

Normal control charts with nonparametric safeguard

Willem Albers and Wilbert C.M. Kallenberg

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

Abstract Parameter estimation causes a considerable stochastic error in standard Shewhart charts. This problem can be solved using suitable correction factors. But if the normality assumption itself fails, in addition a nonvanishing model error will occur. By then, a nonparametric alternative, such as the recently proposed MIN chart, might be a better idea throughout. However, for those reluctant to give up on the Shewhart chart, a third possibility is offered here. One sticks with this traditional chart as long as the data suggest that the resulting model error is acceptable. Only if this not the case, the MIN chart kicks in and as such serves as a nonparametric safeguard.

Keywords and phrases: Statistical Process Control, Phase II Control Limits, Minimum, Maximum

2000 Mathematics Subject Classification: 62 G 32, 62 P 30

1 Introduction and motivation

When using standard Shewhart \bar{X} control charts, it is typically assumed that the observations come from a normal distribution. Even within that simple setup, a problem remains: mean and variance of this normal distribution are almost always unknown and need to be estimated. However, note that the probability of getting a signal while the process is in control (IC), will typically have a very small value, like 0.001 or 0.002. Consequently, this estimation step causes relative errors which are far too large to be ignored and corrections should be applied to such standard charts in order to keep matters in control in this respect as well. By now, this problem is widely recognized. For brevity we just refer to Albers and Kallenberg (AK for short) (2005b), which provides a recent non-technical review of the results available, as well as additional references. In that paper new and simple corrections are presented, based on already existing correction factors for traditional charts. But now a clear link is made to the actual performance of the resulting chart.

Hence under normality the problems have been repaired successfully. However, quite often also the normality assumption itself is questionable. And once more, because we are dealing with tail probabilities, the effects on the performance of the chart can be devastating. In fact, matters are even worse here. The estimation step mentioned above causes a stochastic error (SE), which in principle vanishes if our sample size is sufficiently large. The 'only' problem is that this will typically not be the case in practice, so we do need the corrections. But if the underlying distribution is not normal and we act as if it is, we are simply estimating the wrong quantiles in our control limits. The resulting model error (ME) will be there, no matter how many observations we have at our disposal. Moreover, such ME's are also typically too large to be ignored. In fact, they can easily inflate an intended false alarm probability from e.g. 0.001 to 0.005 (see Albers, Kallenberg and Nurdiani (AKN for short) (2004) for explicit examples). Even the most liberal of practitioners would object to getting signals during IC on the average once every 200 trials if the idea was once in a 1000 trials.

An attractive alternative possibility is to use nonparametric control charts. This topic by now has also received considerable attention in the literature, see e.g. Bakir and Reynolds (1979), Bakir (2004, 2006), Chakraborti et al. (2001), Chakraborti et al. (2004), Chakraborti and Eryilmaz (2006), Qiu and Hawkins (2001, 2003), as well as AK (2004c). Essentially, the idea in the latter paper is that by estimating the quantiles required for the control limits in a nonparametric way, the ME will nicely vanish. However, the drawback obviously may be a dangerous further increase of the already large SE. For example, if we need to use the upper and/or lower 0.001-quantile, but our sample size is only say 150, we clearly have a problem. For a purpose like that, several thousands of observations seem to be needed.

To solve this problem, AK (2006, 2007) propose a new and simple nonparametric chart called MIN. Rather than the subgroup averages, it uses the subgroup minima to decide whether an out of control (OoC) signal should be given. (This holds for the upper limit; if we deal with a two-sided chart, the subgroup maxima are used in connection with the lower limit.) The advantage of MIN is that it requires much less extreme quantiles in setting the control limits. As a result the SE is brought down to a level comparable to the SE for the ordinary \bar{X} chart. Since the latter chart may have a large ME, whereas MIN, being truly nonparametric, has none, it clearly follows that from the point of view of validity MIN is preferable.

On the other hand, there is of course the efficiency aspect: it is nice that MIN is so well-behaved during IC, but its performance during OoC might be poor compared to that of \bar{X} . Fortunately, as is demonstrated in AK (2006), this is not the case. Of course, if normality does hold, the \bar{X} chart, being optimal under normality, is best, but the detection power of MIN is still quite reasonable in comparison. Moreover, if normality does not hold, the optimality of \bar{X} is lost and it is indeed easy to come up with examples where in fact MIN beats \bar{X} . Hence MIN indeed is a serious competitor of \bar{X} and so it is a good idea to simply use it always, see AK (2007).

Nevertheless, there may be some reservations here. If normality does hold, even in the far tail, then \bar{X} is optimal and it is a pity to drop it. Moreover, it is well known and stan-

dard, and in that sense may feel more comfortable than something new, however simple that is. But on the other hand, the above has made clear that it is dangerous and unwise to stick to \bar{X} no matter what the data tell. So why can't we eat our cake and have it, i.e. use \bar{X} as long as it is safe, and only shift to MIN if we really have to? In this paper we will show how this can be achieved. The idea will indeed be to use a data driven procedure: let the data tell which choice should be made among \bar{X} and MIN.

In fact, the required selection rule is obtained along the same lines as in AKN (2006). In that paper, a data driven procedure is presented which for the case of individual observations (i.e. group size 1) chooses among the X -chart, a parametric chart and a non-parametric chart. The parametric chart from this triple forms an intermediate possibility between the extremes of rigidly assuming normality on one hand and assuming nothing on the other. A third parameter is added for this chart to the two from the normal model, thus offering larger flexibility in estimating the required quantile.

In this way, the ME can be substantially reduced compared to the normal model, while the SE does not increase as much as it might do under the nonparametric approach. Hence if the selection rule tells to leave normality, first a parametric intermediate chart comes into the picture. Only if the data are really 'wild', one is forced to drop this possibility as well and then literally nothing is left but to try one's luck with a nonparametric chart. However, as argued above, MIN has such attractive properties that it could very well be used throughout. Hence replacing the individual nonparametric part by MIN it is quite superfluous to add an intermediate parametric chart, to soften the pain of leaving the \bar{X} chart. A simplified version of the selection rule will suffice.

The setup of the paper is as follows. In section 2 the \bar{X} chart is discussed. The MIN chart is the subject of section 3. In section 4 we discuss the selection rule. Throughout examples are given; in section 5 an application to real data is presented.

2 The \bar{X} chart

Consider the traditional Shewhart \bar{X} chart based on subgroups for monitoring the mean of a production process. Note that no particular choice for the size m of these subgroups is optimal. For larger shifts during OoC, smaller values of m are better, and vice versa. Globally speaking, for a wide range of shifts occurring in practice, a value like $m = 3, 4$ or 5 is well-suited (cf. e.g. AK (2006)). An upper limit UL and a lower limit LL are set, after which an OoC signal results as soon as a group average $\bar{Y} = m^{-1} \sum_{i=1}^m Y_i$ falls outside these limits. In this section we assume normality of the observations, with mean μ and variance σ^2 . Then a common choice is $UL = \mu + 3\sigma/\sqrt{m}$ and $LL = \mu - 3\sigma/\sqrt{m}$. More generally, a false alarm rate (FAR) p will result if we replace this factor '3' by ' $u_{p/2}$ ', the upper $p/2$ -quantile of the standard normal distribution. (Technically: $1 - \Phi(u_{p/2}) = p/2$, with Φ the corresponding standard normal distribution function). Hence for example $u_{1/741} = 3$ and $u_{0.001} = 3.09$.

As mentioned in the Introduction, μ and σ usually are unknown and a so-called Phase I sample is needed to estimate these parameters. Hence suppose we have k subgroups, each

of size m , resulting in an overall sample of size $n = km$ in Phase I: X_{ij} , with $i = 1, \dots, k$ and $j = 1, \dots, m$. For μ we obviously will use $\bar{X} = k^{-1} \sum_{i=1}^k \bar{X}_i$, with $\bar{X}_i = m^{-1} \sum_{j=1}^m X_{ij}$. On the other hand, for σ several possibilities exist (cf. AK (2005b), Examples 1-5). Here we shall follow the traditional choice (i.e. Example 5): to begin with, consider the average within sample standard deviation

$$\bar{S} = k^{-1} \sum_{i=1}^k S_i, \text{ where } S_i^2 = (m-1)^{-1} \sum_{j=1}^m (X_{ij} - \bar{X}_i)^2. \quad (1)$$

Next, as $E\bar{S} = c_4(m)\sigma$, with the well-known $c_4(m) = \{2^{1/2}\Gamma(m/2)\}/\{(m-1)^{1/2}\Gamma((m-1)/2)\}$, replace \bar{S} by the unbiased estimator $\sigma^* = \bar{S}/c_4(m)$ for σ . The basic proposal then is to use as estimated limits

$$\widehat{UL} = \bar{X} + \frac{u_{p/2}\sigma^*}{\sqrt{m}} \text{ and } \widehat{LL} = \bar{X} - \frac{u_{p/2}\sigma^*}{\sqrt{m}}. \quad (2)$$

However, the mere use of the customary correction factor $c_4(m)$ in order to make \widehat{UL} and \widehat{LL} unbiased estimators of UL and LL , respectively, is far from sufficient to control the behavior of the estimated chart. The FAR resulting from applying (2) does not need to be close to p , nor does the average run length (ARL) during IC need to be close to $1/p$. In fact, FAR now has become a random variable, as it equals the conditional probability

$$P = Pr(\bar{Y} > \widehat{UL} \text{ or } \bar{Y} < \widehat{LL} | (X_{11}, \dots, X_{km})) = 1 - \Phi(U + u_{p/2}V) + \Phi(U - u_{p/2}V), \quad (3)$$

in which $U = m^{1/2}(\bar{X} - \mu)/\sigma$ and $V = \sigma^*/\sigma$. If the Phase I observations happen to produce somewhat 'unlucky' values for \bar{X} and σ^* , i.e. values for U and/or $(V-1)$ that differ quite a bit from 0, it is clear from (3) that the outcome p^* of P can be much larger than the intended p . Likewise, the $ARL = 1/p^*$ for that sample will be much smaller than $1/p$ and the corresponding chart is likely to produce false alarms much more often than anticipated. Hence we need to apply further corrections than just $c_4(m)$ to ensure that P is sufficiently close to p . As P is stochastic, there is no unique way to do this. Below we shall discuss the two most popular approaches.

2.1 The bias case

The simplest approach is to use the bias criterion: make sure that $EP - p$ is sufficiently small. To achieve this to high precision, it actually suffices to replace (2) by

$$\widehat{UL} = \bar{X} + u_{p/2} \frac{\sigma^*}{\sqrt{m}} \left(1 + \frac{B}{k}\right) \text{ and } \widehat{LL} = \bar{X} - u_{p/2} \frac{\sigma^*}{\sqrt{m}} \left(1 + \frac{B}{k}\right), \quad (4)$$

where

$$B = \frac{1}{2} \{1 + u_{p/2}^2 (c_4^{-2}(m) - 1)\}. \quad (5)$$

To see this, a relatively simple computation suffices. Those not interested can skip it; those interested in technical details ignored here can consult AK (2004a) for a similar derivation. Using (4) instead of (2) in (3) leads to $EP = 1 - E\Phi(U + u_{p/2}(1 + B/k)V) + E\Phi(U - u_{p/2}(1 + B/k)V)$. Note that $EU = E(V - 1) = 0$ and that $\varphi = \Phi'$ satisfies $\varphi'(x) = -x\varphi(x)$. Then a two-step Taylor expansion readily gives $EP \approx p + u_{p/2}\varphi(u_{p/2})\{-2B/k + [\text{var}(U) + u_{p/2}^2\text{var}(V)]\}$. Consequently, the bias is best removed by letting $B = k[\text{var}(U) + u_{p/2}^2\text{var}(V)]/2$. Now $\text{var}(U) = \text{var}(m^{1/2}(\bar{X} - \mu)/\sigma) = m/n = 1/k$. Moreover, $\text{var}(V) = \text{var}(\bar{S}^*/\sigma) = \text{var}(\bar{S}/(c_4(m)\sigma)) = k^{-1}\text{var}(S_1/(c_4(m)\sigma)) = k^{-1}E(S_1^2/(c_4(m)\sigma)^2 - 1) = k^{-1}(c_4^{-2}(m) - 1)$. Hence the choice given in (5) indeed follows.

The interpretation of the effect of (5) on the limits in (4) is straightforward. To make P unbiased, the limits \widehat{UL} and \widehat{LL} from (2) are widened a bit in going to (4) through application of the additional correction factor $(1 + B/k)$. This agrees with the practical observation that for the traditional chart P tends to show a positive bias. Hence some widening of the limits indeed can remove this bias. Moreover, it is clear that, for given m , the additional correction $(1 + B/k)$ will decrease as the total Phase I sample size $n = km$ increases. To give an explicit example, suppose $m = 3$ and $u_{p/2} = 3$ (i.e. $p = 1/370$). As $c_4(3) = 0.886$ we first get $\bar{X} \pm 3\bar{S}/(0.886\sqrt{3}) = \bar{X} \pm 1.95\bar{S}$ as the outcome for the traditional limits from (2). Moreover, since $c_4^{-2}(3) - 1 = 0.273$, we obtain from (5) that $B = 1.73$. Hence the corrected chart uses instead $\bar{X} \pm 1.95(1 + 1.73/k)\bar{S}$. If we moreover choose $n = 150$, it follows that $k = 50$ and thus $\bar{X} \pm 2.02\bar{S}$ will be the final outcome of (4).

Note that quite a few related cases could be treated as well. To avoid repetition, we shall be very brief about these here. The interested reader is referred to AK (2005b) for details. First of all, we may be interested in the one-sided case, rather than the two-sided one. The recipe for that situation turns out to be simple: just use one instead of both the limits in (4) and then (obviously) replace $u_{p/2}$ in both (4) and (5) by u_p if the desired FAR is still p . Next, instead of the average within sample standard deviation \bar{S} used in (1), we can take some other estimator for σ as our starting point and obtain the slightly adapted formulae for that case (cf. again AK (2005b)).

Finally, rather than making the FAR unbiased to high precision, we could also try to do this for the ARL. Hence B in (4) should then be chosen such that $E(1/P) \approx 1/p$. The computation is similar to the one above and results in a negative value of B , i.e. the limits will be narrowed somewhat instead of widened. 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. Hence, although such a negative answer for B is correct, it is doubtful whether one should apply the corresponding correction: $E(1/P)$ does not seem to be a very suitable criterion to judge the behavior of the run length.

2.2 The exceedance case

The bias approach 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, if the bias of P

is made to be small, but its variability is still large, very long runs may be mixed with very short ones, thus producing a more or less correct average. From the point of view of a single application of the chart, this is still not very satisfactory. If we also want to remedy this effect, a second, more strict approach is indicated. We then apply an exceedance criterion, e.g.:

$$Pr(P > p(1 + \varepsilon)) \leq \alpha, \quad (6)$$

for given small positive ε and α . At first sight, (6) may seem a bit difficult to interpret, but actually the idea is quite simple. If a value p , say 0.002, is intended, we accept that the realized p^* will differ from it. But a relative difference of more than say 20% (i.e. $\varepsilon = 0.2$) is considered to be so unpleasant that it should happen only rarely, say in at most 10% (i.e. $\alpha = 0.1$) of the cases.

To ensure that (6) holds to sufficient precision, we propose in analogy to (4) to replace the standard limits from (2) this time by

$$\widehat{UL} = \bar{X} + u_{p/2} \frac{\sigma^*}{\sqrt{m}}(1 + E) \text{ and } \widehat{LL} = \bar{X} - u_{p/2} \frac{\sigma^*}{\sqrt{m}}(1 + E), \quad (7)$$

where

$$E = u_\alpha \left(\frac{c_4^{-2}(m) - 1}{k} \right)^{1/2} - \frac{\varepsilon}{u_{p/2}^2}. \quad (8)$$

Again a relatively simple computation suffices to verify this. (And again as well, those not interested can safely skip it; those wanting more details should consult AK (2004b) this time.) Here $P = 1 - \Phi(U + u_{p/2}(1 + E)V) + \Phi(U - u_{p/2}(1 + E)V) \approx p - 2u_{p/2}\varphi(u_{p/2})\{E + (V - 1)\}$. Hence (6) will approximately hold if $Pr(-2u_{p/2}\varphi(u_{p/2})\{E + (V - 1)\} > \varepsilon p) = \alpha$. As $\varphi(x)/\{1 - \Phi(x)\} \approx x$ for large x , it follows that $p/(2\varphi(u_{p/2})) \approx 1/u_{p/2}$. Together with the asymptotic normality of V , this implies that $E + \varepsilon/u_{p/2}^2$ should be chosen equal to $u_\alpha\{\text{var}(V)\}^{1/2}$ and (8) follows.

As concerns related situations, we remark to begin with that for the one-sided case a similar computation shows that (8) should be replaced by

$$E = u_\alpha \left(\frac{u_p^{-2} + c_4^{-2}(m) - 1}{k} \right)^{\frac{1}{2}} - \frac{\varepsilon}{u_p^2}. \quad (9)$$

in order to satisfy (6). If we prefer, just as with the bias criterion, that the one-sided case only requires replacing $u_{p/2}$ by u_p , we should not start with (6) in the two-sided case, but instead apply the following, slightly different type of exceedance criterion:

$$Pr\left(P_U > \frac{p}{2}(1 + \varepsilon)\right) \leq \alpha, \text{ and } Pr\left(P_L > \frac{p}{2}(1 + \varepsilon)\right) \leq \alpha, \quad (10)$$

where $P_U = 1 - \Phi(U + u_{p/2}V)$ and $P_L = \Phi(U - u_{p/2}V)$, and hence $P = P_U + P_L$ in (3). Note that (6) is more strict in the sense that an unpleasant event has probability at most

α , whereas in (10) approximately 2α is allowed. On the other hand, (10) is the more strict criterion in the sense that both sides are controlled simultaneously, while (6) only controls the sum P . When using (10), it is indeed immediate that one- and two-sided case are essentially the same and only a change of u_p by $u_{p/2}$ is involved. Next observe that fortunately the ARL can be handled in the exceedance case without further effort: the relevant criterion $Pr(1/P < (1/p)(1 - \varepsilon)) \leq \alpha$ simply equals $Pr(P > p/(1 - \varepsilon)) \leq \alpha$, which is nothing but (6) with ε replaced by $\varepsilon/(1 - \varepsilon)$. For (10), clearly a similar remark holds.

Note that in the correction $(1 + E)$ we have a rate $k^{-1/2}$, whereas in $(1 + B/k)$ from (4) we have a rate k^{-1} . Hence the exceedance correction is indeed more strict in the sense that it tends to 0 much slower than the bias correction as n (and thus k) becomes large. Moreover, from (8) it is evident that E decreases as α and/or ε increase, as should be the case. For $\alpha = 1/2$ and $\varepsilon = 0$ we get $E = 0$, corresponding to the first order result $Pr(P > p) \approx 1/2$. Typically, the first term in (8) will dominate the second: E is positive and the limits will again be moved outward somewhat. For an explicit example, let again $m = 3$ and $u_{p/2} = 3$. As $c_4^{-2}(3) - 1 = 0.273$, we obtain through (8) that the corrected chart in (7) uses $\bar{X} \pm 1.95(1 + 0.523u_\alpha/k^{1/2} - \varepsilon/9)\bar{S}$. Letting $\alpha = 0.1$, $\varepsilon = 0.2$ reduces $E = 0.522u_\alpha/k^{1/2} - \varepsilon/9$ to $0.67/k^{1/2} - 0.02$. Using $k = 50$ again then gives $E = 0.07$ and thus $\bar{X} \pm 2.10\bar{S}$ will be the final outcome of (7).

In AK (2004a,b) the effects of the corrections through B and E are amply studied, both theoretically and through extensive simulations. The conclusions are that both work very well and indeed achieve the goals set during IC. Clearly, widening the control limits will not only correct the IC behavior, but also slightly lower the detection power during OoC. Fortunately, for the bias case the effects are negligible, while for the exceedance case these are quite small.

But note that, apart from the magnitude of these effects, it would be quite mistaken to consider these in themselves as drawbacks of the procedures proposed. What actually happens is that the common 'cheating' by the traditional chart is exposed: when an ARL of 500 during IC is promised, but a realization of say 100 is effected, there will certainly be a gain in detection power during OoC as well. But such a gain is 'stolen' and correcting for it is not a matter of loss but of setting things straight. This point sometimes seems to be difficult to absorb, so allow us to draw a parallel with basic hypothesis testing. A test is meant to have an α of 5%, but suppose its application requires several approximations and/or estimations. Careful study subsequently reveals that this may actually lead to values of α well above 10%, which is not really acceptable. Hence corrections are derived which are fortunately demonstrated to readjust the size of the test to the required 5%. Clearly, there is no denying that such corrections will lower the power of the test. However, nobody would dare to complain about it.

3 The MIN chart

As argued in the Introduction and amply demonstrated in AKN (2004) and AK (2006), we are in serious trouble with the \bar{X} chart once the normality assumption fails. An attractive first thought is to keep using the subgroup averages, and instead estimate the upper and/or lower limits in a nonparametric way. This poses interesting theoretical problems and inspires a rather delicate analysis of the tail behavior of empirical distribution functions for convolutions in AK (2005a). Unfortunately, for the present purpose the results remain disappointing. The estimation step still requires an uncomfortably large number of Phase I observations.

Hence other functions of Y_1, \dots, Y_m need to be considered. Let us concentrate to begin with on setting an upper limit UL . At first sight, using the maximum of the Y_i might seem a good idea. Indeed, earlier proposals to use order statistics all went into this direction (see AK (2006) for references). But on second thoughts, this turns out to be not so bright at all. Essentially, it merely produces a delayed individual chart: if $\max(Y_1, \dots, Y_m)$ exceeds UL , then some Y_i for $i \leq m$ already does this as well and we should have stopped then. On the other hand, using the group minimum in connection with UL turns out to work quite well: in AK (2006) it is demonstrated that this MIN chart is a serious competitor to the traditional \bar{X} chart. It is truly nonparametric, so it has no ME, while \bar{X} (quite seriously!) does. Next, its SE is comparable to that of \bar{X} . Moreover, in terms of OoC performance, the competition ends in a draw: under normality \bar{X} (obviously) wins, but outside the normal model the opposite can easily occur. Finally, \bar{X} is easy to use, but the same definitely holds for MIN.

This final point immediately becomes clear when we define the basic MIN chart. Just consider all $n = km$ Phase I observations together and let $X_{(1)}, \dots, X_{(n)}$ be the corresponding order statistics. Let $[y]$ denote the largest integer $\leq y$, then the one-sided proposal simply is to produce a signal as soon as

$$\min(Y_1, \dots, Y_m) > \widehat{UL} = X_{(n-r)}, \text{ with } r = [np^{1/m}]. \quad (11)$$

The very simple underlying idea is that the order statistic occurring in (11) provides the natural nonparametric estimator for the upper $p^{1/m}$ -quantile of the underlying distribution. As long as the process is IC, the Y_i come from that same distribution. Hence the probability that m consecutive Y_i 's all exceed this quantile simply equals $(p^{1/m})^m = p$. But such a $p^{1/m}$ -quantile is much less extreme than the p -quantile and hence much better to estimate. E.g for $m = 3$ we go from $p = 0.001$ to $p^{1/m} = 0.1$. As mentioned before, with $n = 150$, estimating the first quantile is beyond reach. But the second definitely is not!

The two-sided version of the basic proposal is also easily formulated: a signal occurs as soon as either

$$\min(Y_1, \dots, Y_m) > \widehat{UL} = X_{(n-r)} \text{ or } \max(Y_1, \dots, Y_m) < \widehat{LL} = X_{(r+1)}, \quad (12)$$

where now $r = [n(p/2)^{1/m}]$. Clearly, both minimum and maximum figure in this two-sided version. Nevertheless, we stick to the term MIN-chart, to stress the fact that the minimum

is used where at first the maximum was considered. Also observe that (12) forms the direct competitor of the traditional Shewhart \bar{X} chart as given in (2). Just as in section 2, the next step is to fine tune the basic proposal with respect to the bias or exceedance criterion.

3.1 The bias case

Here as well a simple computation will be all that is needed. But again those not interested in details can jump to (13) below; those wanting to know more should consult AK (2007). To begin with, note that for the one-sided case we obtain from (11) that $P = Pr(\min(Y_1, \dots, Y_m) > X_{(n-r)}) = \{1 - F(X_{(n-r)})\}^m$, where F is the underlying continuous distribution function. But during IC this is distributed as $U_{(r+1)}^m$, where $U_{(1)}, \dots, U_{(n)}$ are order statistics from a sample of size n from the uniform distribution on $(0, 1)$. We denote this by

$$P \cong U_{(r+1)}^m. \quad (13)$$

Note the difference between (13) and (3): the expression in (13) reflects that MIN is nonparametric, as it does not depend on F , but the expression in (3) does. (In fact, it assumes that $F = \Phi$.) It is not difficult to show that $EP = \Pi_{j=1}^m (r+j)/(n+j) = \binom{r+m}{m} / \binom{n+m}{m}$, from which it is straightforward that P will typically have a positive bias. Hence the remedy here as well is to move $\widehat{UL} = X_{(n-r)}$ a bit outward.

The easiest way to do that is by shifting to neighboring order statistics like $X_{(n-r+1)}$ or $X_{(n-r+2)}$. We shall apply a shift $\tilde{s} = s + \lambda$ with $s = [\tilde{s}]$ and $\lambda = \tilde{s} - s$. Let V be an independent indicator variable with $P(V = 1) = 1 - P(V = 0) = \lambda$. Then replace \widehat{UL} from (11) by

$$\widehat{UL} = (1 - V)X_{(n-r+s)} + VX_{(n-r+s+1)}, \quad (14)$$

where s and λ are such that

$$\binom{r-s-1+m}{m} < p \binom{n+m}{m} \leq \binom{r-s+m}{m}, \lambda = \frac{\binom{r-s+m}{m} - p \binom{n+m}{m}}{\binom{r-s+m}{m} - \binom{r-s-1+m}{m}}. \quad (15)$$

The proposal from (14) and (15) produces exact unbiasedness; if we are satisfied with an approximate result, several simplifications exist. In the first place, using a randomized limit like (14) could be considered awkward in practice, so we might prefer the deterministic variant

$$\widehat{UL} = (1 - \lambda)X_{(n-r+s)} + \lambda X_{(n-r+s+1)}. \quad (16)$$

Moreover, it can be verified that the shift $\tilde{s} \approx m/2 + 1/2 - \delta$, where $\delta = np^{1/m} - r = np^{1/m} - [np^{1/m}]$ (cf. (11)). Hence e.g. for m odd, we have

$$\widehat{UL} = (np^{1/m} - r)X_{(n-r+(m-1)/2)} + (1 - np^{1/m} + r)X_{(n-r+(m+1)/2)}. \quad (17)$$

A final simplification is to concentrate on the average value $\delta = 1/2$. Then (17) reduces to $\{X_{(n-r+(m-1)/2)} + X_{(n-r+(m+1)/2)}\}/2$ while for m even we simply get $\widehat{UL} = X_{(n-r+m/2)}$.

The two-sided case easily follows. Now p is replaced by $p/2$ everywhere, e.g. leading to $r = \lceil n(p/2)^{1/m} \rceil$. The formulae for \widehat{LL} follow by symmetry, e.g. (14) becomes $\widehat{LL} = (1 - V)X_{(r+1-s)} + VX_{(r-s)}$. The simplicity of the procedure is demonstrated nicely by looking at an example. Just as in section 2.1, let $m = 3$, $p = 1/370$ and $n = 150$. Hence $(p/2)^{1/m} = 0.111$ and thus $n(p/2)^{1/m} = 16.58$, i.e. $r = 16$ and $\delta = 0.58$. Using (15), the basic proposal $\widehat{UL} = X_{(134)}$ is replaced in (16) by $0.82X_{(135)} + 0.18X_{(136)}$. From (17) we obtain $0.58X_{(135)} + 0.42X_{(136)}$, and the simplest proposal is to use $\{X_{(135)} + X_{(136)}\}/2$. Likewise, $\widehat{LL} = X_{(17)}$ is modified into $0.58X_{(16)} + 0.42X_{(15)}$, or even $\{X_{(16)} + X_{(15)}\}/2$. Just as with the \bar{X} chart, we refrain from deriving the bias correction for $1/P$: here as well it means moving inward, which still does not seem an attractive idea.

3.2 The exceedance case

By now, our approach will be clear. Some simple steps suffice to show equality in (6) for the one-sided and in (10) for the two-sided case. The result is given in (19) and the options once more are to go there directly, to read the sketch given here or to check the details from AK (2007). From (13) it follows that for the one-sided case the probability in (6) equals

$$Pr(U_{(r+1)} > \{p(1 + \varepsilon)\}^{1/m}) = B(n, \{p(1 + \varepsilon)\}^{1/m}, r), \quad (18)$$

where $B(n, p^*, j)$ stands for the cumulative binomial probability $Pr(Z \leq j)$ with $Z \text{ bin}(n, p^*)$. Clearly, shifting r to some $r - \tilde{k}$ will lower the value obtained in (18). As the shift \tilde{s} increases, eventually the desired level α is reached. Hence, here as well use (14), but now with s and λ such that

$$B(n, \{p(1 + \varepsilon)\}^{1/m}, r - s - 1) < \alpha \leq B(n, \{p(1 + \varepsilon)\}^{1/m}, r - s), \quad (19)$$

$$\lambda = \frac{B(n, \{p(1 + \varepsilon)\}^{1/m}, r - s) - \alpha}{B(n, \{p(1 + \varepsilon)\}^{1/m}, r - s) - B(n, \{p(1 + \varepsilon)\}^{1/m}, r - s - 1)}.$$

In this case it can be verified that

$$\tilde{s} \approx u_\alpha \{np^{1/m}(1 - p^{1/m})\}^{1/2} - \varepsilon kp^{1/m} \approx u_\alpha \left\{ r \left(1 - \frac{r}{n} \right) \right\}^{1/2} - \frac{\varepsilon r}{m}. \quad (20)$$

Hence r in $\widehat{UL} = X_{(n-r)}$ is replaced by approximately $r(1 + \varepsilon/m) - u_\alpha \{r(1 - r/n)\}^{1/2}$, typically a more prominent change than in the bias case, where r is replaced by approximately $r - m/2$.

If we take (10) as our two-sided criterion, we only need to replace u_p by $u_{p/2}$. To continue with the example above: we have $m = 3$, $p = 1/370$ and $n = 150$. As $r = 16$, we get from (20) that the shift $\tilde{s} \approx 3.78u_\alpha - 5.33\varepsilon$. For $\alpha = 0.1$, $\varepsilon = 0.2$, this results in $\tilde{s} \approx 3.78$, i.e. the advice to use $\widehat{UL} = 0.22X_{(137)} + 0.78X_{(138)}$ and $\widehat{LL} = 0.22X_{(14)} + 0.78X_{(13)}$.

4 The selection rule

The picture so far looks as follows: if normality does hold, the \bar{X} chart from section 2 has: (i) no ME; (ii) a SE which is considerable, but can be controlled through corrections like (5) or (9); (iii) a detection power which is optimal. However, if normality fails, as it often does (especially in the tails), its ME can become huge (e.g. 5 times p) and its optimality under OoC is lost as well. Hence a very sensible alternative is the nonparametric MIN chart from section 3, which has: (i) no ME for all F (being nonparametric); (ii) a SE comparable to that of the \bar{X} chart, which hence is correctible as well (cf. e.g. (14), (15) and (19)); a detection power which is reasonable under normality and competitive elsewhere (i.e. sometimes better than \bar{X} , sometimes not).

The options now are to: (i) always use \bar{X} ; (ii) always use MIN; (iii) let the data decide between the two. The first option really does not seem to be a very bright idea: one has to be pretty sure of the normality, especially in the far tails, and quite often it is not that clear on what grounds this belief rests. The second one offers a robust choice and as such is just fine. But it remains tempting to try the third option: just act 'normally' as long as possible; if the data really protest, it is time enough to go for the robust alternative after all.

Hence we need an appropriate selection rule. This may look pretty straightforward: just apply a standard goodness-of-fit test. The point, however, is that such tests, just as preliminary visual inspections of the data, are primarily influenced by the behavior of the bulk of the data. Quite often, the normality assumption is reasonably well met in the middle of the distribution and serious deviations only start to show once we go farther into the tail. But that is precisely where we are operating in the present context. Consequently, we should not use a standard approach, but instead find a selection rule which is tuned to the problem at hand.

Bearing this in mind, we look more closely at the question what precisely should prompt us to reject the initial choice for the \bar{X} chart. From the summary above it is clear that the proper criterion should be based on the ME resulting from the default choice. As long as this is not too large compared to p , we can stick with this standard chart, but beyond a certain level we will feel obliged to switch. Next, we observe that a good way to judge the ME is to compare the behavior of the $(X - \mu)/\sigma$ at hand to that of a standard normal Z , with special emphasis on the far tails. In view of this, our proposal for the upper tail now is to base the decision on

$$\frac{X_{(n)} - \bar{X}}{\sigma^*} \tag{21}$$

the standardized maximum of the Phase I sample. For the lower tail, this obviously means using $(\bar{X} - X_{(1)})/\sigma^*$.

The remaining issue is to determine the cut-off points. Note that we use the plural form: we should not only drop the \bar{X} chart if the quantity in (21) is too large, but also if it is too small. In the latter case, the underlying distribution has a lighter tail and there seems to be no problem. The ME will be negative and the chart will only be conservative.

However, this will lead to unnecessary loss of detection power during OoC. On the other hand, deviations in the direction of heavier tailed distributions are indeed regarded as more serious, so it is probably a good idea to react sooner if that is suspected.

For details of the derivations involved we refer to AKN (2006); here we sketch the idea and give the results. Let $Z_{(1)}, \dots, Z_{(n)}$ be the order statistics corresponding to a sample of size n from Φ and let the constants a_1, a_2, \dots such that $a_n \rightarrow 0$ as $n \rightarrow \infty$. Then

$$Pr(Z_{(n)} > u_{a_n/n}) = 1 - \left(1 - \frac{a_n}{n}\right)^n \approx 1 - \exp(-a_n) \approx a_n. \quad (22)$$

Hence if normality holds after all, (22) implies that the standardized maximum from (21) will exceed this $u_{a_n/n}$ approximately with probability a_n . In other words, using $u_{a_n/n}$ as upper cut-off point causes abandoning unnecessarily the \bar{X} chart in about a fraction a_n of the cases. In AKN (2006) it is further advocated to choose the error probability $a_n = c_U/\sqrt{n}$ for some constant c_U . Likewise, we obtain for a sequence b_1, b_2, \dots such that $b_n \rightarrow \infty$ and $b_n^2/n \rightarrow 0$ as $n \rightarrow \infty$ that

$$Pr(Z_{(n)} < u_{b_n/n}) = \left(1 - \frac{b_n}{n}\right)^n \approx \exp(-b_n), \quad (23)$$

and thus a lower cut-off point $u_{b_n/n}$ gives an error probability under normality of about $\exp(-b_n)$. Taking $b_n = \{\log(n/c_L^2)\}/2$ for some constant c_L , as suggested in AKN (2006), produces an error probability c_L/\sqrt{n} .

Summarizing, with regard to the upper tail, the selection rule stays with the \bar{X} chart whenever

$$u_{\{\log(n/c_L^2)\}/(2n)} \leq -\frac{X_{(n)} - \bar{X}}{\sigma^*} \leq u_{c_U/(n\sqrt{n})}. \quad (24)$$

For the lower tail, just replace $(X_{(n)} - \bar{X})/\sigma^*$ in (24) by $(\bar{X} - X_{(1)})/\sigma^*$ in (24). Note that both tails can and should be handled separately. It is quite conceivable that at one end the normality assumption really fails, while at the other it is just fine. Then at one side MIN should replace \bar{X} , while at the other it is still O.K. to use \bar{X} . To conclude we give a simple example. Suppose, as before, that $n = 150$. Take e.g. $c_U = 1$ and $c_L = 1/2$, corresponding to error probabilities $1/\sqrt{n} = 0.08$ and $1/(2\sqrt{n}) = 0.04$, respectively. Note the unbalance, reflecting the aspect that heavier tails are more of a threat than lighter tails. Hence the upper \bar{X} chart will be used with probability 0.88, and this occurs whenever $2.03 = u_{0.0213} \leq (X_{(150)} - \bar{X})/\sigma^* \leq u_{0.000544} = 3.27$. For the two-sided chart, this probability becomes 0.77, as in this case it is also required that $2.03 \leq (\bar{X} - X_{(1)})/\sigma^* \leq 3.27$.

5 Application

At each step we have already given brief examples. Here we shall use these on a real life application 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 (2006) 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 (2006)) 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. However, within the present setup, we are able to successfully implement a data driven approach already for samples of a much more common size. To demonstrate this we shall, instead of using the full first sample, only take 150 observations, i.e. we let $n = 150$. If we select once more $m = 3$ for the group size, we thus have $k = 50$ subgroups in our Phase I sample. Moreover, letting $p = 1/370$ (i.e. $u_{p/2} = 3$), according to section 2.1 we arrive at $\bar{X} \pm 1.95\bar{S}$ for the traditional limits from (2) and at $\bar{X} \pm 2.02\bar{S}$ for the bias corrected proposal from (4). Subsequently, for $\alpha = 0.1$, $\varepsilon = 0.2$, section 2.2 gives $\bar{X} \pm 2.10\bar{S}$ for the exceedance case proposal from (7). The realizations of \bar{X} and \bar{S} for the actual Phase I sample at hand turn out to be $\bar{x} = 43.03$ and $\bar{s} = 2.51$, leading to the following three sets of limits (38.12, 47.93), (37.95, 48.10) and (37.76, 48.29). Hence for the remaining sample (exclusive the last two observations), a signal will occur whenever the average of one of the 506 triples contained in these 1518 observations falls outside these limits.

Next we consider MIN. From section 3.1 we know that for the values considered we have $n(p/2)^{1/m} = 16.58$, i.e. $r = 16$. Hence the basic proposal is to take $\widehat{UL} = X_{(134)}$ and the simplest proposal for bias correction is to use $\{X_{(135)} + X_{(136)}\}/2$. Likewise, we have $\widehat{LL} = X_{(17)}$ or $\{X_{(16)} + X_{(15)}\}/2$. If instead we want to control exceedance probabilities, section 3.2 produces for $\alpha = 0.1$, $\varepsilon = 0.2$ that $\widehat{UL} = 0.22X_{(137)} + 0.78X_{(138)}$ and $\widehat{LL} = 0.22X_{(14)} + 0.78X_{(13)}$ should be used. For the present case, we find as realizations $x_{(134)} = 46.38$, $x_{(135)} = 46.39$, $x_{(136)} = 46.50$, $x_{(137)} = 46.55$ and $x_{(138)} = 46.76$, while $x_{(17)} = 39.09$, $x_{(16)} = 38.65$, $x_{(15)} = 38.60$, $x_{(14)} = 38.59$ and $x_{(13)} = 38.55$. This leads to the three sets of limits (39.09, 46.38), (38.63, 46.45) and (38.56, 46.71). For the remaining sample, a signal will occur whenever the minimum of one of the 506 triples exceeds the upper limit used or its maximum falls below the corresponding lower limit.

Hence both \bar{X} and MIN are now ready for use in this actual application. It remains to see which choice the selection rule from section 4 suggests. (Of course, we could have saved ourselves some work by doing this to begin with, but for illustrative purposes it seemed better to first mention the results for both charts on both sides.) Following the example given, we should use \bar{X} at the upper end as long as $2.03 \leq (X_{(150)} - \bar{X})/\sigma^* \leq 3.27$ and at the lower end as long as $2.03 \leq (\bar{X} - X_{(1)})/\sigma^* \leq 3.27$. As mentioned above, here $\bar{x} = 43.03$

and $\bar{s} = 2.51$, and thus $\sigma^* = \bar{s}/0.886 = 2.83$. Moreover, $x_{(150)} = 51.66$ and $x_{(1)} = 25.45$, and thus $(x_{(150)} - \bar{x})/\sigma^* = 3.05$ and $(\bar{x} - x_{(1)})/\sigma^* = 6.20$. Hence \bar{X} can indeed be used in the right tail, but in the left tail MIN is to be preferred.

Inspection of the 506 triple-averages shows that in the right tail the number of signals equals 1 for all three charts, see Figure 1. In the left tail the MIN chart is used. Since we are dealing with the lower limit, the maxima of the 506 triples are calculated. Figure 2 shows these maxima and it is seen that they give 3 signals when the basic chart is used. This is reduced to 1 signal as well for the bias-corrected chart as for the exceedance-corrected chart.

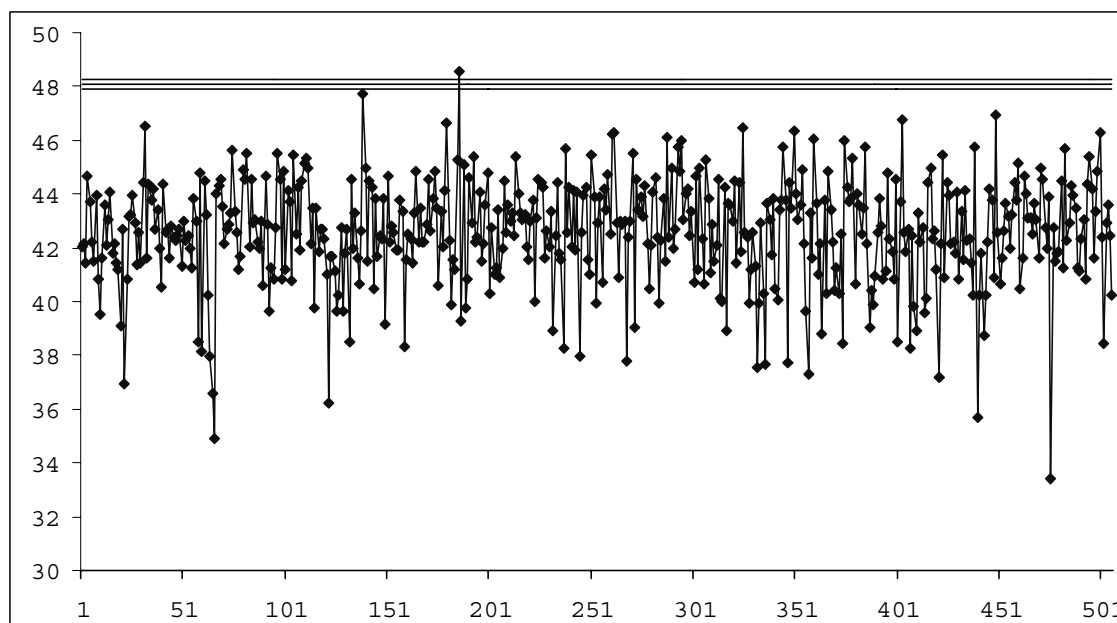


Figure 1: Normal-MIN Chart with $\widehat{UL} = 47.93$ (uncorrected), $\widehat{UL} = 48.10$ (bias corrected) and $\widehat{UL} = 48.29$ (exceedance corrected).

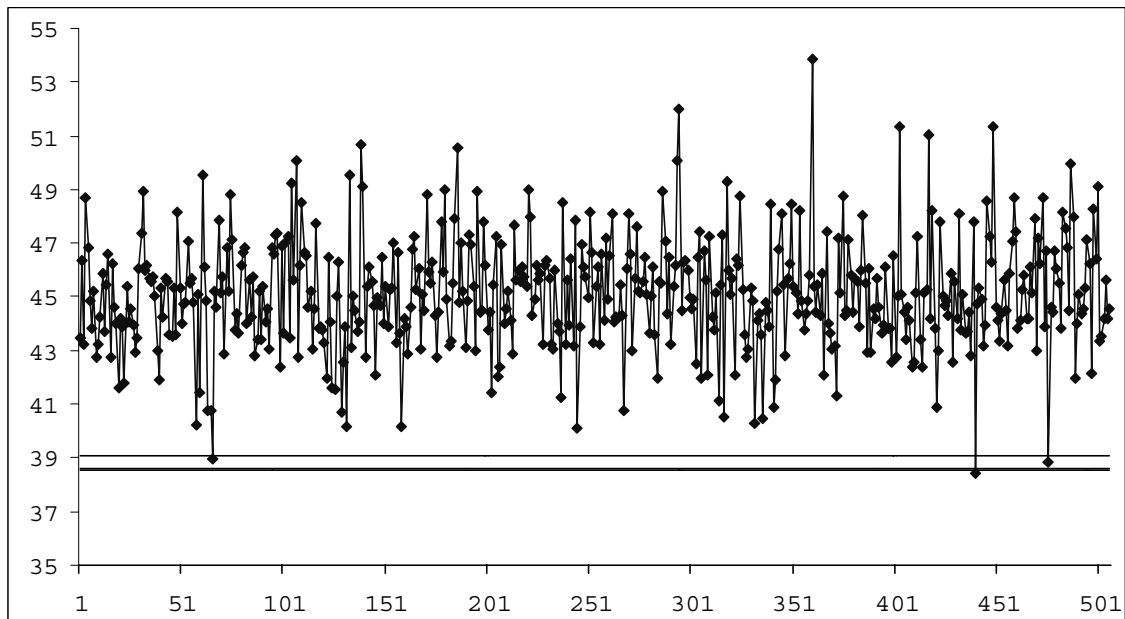


Figure 2: Normal-MIN Chart with $\widehat{LL} = 39.09$ (uncorrected), $\widehat{LL} = 38.63$ (bias corrected) and $\widehat{LL} = 38.56$ (exceedance corrected).

References

- 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. (2005a). Tail behavior of the empirical distribution function of convolutions. *Math. Methods Statist.* **14**, 133-162.
- Albers, W. and Kallenberg, W. C. M. (2005b). New corrections for old control charts. *Quality Engineering* **17**, 467-473.
- Albers, W. and Kallenberg, W. C. M. (2006). Alternative Shewhart-type charts for grouped observations. *Metron* **LXIV** (3).
- Albers, W. and Kallenberg, W. C. M. (2007). Minimum control charts. *To appear in J. Statist. Plann. Inference*.
- Albers, W. and Kallenberg, W. C. M. and Nurdiati, S. (2004). Parametric control charts. *J. Statist. Plann. Inference* **124**, 159-184.
- Albers, W., Kallenberg, W. C. M. and Nurdiati, S. (2006). Data driven choice of control charts. *J. Statist. Plann. Inference* **136**, 909-941.
- Alloway, J. A., Jr. and Raghavachari, M. (1991). Control chart based on the Hodges-Lehmann estimator. *J. Quality Technol.* **23**, 336-347.

- Bakir, S. T. (2004). A distribution-free Shewhart quality control chart based on signed-ranks. *Quality Engineering*, **16**, 611-621.
- Bakir, S. T. (2006). Distribution-free quality control charts based on signed-rank-like statistics. *Comm. Statist. Theory Methods* **35**, 743-757.
- 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. and Eryilmaz, S. (2006). A nonparametric Shewhart-type signed-rank control chart based on runs. *To appear in Comm. Statist.*
- Chakraborti, S., van der Laan, P. and Bakir, S. T. (2001). Nonparametric control charts: an overview and some results. *J. Quality Technol.* **33**, 304-315.
- Chakraborti, S., van der Laan, P. and van de Wiel, M. A. (2004). A class of distribution-free control charts. *J. Roy. Statist. Soc. C* **53**, 443-462.
- Qiu, P. and Hawkins, D. (2001). A rank based multivariate CUSUM procedure. *Technometrics* **43**, 120-132.
- Qiu, P. and Hawkins, D. (2003). A nonparametric multivariate cumulative sum procedure for detecting shifts in all directions. *J. Roy. Statist. Soc. D* **52**, 151-164.