New Corrections for Old Control Charts — Source link

Willem Albers, Wilbert C.M. Kallenberg

Institutions: University of Twente

Published on: 01 Jul 2005 - Quality Engineering (University of Twente, Department of Applied Mathematics)

Topics: Shewhart individuals control chart, Chart, Control chart and Statistical process control

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. 1694

New corrections for old control charts

W. Albers and W.C.M. Kallenberg

December, 2003

ISSN 0169-2690

New corrections for old control charts

Willem Albers* and Wilbert C.M. Kallenberg

Faculty of Electrical Engineering, Mathematics and Computer Science
Department of Applied Mathematics
University of Twente
P.O. Box 217, 7500 AE Enschede
The Netherlands

Abstract When using standard control charts, typically several parameters need to be estimated. For the usual sample sizes, this is known to affect the performance of the chart. Here we present simple corrections to solve this problem. As a basis we use existing factors which are widely used for the traditional charts. The advantage of the new proposals is that a clear link is made to the actual performance characteristics of the chart.

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

2000 Mathematics Subject Classification: 62 F 12, 62 F 25, 62 P 30.

1 Introduction

Application of standard control charts almost always requires estimation of the underlying mean and standard deviation. The impact of this step is larger than usually thought (see e.g. Montgomery and Woodall (1999), p. 379). Hence it seems worthwhile to investigate how the performance of the charts is affected, and what can be done about it. This task has been executed in some detail in the first part of a series of papers by Albers and Kallenberg (2004a, b, c) and by Albers, Kallenberg and Nurdiati (2002, 2003, 2004). In the present paper we give a non-technical review of this first part, in which the classical assumption of normality still holds. Moreover, here we devote attention to an aspect we ignored till now. In the papers quoted we merely used the sample mean and the sample standard deviation as estimators. However, in practice quite a few other choices

*Corresponding author: tel: +31534893816; fax: +31534893069; e-mail: w.albers@math.utwente.nl.
This research was supported by the Technology Foundation STW, applied science division of NWO and the technology programme of the Ministry of Economic Affairs.

occur which are used (even more) widely (see e.g. Does and Schriever (1992) and Roes (1995)). Extension of our previous results to these cases is relatively simple, but substantially widens their applicability.

In section 2 the problem is described, while in the next section possible ways to solve it are discussed. These solutions come in the form of suitable correction factors. Unlike the traditional ones, such as c_4 and d_2, these factors present a clear link to the control chart performance. Section 4 aims at bias removal, not for the estimators themselves, but for the performance of the control chart as a whole. Mild corrections are presented which ensure that the behavior of the control chart is acceptable on the average (i.e. in a long series of applications). In section 5 explicit examples in this category are given for the various types of estimators used in practice. Next more stringent factors are introduced, again meant to correct the behavior of the chart, but now with respect to a single application. Such exceedance probability corrections are the subject of section 6 and the corresponding examples for this situation are presented in section 7. Finally, in section 8 the conclusions are presented.

2 The problem

We 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 given and an out-of-control signal is produced as soon as an observation falls outside these limits. The observations are supposed to come from a normal distribution with mean μ and standard deviation σ. This leads for example to the common '3σ'-limits: $UL = \mu + 3\sigma$ and $LL = \mu - 3\sigma$. As long as the process is in control, a false alarm will then occur on the average once every 370 observations. If desired, this 370 can be replaced by the 'nice' value 500 by using a factor 3.09 rather than just 3 in front of σ. More in general, if we choose some very small p, an average run length (ARL) $1/p$ will result if the factor used is chosen as $u_{p/2}$, which is the point that is exceeded by a standard normal variable with probability $p/2$. (More technically, $1 - \Phi(u_{p/2}) = p/2$, where Φ is the standard normal distribution function.) Hence for example $u_{1/740} = 3$ and $u_{0.001} = 3.09$.

Typically the parameters μ and σ are unknown and a sample of n so-called Phase I observations $X_1, \ldots, X_n$ is used to obtain estimated limits. E.g. we then have as '3σ'-limits

$$\widehat{UL} = \hat{\mu} + 3\hat{\sigma} \text{ and } \widehat{LL} = \hat{\mu} - 3\hat{\sigma}, \qquad (2.1)$$

where typically $\hat{\mu} = \overline{X}$ and for example $\hat{\sigma} = S$, with $S^2 = \sum(X_i - \overline{X})^2/(n-1)$. These estimated limits are then used during Phase II, the monitoring phase. The problem now is that this estimation step affects the performance of the resulting control chart and that it will take at least $n = 300$ before such effects can be safely ignored (see e.g. Quesenberry (1993) and Roes (1995), chapter 2). But in practice one usually deals with smaller samples and thus the impact of using control limits which are in fact random variables should be studied (cf. again Woodall and Montgomery (1999), p. 379).

3 The approach

To figure out what we are going to do about the problem, several steps need to be taken. First of all, we have to decide what characteristics of the chart are important to us. Next we need to determine how to measure possibly unpleasant behavior in terms of these characteristics. Then we can start looking for ways to correct matters.

As concerns the characteristics, from section 2 we see that of interest are the probability p (e.g. $\frac{1}{370}$ or $\frac{1}{500}$) of having a false alarm during in-control and, even more so, the corresponding ARL, which equals $1/p$ (e.g. 370 or 500). Note that through the estimation step (2.1) these fixed quantities are replaced by random variables. In fact, let X_{n+1} be a new observation from Phase II, then the false alarm rate (FAR) becomes the conditional probability

$$P = Pr(X_{n+1} > \widehat{UL} \text{ or } X_{n+1} < \widehat{LL} | (X_1, \ldots, X_n)). \tag{3.1}$$

Hence although we simply write P, we should bear in mind that actually $P = P(\hat{\mu}, \hat{\sigma})$. For instance, when taking $\widehat{UL}$ and $\widehat{LL}$ from (2.1), we get $P(\hat{\mu}, \hat{\sigma}) = 1 - \Phi((\hat{\mu} - \mu)/\sigma + 3(\hat{\sigma}/\sigma)) + \Phi((\hat{\mu} - \mu)/\sigma - 3(\hat{\sigma}/\sigma))$. If the Phase I observations now happen to produce somewhat 'unlucky' values of $\hat{\mu}$ and $\hat{\sigma}$, we may for example wind up with an outcome p^* of P which is much larger than the intended p. Hence the $ARL = 1/p^*$ for this application of the control chart will also fall short of the expected $1/p$ to an unpleasant extent. The chart is then likely to produce a false alarm much sooner than prescribed.

Consequently, unpleasant behavior means a P which tends to produce outcomes too far away from the intended p. As P is random, there is no unique way to express the discrepancy. Two popular approaches are as follows. The first is the mildest one: look at the expected value EP and compare it to p. If the bias $(EP - p)$ is sufficiently small, the chart is considered to be O.K. Similarly, we may compare $E(1/P)$ with $1/p$. This is not equivalent to comparing EP with p, but belongs to the same category of a bias criterion. Note that this is satisfactory if we are interested in the behavior over a considerable period, i.e. we consider a long series of applications of the chart. However, suppose the bias of P is indeed small, but its variability is still large. Then the average behavior may be more or less O.K., but very long runs are mixed with very short ones. From the point of view of a single application of the chart, this is still unfortunate. So then the second criterion is more appropriate: look at the probability that P exceeds p by more than a given percentage (for example 25). If this exceedance probability is sufficiently small (e.g. 20%), the chart is O.K. in this more strict way.

As an example, let $p = 0.002$, so the intended $ARL = 500$. If we use the corresponding limits $\hat{\mu} \pm 3.09\hat{\sigma}$, we may e.g. want to require that EP is between 0.0016 and 0.0024. Similarly, we may ask that $E(1/P)$ is between 400 and 600. Alternatively, our condition can be that P exceeds 0.0025 in at most 20% of the realizations of the chart. Note that this is equivalent with requiring that the ARL is less than 400 in at most 20% of the cases. Hence by now it is clear what characteristics we are watching and how we would like these to behave. As P converges to p as $n \to \infty$, any wish in this respect can obviously be fulfilled by taking n sufficiently large. However, the convergence unfortunately is so slow that unrealistically large values for n result. Even for the mild bias criterion, $n \geq 300$ (cf. section 2) is typically needed, while the stricter exceedance probability approach sooner requires a value of n well over 500.

Hence the question is how to solve this problem. Our suggestion is to apply suitable corrections to the chart. In this way, acceptable behavior can be attained for common sample sizes. Of course, such corrections should be easy to understand and to implement. A particularly simple approach is to adapt the choice of the factor in front of $\hat{\sigma}$ in the control limits. An additional advantage is that this kind of operation is already in common use. Above we mentioned the possible replacement of 3 by 3.09. Moreover, instead of simply using $\hat{\sigma} = S$, it is also quite common to use a corrected version like $\hat{\sigma} = S/c_4$. However, the corrections we propose are superior in the sense that these offer a direct link with the desired improvement of control chart performance.

4 The bias case: general

It is customary to use $\hat{\mu} = \overline{X}$ as an unbiased estimator for μ, which we shall do as well. For $\hat{\sigma}$, a variety of choices is available. A rather common feature among these choices is that a correction factor is applied to transform an arbitrary $\hat{\sigma}$ into an unbiased estimator σ^* of σ. Obviously, for that purpose just let

$$\sigma^* = \frac{\hat{\sigma}}{A}, \text{ with } A = \frac{E\hat{\sigma}}{\sigma}. \tag{4.1}$$

For example, $\hat{\sigma} = S$ becomes $\sigma^* = S/c_4$, where $c_4 = (ES)/\sigma$. However, (4.1) really is a rather halfhearted solution, since in this way only the bias $\widehat{UL}$ is removed. But that is no purpose in itself. As we saw above, what counts is removal of the bias in P or in $1/P$.

To achieve this to high precision, we propose limits of the following form

$$\text{one-sided: } \widehat{UL} = \overline{X} + u_p\sigma^*\left\{1 + \frac{B^{(1)}}{n}\right\} \text{ or } \widehat{LL} = \overline{X} - u_p\sigma^*\left\{1 + \frac{B^{(1)}}{n}\right\},$$

$$\text{two-sided: } \widehat{UL} = \overline{X} + u_{p/2}\sigma^*\left\{1 + \frac{B^{(2)}}{n}\right\} \text{ and } \widehat{LL} = \overline{X} - u_{p/2}\sigma^*\left\{1 + \frac{B^{(2)}}{n}\right\}. \tag{4.2}$$

The corrections $B^{(1)}$ and $B^{(2)}$ depend on the aim (FAR or ARL), on the estimator $\sigma^* = \hat{\sigma}/\{E\hat{\sigma}/\sigma\}$ and on p. However, they do not depend on n.

To simplify the presentation, we first consider the one-sided case, where originally an out-of-control signal is produced when a new observation is larger than $\widehat{UL} = \overline{X} + u_p\sigma^*$. Now a relatively simple computation essentially suffices. (Those not interested can safely skip it and go directly to the result in (4.5); on the other hand, those wanting to know more about the technical niceties ignored here are referred to Albers and Kallenberg (2004a)).

Introduce for a moment $\widehat{UL}_c = \overline{X} + u_p\sigma^*(1+c)$, where c is some small constant yet to be determined. Then (cf. (3.1)) $P = Pr(X_{n+1} > \widehat{UL}_c|(X_1,\ldots,X_n)) = 1 - \Phi((\overline{X}-\mu)/\sigma + u_p(1+c)\sigma^*/\sigma)$. Next, a two-step Taylor expansion of EP gives $EP - p \approx \phi(u_p)\{-cu_p + \frac{1}{2}u_p[\text{var}(\overline{X}/\sigma) + u_p^2\text{var}(\sigma^*/\sigma)]\}$, where $\phi = \Phi'$. Consequently, using (4.1) and the fact that $\text{var}(\overline{X}/\sigma) = 1/n$, it follows that the bias is removed best by letting $c = \frac{1}{2}[1/n + u_p^2\text{var}(\hat{\sigma}/E\hat{\sigma})]$. Finally, the resulting $u_p\sigma^*(1+c)$ approximately equals $u_p\sigma^*\{1+\frac{1}{2}(1+u_p^2\tau^2)/n\}$, where

$$\tau^2 = \lim_{n\to\infty}\left\{n\text{var}\left(\frac{\hat{\sigma}}{E\hat{\sigma}}\right)\right\}, \tag{4.3}$$

and consequently $B^{(1)}$ in (4.2) for the FAR is seen to satisfy

$$B^{(1)} = \tfrac{1}{2}[1 + u_p^2\tau^2]. \tag{4.4}$$

Hence to make P, rather than $\widehat{UL}$, unbiased, the upper limit has to be increased somewhat. This is in line with the observation that P tends to have a positive bias.

For $1/P$ a similar, slightly more complicated, computation leads to the result in (4.4) with the factor '$\frac{1}{2}$' replaced by '$\frac{1}{2}\{1 - 2\phi(u_p)/(pu_p)\}$'. Fortunately, as $\phi(x)/(1-\Phi(x)) \approx x$ for large x, this expression virtually equals '$-\frac{1}{2}$'. Hence for $B^{(1)}$ in case of the ARL we turn out to propose precisely the opposite from the $B^{(1)}$ for the FAR. This is in line with the well-known phenomenon that, contrary to simple intuition, both P and $1/P$ have a positive bias. The occurrence of (very) small values of P leads to (very) large values of $1/P$, which inflate the expectation. Also note that this situation illustrates the point that corrections should not be applied automatically, but only with a clear idea about what these are supposed to achieve.

As mentioned above, for simplicity of presentation we started with the one-sided case. To complete the picture, we now return to the two-sided case, for which the computations are a bit lengthier, but otherwise completely similar. For P the result is particularly straightforward: there $B^{(2)}$ just equals $B^{(1)}$ from (4.4) with $u_{p/2}$ rather than u_p. In case of $1/P$ the transition from the one-sided case to the two-sided one is slightly more complicated. However, the final result remains quite simple: take $B^{(1)}$ from (4.4), again replace u_p in (4.4) by $u_{p/2}$, but in addition replace the '+' by a '−' as well. In summary, our recommendation is to apply (4.2) with the B's according to the following table, with τ^2 as in (4.3),

Aim	One-sided case	Two-sided case
FAR	$B^{(1)} = \frac{1}{2}(1+u_p^2\tau^2)$	$B^{(2)} = \frac{1}{2}(1+u_{p/2}^2\tau^2)$
ARL	$B^{(1)} = -\frac{1}{2}(1+u_p^2\tau^2)$	$B^{(2)} = \frac{1}{2}(1-u_{p/2}^2\tau^2)$

(4.5)

Quite often one will compare the one-sided situation for some given value of p (e.g $p = 0.001$) to the two-sided case for twice that value, i.e. for $2p$. Note that then we simply have for the FAR that $B^{(2)} = B^{(1)}$, and thus the one- and two-sided limits in (4.2) become identical; for the ARL the minor modification $B^{(2)} = B^{(1)} + 1$ is all that is needed. For briefness' sake we will typically focus on this situation in the examples that follow.

5 The bias case: examples

It remains to make correction factors such as the ones in (4.1)-(4.5) completely explicit. This mainly requires specifying a choice for $\hat{\sigma}$. In the first few examples we shall consider the case where no grouping occurs. Hence we simple have one sample of consecutive individual observations $X_1, \ldots, X_n$.

Example 1. *Sample variance*: let $\hat{\sigma} = S$, with

$$S^2 = \frac{1}{(n-1)} \sum_{i=1}^{n} (X_i - \overline{X})^2. \tag{5.1}$$

From the point of view of efficiency, (5.1) is the best possible choice. Consequently, we will see that the corrections required in this example are smaller than those in the ones that follow. This reflects the fact that the variation of S is minimal and thus minimal adaptations suffice to take its effect into account.

Here A from (4.1) equals $c_4(n) = (ES)/\sigma$, which is widely used and tabulated. For the present case, it suffices to use that $c_4 = c_4(n) \approx 1 - 1/(4n)$. Moreover, for the choice under consideration, $\mathrm{var}(\hat{\sigma}/E\hat{\sigma}) \approx 1/(2n)$ and thus τ^2 from (4.3) becomes $\frac{1}{2}$. Hence, using again a certain value p in the one-sided case, together with $2p$ for the two-sided case, (4.5) reduces to

$$FAR: B^{(1)} = B^{(2)} = \tfrac{1}{4}(u_p^2 + 2),\ ARL: B^{(1)} = -\tfrac{1}{4}(u_p^2 + 2),\ B^{(2)} = -\tfrac{1}{4}(u_p^2 - 2). \tag{5.2}$$

As $\sigma^* = S/c_4 \approx S(1 + 1/(4n))$, it readily follows from (5.2) that e.g. for the FAR the limits in (4.2) now boil down to $\overline{X} \pm u_p S\{1 + (u_p^2 + 3)/(4n)\}$. If we select once more the traditional $u_p = 3$ (corresponding to $p = 0.00135$, $1/p = 740$), we thus arrive at the following very simple corrected version of the customary '3σ'-bounds (cf. (2.1))

$$\widehat{UL} = \overline{X} + 3S\left\{1 + \frac{3}{n}\right\} \text{ and } \widehat{LL} = \overline{X} - 3S\left\{1 + \frac{3}{n}\right\}. \tag{5.3}$$

For the ARL, the factor $\{1 + 3/n\}$ in view of (5.2) becomes $\{1 - 2.5/n\}$ and $\{1 - 1.5/n\}$ for the one- and the two-sided case, respectively.

The effects of $B^{(1)}$ for both P and $1/P$ in this example have been studied extensively in Albers and Kallenberg (2004a). The conclusion, supported by massive simulations, is that these corrections indeed work very well, producing acceptable small biases already for $n \geq 40$. Moreover, the effect on the out-of-control behavior is almost negligible. This is a major improvement compared to the generally advised $n \geq 300$ for the uncorrected case, especially as sample sizes in practice will quite often lie in the range (40,300). A final remark is that in Albers and Kallenberg (2004a) the proposals also have been illustrated by means of a real data example involving charge weights of an insecticide dispenser.

Example 2. *Moving Range*: let $\hat{\sigma} = MR$, with

$$MR = \frac{1}{n-1} \sum_{j=1}^{n-1} |X_{j+1} - X_j|. \tag{5.4}$$

Especially for the case of individual observations, (5.4) is the classical choice in practice. It is felt to be less sensitive than $\hat{\sigma} = S$ from (5.1) to trends. However, in the presence of serial correlation, the situation is reversed and S is considered to be more robust than MR.

For the case of the moving range A is given by the - also widely used - correction $d_2 = 2\pi^{(-1/2)} \approx 1.128$. Moreover, $\mathrm{var}(\hat{\sigma}/E\hat{\sigma}) \approx (\frac{2}{3}\pi + 3^{1/2} - 3)/n \approx 0.826/n$ (cf. e.g. Roes (1995), p. 54, but note that in the first line of the formula $[d_2(2)]^2$ must be omitted) and thus

$$\tau^2 = 0.826. \tag{5.5}$$

Clearly, using this value instead of the previous one, an analogue of (5.2) can be readily computed from (4.5). Taking the special case of $u_p = 3$ again, we obtain

$$FAR : B^{(1)} = B^{(2)} = 4.22, \quad ARL : B^{(1)} = -4.22, \quad B^{(2)} = -3.22, \tag{5.6}$$

giving e.g. for the FAR as corrected '3σ'-recipe the advice to use $\overline{X} \pm 2.66MR\{1+4.22/n\}$ (cf. (5.3)).

Example 3. *Interquartile Range*: let $\hat{\sigma} = IQR$, where

$$IQR = X_{(n+1-[n/4])} - X_{([n/4])}. \tag{5.7}$$

Here $X_{(1)}, \ldots, X_{(n)}$ are the order statistics of $X_1, \ldots, X_n$ and $[y]$ denotes the largest integer $\leq y$. This choice in (5.7) is another robust alternative to S.

Let Φ^{-1} denote the inverse of Φ, then its $A \approx \Phi^{-1}(\frac{3}{4}) - \Phi^{-1}(\frac{1}{4}) = 1.35$, while $\mathrm{var}(IQR) \approx 1/\{4n[\phi(\Phi^{-1}(\frac{3}{4}))]^2\} = 2.48/n$, leading to $\mathrm{var}(IQR/EIQR) \approx 1.36/n$. Hence

$$\tau^2 = 1.36, \tag{5.8}$$

which for $u_p = 3$ gives

$$FAR : B^{(1)} = B^{(2)} = 6.62, \quad ARL : B^{(1)} = -6.62, \quad B^{(2)} = -5.62. \tag{5.9}$$

Note that in going from Example 1 to Example 3, the values of τ^2 (cf. (5.5) and (5.8)), and thus the corrections (cf. (5.6) and (5.9)) indeed become larger, reflecting the fact that the estimators used for $\hat{\sigma}$ become less efficient.

Next we turn to the case of grouped variables. The initial sample of size n now actually consists of k sub-samples, each having size m (typically $m = 5$). Hence we have X_{ij}, with $i = 1, \ldots, k$ and $j = 1, \ldots, m$. For $\hat{\mu}$ we still use $\overline{X}$, which now can be written as

$k^{-1}\sum_{i=1}^{k}\overline{X}_i$, with $\overline{X}_i = (\sum_{j=1}^{m} X_{ij})/m$. As concerns $\hat{\sigma}$, besides $\hat{\sigma} = S$ from (5.1), several variants are used as well. These may even be more popular, among others to exclude trend effects, cf. the remarks under Example 2. We mention:

Example 4. *Within sample variance*: let $\hat{\sigma} = \tilde{S}$, with

$$\tilde{S}^2 = \frac{\sum_{i=1}^{k} S_i^2}{k}, \tag{5.10}$$

where $S_i^2 = \sum_{j=1}^{m}(X_{ij} - \overline{X}_i)^2/(m-1)$. It is well-known that the $(m-1)S_i^2/\sigma^2$ are $\chi^2(m-1)$-distributed, and that thus $(n-k)\tilde{S}^2/\sigma^2$ is $\chi^2(n-k)$-distributed. Consequently, $\mathrm{var}(\tilde{S}/E\tilde{S}) \approx 1/(2(n-k))$ and thus

$$\tau^2 = \lim_{n\to\infty} \frac{n}{2(n-k)} = \frac{m}{2(m-1)}. \tag{5.11}$$

Hence the B's from (5.2) are also inflated by a factor approximately equal to $m/(m-1)$. For $m = 2$ this is still quite large, but for the rather customary $m = 5$ it has already gone down to 1.25.

Example 5. *Average within sample standard deviation*: let $\hat{\sigma} = \overline{S}$, where

$$\overline{S} = \frac{\sum_{i=1}^{k} S_i}{k}, \tag{5.12}$$

in which S_i^2 is as in (5.10). This choice produces the classical $\overline{X}$-chart for grouped observations. For this traditional choice $A = c_4(m)$ and, since m is quite small, the approximation $1 - 1/(4m)$ is no longer adequate.

Instead, tabulated values based on the exact result $c_4(m) = \{2^{1/2}\Gamma(m/2)\}/\{(m-1)^{1/2}\Gamma((m-1)/2)\}$ are in common use. For example, $c_4(3) = 0.886$, $c_4(4) = 0.921$ and $c_4(5) = 0.940$. For the variance we obtain that $\mathrm{var}(\overline{S}/E\overline{S}) = (c_4^{-2}(m) - 1)/k$. Hence

$$\tau^2 = m\{c_4^{-2}(m) - 1\}, \tag{5.13}$$

which e.g. equals 0.82, 0.71 and 0.66 for $m = 3, 4, 5$, respectively. In comparison, in Example 1 we found $\tau^2 = \frac{1}{2}$, while in Example 4 (see (5.11)) we had $\tau^2 = m/\{2(m-1)\}$, producing 0.75, 0.67 and 0.63 for $m = 3$, 4 and 5, respectively. Hence the present choice slightly further inflates the variance and thus the correction effects. For example, for $u_p = 3$ and $m = 5$ we now have

$$FAR: B^{(1)} = B^{(2)} = 3.46, \quad ARL: B^{(1)} = -3.46, \quad B^{(2)} = -2.46,$$

and thus e.g. for the FAR we obtain as corrected '3σ'-recipe: $\overline{X} \pm 3.19\overline{S}\{1 + 3.46/n\}$. Note the difference compared to $\overline{X} \pm 3S\{1 + 3/n\}$ from (5.3).

6 The exceedance case: general

In this section we will no longer be satisfied with correcting the average behavior. Instead we want to limit the probability of nasty results in each single application of the chart. Hence, for example we require that the intended FAR p (e.g. 0.002 in the two-sided case) should not be exceeded by the outcome of P by more than a fraction ε (e.g. 0.25, which means $P > 0.0025$) in more than $100\alpha\%$ (e.g. 20%) of the applications. In formula,

$$Pr(P > p(1+\varepsilon)) \le \alpha. \tag{6.1}$$

Likewise, we can also stipulate that unpleasant values of the ARL $1/P$ should be sufficiently rare, leading to $Pr((1/P) < (1/p)(1-\varepsilon)) = \alpha$. For example, if $p = 0.002$, $\varepsilon = 0.20$ and $\alpha = 0.20$, this means that, as the intended ARL should be 500, the risk is at most 20% that the estimation effects will lower this value to less than 400. Fortunately, this second criterion is essentially equivalent to the first. It clearly boils down to $Pr(P > p/(1-\varepsilon)) \le \alpha$, which is nothing but (6.1) with ε replaced by $\varepsilon/(1-\varepsilon)$ (e.g. $\varepsilon = 0.20$ gives $\varepsilon/(1-\varepsilon) = 0.25$). Hence it suffices to study the criterion in (6.1) for P.

In analogy to (4.2), we propose limits of the following form:

$$\text{one-sided}: \widehat{UL} = \overline{X} + u_p\sigma^*\{1+E^{(1)}\} \text{ or } \widehat{LL} = \overline{X} - u_p\sigma^*\{1+E^{(1)}\},$$

$$\text{two-sided}: \widehat{UL} = \overline{X} + u_{p/2}\sigma^*\{1+E^{(2)}\} \text{ and } \widehat{LL} = \overline{X} - u_{p/2}\sigma^*\{1+E^{(2)}\},$$

with corrections $E^{(1)}$ and $E^{(2)}$ to be determined below. To achieve equality in (6.1) to sufficient precision, again a relatively simple computation essentially suffices. (And again as well, those not interested can safely skip it and go directly to the result, which is contained in (6.2) and (6.3); for technical niceties this time consult Albers and Kallenberg (2004b)). Just as in the bias case, the starting point is the one-sided situation with $\widehat{UL}_c = \overline{X} + u_p\sigma^*(1+c)$, with c some small constant yet to be determined. From (3.1) it follows that $P = Pr(X_{n+1} > \widehat{UL}_c|(X_1,\ldots,X_n)) \approx 1 - \Phi(u_p(1+c)+W)$, where $W = (\overline{X}-\mu)/\sigma + u_p\{(\sigma^*/\sigma)-1\}$. Now a one-step Taylor expansion of P already suffices to give $P - p \approx -\phi(u_p)(W + cu_p)$. Consequently, approximate equality in (6.1) can be achieved by setting $Pr(-W > \varepsilon p/\phi(u_p) + cu_p) = \alpha$. In view of the asymptotic normality of W, this implies that $\varepsilon p/\phi(u_p) + cu_p = u_\alpha\{\text{var}W\}^{1/2}$. Using again that $\phi(u_p)/p \approx u_p$ and $\text{var}((\overline{X}-\mu)/\sigma) = 1/n$, we arrive at $c = u_\alpha\{\text{var}(\hat{\sigma}/E\hat{\sigma}) + 1/(nu_p^2)\}^{1/2} - \varepsilon/u_p^2$. Finally, in view of (4.3), the resulting $u_p\sigma^*(1+c)$ can now be written as $u_p\sigma^*\{1+E^{(1)}\}$, where

$$E^{(1)} = E^{(1)}(\varepsilon) = \frac{u_\alpha(\tau^2 + u_p^{-2})^{1/2}}{n^{1/2}} - \frac{\varepsilon}{u_p^2}. \tag{6.2}$$

As already mentioned above, adaptation to the case of $1/P$ is particularly simple here: just use $\varepsilon/(1-\varepsilon)$ in (6.2) for the ARL situation. For the two-sided case, the computation is again similar and quite simple. The result for the FAR is

$$E^{(2)} = E^{(2)}(\varepsilon) = \frac{u_\alpha\tau}{n^{1/2}} - \frac{\varepsilon}{u_{p/2}^2}, \tag{6.3}$$

from which once more the ARL result follows by using $\varepsilon/(1-\varepsilon)$ rather than ε.

To get some feeling for the impact of the various E's, we add some explanatory remarks. First of all, note that using (4.5), we can rewrite (6.2) into

$$E^{(1)} = \frac{u_\alpha}{u_p}\frac{[2B^{(1)}]^{1/2}}{n^{1/2}} - \frac{\varepsilon}{u_p^2}.$$

This makes it particularly easy to compare the present correction factor $\{1+E^{(1)}\}$ to the bias correction factor $\{1+B^{(1)}/n\}$ from (4.2). Observe that $E^{(1)}$ needs much larger n to become close to 0 than $B^{(1)}/n$ (technically: the rate is $n^{-1/2}$ rather than n^{-1}), which reflects the fact that correcting with respect to exceedance probability is more stringent than with respect to bias. As concerns the role of α, note that (6.1) becomes less severe as α increases. In that case, u_α decreases and hence so does $E^{(1)}$: the upper bound $\widehat{UL} = \overline{X} + u_p\sigma^*\{1+E^{(1)}\}$ becomes smaller, as should be the case.

Likewise, increasing ε in (6.1) also makes the criterion less strict and accordingly lowers the upper bound. The limiting case $\alpha = \frac{1}{2}$ and $\varepsilon = 0$ neatly gives back $E^{(1)} = 0$, reflecting the fact that $Pr(P > p)$ indeed equals $\frac{1}{2}$ to first order and thus no further correction of $\sigma^* = \hat{\sigma}/A$ beyond A is needed then. Typically, as we will see in more detail in the examples from the next section, the positive u_α-term dominates the negative ε-term. Hence the E's will in general be positive: protection in the sense of (6.1) against variability effects requires moving $\widehat{UL}$ and/or $\widehat{LL}$ outwards. Also note that, unlike in the bias case, this holds for both the FAR and the ARL.

7 The exceedance case: examples

Not surprisingly, here we shall simply revisit the five examples considered in section 5. To avoid repetition, we shall be quite brief. As in section 5, we take p in the one-sided case and $2p$ in the two-sided case.

Example 1 (continued). *Sample variance*: let $\hat{\sigma} = S$, with S^2 as given by (5.1). Hence $\tau^2 = \frac{1}{2}$. For the by now customary $u_p = 3$ we get $E^{(1)} = u_\alpha\{11/(18n)\}^{1/2} - \varepsilon/9$ and $E^{(2)} = u_\alpha/(2n)^{1/2} - \varepsilon/9$. Specializing further to e.g. the earlier used $\alpha = 0.20$ and $\varepsilon = 0.20$ leads to $E^{(1)} = 0.66/n^{1/2} - 0.02$ and $E^{(2)} = 0.60/n^{1/2} - 0.02$. As $\varepsilon/(1-\varepsilon) = 0.25$, just replace '0.02' by '0.03' to get the results for the ARL.

Note that these completely explicit examples also nicely illustrate the difference in behavior between the correction for bias and that for exceedance probability. In the former case in Example 1 a typical factor roughly looked like $1 + 3/n$. For $n = 40$, this equals 1.08 and thus still differs noticeably from 1. Indeed, as mentioned, these factors result in a substantial bias reduction for such sample sizes. For $n \geq 300$, bias correction no longer seems needed. Indeed, by then this factor has dwindled to 1.01 and not much effect can be expected anymore. In the present exceedance case, we are dealing with factors like $1 + 0.66/n^{1/2}$, which for $n = 300$ still equals 1.04. As concerns 1.01, note that $n = \{0.66/(1.01-1)\}^2 = 4356$ is required to precisely reach this value.

The effects of $E^{(1)}$ for both the FAR and the ARL for this example have been studied in detail in Albers and Kallenberg (2004b). It turns out that these corrections also work very well, are indeed much larger than in the bias case and consequently remain necessary for much larger n. Note that the price to be paid for this stronger form of protection will unavoidably be some loss in performance during out-of-control. Clearly, considerably widening the limits means that the probability of detection during out-of-control will be reduced somewhat as well. Further illustration is provided in Albers and Kallenberg (2004b) by revisiting the aforementioned real data example involving charge weights of an insecticide dispenser.

Example 2 (continued). *Moving Range*: let $\hat{\sigma} = MR$, with MR as in (5.4). Again $\tau^2 = 0.826$ (cf. (5.5)). Hence, just as above, for $u_p = 3$ we evaluate $E^{(1)} = 0.97u_\alpha/n^{1/2} - \varepsilon/9$ and $E^{(2)} = 0.91u_\alpha/n^{1/2} - \varepsilon/9$. For $\alpha = \varepsilon = 0.20$, this specializes to $E^{(1)} = 0.81/n^{1/2} - 0.02$ and $E^{(2)} = 0.76/n^{1/2} - 0.02$.

Example 3 (continued). *Interquartile Range*: let $\hat{\sigma} = IQR$, with IQR as in (5.7). Here $\tau^2 = 1.36$ (cf. (5.8)) and thus, for $u_p = 3$, we find $E^{(1)} = 1.21u_\alpha/n^{1/2} - \varepsilon/9$ and $E^{(2)} = 1.17u_\alpha/n^{1/2} - \varepsilon/9$. For $\alpha = \varepsilon = 0.20$, this boils down to $E^{(1)} = 1.02/n^{1/2} - 0.02$ and $E^{(2)} = 0.98/n^{1/2} - 0.02$.

Example 4 (continued). *Within sample variance*: let $\hat{\sigma} = \tilde{S}$, with $\tilde{S}^2$ as in (5.10). Here $\tau^2 = m/\{2(m-1)\}$ (cf. (5.11)). Hence for $u_p = 3$ e.g. $E^{(2)} = [m/\{2(m-1)\}]^{1/2}u_\alpha/n^{1/2} - \varepsilon/9$.

Example 5 (continued). *Average within sample standard deviation*: let $\hat{\sigma} = \overline{S}$, with $\overline{S}$ as in (5.12). Then $\tau^2 = m(c_4^{-2}(m) - 1)$ (cf. (5.13)), which e.g. for $m = 5$ leads to $\tau^2 = 0.66$. Consequently, $u_p = 3$ then gives $E^{(1)} = 0.88u_\alpha/n^{1/2} - \varepsilon/9$ and $E^{(2)} = 0.81u_\alpha/n^{1/2} - \varepsilon/9$. For $\alpha = \varepsilon = 0.20$, this boils down to $E^{(1)} = 0.74/n^{1/2} - 0.02$ and $E^{(2)} = 0.68/n^{1/2} - 0.02$.

8 Conclusions

We have introduced a number of modifications for traditional correction factors such as c_4 and d_2. The added value of the new proposals is that these provide a clear link to the actual performance of the chart. Performance characteristics considered are the FAR p and the ARL $1/p$, and corrections are aimed at either reducing bias or exceedance probabilities. The rather mild bias corrections typically modify the traditional correction through an additional factor $\{1 \pm B/n\}$, where n is the sample size and B is some constant. The exceedance corrections are more stringent and roughly boil down to an additional factor $\{1 + C/n^{1/2}\}$, for some constant C.

Note that the results obtained can be used in (at least) two ways. One is quite direct: just specify a performance goal and evaluate the corresponding correction to be applied. Another is more global and uses the material to perform relevant sensitivity analyses. Insight is provided into the sensitivity of the performance characteristics to variation (e.g.

due to estimation effects) in the limits used. Hence if a practitioner experiences a rather large number of unexpectedly short runs in the application of a standard chart, he or she can now obtain an idea of whether something is wrong, or whether this is just 'all in the game' with the traditional approach. In the latter case, it can be figured out what to do about it by using one of the new proposals.

References

Albers, W. and Kallenberg, W.C.M. (2004a). Estimation in Shewhart control charts: effects and corrections. To appear in *Metrika*.

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. To appear in *J. Applied Statistics*.

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. (2003). Normal, parametric and nonparametric control charts: a data driven choice. Technical Report 1674, University of Twente.

Albers, W., Kallenberg, W.C.M. and Nurdiati, S. (2004). Parametric control charts. To appear in *J. Statist. Plann. Inference*.

Does, R.J.M.M. and Schriever, B.F. (1992). Variables control chart limits and tests for special causes. *Statist. Neerl.* **46**, 229-245.

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. (1999). *Shewhart-type Charts in Statistical Process Control.* Ph.D.-thesis, University of Amsterdam.

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