\alpha \sigma_{W_1^*} - \varepsilon$ (cf. (6.2)) will produce approximately $P(W_1^*(\kappa^*) > \varepsilon) = \alpha$.

By using a one-step Taylor expansion in (6.5) it is clear that $\sigma_{W_1^*} \approx \sigma_{V^*}\varphi(u_{mp})/(mp)$. Using the fact that $\sigma_{V^*}^2 \approx (u_{mp}^2 + 2m)/(2n)$, whereas $\varphi(u_{mp})/(mp) = g_\Phi(u_{mp})$, we arrive at

$$\sigma_{W_1^*} = g_\Phi(u_{mp})\left\{\frac{u_{mp}^2 + 2m}{2n}\right\}^{1/2}. \tag{6.6}$$

In passing observe that the above mentioned correction c to u_{mp} thus equals $\kappa^*/g_\Phi(u_{mp}) = u_\alpha\{(u_{mp}^2 + 2m)/(2n)\}^{1/2} - \varepsilon/g_\Phi(u_{mp})$, which agrees with the result for $m = 1$ in AK (2004b) when using the further approximation $g_\Phi(u_{mp}) \approx u_{mp}$. More important, however, is that we can now compare (6.4) and (6.6). Again consider an example:

Example 6.2. As in Example 6.1, let $p = 0.001$ and $n = 100$. For $m = 1$, 2, 3, 4 and 5, we now obtain $\sigma_{W_1^*} = 0.81$, 0.79, 0.79, 0.81 and 0.83, respectively. Here we just have a stable pattern. Consequently, while for $m = 1$ the nonparametric chart has a dramatically larger variability than the normal one (a coefficient 9.80 as opposed to 0.81), the difference almost vanishes as soon as $m > 1$ (e.g. 1 versus 0.79, 0.80 versus 0.79, etc.). Note that this observation does not really depend on the particular choice of n we have used. In fact, let e.g. $m = 3$, then (cf. (6.3)) $m(mp)^{-1/(2m)} = 7.9$ and thus $\sigma_{W_1} \approx 7.9n^{-1/2}$, while (6.6) produces $\sigma_{W_1^*} \approx 7.9n^{-1/2}$ as well. □

Hence the pleasant conclusion of the above is that going to grouped observations and subsequently applying the MIN-chart has indeed reduced the SE to a large extent. In fact, the variability has even become comparable to that of the usual normal chart. Consequently, the disadvantage of working nonparametrically seems to have been remedied successfully. On the other hand, the distinct advantage of MIN is still unaffected: it completely avoids making a ME. The normal chart, which has been operating in the above examples under normality only, can easily go very wrong as soon as we leave this normality behind. That also holds for the grouped case, with $m = 3$ or even $m = 5$. Such numbers are still too small to safely rely on the CLT, especially as we are dealing with tail probabilities. If desired, see AK (2004d), section 4, for specific examples.

References

Albers, W. and Kallenberg, W.C.M. (2003). New corrections for old control charts. Technical Report 1694, University of Twente. To appear in *Quality Engineering.*

Albers, W. and Kallenberg, W.C.M. (2004a). Estimation in Shewhart control charts: effects and corrections. *Metrika* **59**, 207-234.

Albers, W. and Kallenberg, W.C.M. (2004b). Are estimated control charts in control? *Statistics* **38**, 67-79.

Albers, W. and Kallenberg, W.C.M. (2004c). Empirical nonparametric control charts: estimation effects and corrections. *J. Appl. Statist.* **31**, 345-360.

Albers, W. and Kallenberg, W.C.M. (2004d). Alternative Shewhart-type charts for grouped observations. Technical Report 1717, University of Twente.

Albers, W. and Kallenberg, W.C.M. (2004e). Tail behavior of the empirical distribution function of convolutions. Technical Report 1740, University of Twente.

Albers, W., Kallenberg, W.C.M. and Nurdiati, S. (2002). Exceedance probabilities for parametric control charts. Technical Report 1650, University of Twente.

Albers, W., Kallenberg, W.C.M. and Nurdiati, S. (2004a). Parametric control charts. *J. Statist.Plann. Inference* **124**, 159-184.

Albers, W., Kallenberg, W.C.M. and Nurdiati, S. (2004b). Data driven choice of control charts. To appear in *J. Statist. Plann. Inference.*

Alloway, J.A. and Raghavachari, M. (1991). Control chart based on the Hodges-Lehmann estimator. *J. Qual. Technol.* **23**, 336-347.

Bakir, S.T. and Reynolds, M.R.Jr. (1979). A nonparametric procedure for process control based on within-group ranking. *Technometrics* **21**, 175-183.

Chakraborti, S. (2000). Run length, average run length and false alarm rate of Shewhart $\overline{X}$ chart: exact derivations by conditioning. *Commun. Statist. Simul. Comput.* **29**, 61-81.

Chakraborti, S., van der Laan, P. and Bakir, S.T. (2001). Nonparametric control charts: an overview and some results. *J. Qual. Technol.* **33**, 304-315.

Chan, L.K., Hapuarachchi, K.P. and Macpherson, B.D. (1988). Robustness of $\overline{X}$ and R charts. *IEEE Trans. Reliability* **37**, 117-123.

Chen, G. (1997). The mean and standard deviation of the run length of $\overline{X}$ charts when control limits are estimated. *Statist. Sinica* **7**, 789-798.

Ghosh, B.K., Reynolds, M.R.Jr. and Hui, Y.V. (1981). Shewhart $\overline{X}$-charts with estimated process variance. *Commun. Statist. Theory Methods* **10**, 1797-1822.

Hackl P. and Ledolter J. (1991). A control chart based on ranks. *J. Qual. Technol.* **23**, 117-124.

Hackl, P. and Ledolter, J. (1992). A new nonparametric quality control technique. *Commun. Statist. Simul. Comput.* **21**, 423-443.

Ion, R.A., Does, R.J.M.M. and Klaassen, C.A.J. (2000). A comparison of Shewhart control charts based on normality, nonparametrics, and extreme-value theory. Report 00-8, University of Amsterdam.

Pappanastos, E.A. and Adams, B.M. (1996). Alternative designs of the Hodges-Lehmann control chart. *J. Qual. Technol.* **28**, 213-223.

Quesenberry , C.P. (1993). The effect of the sample size on estimated limits for $\overline{X}$ and X control charts. *J. Qual. Technol.* **25**, 237-247.

Roes, C.B. (1995). *Shewhart-type Charts in Statistical Process Control*. Ph.D.-thesis, University of Amsterdam.

Willemain, T.R. and Runger, G.C. (1996). Designing control charts using an empirical reference distribution. *J. Qual. Technol.* **28**, 31-38.

Woodall, W.H. and Montgomery, D.C. (1999). Research issues and ideas in statistical process control. *J. Qual. Technol.* **31**, 376-386.