Metrika (2009) 70:111–130
DOI 10.1007/s00184-008-0184-5

CUMIN charts

Willem Albers · Wilbert C. M. Kallenberg

Received: 2 July 2007 / Published online: 26 March 2008

Abstract Classical control charts are very sensitive to deviations from normality. In this respect, nonparametric charts form an attractive alternative. However, these often require considerably more Phase I observations than are available in practice. This latter problem can be solved by introducing grouping during Phase II. Then each group minimum is compared to a suitable upper limit (in the two-sided case also each group maximum to a lower limit). In the present paper it is demonstrated that such *MIN* charts allow further improvement by adopting a sequential approach. Once a new observation fails to exceed the upper limit, its group is aborted and a new one starts right away. The resulting *CUMIN* chart is easy to understand and implement. Moreover, this chart is truly nonparametric and has good detection properties. For example, like the *CUSUM* chart, it is markedly better than a Shewhart X-chart, unless the shift is really large.

Keywords Statistical process control · Phase II control limits · Order statistics · CUSUM-chart

1 Introduction and motivation

By now it is well-known that standard control charts for controlling the mean of a production process, such as the Shewhart or *CUSUM* chart (see, e.g., Page 1954; Lorden 1971), are highly sensitive to deviations from normality (see, e.g., Chan et al. 1988; Pappanastos and Adams 1996; Hawkins and Olwell 1998, p. 75, Albers and Kallenberg 2004, 2005b). Let us take the Shewhart X-chart for individual observations

W. Albers (✉) · W. C. M. Kallenberg
Department of Applied Mathematics, University of Twente,
P.O. Box 217, 7500 AE Enschede, The Netherlands
e-mail: w.albers@utwente.nl

(which we shall denote by *IND*) as a starting point. Here an out-of-control (*OoC*) signal immediately occurs once an incoming observation falls above an upper limit UL or below a lower limit LL. While the process is in-control (IC), the false alarm rate (*FAR*) should equal some small p, like $p = 1/1,000$ or $1/500$. Even if we assume that the observations come from a normal distribution, typically its parameters are unknown. An initial sample of size n (the so-called Phase I observations) is then needed already to estimate these parameters and subsequently the UL and LL. Conditional on the n Phase I observations, the *FAR* of the corresponding estimated chart now also is a random variable (rv) P_n, and this P_n shows considerable variation around the intended p. In fact, quite large values of n are required before this stochastic error (*SE*) becomes negligible. Just see Albers and Kallenberg (AK for short) (2005a), which provides a recent non-technical review of the results available, as well as additional references.

However, if normality fails, we actually estimate the wrong control limits and P_n is not even consistent for p anymore. In addition to the *SE*, we thus have a nonvanishing model error (*ME*). A first remedy is to consider wider parametric families, i.e., to better adapt the distribution used to the data at hand by supplying (and estimating) more than just two parameters. In this way, this *ME* can often be reduced substantially, be it at the cost of a somewhat further increase of the *SE* (see, e.g., Albers et al. 2004). The natural endpoint in this respect is a fully nonparametric approach: see, e.g., Bakir and Reynolds (1979), Bakir (2006), Chakraborti et al. (2001, 2004), Qiu and Hawkins (2001), and Qiu and Hawkins (2003), as well as Albers and Kallenberg (2004). In the latter paper the control limits are simply based on empirical quantiles, i.e., appropriate order statistics, of the initial sample. In this way, the *ME* is indeed removed completely, but the price will typically be a huge *SE*, unless n is very large. By way of example, consider a customary value like $n = 100$ and then realize the difficulty of subsequently estimating the upper and lower $1/1,000$-quantiles in a nonparametric way. Hence, as each type of chart has its own potential drawback, a sensible overall approach thus is to adopt a data driven method (see Albers et al. 2006): let the data decide whether it is safe to stick to a normality based chart, or, if not, whether estimating an additional parameter offers a satisfactory solution. If neither is the case, a nonparametric approach is called for, which will be fine if n is sufficiently large.

Consequently, what does remain is the need for a satisfactory nonparametric procedure for ordinary n. This problem has subsequently been successfully addressed by Albers and Kallenberg (2006, 2008). The idea is to group the observations during the monitoring phase. Hence the decision to give a signal is no longer based on a single incoming observation, but instead on a group of size m, with $m > 1$ (with $m = 1$ we are back in the boundary case *IND*). The question which choice is best, is more complicated than it might seem at first sight, even if we restrict attention (as is quite customary) to *OoC* behavior characterized by a shift d. In fact, it is twofold: (i) what m should we take, and (ii) which group statistic? Consequently, this problem is dealt with first in Albers and Kallenberg (2006) for the case of known, not necessarily normal, underlying distributions. Afterwards, the estimation aspects—which form the very motivation to consider grouping at all—are the topic of Albers and Kallenberg (2008).

Because of its optimality under normality, the obvious group statistic is the average, or equivalently, the sum. The corresponding chart is nothing but a Shewhart $\overline{X}$-chart chart, which we will denote by *SUM* (or occasionally by $SUM(m)$) here. It is easily

verified that the optimal value of m decreases in d. In fact, for larger d, $SUM(1) = IND$ is best, but for a wide range of d-values of practical interest, a choice of m between say 2 and 5 will provide better performance. Incidentally, this is in line with the observed superiority of *CUSUM* over *IND* for d not too large; we will come back to this point in Sect. 3. However, we should realize that all of the foregoing assumes normality; once this assumption is abandoned, *SUM* is no longer optimal. Even worse, it is also difficult to adapt it to the nonparametric case. Approximations based on the central limit theorem are simply not at all reliable, as m is small and we are dealing with the tails. Moreover, a direct approach (see Albers and Kallenberg 2005b) leads to interesting theoretical insights into the tail behavior of empirical distribution functions for convolutions, but does not help much as far as practical implementation is concerned: the estimation still requires an n which is typically too large.

Consequently, there remains a definite need to consider alternative choices for the group statistic. Now a very good idea turns out to be using the minimum of the m observations in the group in connection with some upper limit (and thus the group maximum with a lower limit). The corresponding chart we have called *MIN* (see Albers and Kallenberg 2006). Just like *SUM*, it beats *IND*, unless d becomes quite large. Of course, under normality it is (somewhat) less powerful than *SUM*, but outside the normal model, the roles can easily be reversed. Hence, even for known distributions, *MIN* is a serious competitor for *SUM*. However, as soon as we drop this artificial assumption, the attractiveness of *MIN* becomes fully apparent. For, as we just argued, in this nonparametric setting *SUM* can easily lead to a large *ME* if we continue to assume normality, while its nonparametric adaptation is no success. On the other hand, the nonparametric version of *IND* is simple, but has a huge *SE* unless n is very large. In fact, this was what prompted us to consider grouping.

Hence with both *SUM* and *IND* we run into trouble. However, *MIN* has a straightforward nonparametric adaptation, and hence $ME = 0$, just like the nonparametric *IND*. Moreover, unlike *IND*, it turns out to have an *SE* which is quite well-behaved and comparable to that of the normal *SUM* chart. The intuitive explanation is actually quite simple: application of *MIN* requires estimation of much less extreme quantiles than *IND* or *SUM*. Take e.g., $m = 3$, then the upper $1/10$-quantile is exceeded by a group minimum with probability $(1/10)^3 = 1/1,000$, which is the same small value as before. But estimating an upper $1/10$-quantile on the basis of a sample of size $n = 100$ is quite feasible, i.e., leads to a very reasonable *SE*. Hence (only) for *MIN*, both *ME* and *SE* are under control! As a consequence, the conclusion from Albers and Kallenberg (2008) is quite positive towards this new chart: it is easy to understand and to implement, it is truly nonparametric and its power of detection is comparable to that of the standard, normality based, charts using sums.

After this favorable conclusion, the question arises whether there is room for further improvement. Specifically, having mentioned the *CUSUM* chart before, and having remarked that for not too large shifts this chart is superior compared to *IND*, the idea suggests itself that a cumulative or sequential version of *MIN* might serve this purpose. In the present paper we shall demonstrate that this is indeed the case. Not surprisingly, we will call the corresponding proposal a *CUMIN* chart. In Sect. 2 we will introduce these charts in a systematic manner, taking once more the case of a known underlying distribution as our starting point (cf. Albers and Kallenberg 2006). The focus will

be on demonstrating that *CUMIN* remains quite easy to understand and implement. Section 3 is devoted to studying the performance during *OoC* and comparing it to that of its competitors. In Sect. 4 the artificial assumption of known underlying distribution is abandoned and it is shown how the estimated version of the chart is obtained.

2 Definition and basic properties of *CUMIN*

Let X be a random variable (rv) with a continuous distribution function (df) F. As announced, we shall begin by assuming that F is known. Hence for now, there is no Phase I sample: we start immediately with the monitoring phase for the incoming $X_1, X_2, \ldots$. For ease of presentation, we shall mainly concentrate on the one-sided case; only occasionally we shall consider the two-sided case, which can be treated in a completely similar fashion. (Merely keep in mind to switch from *(CU)MIN* to *(CU)MAX* at the lower control limit.) First consider *IND*, the individual case with $m = 1$. Hence for given p, we need UL such that $P(X > UL) = p$ during *IC*. For any df H we write $\overline{H} = 1 - H$ and H^{-1} and $\overline{H}^{-1}$ for the respective inverse functions, and thus $UL = F^{-1}(1-p) = \overline{F}^{-1}(p)$.

Next we move on to the grouped case, where $m > 1$ and consider for the first group

$$T = T(m) = min\,(X_1, \ldots, X_m) \tag{2.1}$$

as our control statistic for the upper *MIN* chart. (Here and in what follows we add '(m)' to the quantities we define when needed to avoid confusion, but often we use the abbreviated notation.) As in this case $P(T > UL) = \overline{F}(UL)^m$ during IC, it follows that a fair comparison to *IND* is obtained by choosing $UL = UL(m) = \overline{F}^{-1}((mp)^{1/m})$, leading to $FAR = P(T > UL) = mp$. To see this, note that in this way the average run length (*ARL*) will be $m/FAR = 1/p$, which thus agrees with the *ARL* of *IND* based on $UL = \overline{F}^{-1}(p)$. During *OoC*, we consider a shift $d > 0$, i.e., the X_i will have df $F(x-d)$. Thus we immediately have that in this case we obtain for the *ARL* of *MIN* that

$$ARL_M(m,d) = \frac{m}{P(T > UL)} = \frac{m}{\{\overline{F}(UL-d)\}^m} = \frac{m}{\{\overline{F}(\overline{F}^{-1}((mp)^{1/m}) - d)\}^m}. \tag{2.2}$$

Clearly, $ARL_M(m,0) = 1/p$ again. Moreover, by looking at $ARL_M(1,d) - ARL_M(m,d)$ and/or $ARL_M(m,d)/ARL_M(1,d)$, we can compare the performance of *MIN* to that of *IND*. As demonstrated in Albers and Kallenberg (2006), the conclusion is that *MIN* is better than *IND* for a wide range of d values of practical interest. Only for large d, *IND* is best.

Note that the above holds for arbitrary F, and not just for the normal case. For the sake of comparison, we shall now also briefly consider the *SUM* chart (i.e., the Shewhart $\overline{X}$-chart). However, here normality is more or less required: for general F, we wind up with rather intractable convolutions. So let Φ denote the standard normal *df* and suppose that $F(x) = \Phi((x-\mu)/\sigma)$. Actually, since we are in the case of

known F, we can take $\mu = 0$ and $\sigma = 1$ without loss of generality, and thus $F = \Phi$. In the case of *SUM*, we replace T in (2.1) by the standardized *SUM* of the first group $X_1, \ldots, X_m$:

$$T = T(m) = m^{-1/2} \sum_{i=1}^{m} X_i = m^{1/2} \overline{X}. \tag{2.3}$$

Clearly, T then has df Φ as well and thus the choice $UL = \overline{\Phi}^{-1}(mp)$ will produce the desired $ARL = 1/p$ for $F = \Phi$. It is also straightforward that under $\Phi(x - d)$

$$ARL_S(m, d) = \frac{m}{\overline{\Phi}(\overline{\Phi}^{-1}(mp) - m^{1/2}d)}. \tag{2.4}$$

Again under $F = \Phi$, studying $ARLS_S(1, d) - ARL_S(m, d)$ and/or $ARLS_S(m, d)/ARL_S(1, d)$ makes sense for comparing the performance of *SUM* and *IND*. Once more the resulting picture is that *IND* is preferable only for rather large d (see Albers and Kallenberg 2006 for details). Likewise $ARL_M(m, d) - ARL_S(m, d)$ and/or $ARL_S(m, d)/ARL_M(m, d)$ can be studied in order to compare *SUM* and *MIN* (cf. Albers and Kallenberg 2006 again).

In the above we have introduced and described *IND*, *MIN* and *SUM*. Now we are in a position to move on to the cumulative or sequential approach. As announced in the Introduction, the idea is actually quite simple. Just look at the *MIN* chart for some given m. Then each time a complete group of size m is assembled, its minimum value T from (2.1) is computed and this T is subsequently compared to $UL = \overline{F}^{-1}((mp)^{1/m})$. But of course, as soon as an observation occurs within such a group which falls below this *UL*, it makes no sense to complete that group and we could as well stop right away. The next observation will then be the first of a new attempt. This idea leads to the following definition of a sequential *MIN* procedure:

"Give an alarm at the 1st time m consecutive observations all exceed some UL" (2.5)

In other words, this *CUMIN* chart is an accelerated version of *MIN*: before the final successful attempt to get m consecutive $X_i > UL$, the failed ones are broken of as soon as possible, rather than letting these all reach length m as well.

The proposal in (2.5) is inspired by the representation of *CUSUM* which can be found, e.g., in Page (1954) and Lorden (1971). The alternative form of *CUSUM* from, e.g., Lucas (1982) leads to an alternative for (2.5) as well. Let $I(A)$ be the indicator function of the set A and set $S_0 = 0$. Consider $S_i = I(\{X_i > UL\})(1 + S_{i-1})$, $i = 1, 2, \ldots$, and give an alarm as soon as $S_k \geq m$ for some k.

Next we shall investigate the properties of *CUMIN*. In (2.5) we have deliberately been a bit vague ('some *UL*'). Indeed, the *UL* for *CUMIN*, say $\overline{F}^{-1}(\tilde{p})$, will have to be different from $\overline{F}^{-1}((mp)^{1/m})$, the *UL* of *MIN*. As *CUMIN* reacts more quickly than *MIN*, it is evident that its *UL* will have to be somewhat larger, i.e., $\tilde{p} < (mp)^{1/m}$ will hold. To find this $\tilde{p}$ exactly, a bit more effort is required. First let us introduce some notation. By 'Y *is* $G(\theta)$' we will mean that the rv Y has a geometric distribution with

parameter θ, and thus that $P(Y = k) = \theta(1-\theta)^{k-1}$, for $k = 1, 2, \ldots$. Moreover, by 'Z *is* $G_m(\theta)$' we will mean that the rv Z has an m-truncated geometric distribution with parameter θ, which is defined through $P(Z = k) = P(Y = k|Y \leq m), k = 1, \ldots, m$, where Y is $G(\theta)$. Clearly, $G_\infty = G$ again. Finally, let RL denote the run length of a chart (and thus $E(RL) = ARL$). Then we have the following result.

Lemma 2.1 *For the CUMIN chart defined in (2.5), with* $UL = \overline{F}^{-1}(\tilde{p})$, *the run length is distributed as*

$$RL = m + \sum_{i=1}^{V-1} B_i, \tag{2.6}$$

where $V, B_1, B_2, \ldots,$ *are independent rv's and moreover* V *is* $G(\tilde{p}^m)$ *and the* B_i *are* $G_m(1-\tilde{p})$. *Consequently,*

$$E(RL) = \frac{1-\tilde{p}^m}{(1-\tilde{p})\tilde{p}^m} = \frac{1}{1-\tilde{p}}\left(\frac{1}{\tilde{p}^m} - 1\right),$$

$$var(RL) = \frac{1-\tilde{p}^m}{\{(1-\tilde{p})\tilde{p}^m\}^2}\left\{1 + \frac{\tilde{p}^m\{\tilde{p} - 2m(1-\tilde{p})\}}{1-\tilde{p}^m}\right\}. \tag{2.7}$$

Before proving Lemma 2.1 we present the following general result on m-truncated distributions.

Lemma 2.2 *Let* $B_1^*, B_2^*, \ldots,$ *be independent and identically distributed (iid) rv's with* $P(B_1^* > m) > 0$ *and df* H. *Let* $V = min\{k : B_k^* > m\}$. *Then, conditional on* $V = v$, *the rv's* $B_1^*, \ldots, B_{v-1}^*$ *are iid with df* H_m *given by*

$$H_m(b) = \frac{H(b)}{H(m)} \; for\; b \leq m \; and \; H_m(b) = 1 \; for \; b > m.$$

Moreover, there exist rv's $B_1, B_2, \ldots,$ *such that* $V, B_1, B_2, \ldots,$ *are independent,* B_i *has df* H_m *and for each function* g *the rv's* $g(B_1^*, \ldots, B_{V-1}^*)$ *and* $g(B_1, \ldots, B_{V-1})$ (*with* g *equal to some constant if* $V = 1$) *have the same distribution.*

Proof By definition of V, the event $\{V = v\} = \{B_1^* \leq m, \ldots, B_{v-1}^* \leq m, B_v^* > m\}$. Hence, we obtain for $b_1, \ldots, b_{v-1} \leq m$, using the independence of $B_1^*, B_2^*, \ldots,$

$$\begin{aligned} P(B_1^* \leq b_1, \ldots, B_{v-1}^* \leq b_{v-1}|V = v) &= \frac{P(B_1^* \leq b_1, \ldots, B_{v-1}^* \leq b_{v-1}, B_v^* > m)}{P(B_1^* \leq m, \ldots, B_{v-1}^* \leq m, B_v^* > m)} \\ &= \Pi_{i=1}^{v-1}\left\{\frac{P(B_i^* \leq b_i)}{P(B_i^* \leq m)}\right\} = \Pi_{i=1}^{v-1} H_m(b_i) \end{aligned}$$

and the first result easily follows. Define rv's $B_1, B_2, \ldots,$ such that $V, B_1, B_2, \ldots,$ are independent and B_i has df H_m. Note that H_m, the conditional df of $B_1^*, \ldots, B_{v-1}^*$

given $V = v$, does not depend on v, and hence the B_i can be defined as above. Now we have for any x

$$\begin{aligned} P(g(B_1^*, \ldots, B_{V-1}^*) \le x) &= \sum_{v=1}^{\infty} P(g(B_1^*, \ldots, B_{v-1}^*) \le x | V = v) P(V = v) \\ &= \sum_{v=1}^{\infty} P(g(B_1, \ldots, B_{v-1}) \le x) P(V = v) \\ &= \sum_{v=1}^{\infty} P(g(B_1, \ldots, B_{v-1}) \le x, V = v) \\ &= P(g(B_1, \ldots, B_{V-1}) \le x). \end{aligned}$$

□

Proof of Lemma 2.1. Consider two forms of blocks of experiments for the sequence $X_1, X_2, \ldots$. The first one is related to the *MIN* chart and consists of fixed blocks of size $m : W_1 = (X_1, \ldots, X_m)$, $W_2 = (X_{m+1}, \ldots, X_{2m}), \ldots$. Obviously, $W_1, W_2, \ldots$ are iid. The second one concerns the *CUMIN* chart. The first block now ends with the first $X_i \le UL$. This gives W_1. The second block starts with the next X and ends with the second $X_i \le UL$. This produces W_2, and so on. Again, $W_1, W_2, \ldots$ are iid. In both situations the experiment W_i is called successful if at least m X's in W_i satisfy $X_i > UL$. Hence the probability of success in experiment W_i equals $\theta = \tilde{p}^m$ in either situation. Let V be the waiting time till the first successful experiment W_i, then V is indeed $G(\tilde{p}^m)$. For the *MIN* chart we simply have $RL = mV$ and $E(RL) = m/\tilde{p}^m$ shows that in that case choosing $\tilde{p} = (mp)^{1/m}$ indeed produces $E(RL) = ARL = 1/p$.

For the second situation define B_i^* as the length of the vector W_i. Since $W_1, W_2, \ldots,$ are iid, the rv's $B_1^*, B_2^*, \ldots,$ are also iid. Furthermore, the experiment W_i is successful if $B_i^* > m$ and hence $V = min\{k : B_k^* > m\}$. In view of (2.5) we have that $RL = m + \sum_{i=1}^{V-1} B_i^*$. The first part of Lemma 2.1 now follows by application of Lemma 2.2 with $g(B_1, \ldots, B_{V-1}) = m + \sum_{i=1}^{V-1} B_i$, noting that B_i^* is the first time that we get $X \le UL$ and thus B_i^* is $G(1 - \tilde{p})$.

To obtain the moments in (2.7), let Y be $G(\theta)$ and Z be $G_m(\theta)$. For $r = 1, 2, \ldots,$ we observe that the memoryless property of the geometric distribution produces $E(Y + m)^r = \sum_{k=1}^{\infty}(k+m)^r P(Y = k+m | Y > m) = \sum_{k=m+1}^{\infty} k^r P(Y = k)/P(Y > m) = \{EY^r - EZ^r P(Y \le m)\}/P(Y > m)$ and thus $EZ^r = \{EY^r - E(Y + m)^r P(Y > m)\}/P(Y \le m)$. For $r = 1$ this gives $EZ = EY - mP(Y > m)/P(Y \le m) = 1/\theta - m(1-\theta)^m/\{1 - (1-\theta)^m\}$. Hence $E(RL) = m + E(V - 1)EB = m + (1/\tilde{p}^m - 1)\{1/(1-\tilde{p}) - m\tilde{p}^m/(1-\tilde{p}^m)\}$ and the first result in (2.7) follows. Moreover, applying the result above for $r = 2$ as well leads to $var(Z) = var(Y) - m^2 P(Y > m)/\{P(Y \le m)\}^2 = (1-\theta)/\theta^2 - m^2(1-\theta)^m/\{1-(1-\theta)^m\}^2$. It remains to use that $var(RL) = (EB)^2 var(V) + var(B)(EV - 1)$ in order to obtain the second result in (2.7). □

Remark 2.1 $E(RL)$ can also be obtained by applying renewal theory (see, e.g., Ross 1996). Instead of (2.6), use the representation $RL = m - C_V + \sum_{i=1}^{V} C_i$, where the C_i are simply $G(1-\tilde{p})$. As $EC_V = m + 1/(1-\tilde{p})$, while Wald's equation gives $E(\sum_{i=1}^{V} C_i) = EVEC_1 = 1/\{\tilde{p}^m(1-\tilde{p})\}$, the first line in (2.7) again follows. □

From (2.7) it follows that $ARL = 1/p$ will result if $\tilde{p}$ is chosen such that

$$\frac{(1-\tilde{p})\tilde{p}^m}{1-\tilde{p}^m} = p, \tag{2.8}$$

As p is very small, $\tilde{p}^m$ will be of the order p, and hence as a first approximation we have $\tilde{p}^m \approx p/(1-p^{1/m})$, i.e.,

$$\tilde{p} \approx \left(\frac{p}{1-p^{1/m}}\right)^{1/m}. \tag{2.9}$$

This already is quite accurate; if desired, (2.9) can be replaced by $\tilde{p} \approx \{p/(1-[p/(1-p^{1/m})]\}^{1/m})\}^{1/m}$, which is very precise. Note that the interpretation of (2.9) is still rather simple: the failed sequences of fixed length m for *MIN* are replaced by sequences of expected length approximately $1/(1-\tilde{p})$ for *CUMIN*. Hence the total expected length changes from $m/\tilde{p}^m$ to about $1/\{(1-\tilde{p})\tilde{p}^m\}$ and thus the former solution $(mp)^{1/m}$ becomes (2.9). Indeed, $1/(1-p^{1/m})$ is considerably smaller than m : for $p = 0.001$, e.g., 1.11 for $m = 3$ and 1.46 for $m = 6$.

Next we note that the fact that $\tilde{p}^m$ is of order p implies in view of (2.7) that $var(RL) \approx 1/\{(1-\tilde{p})\tilde{p}^m\}^2$. This leading term is essentially due to $(EB)^2 var(V)$; the second part $var(B)(EV-1)$ of $var(RL)$ just gives a lower order contribution. In other words, the RL of *CUMIN* behaves to first order as $V/(1-\tilde{p})$ (cf. the RL of *MIN* which exactly equals mV). Moreover, if $\tilde{p}$ satisfies (2.8), it follows that $var(RL) \approx 1/p^2$. Hence the simple conclusion is that the RL of the *CUMIN* chart from Lemma 2.1 with $\tilde{p}$ selected such that (2.8) holds, behaves like a $G(\tilde{p}^m)/(1-\tilde{p})$ rv. By way of illustration, we give:

Example 2.1 For $p = 0.001$ and $m = 3$ we obtain that $\tilde{p} = 0.103677$ and $\tilde{p}^m = 0.001114$. The approximation from (2.9) leads to $\tilde{p} = 0.103574$ and $\tilde{p}^m = 0.001111$, which produces 0.000997 rather than $p = 0.001$ in (2.8). The refinement below (2.9) gives $\tilde{p} = 0.103712$ and $\tilde{p}^m = 0.001116$, which gives 0.001001 in (2.8). (We have dragged along more digits than would be useful in practice, just to show the differences.) Roughly speaking, the RL behaves like 10/9 times a $G(1/900)rv$.

If we choose instead $m = 6$, the results become $\tilde{p} = 0.338708$ and $\tilde{p}^m = 0.001510$. The approximation from (2.9) then leads to $\tilde{p} = 0.336911$ and $\tilde{p}^m = 0.001462$, which produces 0.000971 rather than $p = 0.001$ in (2.8). The refinement below (2.9) leads to $\tilde{p} = 0.338640$ and $\tilde{p}^m = 0.001508$, and 0.000999 as the result of (2.8). Here RL is roughly 3/2 times a $G(3/2000)$ rv. □

We summarize the previous discussion with the following formal result.

Lemma 2.3 *Let $\tilde{p}$ be defined by (2.8) and let V be $G(\tilde{p}^m)$. Then, for $p \to 0$,*

$$E(RL) = E\left(\frac{V}{1-\tilde{p}}\right) - \frac{1}{1-\tilde{p}} = E\left(\frac{V}{1-\tilde{p}}\right)(1+O(p))\,, \tag{2.10}$$

$$var(RL) = var\left(\frac{V}{1-\tilde{p}}\right)\left\{1+\tilde{p}^m\,\frac{\tilde{p}-2m(1-\tilde{p})}{1-\tilde{p}^m}\right\} = var\left(\frac{V}{1-\tilde{p}}\right)(1+O(p))\,. \tag{2.11}$$

Proof Let $h(x) = (1-x)x^m/(1-x^m)$, then $h(\tilde{p}) = p$. For any ε we obtain that $lim_{p\to 0} h(p^{1/m}(1+\varepsilon))/p = (1+\varepsilon)^m$ and hence

$$\tilde{p} = p^{1/m}(1+o(1)) \tag{2.12}$$

as $p \to 0$. As V is $G(\tilde{p}^m)$, it follows that $E(V/(1-\tilde{p})$ equals

$$\frac{1}{\tilde{p}^m(1-\tilde{p})} = \frac{1-\tilde{p}^m}{\tilde{p}^m(1-\tilde{p})} + \frac{1}{1-\tilde{p}} = E(RL) + \frac{1}{1-\tilde{p}} = \frac{1}{p} + O(1)$$

as $p \to 0$ and thus (2.10) holds. Likewise, the definition of V implies that $var(V/(1-\tilde{p})) = (1-\tilde{p}^m)/\{(1-\tilde{p})\tilde{p}^m\}^2$. Now (2.11) follows from (2.7) by noting that $\tilde{p}^m\{\tilde{p} - 2m(1-\tilde{p})/(1-\tilde{p}^m\} = \tilde{p}^m\{-2m + O(\tilde{p})\} = O(p)$. □

3 Out-of-control behavior

In this section we shall study the *OoC* behavior of *CUMIN* and compare it to that of its competitors. For *MIN* and *SUM*, the *ARL* during *OoC* has already been given in (2.2) and (2.4), respectively. Lemma 2.1 continues to hold in the *OoC* case if we replace $\tilde{p}$ by $\overline{F}(\overline{F}^{-1}(\tilde{p}) - d)$. In view of (2.7) we now obtain for *CUMIN* that

$$ARL_{CM}(m,d) = \left\{\frac{1}{(\overline{F}(\overline{F}^{-1}(\tilde{p}) - d))^m} - 1\right\}\frac{1}{F(\overline{F}^{-1}(\tilde{p}) - d)}, \tag{3.1}$$

where $\tilde{p} = \tilde{p}(m)$ is the solution of (2.8), as given approximately by (2.9). Hence we have $ARL_{CM}(m,0) = 1/p$ again for all F (just like *MIN*, cf. (2.2)), and not just for $F = \Phi$ (like *SUM*, cf. (2.4)).

Note that we have made only explicit in (3.1) the dependence of the *ARL* on m and d. To achieve full generality, we should of course write $ARL_{CM}(p,m,d,F)$. However, to avoid an unnecessarily lengthy exposition, we shall not pursue the dependence on p and F in detail. For p the reason is quite simple: it really suffices to concentrate on a single representative value, like the case $p = 0.001$ from our examples. The values used in practice will be of a similar order of magnitude and it can be verified that for such values the conclusions about the behavior of the function from (3.1) will be qualitatively the same. As concerns F, the situation is a bit more complicated. In principle, it would be quite interesting to see how (3.1) behaves for a variety of F's.

However, as most of the competitors (*IND*, *SUM*, *CUSUM*) are only valid under the single option $F = \Phi$, there is little to compare to outside normality. For that reason only, we will restrict attention to $F = \Phi$ for our *CUMIN* as well. Hence, as indicated in (3.1), in what follows we concentrate on m and d.

The first question of interest (cf. Sect. 1) is of course: what m should we take? As mentioned, the answer depends on d: the larger d, the smaller m should be. To be a bit more specific, for really large d, like $d = 3$, it is best to simply let $m = 1$, i.e., to use *IND*. For values in an interval around the typical choice $d = 1$ (cf. e.g., Ryan 1989, p.107), a simple rule of thumb for the optimal value of m is:

$$m_{opt} \approx \frac{17}{1 + 2d^2}. \tag{3.2}$$

As d increases from 1/2 to 3/2 in steps of 1/4, the rule in (3.2) indeed produces the corresponding correct values of m_{opt}: 11, 8, 6, 4 and 3. For values of d even smaller than 1/2, the optimal value of m rises sharply. However, the function in (3.1) then remains quite flat over a wide range of m-values, so there seems to be no need to consider m larger than 10. All in all, a simple advice for use in practice could be:

- *Use* $m = 1$, *i.e.,* IND, *only if the supposed* d *is really large* ($d \approx 3$).
- *In all other cases, considerable improvement w.r.t. IND is possible.*
- *If* d *is supposed to be moderately large* ($\approx$ 3/2 *or* 2), $m = 3$ *is suitable.* (3.3)
- *For somewhat smaller* d (≈ 1), $m = 6$ seems fine.
- *For really small* d (1/2 or below), $m = 10$ should do.

Do remember that this advice is tuned at $p = 0.001$ and $F = \Phi$. For different p we might get slightly different results; for (quite) different F in principle (quite) different behavior could be advisable. However, if a specific interest arises for a given F, a suitable analog of (3.2) can easily be found through (3.1) along the same lines.

It should be stressed that the resulting picture about the relation between d and m is by no means typical for *CUMIN*. In fact, expressions (2.2) and (2.4) lead to completely similar results for *MIN* and *SUM*, respectively. From (2.2) we obtain as an analog to (3.2) for *MIN* that $m_{opt} \approx 1,000/(75+80d^2)$ for $1/2 \leq d \leq 3/2$, while (2.4) produces $m_{opt} \approx 40/(1 + 4d^2)$ for *SUM* and these values of d, e.g., for $d = 1$, $m_{opt} = 6$ for *MIN* and $m_{opt} = 8$ for *SUM*. Hence, as already stated before, both *SUM* and *MIN* also beat *IND* for smaller values of d. In fact, detailed information on the relation between *IND*, *SUM* and *MIN* was already presented in AK (2006). Here we just present a single but representative example.

Example 3.1 From Albers and Kallenberg (2006) we quote that for $p = 0.001$ and $F = \Phi$, at $d = 1$ the *ARL* of the individual chart equals 54.6. Suppose we had decided to use $m = 3$, then this result is improved with 26.7 by taking *MIN*, yielding $ARL = 27.9$; the further improvement when using *SUM* is much less: 8.5, giving $ARL = 19.4$. (That the overall winner here is *SUM* is of course by virtue of the choice $F = \Phi$; outside normality, *MIN* can be the winner; see Albers and Kallenberg (2006) for examples.) If we now in addition suppose that we did not simply use $m = 3$, but in fact had guessed correctly and selected m_{opt} in either case, the picture is modified

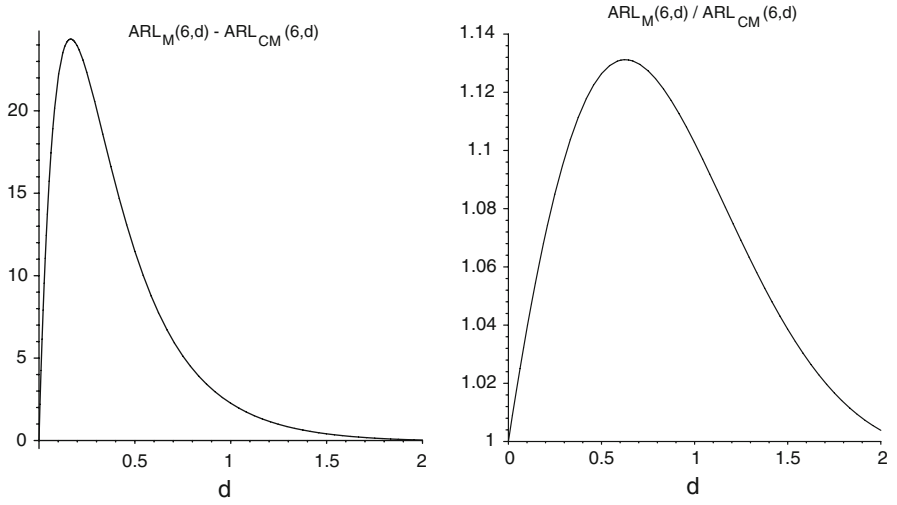

Fig. 1 Comparison of CUMIN to MIN

as follows. For *MIN*, we then apply $m = 6$, leading to $ARL = 24.3$, while *SUM* uses $m = 8$, leading to $ARL = 12.1$. Indeed some further improvement, but note that the discretization effect will be larger for these higher m-values (cf. the remark following Example 3.1 (cont.) below). □

In view of the already existing comparison results just mentioned, here we can focus on the comparison of *CUMIN* to *MIN*. This can be done in the same way as described already in Sect. 2 for the other charts. Here use (3.1) together with (2.2) and then look at $ARL_M(m, d) - ARL_{CM}(m, d)$ and/or $ARL_{CM}(m, d)/ARL_M(m, d)$. In Fig. 1 a representative picture is given for $m = 6$, which is the optimal value for both *CUMIN* and *MIN* for $d = 1$.

Hence indeed *CUMIN* forms a useful further improvement over *MIN*. For $m = 3$, the picture looks completely similar. To present some actual values, we have:

Example 3.1 (cont.) Above we found for the given choice $m = 3$ at a realized $d = 1$ an *ARL* of 54.6 for *IND* and of 19.4 for *SUM*. Most of this gap was bridged by *MIN* with a value 27.9; now we can offer a further reduction through *CUMIN* to 24.8. The luckiest choice of m for the realized value $d = 1$ would have been $m = 6$ for both *MIN* and *CUMIN*, leading to realizations for the *ARL* of 24.3 and 22.0, respectively.

An additional advantage of *CUMIN* over *MIN* that should be mentioned concerns the discrete character of the charts. Typically, the point where a shift occurs will only rarely coincide precisely with the start of a new group. Hence it is quite likely that the impact of the process going *OoC* will be delayed until the present group has ended. Clearly, this effect will be more pronounced for procedures such as *MIN* and *SUM*, with groups of fixed size m, than for the more quickly reacting *CUMIN*. Especially for small d, and thus large m, this effect is not negligible.

To complete the picture, it remains to add some comparison to *CUSUM* as well. However, let us first point out some confusion which might arise here, due to the fact that the notion of grouped data is used in various ways. Quite often, data used for control charting occur already in subgroups of sizes, e.g., 3, 4 or 5. The corresponding subgroup averages are then used and a Shewhart $\overline{X}$-chart is applied, rather than a Shewhart X-chart for individual observations. This sounds as if, in our terminology, *SUM* is used instead of *IND*. However, this does not necessarily have to be the case. Consider, e.g., Ryan (1989), Sect. 5.3, where the *CUSUM* procedure is compared to the Shewhart $\overline{X}$-chart. An example involving subgroups of size 4 is used and it is rightfully concluded that, e.g., for $d = 1$ the *CUSUM* chart really is much better. The question; however, is: much better than what? The point is that in this example the shift d is given in units of $\sigma_{\overline{X}}$ and not of σ_X. Hence, in our terminology, the $\overline{X}_i$ are used as individual observations again, and the comparison is between *CUSUM* and *IND*, and not between *CUSUM* and *SUM*. If the appropriate $\overline{X}_i$ in their turn are collected into groups according to our setup, the gap in performance would be much smaller. To illustrate this qualitative explanation, we have the following example.

Example 3.2 Ryan (1989) gives in Table 5.6 an *ARL* of 10.4 for the *CUSUM* chart with $d = 1$ ($k = 0.5$) and $h = 5$. In comparison, he mentions that the $\overline{X}$-chart scores the much larger 43.96. Indeed, this latter value is the *ARL* of *IND* for $d = 1$ and $p = 0.00135 = \overline{\Phi}(3)$, used in the customary '3σ'-chart. As according to Table 5.6 the two-sided *CUSUM* chart in question has $ARL = 465$ during IC, the appropriate p to use would be 1/930. In that case *IND* even requires an $ARL = 51.8$ for $d = 1$. However, suppose we would have used *SUM* with $m = 8$ (which is m_{opt} for $d = 1$ and the present value $p = 1/930$ as well). Then it follows from (2.4) that the corresponding *ARL* is merely 11.9, which indeed is much closer to *CUSUM*'s 10.4 than *IND*'s 51.8. Admittedly, this result looks extremely nice because we (more or less) took m_{opt} in *SUM*. But take, e.g., $d = 1/2$ instead of $d = 1$, then the *ARL*'s rise for *CUSUM* to 38.0 and for *IND* to 196. In this situation, $m = 8$ is not at all optimal anymore for *SUM*. Nevertheless, the $SUM(8)$ chart has $ARL = 48.0$ for $d = 1/2$, which still largely bridges the gap between 196 and 38.0.

Hence the resulting picture is as follows. For a wide range of d values, an (often substantial) improvement over *IND* is offered by *MIN*. This chart in its turn is further improved by its sequential analogue *CUMIN*, both directly (cf. Fig. 1) and because of the discrete character of the charts. For the sum-based procedures the situation actually is completely analogous. First *IND* is substantially improved by *SUM*, which in its turn is further improved by *CUSUM*. When focusing on the case $F = \Phi$, sum-based charts are obviously better than min-based ones. But always bear in mind that this superiority rests on this normality assumption, which is often quite questionable, especially in the tails. If normality fails, both *SUM* and *CUSUM* run into trouble. For known $F \neq \Phi$, they are awkward to handle, whereas for the min-based charts Φ plays no special role at all (cf. (2.2) and (3.1)). And when F is unknown, *SUM* and *CUSUM* (cf. Hawkins and Olwell 1998, p.75) may lead to a considerable *ME*. In case of *IND*, see, e.g., Table 1 on p. 173 of Albers et al. (2004). Various nonnormal distributions are considered here, such as the normal power family, based on $|Z|^{1+\gamma} sign(Z)$, with Z standard normal and $\gamma > -1$. For $\gamma = 1/2$, and $p = 0.001$, we have $ME = 5.6p$, while for $\gamma = 1$

Table 1 *ARL*'s of five charts for $p = 1/930$ and various values of d

d	0	1/4	1/2	3/4	1	3/2	2
IND	930	415	196	98.0	51.8	17.1	7.01
$MIN(6)$	930	257	97.5	43.7	23.6	10.7	7.38
$CUMIN(6)$	930	236	86.8	38.9	21.5	10.3	7.35
$SUM(8)$	930	170	48.0	20.1	11.9	8.26	8.00
CUSUM	930	139	38.0	17.0	10.4	5.75	4.01

we even obtain $ME = 9.4p$. For a Student(6)-df, we get $ME = 3.6p$, while Tukey's λ family (based on $\{U^\lambda - (1-U)^{1-\lambda}\}$ with U uniform on (0,1)) produces $ME = 4.7p$ for $\lambda = -0.1$. On the other hand, when F is unknown both *MIN* and *CUMIN* allow a rather straightforward nonparametric adaptation by using appropriate order statistics from an initial sample. In case of *MIN* this has been shown in Albers and Kallenberg (2008); for *CUMIN* we shall demonstrate it in Sect. 4. But before doing so, we shall conclude the present section by giving a representative example of *ARL*'s for the five charts considered so far.

Example 3.2 (cont) Above we already used Table 5.6 from Ryan (1989) for making some illustrative comparisons between *IND*, *CUSUM* and $SUM(8)$ (using that at $d = 1$ for the latter chart $m_{opt} = 8$). Now we add $MIN(6)$ and $CUMIN(6)$ to the picture (as at $d = 1$ in either case we have $m_{opt} = 6$) and we consider a somewhat wider range of d-values. The result is given in Table 1 above.

Indeed, especially for the smaller d, a wide gap exists between *IND* and *CUSUM*, which is bridged to a large extent by *MIN* and even better by *CUMIN*.

The improvement of *CUMIN* over *MIN*, illustrated in Fig. 1 , can be explained and generalized by Lemma 3.1 below. The condition in this lemma concerns the behavior of $f/\overline{F}$ in the tail and is e.g., satisfied for the standard normal distribution, as is shown in Lemma 3.2. Under this tail condition, ARL_{CM} is smaller than ARL_M for sufficiently small p and d. This holds for each m. Let m_M be the m_{opt} for *MIN* and m_{CM} the one for *CUMIN*. Then, for sufficiently small p and d, $ARL_{CM}(m_{CM}, d) \leq ARL_{CM}(m_M, d) < ARL_M(m_M, d)$ and hence the improvement of *CUMIN* over *MIN* continues to hold for the optimal choices of m, even if these are different for *MIN* and *CUMIN*.

Lemma 3.1 *Assume that $h(x) = f(x)/\overline{F}(x)$ is increasing in the tail in the following sense: there exists a normalizing function $z(p) > 0$ such that, if $c(p) \to c > 1$*

$$\lim_{p\to 0} \left\{ 1 - \frac{h(\overline{F}^{-1}(c(p)p))}{h(\overline{F}^{-1}(p))} \right\} z(p) > 0, \tag{3.4}$$

$$\lim_{p\to 0} pz(p) = 0. \tag{3.5}$$

Then, for each $m \geq 2$,

$$\lim_{p\to 0} \lim_{d\to 0} \left\{ \frac{ARL_{CM}(m,d)}{ARL_M(m,d)} - 1 \right\} \{dh(\overline{F}^{-1}(\tilde{p}))\}^{-1} z(\tilde{p}) < 0.$$

Proof Taylor expansion of $ARL_{CM}(m,d)$, given in (3.1), and application of $ARL_{CM}(m,0) = (1-\tilde{p})^{-1}(\tilde{p}^{-m}-1)$, cf. (2.7), yields as $d \to 0$

$$ARL_{CM}(m,d) = ARL_{CM}(m,0) - \frac{mdh(\overline{F}^{-1}(\tilde{p}))}{(1-\tilde{p})\tilde{p}^m} + d\left(\frac{1}{\tilde{p}^m} - 1\right)\frac{\tilde{p}h(\overline{F}^{-1}(\tilde{p}))}{(1-\tilde{p})^2}$$
$$+ O(d^2) = ARL_{CM}(m,0)\{1 - mdk(\tilde{p}) + O(d^2)\},$$

where $k(\tilde{p}) = h(\overline{F}^{-1}(\tilde{p}))[1 + \tilde{p}^m/(1-\tilde{p}^m) - \tilde{p}/((1-\tilde{p})m)]$. By Taylor expansion of $ARL_M(m,d)$, as given in (2.2), we get

$$ARL_M(m,d) = ARL_M(m,0) - m^2 d\overline{F}(\overline{F}^{-1}((mp)^{1/m}))^{-m-1} f(\overline{F}^{-1}((mp)^{1/m})) + O(d^2)$$
$$= ARL_(m,0)\{1 - mdh(\overline{F}^{-1}((mp)^{1/m})) + O(d^2)\}$$

as $d \to 0$. Since $ARL_{CM}(m,0) = ARL_M(m,0) = p^{-1}$, we obtain

$$\frac{ARL_{CM}(m,d)}{ARL_M(m,d)} = \frac{1 - mdk(\tilde{p})}{1 - mdh(\overline{F}^{-1}((mp)^{1/m}))} + O(d^2)\}$$
$$= 1 - md\{k(\tilde{p}) - h(\overline{F}^{-1}((mp)^{1/m}))\} + O(d^2)\}$$

as $d \to 0$. Hence we get

$$\lim_{d\to 0}\left\{\frac{ARL_{CM}(m,d)}{ARL_M(m,d)} - 1\right\}d^{-1} = -m\{k(\tilde{p}) - h(\overline{F}^{-1}((mp)^{1/m}))\}. \quad (3.6)$$

Define $c(\tilde{p}) = (mp)^{1/m}\tilde{p}^{-1}$. (Note that p can be considered as a function of $\tilde{p}$ and vice versa.) In view of (2.12) we have that $lim_{p\to 0}\, c(\tilde{p}) = m^{1/m} > 1$. According to the condition on h there exists a function z with $z(\tilde{p}) > 0$ such that

$$\lim_{p\to 0}\left\{1 - \frac{h(\overline{F}^{-1}((mp)^{1/m}))}{h(\overline{F}^{-1}(\tilde{p}))}\right\}z(\tilde{p}) > 0$$

and $lim_{p\to 0}\, \tilde{p}z(\tilde{p}) = 0$. Together with (3.6) and the definition of $k(\tilde{p})$ we obtain

$$\lim_{p\to 0}\lim_{d\to 0}\left\{\frac{ARL_{CM}(m,d)}{ARL_M(m,d)} - 1\right\}\{dh(\overline{F}^{-1}(\tilde{p}))\}^{-1}z(\tilde{p})$$
$$= \lim_{p\to 0} - mz(\tilde{p})\left\{1 + \frac{\tilde{p}^m}{1-\tilde{p}^m} - \frac{\tilde{p}}{(1-\tilde{p})m} - \frac{h(\overline{F}^{-1}((mp)^{1/m}))}{h(\overline{F}^{-1}(\tilde{p}))}\right\}$$
$$= \lim_{p\to 0} - mz(\tilde{p})\left\{1 - \frac{h(\overline{F}^{-1}((mp)^{1/m}))}{h(\overline{F}^{-1}(\tilde{p}))}\right\} < 0$$

as was to be proved. □

We check the conditions on h in case where $F = \Phi$.

Lemma 3.2 *For the standard normal distribution* $h(x) = \varphi(x)/\overline{\Phi}(x)$ *is increasing in the sense of* (3.4) *and* (3.5).

Proof The behavior of Φ in the tail is given by the following expansion for large quantiles:

$$\overline{\Phi}^{-1}(q) = (2|\log q|)^{1/2}[1 - k_1(q) + o(|\log q|^{-1})],$$

as $q \to 0$, where $k_1(q) = (2|\log q|)^{-1}\{\log(2|\log q|) + \log(2\pi)\}/2$.

Furthermore use that $h(x) = x[1 + x^{-2}\{1 + o(1)\}]$ as $x \to \infty$. Let $c(p) \to c > 1$ as $p \to 0$. Then we obtain, as $p \to 0$, that $h(\overline{\Phi}^{-1}(c(p)p))/h(\overline{\Phi}^{-1}(p))$ equals

$$\begin{aligned}&\frac{\overline{\Phi}^{-1}(c(p)p)}{\overline{\Phi}^{-1}(p)}\left\{\frac{1 + [\overline{\Phi}^{-1}(c(p)p)]^{-2}(1 + o(1))}{1 + [\overline{\Phi}^{-1}(p)]^{-2}(1 + o(1))}\right\}\\ &= k_0(p)\left\{\frac{1 - k_1(c(p)p) + o(|\log p|^{-1})}{1 - k_1(p) + o(|\log p|^{-1})}\right\}k_2(p)(1 + o(1)),\end{aligned}$$

in which $k_0(p) = \{|log(c(p)p)|/|\log p|\}^{1/2}$ and $k_2(p) = \{1+(2|\log(c(p)p)|)^{-1}\}\{1+(2|\log p|)^{-1}\}$. For the various k_i we have the following results:

$$\begin{aligned}k_0(p) &= \left\{\frac{-\log c(p) + |\log p|}{|\log p|}\right\}^{1/2} = 1 - \frac{1}{2}\frac{\log c}{|\log p|} + o(|\log p|^{-1}),\\ \frac{1 - k_1(c(p)p))}{1 - k_1(p))} &= [1 - k_1(c(p)p))][1 + k_1(p)] + o(|\log p|^{-1}) = 1 + o(|\log p|^{-1}),\\ k_2(p) &= 1 + o(|\log p|^{-1}),\end{aligned}$$

and thus, as $p \to 0$,

$$\frac{h(\overline{\Phi}^{-1}(c(p)p))}{h(\overline{\Phi}^{-1}(p))} = 1 - \frac{1}{2}\left(\frac{\log c}{|\log p|}\right) + o(|\log p|^{-1}).$$

Now define $z(p) = |\log p|$, then the limit in (3.4) equals $(\log c)/2$. As $c > 1$, this is indeed positive. Moreover, (3.5) holds as well. □

4 The nonparametric chart

In Sects. 2 and 3 we have worked under the assumption of known F. This was very useful in order to demonstrate the properties and performance of *CUMIN* and to compare it to its various competitors. However, by now we should drop this artificial assumption again and return to our main case of interest. There the normality assumption is not to be trusted, especially in the tail area we are dealing with, and a nonparametric approach is desired. Hence a Phase I sample $X_1, \ldots, X_n$ is needed again and will be used to obtain an estimated $\widehat{UL}$ (and, for the two-sided case, an estimated $\widehat{LL}$).

Assume that F is continuous and let $F_n(x) = n^{-1}\#\{X_i \le x\}$ be the empirical df and F_n^{-1} the corresponding quantile function, i.e., $F_n^{-1}(t) = \inf\{x|F_n(x) \ge t\}$. Then it follows that $F_n^{-1}(t)$ equals $X_{(i)}$ for $(i-1)/n < t \le i/n$, where $X_{(1)} < \cdots < X_{(n)}$ are the order statistics corresponding to $X_1, \ldots, X_n$. Hence, letting $\overline{F}_n^{-1}(t) = F_n^{-1}(1-t)$, we get for the nonparametric *IND* that a signal occurs if for a single new observation Y we have

$$Y > \widehat{UL}, \quad \text{with } \widehat{UL} = \overline{F}_n^{-1}(p) = X_{(n-r)}, \tag{4.1}$$

where $r = [np]$, with $[y]$ the largest integer $\le y$. Note that for $p = 0.001$ this r will remain 0, and thus $\widehat{UL}$ will equal the maximum of the Phase I sample, until n is at least 1,000. Details on this chart, as well as suitably corrected versions, can be found in Albers and Kallenberg (2004). For the grouped case, after Phase I, we have a new group of observations $Y_1, \ldots, Y_m$ and consider $T = min(Y_1, \ldots, Y_m)$ for *MIN* (cf. (2.1)). In analogy to (4.1), the estimation step for the nonparametric version of *MIN* leads to

$$T > \widehat{UL}, \quad \text{with } \widehat{UL} = \overline{F}_n^{-1}((mp)^{1/m}) = X_{(n-r)}, \tag{4.2}$$

with this time $r = [n(mp)^{1/m}]$. For $p = 0.001$, $m = 3$ and $n = 100$, we e.g., obtain $r = 14$ and we are dealing with $X_{(86)}$, which is much less extreme than the sample maximum $X_{(100)}$. Details and corrected versions for this chart are given in Albers and Kallenberg (2008).

In view of (4.1) and (4.2), it is clear how to obtain a nonparametric adaptation of *CUMIN*. In Sect. 2 , we replaced $\overline{F}^{-1}((mp)^{1/m})$ by $\overline{F}^{-1}(\tilde{p})$ and thus (2.5) will now become:

$$\begin{aligned} &\text{“}Give\ an\ alarm\ at\ the\ 1st\ time\ m\ consecutive \\ &observations\ all\ exceed\ \overline{F}_n^{-1}(\tilde{p}) = X_{(n-r)}\text{”}, \end{aligned} \tag{4.3}$$

with $r = [n\tilde{p}]$ here, in which $\tilde{p}$ is defined through (2.8) as a function of p and m (see also (2.9)). For $p = 0.001$, $m = 3$ and $n = 100$ we find $r = 10$ (see Example 2.1) and thus $X_{(90)}$, which again is much less extreme than $X_{(100)}$.

Using stochastic limits in (4.1)–(4.3) means that the fixed *ARL*'s from the case of known F now have become stochastic. From (2.2) together with (4.2), we immediately get for *MIN* that, conditional on $X_1, \ldots, X_n$,

$$ARL_M(m,d) = \frac{m}{\{\overline{F}(\overline{F}_n^{-1}((mp)^{1/m}) - d)\}^m}. \tag{4.4}$$

Let $U_{(1)} < \cdots < U_{(n)}$ denote order statistics for a sample of size n from the uniform df on (0,1), then it readily follows from (4.4) that during IC

$$ARL_M(m,0) \cong \frac{m}{\{U_{(r+1)}\}^m}, \tag{4.5}$$

with '$\cong$' denoting 'distributed as' and $r = [n(mp)^{1/m}]$. Hence indeed *MIN* and *IND* (which is the case $m = 1$ in (4.4) and (4.5)) are truly nonparametric. Moreover,

$\{U_{(r+1)}\}^m \to^P mp$ as $n \to \infty$ and thus $ARL_M(m,0) \to^P 1/p$: there is no *ME* and the *SE* tends to 0. However, as mentioned in the Introduction, this convergence is quite slow and for $m = 1$ the *SE* of the corresponding *IND* is huge, unless n is very large. The explanation is that the relevant quantity of course is the relative error

$$W_M = \frac{ARL_M(m,0)}{\left(\frac{1}{p}\right)} - 1 \cong \frac{mp}{\{U_{(r+1)}\}^m} - 1, \tag{4.6}$$

which for $m = 1$ indeed shows a very high variability. As is demonstrated in Albers and Kallenberg (2008), using $m > 1$, i.e., a real *MIN* chart, dramatically reduces this variability. In fact, from $m = 3$ on, the resulting *SE* is roughly the same as that of the Shewhart $\overline{X}$-chart.

For *CUMIN* we obtain along the same lines through (3.1) and (4.3) that

$$ARL_{CM}(m,d) = \left\{ \frac{1}{(\overline{F}(\overline{F}_n^{-1}(\tilde{p}) - d))^m} - 1 \right\} \frac{1}{F(\overline{F}_n^{-1}(\tilde{p}) - d)}, \tag{4.7}$$

and thus that during *IC*

$$ARL_{CM}(m,0) \cong \left\{ \frac{1}{\{U_{(r+1)}\}^m} - 1 \right\} \frac{1}{(1 - U_{(r+1)})}, \tag{4.8}$$

where $r = [n\tilde{p}]$, with $\tilde{p}$ as in (2.8). Obviously, about the relative error $W_{CM} = ARL_{CM}(m,0)/(1/p) - 1$, completely similar remarks can be made as about W_M from (4.6). Hence, just like *MIN*, *CUMIN* has no *ME* and a *SE* which is as well-behaved as that of a Shewhart $\overline{X}$-chart for $m \geq 3$.

This actually already concludes the discussion of the simple basic proposal (4.3) for the nonparametric version of *CUMIN*. However, the following should be noted. The fact that for $m \geq 3$ the *SE* is no longer huge but comparable to that of an ordinary Shewhart $\overline{X}$-chart, is gratifying of course. But on the other hand, such an *SE* is still not negligible. In fact, at the very beginning of the paper we remarked that quite large values of n are required before this will be the case, even for the most standard types of charts. Hence it remains worthwhile to derive corrections to bring such stochastic character under control. This has e.g., been done for both normal and nonparametric *IND*, as well as for nonparametric *MIN* (see Albers and Kallenberg 2005a, 2004, 2008, respectively). Here we shall address this point for *CUMIN* as well. However, to avoid repetition, we shall not go into full detail about all possible types of corrections. For that purpose we refer to the papers just mentioned.

The idea behind the desire for corrections is easily made clear by means of an example. For our typical value $p = 0.001$, during IC the intended $ARL_{CM} = 1/p = 1,000$. However, the estimation step results in the stochastic version given by (4.8), rather than in a fixed value such as 1,000. On the average, the result from (4.8) will be close to this target value 1,000, but its actual realizations for given outcomes $x_1, \ldots, x_n$ may fluctuate quite a bit around this value. The larger the *SE*, the larger this variation will be. To some extent, such variation is acceptable, but it should only rarely exceed

certain bounds, e.g., a value below 800 should occur in at most 20% of the cases. Hence what we in fact want is a bound on an exceedance probability like:

$$P\left(ARL_{CM}(m,0) < \frac{1}{\{p(1+\varepsilon)\}}\right) \leq \alpha, \tag{4.9}$$

for given small, positive ε and α. In the motivating example, $\varepsilon = 0.25$ and $\alpha = 0.2$. Note that (4.9) can also be expressed as $P(W_{CM} < -\tilde{\varepsilon}) \leq \alpha$, with $\tilde{\varepsilon} = \varepsilon/(1+\varepsilon) \approx \varepsilon$.

First we shall give expressions for the exceedance probability in (4.9) for the uncorrected version of the chart.

Lemma 4.1 *Let* $h(x) = (1-x)x^m/(1-x^m)$ *and* $\tilde{p}_\varepsilon = h^{-1}(p(1+\varepsilon))$ *(and thus* $\tilde{p}_0 = \tilde{p} = h^{-1}(p)$*). Let* $B(n, p^*, j)$ *stand for the cumulative binomial probability* $P(Z \leq k)$ *with* Z *bin*(n, p^*)*. Then*

$$P\left(ARL_{CM}(m,0) < \frac{1}{p(1+\varepsilon)}\right) = B(n, \tilde{p}_\varepsilon, r) \rightarrow$$

$$\Phi\left(\frac{(r+1/2-n\tilde{p}_\varepsilon)}{\{n\tilde{p}_\varepsilon(1-\tilde{p}_\varepsilon)\}^{1/2}}\right) \approx \Phi\left(-\frac{\varepsilon}{m}\left\{\frac{n\tilde{p}}{1-\tilde{p}}\right\}^{1/2}\right), \tag{4.10}$$

where the first step is exact, the second holds for $n \rightarrow \infty$ *and the last one moreover is meant for* ε *small.*

Proof From (4.8) it is immediate that $ARL_{CM}(m,0) = 1/h(U_{(r+1)})$ and thus that the probability in (4.9) equals $P(h(U_{(r+1)}) > p(1+\varepsilon)) = P(U_{(r+1)} > \tilde{p}_\varepsilon)$. Now there is a well-known relation between beta and binomial distributions: $P(U_{(i)} > p) = B(n, p, i-1)$ and thus the first result in (4.10) follows. The second step is nothing but the usual normal approximation for the binomial distribution. As $r = [n\tilde{p}]$, we have $r + 1/2 \approx n\tilde{p}$, while $\tilde{p}_\varepsilon \approx \tilde{p}(1+\varepsilon)^{1/m}$ and therefore $r+1/2-n\tilde{p}_\varepsilon \approx n\tilde{p}\{1-(1+\varepsilon)^{1/m}\} \approx -\varepsilon n\tilde{p}/m$. □

The result from (4.10) readily serves to illustrate the point that the *SE* is not negligible and corrections are desirable.

Example 4.1 Once more let $p = 0.001$, $m = 3$ and $n = 100$ and, just as above, choose $\varepsilon = 0.25$. From Example 2.1 we have that $\tilde{p} = 0.1037$ and thus $r = 10$; likewise we obtain that $\tilde{p}_{0.25} = h^{-1}(0.00125) = 0.1120$. Hence the exact exceedance probability in this case equals $B(100, 0.1120, 10) = 0.428$, whereas the two approximations from (4.10) produce 0.412 and 0.388, respectively. Consequently, in about 40% of the cases the *ARL* will produce a value below 800, which percentage is well above the value $\alpha = 0.2$ used above. □

A corrected version can be given in exactly the same way as for *MIN* in Albers and Kallenberg (2008). In order to satisfy (4.9), essentially $X_{(n-r)}$ in (4.3) is replaced by a slightly more extreme order statistic $X_{(n+k-r)}$, for some nonnegative integer k. To be more precise, equality in (4.9) can be achieved by randomizing between two such shifted order statistics. Let V be independent of $(X_1, \ldots, X_n, Y_1, \ldots)$, with

$P(V = 1) = 1 - P(V = 0) = \lambda$. Then replace $X_{(n-r)}$ in (4.3) by

$$\widehat{UL}(k, \lambda) = (1 - V)X_{(n+k+1-r)} + VX_{(n+k-r)}. \tag{4.11}$$

Let $b(n, p^*, j)$ stand for the binomial probability $P(Z = j)$, with Z bin(n, p^*), then:

Lemma 4.2 *Equality in* (4.9) *will result by selecting* k *and* λ *in* (4.11) *such that*

$$B(n, \tilde{p}_\varepsilon, r - k - 1) \le \alpha < B(n, \tilde{p}_\varepsilon, r - k), \quad \lambda = \frac{(\alpha - B(n, \tilde{p}_\varepsilon, r - k - 1))}{b(n, \tilde{p}_\varepsilon, r - k)}. \tag{4.12}$$

Moreover, for large n, *approximately* $k = [k_i]$ *and* $1 - \lambda = k_i - [k_i]$, $i = 1, 2$, *where*

$$k_1 = u_\alpha\{n\tilde{p}_\varepsilon(1 - \tilde{p}_\varepsilon)\}^{1/2} + \{r + 1/2 - n\tilde{p}_\varepsilon\} \approx k_2 = u_\alpha\{n\tilde{p}(1 - \tilde{p}\}^{1/2} - \frac{\varepsilon n\tilde{p}}{m}, \tag{4.13}$$

with k_2 *meant for* ε *small. Equivalently,* $k_2 \approx u_\alpha\{r(1 - r/n)\}^{1/2} - \varepsilon r/m$.

Proof In view of (4.11), in combination with (4.9) and (4.10), it is immediate that $P(ARL_{CM}(m, 0) < 1/\{p(1 + \varepsilon)\}) = \{(1 - \lambda)P(U_{(r-k)} > \tilde{p}_\varepsilon) + \lambda P(U_{(r-k+1)} > \tilde{p}_\varepsilon)\} = \{(1 - \lambda)B(n, \tilde{p}_\varepsilon, r - k - 1) + \lambda B(n, \tilde{p}_\varepsilon, r - k)\} = B(n, \tilde{p}_\varepsilon, r - k - 1) + \lambda b(n, \tilde{p}_\varepsilon, r - k)$, from which (4.12) follows. Arguing as in Lemma 4.1, we have that $B(n, \tilde{p}_\varepsilon, r - k) \to \Phi((r - k + 1/2 - n\tilde{p}_\varepsilon)/\{n\tilde{p}_\varepsilon(1 - \tilde{p}_\varepsilon)\}^{1/2})$. Equating this to the desired boundary value $\Phi(-u_\alpha) = \alpha$ gives (4.13) for k_1. The result for k_2 follows likewise. □

Example 4.1 (cont.) Again $p = 0.001$, $n = 100$ and $m = 3$, leading to $r = 10$, and $\varepsilon = 0.25$. We obtain for $B(100, 0.1120, 10 - j)$ the outcomes 0.428, 0.305 and 0.199 for $j = 0$, 1 and 2 respectively. Hence if $X_{(90)}$ is replaced by $X_{(92)}$, the percentage of *ARL*'s below 800 is indeed reduced to less than 20. Equality in (4.9) for $\alpha = 0.2$ results according to (4.12) by letting $k = 1$ and $\lambda = 0.01$, i.e., by using $X_{(91)}$ rather than $X_{(92)}$ in 1% of the cases. The approximations from (4.13) produce $k_1 = 1.95$ and $k_2 = 1.69$, respectively. Hence indeed $k = 1$ in either case, while $\lambda = 0.05$ and 0.31, respectively. □

References

Albers W, Kallenberg WCM (2004) Empirical nonparametric control charts: estimation effects and corrections. J Appl Stat 31:345–360

Albers W, Kallenberg WCM (2005a) New corrections for old control charts. Qual Eng 17:467–473

Albers W, Kallenberg WCM (2005b) Tail behavior of the empirical distribution function of convolutions. Math Methods Stat 14:133–162
Albers W, Kallenberg WCM (2006) Alternative Shewhart-type charts for grouped observations. Metron LXIV(3):357–375
Albers W, Kallenberg WCM (2008) Minimum control charts. J Stat Plan Inference 138:539–551
Albers W, Kallenberg WCM, Nurdiati S (2004) Parametric control charts. J Stat Plan Inference 124: 159–184
Albers W, Kallenberg WCM, Nurdiati S (2006) Data driven choice of control charts. J Stat Plan Inference 136:909–941
Bakir ST, Reynolds MR Jr (1979) A nonparametric procedure for process control based on within-group ranking. Technometrics 21:175–183
Bakir ST (2006) Distribution-free quality control charts based on signed-rank-like statistics. Commun Stat Theory Methods 35:743–757
Chakraborti S,van der Laan P, Bakir ST (2001) Nonparametric control charts: an overview and some results. J Qual Technol 33:304–315
Chakraborti S,van der Laan P,van de Wiel MA (2004) A class of distribution-free control charts. J Royal Stat Soc Ser C 53:443–462
Chan LK, Hapuarachchi KP, Macpherson BD (1988) Robustness of $\overline{X}$ and R charts. IEEE Trans Reliability 37:117–123
Hawkins DM, Olwell DH (1998) Cumulative *SUM* Charts and charting for quality improvement. Springer, New York
Lorden G (1971) Procedures for reacting to a change in distribution. Ann Math Stat 42:1897–1908
Lucas JM (1982) Combined Shewhart-CUSUM quality control schemes. J Qual Technol 14:51–59
Page ES (1954) Continuous inspection themes. Biometrika 41:100–115
Pappanastos EA, Adams BM (1996) Alternative designs of the Hodges–Lehmann control chart. J Qual Technol 28:213–223
Qiu P, Hawkins D (2001) A rank based multivariate *CUSUM* procedure. Technometrics 43:120–132
Qiu P, Hawkins D (2003) A nonparametrice multivariate cumulative sum procedure for detecting shifts in all directions. J Royal Statist Soc, Ser d 52:151–164
Ross SM (1996) Some results for renewal processes, 2nd edn. Wiley, New York
Ryan TP (1989) Statistical methods for quality improvement. Wiley, New York