Memorandum 1919 (April 2010). ISSN 1874-4850. Available from: http://www.math.utwente.nl/publications
Department of Applied Mathematics, University of Twente, Enschede, The Netherlands

Empirical nonparametric control charts for high-quality processes

Willem Albers

Department of Applied Mathematics

University of Twente

P.O. Box 217, 7500 AE Enschede

The Netherlands

Abstract. For attribute data with (very) small failure rates often control charts are used which decide whether to stop or to continue each time r failures have occurred, for some $r \geq 1$. Because of the small probabilities involved, such charts are very sensitive to estimation effects. This is true in particular if the underlying failure rate varies and hence the distributions involved are not geometric. Such a situation calls for a nonparametric approach, but this may require far more Phase I observations than are typically available in practice. In the present paper it is shown how this obstacle can be effectively overcome by looking not at the sum but rather at the maximum of each group of size r.

Keywords and phrases: Statistical Process Control, health care monitoring, geometric charts, average run length, estimated parameters, order statistics

2000 Mathematics Subject Classification: 62P10, 62C05, 62G15

1 Introduction and motivation

High-quality processes are by now a regular phenomenon in industrial settings, due tot the fact that production standards have been increasing over the last few decades. Moreover, they are in fact the norm in the area of health care monitoring, as typical failures such as malfunctioning equipment, surgical errors or recurrence of cancer, should by their very nature be quite rare. Some review papers in this latter field are Sonesson and Bock (2003), Thor et al.(2007) and Shaha (1995). Here application of control charts to improve and maintain quality is strongly advocated.

As concerns the choice of which chart to apply, many authors have argued that for the really small failure probabilities p involved it is advisable to use so-called time-between-events charts. These are typically based on waiting times till $r(r \geq 1)$ failures have occurred. A signal then follows if the corresponding negative binomial random variable $(r.v.)X$ attains a value which is judged to be too small. Many references on such negative binomial charts are given in Albers (2010), as well as a detailed analysis. In particular, it is demonstrated which choice of r is best for a given combination of underlying parameters. Moreover, the problem is tackled how to deal with the fact that p is typically unknown and thus has to be estimated on the basis of a so-called Phase I sample. Simple corrections are derived which control the estimation effects involved. This estimation step is quite important: contrary to what is commonly assumed, its effects are only negligible for very large sample sizes, which in practice are typically not available.

e-mail: w.albers@utwente.nl;
running title: Empirical nonparametric control charts for high-quality processes.

Note the general nature of this estimation issue: almost invariably, control charts have one or more unknown parameters, and these need to be estimated before the monitoring phase can start. For example, in the continuous case of controlling the mean of a process using a Shewhart chart, estimators for the underlying normal mean and variance will be required, and preferably corrections as well for their effects (see Albers and Kallenberg (abbreviated in the sequel to AK) (2004a,b)). However, here it is widely realized that the solution thus obtained can still be quite inadequate, as the underlying normality assumption itself may be unwarranted. The actual distribution can e.g. be heavier tailed, producing false alarm rates (FAR's) which are in fact many times as large as the intended, typically very small, value. This state of affairs inspired Albers, Kallenberg and Nurdiati (2004) to fit a wider parametric family than the normal one, with an additional parameter serving to accommodate tail length.

For the negative binomial charts, exactly the same type of problem occurs. Here it concerns the underlying homogeneity assumption: each and every incoming item is supposed to have the same probability p of being defective. For industrial processes this assumption may usually be quite reasonable, but in medical applications patients will often show large heterogeneity. Typically this will produce an underlying distribution with a larger variance than the negative binomial one. This overdispersion effect is quite well-known (cf. Poortema (1999), Christensen et al. (2003) and Fang (2003)) and in Albers (2009) the problem is attacked along the lines indicated above by fitting a wider parametric family. In addition to the failure rate p, a second parameter is used to accommodate the degree of overdispersion. Indeed, as shown in Albers (2009), the results obtained are far better than those using the basic negative binomial approach. However, neither in Albers, Kallenberg and Nurdiati (2004), nor in Albers (2009), it is pretended that the wider model will achieve a perfect fit. The precise underlying mechanism remains unknown in either case and by using a more flexible model 'only' a better approximation of reality is achieved.

The development sketched above strongly suggests taking a next (and final) step, towards nonparametric control charts. In this way the model error, caused by estimating within a wider but still incorrect model, will nicely vanish altogether. However, the obvious catch here is the trade-off involved w.r.t. the other error involved. This is the stochastic error, which grows if the parametric model is widened and is prone to become unacceptably large in the nonparametric case. Just observe that the FAR's involved will be (very) small, so extreme quantiles need to be estimated in the nonparametric setup. But assembling a Phase I sample of size m requires in the negative binomial case on average mr/p observations. The latter quantity will be very large, even for moderate m. So in a typical situation one might face the need to estimate a 0.001-quantile based on a sample of size 100, which clearly cannot be done in a satisfactory manner.

Hence it may seem that the nonparametric option is useful only on those rare occasions where a multitude of Phase I data exists. However, as shown by Albers (2008), there actually is a way around this obstacle. In fact, the idea is quite simple, as can be (roughly) illustrated through an example. Estimating the quantile at e.g. 0.1 rather than at 0.001 can easily be realized with a stochastic error which does remain acceptable, even for a moderate sample size. This suggests to take 3 observations rather than just 1 and to note that the probability of all of these falling below the 0.1-quantile is $(0.1)^3$, which is again the originally desired 0.001. In other words, the required small FAR can be realized by judging on the basis of a - typically small - group, rather than on the basis of a single observation. This idea is exploited in AK (2008) for the continuous case of controlling the mean and shown to work very well. The resulting chart is truly nonparametric and both its power of detection during out-of-control and its stochastic error during in-control are comparable to that of the customary chart based on averages, while

this latter choice does rely on the dubious assumption of normality.

In the present paper we shall demonstrate how a similar approach can be used for high-quality processes. In Section 2 we introduce the notation required, as well as the necessary background from the negative binomial approach. After the homogeneous case, the overdispersion generalization is described, as well as the further step towards a fully nonparametric approach. As the latter method typically requires too many observations, we turn in section 3 to the new proposal based on (small) groups. Again the starting point is the homogeneous case, where the type of the underlying distribution is known to be negative binomial. This may seem a bit strange, as the very motivation for the present paper is the seriousness of the estimation aspect and the question how to deal with it in a thorough way. However, solving these estimation problems by means of a nonparametric approach is only attractive if the price in terms of loss of detection power is reasonable (and preferably small). Hence it makes sense to first compare for the case of known distributions how the alternative approach holds up compared to the standard negative binomial chart. Once this has been settled, it makes sense to continue in section 4 with the estimation aspects and the description of the nonparametric chart. The impact of the estimation step is analyzed and suitable corrections for the corresponding effect are proposed. For convenience, the actual application of the procedure is summarized in section 5.

2 The negative binomial chart and generalizations

As our starting point we consider the homogeneous case, where $D_1, D_2, \ldots$, is a sequence of i.i.d. r.v.'s, with failure probability $P(D_1 = 1) = 1 - P(D_1 = 0) = p$ during in-control (IC). During the out-of-control (OoC) stage, this p becomes θp, for some $\theta > 1$ and the process should now be stopped as quickly as possible. First consider the negative binomial chart as discussed in Albers (2010). The 'time-between-events' approach referred to in the Introduction implies that no fixed-length blocks of D's are used. Rather we wait till the r^{th} failure occurs, for some $r \geq 1$, and repeat this as long as the process is judged to be IC. Let $X_i, i = 1, 2, \ldots...$ be the successive numbers of D's involved, then these X_i clearly are i.i.d. copies of a negative binomial r.v. $X_{r,p}$ such that

$$P(X_{r,p} = k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}, \tag{2.1}$$

where $k = r, r+1, \ldots$. Here as well as in the sequel, we suppress indices whenever possible, and e.g. write X instead of $X_{r,p}$, unless confusion might occur.

As $\theta > 1$, a signal should result when an r^{th} failure arrives too early. More specific, this happens as soon as an X_i occurs which is at most equal to some well-determined lower limit $n = n_{r,p}$. In Albers (2010) it is proposed to use $F_{r,p}(n) = P(X_{r,p} \leq n) = r\alpha$ as a criterion, for some small $\alpha > 0$. Then the average run length (ARL) during IC has the same value $r/(r\alpha) = 1/\alpha$ for all r, and the negative binomial charts for various r can be compared in a fair and meaningful way. Hence $n = n_{r,p} = F_{r,p}^{-1}(r\alpha)$, the $r\alpha^{th}$ quantile of the negative binomial distribution function $(df)F_{r,p}$. For $r = 1$, the geometric case, $F_{1,p}(n) = 1 - (1-p)^n$ and thus we have the exact result

$$n = n_{1,p} = \frac{\log(1-\alpha)}{\log(1-p)}. \tag{2.2}$$

For $r > 1$, a numerical solution is readily obtained, but in addition a transparent and accurate

approximation helps to see how n depends on r, p and α. Just use that

$$F_{r,p}(n) = P(X_{r,p} \leq n) = P(Y_{n,p} \geq r) \approx P(Z_{np} \geq r), \tag{2.3}$$

where $Y_{n,p}$ is a binomial r.v. with parameters n and p, while Z_{np} is a Poisson r.v. with parameter $\lambda = np$. The Poisson approximation in (2.3) requires n to be large, which will be the case for $r > 1$. Hence $n \approx \lambda/p$, with λ solving $P(Z_\lambda \geq r) = r\lambda$. It is demonstrated in Albers (2010) that this λ in its turn can be approximated quite well by

$$\tilde{\lambda} = \alpha_r(1 + \zeta_r) \text{ with } \zeta_r = \frac{\alpha_r}{(r+1)} + \frac{\frac{1}{2}\alpha_r^2(3r+5)}{\{(r+1)^2(r+2)\}}, \tag{2.4}$$

with $\alpha_r = (r!r\alpha)^{1/r}$, for $p \leq 0.01$, $r \leq 5$ and $\alpha \leq 0.01$, which region is amply sufficient.

During OoC the probability of a signal is given by $F_{r,\theta p}(n_{r,p})$, and hence

$$ARL = \frac{r}{F_{r,\theta p}(n_{r,p})}, \tag{2.5}$$

which decreases from the prescribed value $1/\alpha$ at $\theta = 1$ to a lower limit r as θ becomes very large. Again, a numerical outcome for (2.5) is easily obtained, while in addition an accurate and transparent approximation can be found in Albers (2010) (see Lemma 3.1). Finally, in this homogeneous case the estimation aspect is relatively easy to tackle. Consider a Phase I sample containing m geometric r.v.'s $X_{1,p}$ (cf. (2.1)), let $\overline{X} = m^{-1}\Sigma_{i=1}^m X_i$ and estimate the unknown p by $\hat{p} = 1/\overline{X}$. Next, observe that $\hat{n} = n_{r,\hat{p}} = F_{r,\hat{p}}^{-1}(r\alpha) \approx \lambda/\hat{p} = \lambda\hat{X} \approx \tilde{\lambda}\overline{X}$ (cf.(2.4)) and the estimated version is ready for use. See Albers (2010) for details, as well as results on corrections for the estimation effects involved.

Once the homogeneity assumption is dropped, matters become essentially more complicated, as $X_{r,p}$ from (2.1) will no longer be negative binomial. As mentioned in the Introduction, in Albers (2009) a wider parametric family is introduced in order to accommodate the overdispersion effect. Under the assumed model, this e.g. is shown to lead to replacing $P(Z_{np} \geq r)$ in (2.3) by

$$P(Y_{n^*,p^*} \geq r), \tag{2.6}$$

where $n^* = r + 1 + \tau^{-1}$, $p^* = 1/\{1 + (1 + \tau^{-1}/(np)\}$ and τ is the overdispersion parameter from the model. (For $\tau \to 0$, note that $n^* \to \infty$, $p^* \to 0$, $n^*p^* \to np$, and thus this binomial approximation reduces to the previous Poisson one.) Along these lines, again an approximated lower limit $\tilde{n}_\tau = \tilde{\lambda}/p$ (cf. (2.4)) can be obtained, as well as an estimated version based on $\hat{p}$ and $\hat{\tau}$ obtained from a Phase I sample. See Albers (2009) for details; as demonstrated in that paper, for $\tau > 0$ this overdispersion chart provides a marked improvement over the simple negative binomial one.

Nevertheless, it remains a solution within a wider, but still incorrect, framework. The additional parameter τ simply allows us to get closer to the real underlying structure than within the rigid homogeneous model. Completely removing the yet remaining model error requires a nonparametric approach. Here no assumptions other than of smoothness type (e.g. continuity) are made about the $dfF = F_{r,p}$ of $X_{r,p}$ from (2.1). For a Phase I sample of size m, let $F_m(x) = m^{-1}\#\{X_i \leq x\}$ be the empirical df and F_m^{-1} the corresponding quantile function, i.e. $F_m^{-1}(t) = \inf\{x \mid F_m(x) \geq t\}$. Then it follows that $F_m^{-1}(t)$ equals $X_{(i)}$ for $(i-1)/m < t \leq i/m$, where $X_{(1)} < \cdots < X_{(m)}$ are the order statistics corresponding to the sample. Consequently,

the $r\alpha^{th}$ quantile $n = F^{-1}(r\alpha)$ can now in principle be estimated in a straightforward manner by

$$\hat{n} = F_m^{-1}(r\alpha) = X_{(s)}, \tag{2.7}$$

where $s = \{mr\alpha]$, with $\{y]$ denoting the smallest integer $\geq y$. However, as argued in the Introduction, on the average mr/p observations D_i will be needed. Since p is typically (very) small, mr should be moderate. On the other hand, α should be small as well, and hence we will wind up more often than not with $s = 1$, i.e. with the minimum of the sample as our estimator in (2.7). For e.g. $m = 100$, $r = 3$ and $\alpha = 0.001$ and thus $s = \{0.3] = 1$, this clearly is far from satisfactory. Of course, if e.g. $m = 200$, $r = 5$ and $\alpha = 0.01$, then $s = 10$, and using $X_{(10)}$ may work well. But this latter type of example will typically be rarer in practice than the first.

Summarizing so far, the negative binomial chart (cf. (2.1) and (2.4)) is fine for the homogeneous case, it can be upgraded to an overdispersion chart (cf. (2.6)) to fit a wider model, and it is not really difficult to come up with a nonparametric version for general F (cf. (2.7)). However, the applicability of the latter proposal remains rather limited, as typically too large Phase I samples are needed. So the question how to define a satisfactory nonparametric solution still stands.

3 The MAX-chart

In view of the conclusion of the previous section, an alternative proposal is sought. For this purpose, we adapt the approach presented by AK (2008) for the continuous case. For an upward shift, the $\overline{X}$-chart is replaced there by the MIN-chart: a signal is given not if an incoming group of size r has an $\overline{X} = r^{-1}\Sigma_{j=1}^{r} X_j$ which is deemed too large, but rather when this is the case for $\min(X_1, \ldots, X_r)$. Although using $\overline{X}$ is obviously optimal if the underlying distribution is normal, the loss in detection power sustained by applying the MIN-chart instead stays reasonable, while MIN can be better than $\overline{X}$ once we have a nonnormal distribution. Hence the new proposal provides a reasonable competitor to the $\overline{X}$-chart for the case of known distributions. Once the estimation aspect enters the picture, it decidedly becomes the more attractive choice.

For the present situation observe that in the negative binomial chart the $X_{r,p}$ we use are in fact $\Sigma_{j=1}^{r} X_{1,p,j}$, i.e. it corresponds to the $\overline{X}$-chart above. Since here we are dealing with lower limits, the equivalent of the MIN-chart will now become a MAX-chart, giving a signal if $\max(X_1, \ldots, X_r)$ is too small, where $X_j = X_{1,p,j}$ is the j^{th} geometric random variable in the group. Obviously, in the context of high-quality processes, this type of proposal in itself is far from new. In fact, Chen (1978) already advocated the so-called sets method, according to which a signal results once all of r successive X_j are too small. However, the focus in the present paper is on showing how this type of approach can lead to a satisfactory nonparametric procedure, thus adequately solving the serious underlying estimation issues. In practice these problems are either conveniently ignored completely, or at best the effect of estimating a single parameter is studied. Typically, the latter already turns out to be substantial: e.g. Chen et al. (1997) mention a 30-90% increase in FAR for a 10 per cent bias in $\hat{p}$.

But before embarking on the estimation aspects, we first need to establish that, just as with MIN and $\overline{X}$ in the continuous case, there is only a limited loss in detection power when the negative binomial choice from section 2 is abandoned in favor of MAX. First we derive the lower limit n to be used here. Note that now the requirement for this n will be that during IC

$$P(\max(X_1, \ldots, X_r) \leq n) = r\alpha. \tag{3.1}$$

Hence $F_{r,p}(n) = r\alpha$ from section 2 turns into $\{F_{1,p}(n)\}^r = r\alpha$, implying that $1-(1-p)^n = \{r\alpha\}^{1/r}$ and thus (cf. (2.2)) we have for all r the exact solution

$$n = \frac{(\ \log\ (1-\{r\alpha\}^{\frac{1}{r}})}{\log\ (1-p)}, \tag{3.2}$$

which to first order equals $\{\ \log\ (1-\{r\alpha\}^{1/r})\}/p$. Note that for $r=1$ the MIN-chart, just as the negative binomial one, simply boils down to the geometric chart (cf. (2.2)).

Next, it is immediate that during OoC the alarm rate equals $\{F_{1,\theta p}(n)\}^r = \{1-(1-\theta p)^n\}^r$. In view of (3.2) this leads to replacing the result from (2.5) by

$$ARL = ARL_{r,\theta} = \frac{r}{\left\{1-(1-\{r\alpha\}^{\frac{1}{r}})^{\frac{\log(1-\theta p)}{\log(1-p)}}\right\}^r}, \tag{3.3}$$

which to first order equals $r/\{1-(1-\{r\alpha\}^{1/r})^\theta\}^r$. To be more precise, let $p < p_0$ and $\theta < p_0^{-1}$ for some $p_0 << 1$, then the relative difference $q(\theta) = \{\ \log\ (1-\theta p)/\ \log\ (1-p)\}/\theta - 1$ satisfies $\theta p\{1-\frac{1}{2}p/(1-p_0)^2\} < 2q(\theta) < p\{\theta/(1-\theta p_0)^2 - 1\}$ and hence is very small. Only if $\alpha < p$, which is undesirable anyway, the simplification of $\log\ (1-\theta p)/\ \log\ (1-p)$ to θ might make a nonnegligible difference in the resulting ARL. But otherwise, the dependence of ARL on p is negligible; hence the notation $ARL_{r,\theta}$ in (3.3).

Another observation from (3.3) is that $ARL_{1,\theta} \approx 1/\{1-(1-\alpha)^\theta\} \approx 1/(\theta\alpha)$, implying that the ARL of the simple geometric chart roughly decreases at the same rate θ at which p decreases. In Albers (2010) it was already argued that this rate is quite slow and consequently that increasing r is very worthwhile. To demonstrate this, for $r \geq 2$ and $ARL = ARL_{r,\theta}$ from (2.5) the functions

$$h_r = h_{r,\theta} = \frac{ARL_{1,\theta}}{ARL_{r,\theta}}, \tag{3.4}$$

were introduced. These h_r as functions of θ start at 1, increase towards a maximum, and then slowly decline towards $1/r$. For larger r the peak is higher and it occurs for lower θ. On the other hand, the decline is also faster as r increases. Nevertheless, it still takes quite large θ before h_r hits 1 again, i.e. the geometric chart start to dominate See Figure 3.2 from Albers (2010) for an illustration. Moreover, a simple rule of thumb is presented there for r^{opt}, the value of r which approximately minimizes ARL for given α and θ:

$$r^{opt} = \frac{1}{\{\alpha(2.6\theta+2) + 0.01(4\theta-3)\}}. \tag{3.5}$$

Qualitatively, precisely the same conclusions follow if the h_r in (3.4) are based on the present ARL, as defined in (3.3). To make matters a bit more exact, we obtain the following analogue for Lemma 3.2 from Albers (2010):

Lemma 3.1. *The value $\theta^{max} = \theta_r^{max}$ at which h_r from (3.4) with ARL as in (3.3) reaches its maximum to first order equals*

$$\tilde{\theta}^{max} = \frac{\log\ (1-k^{-1}(\frac{1}{r}))}{\log\ (1-(r\alpha)^{\frac{1}{r}})}, \tag{3.6}$$

where k^{-1} is the inverse of $k(x) = \{-(1-x)\ log\ (1-x)\}/x$.

Proof. From (3.3) and (3.4) it follows that h_r to first order equals $\{1-(1-\{r\alpha\}^{1/r})^{\theta}\}^r/[r\{1-(1-\alpha)^{\theta}\}]$. Hence

$$\frac{\partial(\log h_r)}{\partial\theta} \approx \frac{(1-\alpha)^{\theta}\log(1-\alpha)}{\{1-(1-\alpha)^{\theta}\}} - \frac{r(1-g(r,\theta))\log(1-g(r,\theta))}{\{\theta g(r,\theta)\}}, \tag{3.7}$$

in which $g(r,\theta) = 1-\{1-(r\alpha)^{1/r}\}^{\theta}$. Now the first term on the right-hand side of (3.7) to first order equals $-1/\theta$, which implies that $\partial(\log h_r)/\partial\theta = 0$ requires $k(g(r,\theta)) \approx 1/r$. As $k(x)$ decreases in x, while $g(r,\theta)$ increases in θ, it follows that $g(r,\tilde{\theta}^{max}) = k^{-1}(1/r)$. In view of the definition of g, the result in (3.6) readily follows. □

Example 3.1. It is easily verified that $-\log(1-k^{-1}(1/r))$ equals 1.26, 1.90, 2.34 and 2.66 for $r = 2$, 3, 4, 5, respectively. For e.g. $\alpha = 0.01$ and $r = 3$ we then have $\tilde{\theta}^{max} = 5.12$, while for $r = 5$ we get $\tilde{\theta}^{max} = 3.34$. These values are indeed quite close to the corresponding exact values θ^{max}, which are 5.22 (with corresponding maximum 4.10 for h_3) and 3.37 (with maximum 4.22 for h_5), respectively. Moreover, for the negative binomial case we obtained very similar outcomes in Example 3.2 from Albers (2010): for $\alpha = 0.01$ and $r = 3$ the approximation again is also 5.12, with now an exact value 5.19 and an attained maximum 4.41. For $r = 5$ these figures become 3.34, 3.23 and 4.78, respectively. □

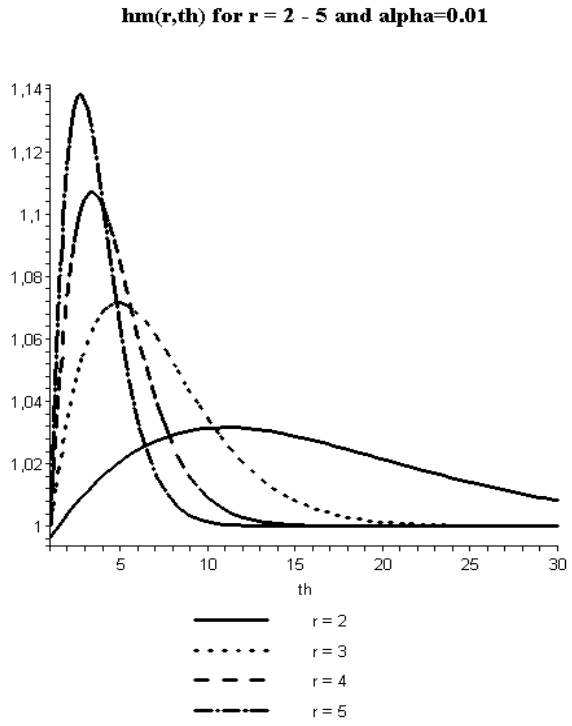

Figure 3.1:

As is illustrated by Example 3.1, considerable improvement over the geometric chart is achieved by both the negative binomial and the MAX-chart, be it with a slightly smaller gain for the latter case. This is not surprising, as in the fixed setup we are using, the negative binomial choice simply is optimal for the homogeneous case (cf. the $\overline{X}$-chart for the normal situation). The point is that, here as well, the loss incurred under these favorable circumstances, remains quite limited. To illustrate this, consider the functions

$$hm_{r,\theta} = \frac{ARL(MAX)_{r,\theta}}{ARL(NB)_{r,\theta}}, \tag{3.8}$$

where we have added 'MAX' to the ARL from (3.3) and 'NB' to the one from (2.5) to avoid confusion. A typical picture of hm_r for various r is presented in Figure 3.1. Indeed the excess over 1 stays very limited. See the Appendix for some additional technical comments on this somewhat intriguing situation.

Hence, taking the negative binomial chart as our benchmark, the additional step towards the MAX-chart is quite reasonable and we can indeed carry on. Of course, in passing we note that 'on the opposite side' of this negative binomial choice, quite a few alternative approaches exist as well, see e.g. Sego et al. (2008). As mentioned there, such procedures (sets method, $CUSUM$, $CUSCORE$, etc.) accumulate information over time and in this way are typically more efficient than fixed methods of Shewhart-type, such as the negative binomial chart. On the other hand, as is also observed in that paper, the latter type of procedure is easier to interpret and implement. For example, the $CUSUM$ approach suffers from numerical complications (Sego et al.(2008), p.1233). Moreover, the fact is duly noted that its implementation requires a reliable estimate of p (see p.1242) as well, but results on the impact of this step in this considerably more complicated context seem to be lacking.

4 The estimated version

Having established in the previous section that the MAX-chart performs quite well even if the homogeneity assumption holds after all, it makes sense to move on to the nonparametric approach for the case of an unknown distribution. In view of (3.1), we need to estimate $n = F^{-1}((r\alpha)^{1/r})$, where now F stands for the *df* of $X = X_{1,p}$ (and not of $X = X_{r,p}$, as was the case in section 2). On the basis of a Phase I sample of size m we obtain in complete analogy to (2.7) that

$$\hat{n} = F_m^{-1}((r\alpha)^{1/r}) = X_{(s)}, \tag{4.1}$$

where this time $s = \{m(r\alpha)^{1/r}]$, with again $\{y]$ denoting the smallest integer $\geq y$. Note that a configuration like $m = 100$, $r = 3$ and $\alpha = 0.001$ now leads to $s = \{14.4] = 15$, and thus the estimation is based on a much less extreme order statistic ($X_{(15)}$, as opposed to $X_{(1)}$ after (2.7)). Also observe that the step in (4.1) actually is all that is needed to define and to implement the nonparametric chart. Once $\hat{n}$ has been obtained, the actual monitoring can start: after each incoming group of r waiting times $Y_1, \ldots, Y_r$, a signal is given if

$$\max(Y_1, \ldots, Y_r) \leq X_{(s)}. \tag{4.2}$$

What remains is to analyze the behavior of the resulting chart. Fortunately, we can be brief here, as this closely parallels the study of the MIN-chart for the continuous case from AK (2008). Hence we shall quite often refer to this paper for (technical) details and further refinements. To begin with, note that the use of a stochastic lower limit $\hat{n}$ implies that the fixed $FAR = r\alpha$ during IC should now be replaced by the random quantity

$$\widehat{FAR} = P(\max(Y_1, \ldots, Y_r) \leq X_{(s)} \mid (X_1, \ldots, X_m)) = F^r(X_{(s)}). \tag{4.3}$$

From (4.3) it is immediate that this $\widehat{FAR}$ is distributed as $(U_{(s)})^r$, where $U_{(1)} < \cdots < U_{(m)}$ are order statistics from a sample of size m from the uniform *df* on $(0, 1)$. Hence during IC indeed no dependence on the underlying F remains and we e.g. simply have

$$E\widehat{FAR} = \Pi_{j=1}^{r} \frac{(s-1+j)}{(m+j)}, \tag{4.4}$$

and thus a relative bias $E\widehat{FAR}/FAR - 1 = \Pi_{j=1}^{r}[(s-1+j)/\{m(r\alpha)^{1/r}(1+j/m)\}] - 1$ results. In AK (2008) it is shown that this quantity behaves like c/m, with c typically nonnegative and for e.g. $\alpha = 0.001$ equal to at most 1000 for $r = 1$ and at most 40 for $r = 3 - 5$. This result shows two points: (i) using groups indeed helps a lot and (ii) a relative error $40/m$ still is not negligible for common sample sizes m. In view of the second remark, it makes sense to propose a correction in order to remove the bias. Quite a few details on this topic can be found in AK (2008). Here we merely give the most simplified version of the resulting proposal: replace $X_{(s)}$ from (4.1) by $X_{(s-r/2)}$ for r even and by $\frac{1}{2}X_{(s-(r+1)/2)} + \frac{1}{2}X_{(s-(r-1)/2)}$ for r odd. In the example with $r = 3$ after (4.1) thus replace $X_{(15)}$ by $\frac{1}{2}X_{(13)} + \frac{1}{2}X_{(14)}$. Hence using a slightly lower order statistic as our $\hat{n}$ effectively removes the positive bias in FAR.

A completely similar bias evaluation and correction exercise can be performed for $\widehat{ARL} = r/\widehat{FAR}$. Here we encounter the usual phenomenon that, contrary to simple intuition, both $\widehat{FAR}$ and $1/\widehat{FAR}$ have a positive bias (which moreover roughly has the same maximum in either case for this nonparametric application, cf. AK (2008)). The explanation is that the occurrence of (very) small values of $\widehat{FAR}$ leads to (very) large values of $1/\widehat{FAR}$, which inflate the expectation. As a consequence, bias removal in this case would require a slightly higher, instead of slightly lower, order statistic than $X_{(s)}$. Although mathematically correct, this type of solution does not seem to be attractive from a practical point of view. Moreover, it illustrates that bias reduction should not be the main and certainly not the only criterion to be considered.

In fact, bias correction merely serves to adjust the behavior of the chart over a long series of separate applications. It remains to deal with the variability in each individual application, which can e.g. be done by looking, for some small $\varepsilon > 0$, at the exceedance probability

$$P(W > \varepsilon), \text{ with } W = \widehat{FAR}/FAR - 1. \tag{4.5}$$

Clearly, the probability in (4.5) equals $P(\widehat{FAR} > FAR(1+\varepsilon))$, which can also be written as $P(\widehat{ARL} < ARL/(1+\varepsilon))$. In other words, a relative error of size ε in $\widehat{FAR}$ corresponds to one of size $-\varepsilon/(1+\varepsilon) \approx -\varepsilon$ in ARL. Hence dealing with (4.5) also settles the variability issue for $\widehat{ARL}$. E.g. for $\alpha = 0.001$ and $\varepsilon = 0.25$ we consider $P(\widehat{FAR} > 0.00125)$ and/or $P(\widehat{ARL} < 800)$: the corresponding event represents a considerable deviation from the intended value and thus it is interesting to know whether its probability is acceptably small.

To study this aspect, we derive, in analogy to (2.3), from (4.3) and (4.5) that

$$P(W > \varepsilon) = P(U_{(s)} > p^*) = P(Y_{m,p^*} < s), \tag{4.6}$$

with $p^* = \{r\alpha(1+\varepsilon)\}^{1/r}$. As $mp^* - (s-1) > \varepsilon m(r\alpha)^{1/r}/r$, the binomial probability in (4.6), and thus the exceedance probability, does tend to 0 as $m \to \infty$. However, especially for small r, the convergence is slow. In fact, Lemma 4.1 in AK (2008) states that for m large and ε small

$$P(W > \varepsilon) \approx \Phi(-\varepsilon m^{1/2} v(r,\alpha)), \tag{4.7}$$

where Φ is the standard normal *df* and $v(r,\alpha) = \{(r\alpha)^{1/r}/(1-(r\alpha)^{1/r})\}^{1/2}/r$. For example, for $r = 1, 2, \ldots, 5$ the factor $v(r, 0.001)$ in (4.7) equals 0.032, 0.108, 0.137, 0.145 and 0.146, respectively. Just as in the bias case, this shows that taking $r > 1$ provides a very substantial improvement, but also that $P(W > \varepsilon)$ remains unpleasantly high for common values like $\varepsilon = 0.25$ and $m = 100$: for $r = 3$ we e.g. obtain the value 0.37. Getting an ARL which falls more than 200 below the intended value 1000 in more than a third of the cases, may very well be considered still unsatisfactory.

Hence for the exceedance criterion as well corrections are called for. For extensive details, consult AK (2008); here we just mention that a desired upper bound β, for some small $\beta > 0$, will approximately hold for $P(W > \varepsilon)$ if the lower bound $X_{(s)}$ is replaced by $X_{(s^*)}$, where

$$s^* = s(1 + \varepsilon/r) - u_\beta\{s(1 - s/m)\}^{1/2}, \tag{4.8}$$

with $u_\beta = \Phi^{-1}(1-\beta)$, the standard normal upper β-quantile. Usually s^* will be no integer, but this is easily remedied by suitable interpolation between the neighboring order statistics $X_{(\{s^*\]-1)}$ and $X_{(\{(s^*\])}$. Note that the corrected version in (4.8) is quite straightforward to interpret: increasing ε lowers $P(W > \varepsilon)$, which makes it easier to satisfy $P(W > \varepsilon) \leq \beta$ and thus s^* increases as well in that case. On the other hand, decreasing β means a more strict bound and s^* will then decrease, as u_β increases. Moreover, the second term in (4.8) will typically dominate the first and therefore s^* will be smaller than s. Only if m is really large (much larger than common in practice), the correction becomes superfluous. For ordinary m, however, it will be somewhat larger than the bias correction. This agrees with the fact that here we are controlling the individual behavior, rather than merely the overall average over a large number of separate applications. E.g. using once more $\alpha = 0.001$, $\varepsilon = 0.25$ and $m = 100$, a value $\beta = 0.2$ produces $s^* = 13.25$. Hence using $\{3X_{(13)} + X_{(14)}\}/4$ reduces the exceeding probability $P(\widehat{ARL} < 800)$ from the value 0.37 calculated above to 0.20.

The final topic to be covered concerns the OoC-behavior of the (un)corrected charts. A detailed and technical discussion of the issues involved can be found in section 5 and 6 of AK (2008). To avoid repetition, we shall be very brief here and mainly state the relevant conclusions. A first point is the comparison between the case of known and that of unknown F. In fact, no real comparison is involved here: if F is unknown, it is unavoidable to apply some kind of estimation step. The only precaution one can take is to select as starting point a chart that compares reasonably well with the standard procedure for the known case. This issue has been addressed extensively in section 3 for the MAX-chart versus the negative binomial one.

The next point concerns the comparison in the unknown case between the corrected and the uncorrected proposals. In a sense, no real comparison is involved here as well. Obviously, lowering $X_{(s)}$ by e.g. using $s - r/2$ rather than s itself, will somewhat lower the detection power during OoC as well. But this advantage of the uncorrected chart is only achieved by 'cheating' with respect to the intended value $r\alpha$ of FAR. Just compare this to the more generally familiar case of hypothesis testing. Suppose two tests are proposed, the first of which is forced to maintain the prescribed significance level, whereas for the second a positive bias in this respect is allowed. Clearly, that advantage can be used to produce a higher power for the latter test, but such a gain will generally be recognized as quite dubious.

Yet another observation is that generally speaking the impact of the estimation step is considerably smaller during OoC than during IC. An explanation for this effect is that the impact of estimation is huge during IC because of the very small probabilities involved. However, during OoC the alarm rate will be considerably larger than the FAR, and the relative error due to estimation thus (much) smaller. Note that this also holds for the impact of the corrections. See e.g. Lemma 5.1 from AK (2008).

5 Summary

For convenience, we summarize the application of the nonparametric chart as discussed in the previous sections:

1. Select a desired in-control $ARL = 1/\alpha$ and a degree of change θ in the average failure probability during OoC that should be optimally protected against.
2. Apply rule of thumb (3.5) to obtain the best r (typically truncate at 5 in practice).
3. Select a size m (e.g. $m = 100$) and collect a Phase I sample of geometric $X_1, \ldots, X_m$.
4. Compute the smallest integer $s \geq m(r\alpha)^{1/r}$ and find the order statistic $X_{(s)}$ (cf. 4.1).
5. Start monitoring: after each r geometric $Y_1, \ldots, Y_r$, signal if $\max(Y_1, \ldots, Y_r) \leq X_{(s)}$.
6. If desired, replace s by $s - r/2$ to remove the bias in FAR (interpolate for r odd.
7. If desired, select small ε and β such that $P(\widehat{FAR} < 1/\{\alpha(1+\varepsilon\}) = \beta$ is acceptable.
8. Replace s by $s^* = s(1+\varepsilon/r) - \Phi^{-1}(1-\beta)\{s(1-s/m)\}^{1/2}$ (cf.(4.8)) to achieve this bound.

Appendix

The behavior of the functions $hm_{r,\theta}$ from (3.7)

To see why these $hm_{r,\theta}$ remain so close to 1, we proceed as follows. From (2.3) and (2.5) it follows that $ARL(NB) \approx r/P(Z_{\theta\lambda} \geq r)$, with λ such that $P(Z_\lambda \geq r) = r\alpha$. In combination with (3.3) this shows that $ARL(MAX) \approx r/\{1 - (1 - \{P(Z_\lambda \geq r)\}^{1/r})^\theta\}^r$. Hence (3.7) leads to

$$hm_{r,\theta} \approx \frac{P(Z_{\theta\lambda} \geq r)}{\{1 - (1 - \{P(Z_\lambda \geq r)\}^{1/r})^\theta\}^r}. \tag{A.1}$$

To ensure that $hm_{r\theta} \approx 1$, it suffices in view of (A.1) to have $g_r(\theta\lambda) \approx \{g_r(\lambda)\}^\theta$, where $g_r(\lambda) = 1 - \{P(Z_\lambda \geq r)\}^{1/r}$. In other words, g_r should satisfy $-\log g_r(\lambda) \approx c_r\lambda$ for some constant c_r. Indeed, using Maple shows that this linearity is nearly perfect for $r = 2$ or 3 and very reasonable for $r = 4$ or 5. This holds true for a wide range of λ (up to 20 for $r \leq 4$ and up to 10 for $r = 5$). The values obtained for c_r are 0.86, 0.75, 0.64 and 0.50 for $r = 2$, 3, 4 and 5, respectively. No obvious further explanation seems available. Note that simple Taylor expansion (which anyhow only holds for λ small) produces that $\log(g_r(\lambda)) \approx -\{P(Z_\lambda \geq r)\}^{1/r} \approx -c_r^*\lambda$ with $c_r^* \approx (r!)^{-1/r}$. This leads to the substantially lower values 0.71, 0.55, 0.45 and 0.38 for $r = 2$, 3, 4 and 5, respectively.

Within the wide range of λ mentioned above, the focus of course is on values such that $P(Z_\lambda \geq r) = r\alpha$ for α small, e.g. $0.001 \leq \alpha \leq 0.01$. Some representative values (cf. Table 1 from Albers (2010)) are $\lambda_r = 0.15$, 0.45, 0.85 and 1.5 for $r = 2$, 3, 4 and 5, respectively. Taking a closer look for these examples using Maple reveals that $\log(g_r(\theta\lambda_r))/\{\theta\log(g_r(\lambda_r))\} \approx 1 + \tilde{c}_r(\theta - 1)$, with $\tilde{c}_r = 0.0030$, 0.012, 0.025 and 0.05, respectively. Again, this approximate linearity holds very well over a wide region of θ (up to 30 for $r = 2$, 3, up to 20 for $r = 4$ and up to 10 for $r = 5$). Clearly, for really large θ (e.g. $\theta \geq 4$), the correction $\tilde{c}_r(\theta - 1)$ becomes substantial, especially if r grows as well. However, do note that for such large θ the g_r have become quite small. Hence in that case $hm_{r,\theta} \approx (1 - g_r(\theta\lambda_r))^r/(1 - \{g_r(\lambda_r)\}^\theta)^r \approx 1 - r\{g_r(\lambda_r)\}^\theta[g_r(\theta\lambda_r)/\{g_r(\lambda_r)\}^\theta - 1] \approx 1 - r\tilde{c}_r(\theta - 1)\{g_r(\lambda_r)\}^\theta\log(g_r(\lambda_r)$. Consequently, the increase due to $\tilde{c}_r(\theta - 1)$ is dampened by the factor $-r\{g_r(\lambda_r)\}^\theta\log(g_r(\lambda_r))$, as can be seen from Figure 3.1. □

Acknowledgement The author is grateful to Wilbert Kallenberg for his contribution leading to the Appendix.

References

Albers,W. (2009). Control charts for health care monitoring under overdispersion. TW-report 1891. To appear in Metrika.

Albers, W. (2010). The optimal choice of negative binomial charts for monitoring high-quality processes. *J. Statist. Planning & Inference* **140**, 214-225.

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. (2008). Minimum control charts. *J. Statist. Planning & Inference* **138**, 539-551.

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

Chen, R. (1978). A surveillance system for congenital malformations. *J. Amer. Statist. Ass.* **73**, 323-327.

Chen, R., Connelly, R.R. and Mantel, N. (1997). The efficiency of the sets and the cuscore techniques under biased baseline rates. *Statist. in Med.* **16**, 1401-1411.

Christensen, A., Melgaard, M., Iwersen, J., and Thyregod, P. (2003). Environmental Monitoring Based on a Hierarchical Poisson-Gamma Model. *J. Qual. Technol.* **35**, 275-285.

Fang, Y. (2003). c-Charts, X-Charts, and the Katz Family of Distributions. *J. Qual. Technol.* **35**, 104-114.

Poortema, K. (1999). On modelling overdispersion of counts. *Statist.Neerl.* **53**, 5-20.

Shaha, S H. (1995). Acuity systems and control charting. *Qual. Manag. Health Care* **3**, 22-30.

Sego, L.H, Woodall, W.H. and Reynolds, M.R. (2008) A comparison of surveillance methods for small incidence rates. *Statist. in Med.* **27**, 1225-1247.

Sonesson, C. and Bock, D. (2003). A review and discussion of prospective statistical surveillance in public health. *J. R. Statist. Soc. A* **166**, 5-21.

Thor, J., Lundberg, J., Ask, J., Olsson, J., Carli, C., Härenstam, K.P. and Brommels, M. (2007). Application of statistical process control in healthcare improvement: systematic review. *Qual. & Safety in Health Care* **16**, 387-399.