Sequential Analysis, 26: 123–151, 2007

ISSN: 0747-4946 print/1532-4176 online
DOI: 10.1080/07474940701246992

Shewhart Control Charts in New Perspective

Willem Albers
Department of Applied Mathematics, University of Twente, Enschede, The Netherlands

Wilbert C. M. Kallenberg
Department of Applied Mathematics, University of Twente, Enschede, The Netherlands

Abstract: The effects of estimating parameters and the violation of the assumption of normality when dealing with control charts are discussed. Corrections for estimating errors and extensions of the normal control chart to parametric and nonparametric charts are investigated. The underlying theory is extensively discussed, including the choice of a suitable parametric family containing the normal family. It turns out that classical contamination families like random or deterministic mixtures do not give a suitable solution here. The so-called normal power family leads to an acceptable family, as it is intimately connected to the problem at hand of modeling and estimating an extreme quantile. When the underlying distribution cannot be modeled sufficiently accurately by the normal power family, the nonparametric control chart comes into the picture. A data-driven procedure makes the choice between the three different charts. When the nonparametric chart turns up, a large number of Phase I observations are needed. When such a large sample size is not available, it may be preferred to replace the individual chart by a grouped one. The new minimum chart is recommended in that case.

Keywords: Bias correction; Empirical quantiles; Exceedance probability; Minimum control chart; Model error; Model selection; Normal power family; Out of control; Phase II control limits; Statistical process control; Stochastic error.

Subject Classifications: 62E20; 62P30; 62G32.

1. INTRODUCTION

Two aspects of standard control charts that have obtained a lot of attention in the last years will be discussed in this paper: the effect of estimating parameters

Received May 4, 2005, Revised August 3, 2005, Accepted November 2, 2005
Recommended by W. Schmid
Address correspondence to W. C. M. Kallenberg, Department of Applied Mathematics, Faculty of Electrical Engineering, Mathematics and Computer Science, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands; Fax: +31-534893069; E-mail: w.c.m.kallenberg@math.utwente.nl

and the assumption of normality. For monitoring the mean, the basic Shewhart $\overline{X}$ chart produces a signal as soon as an incoming new observation exceeds the 3σ upper or lower limit. More precisely, assuming that the new observation X follows a normal distribution with mean μ and standard deviation σ, an upper control limit $UCL = \mu + 3\sigma$ and a lower control limit $LCL = \mu - 3\sigma$ are defined, and a signal occurs when $X > UCL$ or $X < LCL$. (In fact, the new observation X may be the sample mean of a small group of observations. Grouped observations and statistics based on them other than the sample mean will be discussed later on. This is especially of interest when normality fails.) The corresponding probability p of a false alarm producing a signal when the observations are still in control equals 0.0027. Equivalently, as long as the process is in control, a false alarm will occur on average once every 370 observations. That is, the average run length (ARL) equals 370. For simplicity, from now on we focus on the one-sided case of an upper limit only. Two-sided control charts are treated in a similar way. For obtaining more generally a false alarm rate (FAR) equal to p, we simply replace 3 by $u_p = \overline{\Phi}^{-1}(p)$, where $\overline{\Phi} = 1 - \Phi$ and Φ denotes the standard normal distribution function (d.f.).

Obviously, when μ and σ are unknown, UCL cannot be calculated, and often estimates of μ and σ are simply plugged in without further adjustment, although the dangers have been pointed out from time to time in the literature; see, e.g., Ghosh et al. (1981), Quesenberry (1993), Roes (1995), Chen (1997), Woodall and Montgomery (1999), Chakraborti (2000), Nedumaran and Pignatiello (2001), and Albers and Kallenberg (2004a,c). These estimates are based on so-called Phase I observations $X_1, \ldots, X_n$, which are assumed to be in control. While in many statistical problems a sample size of 50–100, say, is already giving rather accurate results, here much larger sample sizes are needed to reduce the relative error adequately due to the fact that we are dealing with extreme quantiles, since p is very small. When such large samples are not available, a correction may be applied to control the control chart behavior. The first complication is that FAR is no longer a number, but a random variable (r.v.), because it depends on the estimates and hence on the Phase I observations $X_1, \ldots, X_n$. Denoting now the conditional FAR, given $X_1, \ldots, X_n$, by $P_n = P_n(X_1, \ldots, X_n)$ it is aimed that P_n is "close" to the intended p.

Two approaches will be discussed, one reducing the bias and the other reducing the exceedance probability. The most obvious first choice of getting P_n close to p is to correct UCL in order that EP_n is close to p. This is similar to the classical statistical approach of reducing the bias of an estimator of an unknown parameter. Note, however, that when, e.g., S^2 is an unbiased estimator of σ^2, the estimator S is not unbiased for estimating σ. Similarly, here a correction for bringing EP_n close to p is not suitable for making $E(1/P_n)$, the expected ARL, close to $1/p$.

The variability of P_n around its expected value is rather large (again unless n is very large). Bias correction is useful with respect to the long-term behavior of the chart in a series of separate applications. But for a single application, controlling an exceedance probability like $P((P_n - p)/p > 0.1)$ by an appropriate correction of UCL is more interesting. So, with this second approach the aim is to correct UCL in such a way that P_n exceeds p by more than 10%, say, only with some small probability.

Errors due to estimation is one aspect, but violating the normality assumption is another one, and often this has an even much larger effect. This has been shown, e.g., by Chan et al. (1988), Pappanastos and Adams (1996), and Albers et al. (2004, 2005). The error due to estimation is called the stochastic error (*SE*), while the error due to a wrong distributional assumption is called the model error (*ME*). To avoid *ME* we might deploy a nonparametric control chart, thus removing *ME* completely. However, the extreme $(1-p)$ quantile should be estimated in that case in a nonparametric way, thus inserting a huge *SE* (unless n is extremely large). A balance between these two extremes is a parametric control chart, where the family of normal distributions is extended to a larger parametric family. Surprisingly, such a seemingly innocent extension reveals itself as a very delicate point. Classical parametric models such as contamination models or Tukey's family lead to insuperable problems as, e.g., estimation comes in. It turns out that the so-called normal power family provides a good intermediate position, where the (needed!) correction for estimating the parameters can be executed.

The three control charts (the normal one, the parametric one, and the nonparametric control chart) are useful tools on their own, each in its own application region. As long as normality holds, we should not take the more complicated parametric chart, where more parameters need to be estimated. Similarly, the nonparametric chart should not be invoked when the parametric chart suffices, thus avoiding an unnecessarily large *SE*. It is hard to see beforehand what the most suitable model is, especially because we are dealing with the extreme tail. Therefore, the data should provide us this information. With this data-driven choice of the type of chart, a combined procedure arises with nice properties.

Although this combined control chart works very well in many cases, there still may be a problem when the data tell us to use the nonparametric control chart and not that many observations are available. Then we still end up with a rather large *SE* and hence an unsatisfactory procedure. Immediate solutions are to either collect additional data or to reduce the *SE* by switching over to a larger p. Both solutions are not really satisfactory, because in both cases the rules of the game are changed.

A more fundamental way to attack this remaining problem (with keeping n and p as they are) is to postpone the decision to deliver a signal until a (typically small) group of new observations has arrived. New questions then arise, like "how does the group size affect the behavior of the chart?" and "what group statistic should one take?". It turns out that in general the chart based on small groups outperforms the individual chart. With respect to the second question, it is seen that the sample average (*AVE*) (being optimal under normality) is neither optimal nor easy to handle in a nonparametric setting. The minimum (*MIN*) of the group is a nice candidate, in the sense that its loss compared to *AVE* when normality holds is small, and outside the normal family its gain is often large. As the observed shift in *MIN* does not need to be very extreme in order to warrant a signal, the estimation step involved automatically also deals with rather modest quantiles and thus leads to a smaller *SE*.

The main attention of the present paper is on the ideas and fundamental theoretical support for the new control charts, taking into account the estimation aspects and the possible lack of normality. For a nontechnical, methodological review on the control charts restricted to the ungrouped case, we refer to Albers and Kallenberg (2005a, 2006b). The paper is organized as follows. Sections 2, 3, and 4 deal with the normal, parametric and nonparametric control charts, respectively. In Section 5 the data-driven choice between them is considered. The last section gives results on the grouped charts.

2. NORMAL CONTROL CHARTS

In this section we consider the normal control chart for the ungrouped case, and hence we assume that the observations $X_1, \ldots, X_n, X_{n+1}$ are independent and identically distributed (i.i.d.) r.v.'s each with a $N(\mu, \sigma^2)$-distribution as long as it concerns the in-control situation. The r.v.'s $X_1, \ldots, X_n$ are the observations belonging to Phase I, on which the estimators of μ and σ are based, while X_{n+1} belongs to Phase II: the monitoring phase. In the out-of-control situation, X_{n+1} has a $N(\mu_1, \sigma^2)$ distribution with $\mu_1 > \mu$, as we restrict attention to *UCL*s.

2.1. In-Control Behavior

If μ and σ are known and $FAR = p$, then $UCL = \mu + u_p\sigma$. As a rule μ and σ are unknown, and we estimate them by the sample mean $\overline{X}$ and the sample standard deviation $S = \sqrt{S^2}$ with $S^2 = (n-1)^{-1}\sum_{i=1}^n (X_i - \overline{X})^2$. The results of this section go through in a similar way for other estimators as well, see, e.g., Albers and Kallenberg (2004c, 2005a), but for simplicity of presentation we consider here $\overline{X}$ and S. This leads to the observed *FAR*, given by

$$P_n = P_n(\overline{X}, S) = P\big(X_{n+1} > \overline{X} + u_p S\big) = \overline{\Phi}\left(\frac{\overline{X}-\mu}{\sigma} + u_p\frac{S}{\sigma}\right).$$

2.1.1. Bias

It is easy to correct *UCL* in terms of unbiasedness: simply replace S by the unbiased estimator $S/c_4(n)$ of σ, where $c_4(n) = \sqrt{2}\Gamma(n/2)/\{\sqrt{n-1}\Gamma((n-1)/2)\}$. However, this is unsatisfactory, since the goal is not to get an unbiased *UCL*, but to remove the bias in P_n or more generally in $g(P_n)$ with, e.g., $g(p) = 1/p$ corresponding to the *ARL*. So we want to correct *UCL* in order that $Eg(P_n)$ is close to $g(p)$. Particular functions g that are of interest are the already-mentioned $g(p) = p$ and $g(p) = 1/p$. Furthermore, the function $g(p) = 1-(1-p)^k$ is of interest; it corresponds to the probability that the run length is at most equal to k. The standard deviation of the run length is represented by $g(p) = \sqrt{1-p}/p$ and its median by $g(p) = (-\log 2)/\log(1-p)$. Note, however, that since p is very small the latter two functions behave like $1/p$ and $(\log 2)/p$, respectively, and hence they are essentially the same as the *ARL*. Because $E(1/P_n)$ is strongly determined by the occurrence of extremely long runs, which are not relevant in practice, Roes (1995, p. 34) remarks that $E(1/P_n)$ does not adequately summarize the run length properties of the chart; see also Quesenberry (1993, p. 242). For a more extensive discussion of the bias of the *ARL* and the corresponding correction, see Albers and Kallenberg (2004c), in particular Remarks 2.1 and 2.4, and for the consequences on the out-of-control behavior, see pp. 228 and 232 in that paper.

If we take $g(p) = p$, exact correction is possible. Since EP_n equals the unconditional probability $P\big(X_{n+1} > \overline{X} + u_p S\big)$ and $(X_{n+1} - \overline{X})/\big(S\sqrt{1+n^{-1}}\big)$ follows a Student distribution with $n-1$ degrees of freedom, exact correction is obtained when replacing u_p by $\sqrt{1+n^{-1}}t_{n-1;p}$. This correction can be found, e.g., in Yang and Hillier (1970), Ghosh et al. (1981), and Quesenberry (1991). Roes et al. (1993) present exact corrections for control charts with several other estimators as well.

To find suitable correction terms for other functions g, an exact correction is not possible and we apply an asymptotic approach. Investigating the limiting behavior of $g(P_n)$ also gives insight in the number of observations needed to get satisfactory results when using no correction at all. Here we restrict attention to the functions $g(p) = p$ and $g(p) = 1/p$, but other functions can be treated similarly; see Theorem 2.2 in Albers and Kallenberg (2004c). Let φ denote the standard normal density. In contrast to the rest of the paper, in the following theorem p is considered to be dependent on n.

Theorem 2.1. *Suppose that* $u_p = u_p(n) \geq 1$ *and that* $u_p = O(n^{1/4})$ *as* $n \to \infty$. *Then we have* (*with* $p = \overline{\Phi}(u_p)$)

$$\frac{EP_n - p}{p} = \frac{u_p\varphi(u_p)\left(u_p^2 + 3\right)}{4pn} + O\left(u_p^8 n^{-2}\right)$$

and

$$\frac{E(1/P_n) - (1/p)}{1/p} = -\frac{u_p\varphi(u_p)\left(u_p^2 + 3\right)}{4pn} + \frac{\{\varphi(u_p)\}^2\left(u_p^2 + 2\right)}{2p^2 n} + O\left(u_p^8 n^{-2}\right)$$

as $n \to \infty$.

Sketch of Proof. By Taylor expansion we get $EP_n \approx p - \varphi(u_p)E\Delta(u_p) + \frac{1}{2}u_p\varphi(u_p) E\Delta^2(u_p)$ with $\Delta(u_p)$ given by

$$\Delta(u_p) = \frac{\overline{X} - \mu}{\sigma} + u_p\left(\frac{S}{\sigma} - 1\right).$$

The result now follows by calculating suitable approximations of $E\Delta(u_p)$ and $E\Delta^2(u_p)$ and a careful treatment of the remainder terms in the Taylor expansion. The result for $E(1/P_n)$ is obtained similarly. □

Note that in order to get a relative error tending to 0, one should restrict attention to $u_p = o(n^{1/4})$ as $n \to \infty$. When u_p is of exact order $n^{1/4}$, we get that the explicit terms and the O terms in Theorem 2.1 are of order $O(1)$. This is still nontrivial, since $p \to 0$ in that case (and even very fast!).

Theorem 2.1 leads to the following approximations:

$$EP_n \approx p + \frac{u_p\varphi(u_p)\left(u_p^2 + 3\right)}{4n},$$
$$E(1/P_n) \approx 1/p - \frac{u_p\varphi(u_p)\left(u_p^2 + 3\right)}{4p^2 n} + \frac{\{\varphi(u_p)\}^2\left(u_p^2 + 2\right)}{2p^3 n}. \tag{2.1}$$

For instance, take $p = 0.001$ (yielding $u_p = 3.09$) and use the right-hand side of (2.1) to calculate the smallest value of n such that $|(EP_n - p)/p| < 0.1$. This results in $n = 326$. Exact calculation using t distributions gives $n = 337$. This shows that the approximation works quite well. It also shows that indeed very many Phase I observations are needed to get an accurate control chart limit when no correction is applied.

The bias can be removed by introducing an appropriate correction term in *UCL*. Theorem 2.1 gives us the tools to derive such correction terms. Note that when changing u_p in *UCL* to a corrected version $u_p + c$ for some correction term c, we have to change in Theorem 2.1 also p into $\overline{\Phi}(u_p + c)$. Obviously, P_n then stands for $P_n(\overline{X}, S) = P(X_{n+1} > \overline{X} + (u_p + c)S)$. The correction term for removing the bias when $g(p) = p$ is obtained from (2.1) by the equation

$$\overline{\Phi}(u_p + c) + \frac{(u_p + c)\varphi(u_p + c)\{(u_p + c)^2 + 3\}}{4n} = \overline{\Phi}(u_p).$$

Ignoring lower-order terms like c^2, cn^{-1}, this simply gives

$$\overline{\Phi}(u_p) - c\varphi(u_p) + \frac{u_p\varphi(u_p)\left(u_p^2 + 3\right)}{4n} = \overline{\Phi}(u_p),$$

and hence

$$c = \frac{u_p\left(u_p^2 + 3\right)}{4n}.$$

Similarly, the correction term when $g(p) = 1/p$ is given by

$$c = \frac{u_p\left(u_p^2 + 3\right)}{4n} - \frac{\varphi(u_p)}{\overline{\Phi}(u_p)} \frac{\left(u_p^2 + 2\right)}{2n}.$$

Taking again $p = 0.001$ and applying the corrected control chart, the smallest value of n such that $|(EP_n - p)/p| < 0.1$ turns out to be $n = 31$. This shows that the sample size needed to get accurate control charts indeed is tremendously reduced and that common sample sizes of Phase I observations are sufficient.

2.1.2. Exceedance Probability

The second criterion to express the closeness of P_n to the prescribed p is the exceedance probability. Rather than worrying about $|(EP_n - p)/p| < 0.1$, we now try to figure out how large $P((P_n - p)/p > 0.1)$ is and which correction is needed to reduce this probability for moderate sample sizes. While in the bias case rather large sample sizes were already needed when no correction was applied, here really huge sample sizes should be available to get the exceedance probability at a reasonable level. For instance, when $p = 0.001$ and $n = 5000$, then $P((P_n - p)/p > 0.1) = 0.203$. In general, we want to find correction terms such that for suitable (small) values of $\varepsilon \geq 0$ and $\alpha > 0$ we get

$$P\left(\frac{g(P_n) - g(p)}{g(p)} > \varepsilon\right) \leq \alpha$$

for increasing (and positive) functions g, like $g(p) = p$, $g(p) = 1 - (1 - p)^k$, and

$$P\left(\frac{g(P_n) - g(p)}{g(p)} < -\varepsilon\right) \leq \alpha$$

for decreasing (and positive) functions g, like $g(p) = 1/p$, $g(p) = \sqrt{1-p}/p$, $g(p) = (-\log 2)/\log(1-p)$. Note that for increasing (and positive) functions g we have

$$P\left(\frac{g(P_n) - g(p)}{g(p)} > \varepsilon\right) = P\left(\frac{P_n - p}{p} > \tilde{\varepsilon}\right)$$

with

$$\tilde{\varepsilon} = \frac{g^{-1}(g(p)(1+\varepsilon)) - p}{p} \tag{2.2}$$

and similarly for decreasing (and positive) functions g: just replace ε by $-\varepsilon$ in (2.2). Hence, we may restrict ourselves without loss of generality to $g(p) = p$.

Writing the corrected UCL as $\overline{X} + (u_p + c)S$ the next theorem gives the exact correction term.

Theorem 2.2. *Let $G_{n-1,\delta}$ stand for the d.f. of the noncentral t distribution with $n-1$ degrees of freedom and noncentrality parameter δ and write $\overline{G}_{n-1,\delta} = 1 - G_{n-1,\delta}$. The correction term*

$$c = n^{-1/2}\overline{G}^{-1}_{n-1,n^{1/2}b}(\alpha) - u_p \tag{2.3}$$

with $b = u_{p(1+\varepsilon)}$ gives

$$P\left(\frac{P_n - p}{p} > \varepsilon\right) = \alpha.$$

Proof. The random FAR P_n with the correction term c in UCL is given by

$$P_n = P_n(\overline{X}, S) = P(X_{n+1} > \overline{X} + (u_p + c)S) = \overline{\Phi}\left(\frac{\overline{X} - \mu}{\sigma} + (u_p + c)\frac{S}{\sigma}\right).$$

Hence, we get

$$\begin{aligned} P\left(\frac{P_n - p}{p} > \varepsilon\right) &= P\left(\frac{\overline{X} - \mu}{\sigma} + (u_p + c)\frac{S}{\sigma} < b\right) \\ &= P\left(-n^{1/2}\frac{\overline{X} - \mu}{\sigma} + n^{1/2}b > n^{1/2}(u_p + c)\frac{S}{\sigma}\right) \\ &= P\left(\frac{-n^{1/2}(\overline{X} - \mu)/\sigma + n^{1/2}b}{S/\sigma} > \overline{G}^{-1}_{n-1,n^{1/2}b}(\alpha)\right) = \alpha, \end{aligned}$$

which completes the proof. □

To get more insight in the nature of the correction term it is useful to derive an approximation to it. The following lemma produces an informative and accurate approximation.

Lemma 2.1. *For the correction term c given in* (2.3) *we have*

$$c = u_{p(1+\varepsilon)} - u_p + u_\alpha\left(\frac{u^2_{p(1+\varepsilon)} + 2}{2n}\right)^{1/2} + O(n^{-1}) \tag{2.4}$$

$$= -\frac{\varepsilon}{u_p} + u_\alpha\left(\frac{u_p^2 + 2}{2n}\right)^{1/2} + R \tag{2.5}$$

with $|R| \le C_1(\varepsilon, p)n^{-1} + C_2(p)\varepsilon^2 + C_3 u_p^{-6}$, *in which* C_1 *depends on* ε *and* p, C_2 *depends on* p *only, and* C_3 *is just a constant, not depending on* p *or* ε.

For the proof of this lemma and more refinements of it we refer to Albers and Kallenberg (2004a). As expected, this correction is much larger than the one for the bias. The latter is of order n^{-1}, while here the order is $n^{-1/2}$. To show the accuracy of the approximations, let $p = 0.001$, $\varepsilon = 0.1$, $\alpha = 0.2$, and $n = 100$; then $P((P_n - p)/p > \varepsilon) = 0.224$ when using approximation (2.4) and 0.228 when applying (2.5). For more details and an extensive discussion on the roles of n, p, ε, and α, we refer to Albers and Kallenberg (2004a).

2.2. Out-of-Control Behavior

In the out-of-control situation the new observation X_{n+1} has an $N(\mu_1, \sigma^2)$ distribution with $\mu_1 > \mu$, as we restrict attention to *UCL*s. For convenience we write $\mu_1 = \mu + d\sigma$ with $d > 0$. Let $p_1 = \overline{\Phi}(u_p - d)$ be the out-of-control rate when the parameters μ and σ are known. By a similar type of argument as in Theorem 2.1, we get as approximation for the expected random out-of-control rate when applying the corrected control chart with $UCL = \overline{X} + (u_p + c)S$ (denoted by $E_d P_n$) the following expression:

$$E_d P_n \approx p_1 - c\varphi(u_p - d) + \frac{u_p \varphi(u_p - d)}{4n} + \frac{(u_p - d)\varphi(u_p - d)\left(2 + u_p^2\right)}{4n}.$$

Clearly, the influence of the correction term c is only in the term $-c\varphi(u_p - d)$. Since p_1 is typically not small (in contrast to p), the effect of the correction term on the out-of-control behavior with respect to relative error is negligible. In fact, the relative error can be approximated well by

$$\begin{aligned}\frac{E_d P_n - p_1}{p_1} &\approx \frac{\varphi(u_p - d)}{\overline{\Phi}(u_p - d)}\left\{-c + \frac{u_p + (u_p - d)(2 + u_p^2)}{4n}\right\} \\ &\approx \frac{4}{5}\{1 + (u_p - d)\}\left\{-c + \frac{u_p + (u_p - d)\left(2 + u_p^2\right)}{4n}\right\},\end{aligned}$$

where we use that $\varphi(x)/\overline{\Phi}(x)$ can be approximated adequately by $4(1 + x)/5$ for $0 \le x \le 3.5$. In the case of exceedance probability, the correction term c is of order $n^{-1/2}$, and thus the n^{-1} term is negligible in that situation. Therefore, we end up with

$$\frac{E_d P_n - p_1}{p_1} \approx -\frac{4}{5}c\{1 + (u_p - d)\} \tag{2.6}$$

for the exceedance case. To illustrate that the influence of this correction term is indeed rather small (even although this correction term is much larger than the one that reduces the bias), take $p = 0.001$, $p_1 = 0.20$ (leading to $d = 2.25$), $\varepsilon = 0.1$, $\alpha = 0.2$, and $n = 100$ (leading to $c = 0.170$ when using equation 2.5); then the right-hand side of (2.6) yields $(E_d P_n - p_1)/p_1 \approx 0.25$, and thus the (only theoretically attainable) value 0.20 is replaced by 0.15. In terms of the *ARL* (for which the relative error result holds as well), we find 6.25 instead of $1/0.20 = 5$. We may conclude that the correction terms do not disturb the behavior of the control charts in the out-of-control situation.

3. PARAMETRIC CONTROL CHARTS

The effect of nonnormality on standard control charts (which assume normality) is very large. It is not unusual that *FAR* is 5 or even 10 times as large as it should be when the true distribution differs from normality. One way of avoiding such errors is to extend the normal family to a larger parametric family containing the normal family as a subfamily. The advantage is, of course, that the true distribution is closer to the supposed distribution (as we have a larger domain of distributions available); the disadvantage might be that we have to estimate more parameters, thus leading to larger *SE*s. As in the normal family we will always take a location parameter μ and a scale parameter σ. Under normality, then, the distribution of $(X-\mu)/\sigma$ is fixed to the standard normal distribution. The extension consists in embedding the standard normal distribution in a family of distributions with one or more additional parameters. Let us call this parameter or vector of parameters γ, its d.f. K_γ, and the corresponding upper p quantile $\overline{K}_\gamma^{-1}(p)$. The estimated uncorrected *UCL* equals

$$\overline{X} + \overline{K}_{\hat{\gamma}}^{-1}(p)S,$$

where $\hat{\gamma}$ is an estimator of γ.

3.1. Model Error and Stochastic Error

We consider the in-control situation. We assume that the observations $X_1, \ldots, X_n, X_{n+1}$ are i.i.d. r.v.'s each with a d.f. F. The total error $P_n - p$ with $P_n = P_n(\overline{X}, S, \hat{\gamma}) = \overline{F}(\overline{X} + \overline{K}_{\hat{\gamma}}^{-1}(p)S)$ can be split up in two parts

$$P_n - p = \left\{\overline{F}(\mu + \overline{K}_\gamma^{-1}(p)\sigma) - p\right\} + \left\{\overline{F}(\overline{X} + \overline{K}_{\hat{\gamma}}^{-1}(p)S) - \overline{F}(\mu + \overline{K}_\gamma^{-1}(p)\sigma)\right\}.$$

The first part is a deterministic term and expresses the error due to model misspecification (and equals 0 if the observations come from the parametric model with $(X_i - \mu)/\sigma$ having the upper p quantile $\overline{K}_\gamma^{-1}(p)$). We call this term the *ME*. The second part deals with the replacement of the unknown parameters by the corresponding estimators and is called the *SE*. The idea behind the parametric model is that for many distributions, *ME* is substantially reduced compared to the *ME* obtained when we deal with the normal control chart. The latter *ME* is called the restrictive model error (*RME*), since it occurs when we have the restriction to normality, and is defined by

$$RME = \overline{F}(\mu + u_p\sigma) - p.$$

3.2. Parametric Models

The following models are candidates for the parametric model. The models are defined in such a way that varying tail behavior can be described. Heavier tails than those of the normal distribution are especially of interest. In terms of high upper quantiles, this means larger values than the normal upper quantiles. The location and scale parameters μ and σ are treated separately, and they are estimated by $\overline{X}$ and S. Therefore, the d.f. K_γ corresponds to an r.v. Z_γ with $EZ_\gamma = 0$ and $var(Z_\gamma) = 1$.

The conditions for an appropriate general model are rather comprehensive. Therefore, several classical ways of extending the normal model turn out to cause (technical) difficulties. In order to make the necessary (bias) corrections, we need to evaluate (first and second) moments of $\overline{X} + S\overline{K}_{\hat{\gamma}}^{-1}(p) - \left(\mu + \sigma\overline{K}_\gamma^{-1}(p)\right)$ up to high precision. This implies that either $\overline{K}_\gamma^{-1}(p)$ should be analytically tractable as a function of γ, or we should have a very precise approximation of $\overline{K}_\gamma^{-1}(p)$ by a simple function of γ.

1. *Random mixture*. In the random mixture model we take $K_\gamma = (1-\gamma)\Phi + \gamma K_1$ with K_1 a (fixed) d.f. with corresponding expectation 0 and variance 1. The r.v. Z_γ can be written as

$$Z_\gamma = (1-W)Z_0 + WZ_1,$$

where W is independent of Z_0 and Z_1, $P(W=1) = 1 - P(W=0) = \gamma$, and Z_0 and Z_1 have d.f.'s Φ and K_1, respectively. The random mixture model looks at first sight like an attractive parametric model, and is indeed very often used as extension of normality. However, normality is a "boundary point" in this model, obtained by taking $\gamma = 0$. Because negative values of γ are meaningless in this model, $\hat{\gamma}$ should be restricted to nonnegative values, which can often only be achieved by adding a suitable indicator function to the definition of the estimator, thus making $\hat{\gamma} = 0$ when its "natural" definition would give negative values. Due to the required precision, this causes great (technical) problems, aggravated by the fact that $\hat{\gamma}$ (and also the indicator function) is tied up with $\overline{X}$ and S. Apart from that, the truncation of negative values also introduces a large artificial bias near $\gamma = 0$, which is also rather unattractive. To be more precise, take for Z_1 a symmetric distribution with fourth moment unequal to $3 = EZ_0^4$, e.g., a standardized Student distribution with six degrees of freedom, giving $EZ_1^4 = 6$. Because $EZ_\gamma^4 = \gamma EZ_1^4 + 3(1-\gamma)$ and hence $\gamma = \left(EZ_\gamma^4 - 3\right)/\left(EZ_1^4 - 3\right)$, we take as initial estimator of γ

$$\frac{n^{-1}\sum_{i=1}^n \left(\frac{X_i - \overline{X}}{S}\right)^4 - 3}{EZ_1^4 - 3},$$

giving $\frac{1}{2}\left\{n^{-1}\sum_{i=1}^n \left(\frac{X_i-\overline{X}}{S}\right)^4 - 3\right\}$ in case of the above-mentioned Student distribution. To get nonnegative values for our estimator, we take

$$\hat{\gamma} = \frac{n^{-1}\sum_{i=1}^n \left(\frac{X_i - \overline{X}}{S}\right)^4 - 3}{EZ_1^4 - 3} 1\left(\frac{n^{-1}\sum_{i=1}^n \left(\frac{X_i - \overline{X}}{S}\right)^4 - 3}{EZ_1^4 - 3} > 0\right),$$

where $1(A) = 1$ if A holds and 0 otherwise. For finding appropriate correction terms, one has to evaluate moments like $E\{(S-1)\hat{\gamma}\}$ up to the needed high precision of order n^{-1}. This is not easy at all. Moreover, under normality we get $\lim_{n\to\infty} P(\hat{\gamma} = 0) = 1/2$, and obviously a large bias is introduced, which should be corrected too. In view of all these kinds of problems, this model is not a suitable model for our purposes.

2. *Deterministic mixture*. Since we are focused on quantiles here, it seems more natural to consider mixtures of quantiles than mixtures of d.f.'s as in the random mixture model; that is, take $K_\gamma^{-1} = c(\gamma)\{(1-\gamma)\Phi^{-1} + \gamma K_1^{-1}\}$ with K_1 a d.f. with corresponding expectation 0 and variance 1 and where $c(\gamma)$ is a normalizing factor such that $var(Z_\gamma) = 1$. The r.v. Z_γ can be written as

$$Z_\gamma = c(\gamma)\{(1-\gamma)Z_0 + \gamma Z_1\}$$

with $Z_0 = \Phi^{-1}(U), Z_1 = K_1^{-1}(U)$, and U a r.v. with a uniform distribution on $(0, 1)$. Note that Z_0 and Z_1 have d.f.'s Φ and K_1, respectively, but that they are anything but independent; in fact these r.v.'s are comonotone. Although K_γ^{-1} is analytically more attractive in the deterministic mixture model than in the random mixture model, unfortunately, the deterministic mixture model suffers from the same problem as the deterministic mixture model for estimating γ. Again normality is a boundary point in this model, which causes great problems and also makes this model impracticable.

3. *Tukey's family*. The r.v. Z_γ is given by

$$Z_\gamma = c(\gamma)\{U^{0.14-\gamma} - (1-U)^{0.14-\gamma}\},$$

where U has a uniform distribution on $(0, 1)$ and $c(\gamma)$ is a normalizing constant such that Z_γ has variance 1. The choice $\gamma = 0$ gives a distribution close to the standard normal distribution, especially for upper t quantiles with t from 0.2 to 0.005; cf. also Chan et al. (1988, p. 118). For $\gamma = 0.14$, we define Z_γ in a continuous way, leading to the logistic distribution. In this model $\overline{K}_\gamma^{-1}(p)$ is simply given by $c(\gamma)\{p^{0.14-\gamma} - (1-p)^{0.14-\gamma}\}$, and $\gamma = 0$ is an interior point of the parameter space. Nevertheless, analytic evaluation of the estimators of the parameters up to the required precision is very difficult, and therefore this model also is not used. The same holds for the generalization of Tukey's family, the so called generalized λ family, introduced by Ramberg and Schmeisser (1972, 1974).

4. *Orthonormal family*. Starting from a uniform distribution, an orthonormal family of densities with respect to the Lebesgue measure on $(0, 1)$ is defined by

$$f(y, \gamma) = c^*(\gamma)\exp\left\{\sum_{j=1}^{k} \gamma_j \pi_j(y)\right\},$$

where $c^*(\gamma)$ is a normalizing constant such that the integral of f equals 1, and where π_j is the jth Legendre polynomial on $(0, 1)$. Let Y be an r.v. having density $f(y, \gamma)$ and let $E(\gamma)$ and $c(\gamma)^{-1}$ be the expectation and standard deviation of $\Phi^{-1}(Y)$. The r.v. Z_γ is given by

$$Z_\gamma = c(\gamma)\{\Phi^{-1}(Y) - E(\gamma)\}.$$

Indeed, again $c(\gamma)$ is a normalizing factor such that $var(Z_\gamma) = 1$. This model offers explicitly the possibility for more than one additional parameter beyond μ and σ. However, if desired, the random and deterministic mixtures obviously can also be taken for more than just two. The orthonormal family on $(0, 1)$ is attractive in the sense that the log-density is approximated in a natural way, which approximation can be made more and more accurate by adding new terms, that is, taking a larger k. Normality ($\gamma = 0$) is an interior point, but $\overline{K}_\gamma^{-1}(p)$ is not easy, and again the estimators are not easily handled. Therefore, this model is not appropriate for our purposes.

5. *Normal power family*. Other distributions than the (standard) normal one are characterized by larger quantiles (when heavier tails occur) or smaller quantiles (when we have a lighter tail). One way to model this, still getting normality as an interior point, is to take as p quantiles $u_p^{1+\gamma}$. This seems to be the most natural approach for our purposes. Values $\gamma > 0$ correspond to heavier tails, and $\gamma < 0$ gives lighter tails. This approach leads to the normal power family, defined by

$$\overline{K}_\gamma^{-1}(p) = c(\gamma)|u_p|^{1+\gamma} sign(u_p), \tag{3.1}$$

where $\gamma > -1$ and where $c(\gamma)$ is a normalizing constant given by

$$c(\gamma) = \{E|Z|^{2(1+\gamma)}\}^{-1/2} = \pi^{1/4} 2^{-(1+\gamma)/2} \Gamma\left(\gamma + \frac{3}{2}\right)^{-1/2}$$

with Z an r.v. with a standard normal distribution. We may also write

$$Z_\gamma = c(\gamma)|Z|^{1+\gamma} sign(Z)$$

for $\gamma > -1$. It turns out that this model is appropriate for our goals, although even here a lot of technical problems should be solved.

At first sight it is surprising that going (a little bit) beyond normality causes immediately such big problems in many parametric models, but on the other hand a lot of requirements have to be fulfilled in order to get a suitable parametric family. From now on our parametric model will be the normal power family. Obviously, the reduction of the RME is very large when in fact F belongs to the normal power family itself. For instance, when $\gamma = 0.75$ we have $RME = 7.9$ and $ME = 0$. Note that $RME = 7.9$ means that in the limit (when $n \to \infty$), FAR is about nine times as large as it should be. This reduction, fortunately, is not restricted to the normal power family itself. Also, for many distributions outside the normal power family a substantial reduction appears. For instance, for the logistic distribution we get $RME = 2.7$ and $ME = 1.3$, while the normal inverse Gaussian $(2, 1.5, 0, 1)$ distribution (cf. Barndorff-Nielsen, 1996) gives $RME = 14.7$ and $ME = 1.9$.

3.3. Estimation

The estimator $\hat{\gamma}$ of γ that we use does not try to fit the distribution globally, but takes into account that we are dealing with the right tail only. This is particularly important for skew distributions like the normal inverse Gaussian $(2, 1.5, 0, 1)$. The estimator is based on the ratio of two quantiles, thus getting rid of $c(\gamma)$. The choice of the quantiles is such that they are in the tail, but not in the very far tail, where we have no observations to estimate them properly. It is seen from (3.1) that

$$\frac{\overline{K}_\gamma^{-1}(0.05)}{\overline{K}_\gamma^{-1}(0.25)} = \left(\frac{u_{0.05}}{u_{0.25}}\right)^{1+\gamma}$$

and hence

$$\gamma = \frac{\log\left(\overline{K}_\gamma^{-1}(0.05)/\overline{K}_\gamma^{-1}(0.25)\right)}{\log(u_{0.05}/u_{0.25})} - 1.$$

Our estimator now becomes

$$\hat{\gamma} = \frac{\log\left(\left(X_{([0.95n+1])} - \overline{X}\right)/\left(X_{([0.75n+1])} - \overline{X}\right)\right)}{\log(u_{0.05}/u_{0.25})} - 1,$$

where $[x]$ denotes the entier of x and $X_{(1)}, \ldots, X_{(n)}$ are the order statistics of $X_1, \ldots, X_n$.

Some large deviation properties of the estimators $\overline{X}$, S, and $\hat{\gamma}$ are presented in the next theorem. They are used in this section, but also in the proof of Theorem 5.1. Furthermore, they are of interest on their own. Note that for $\gamma > 1$, the moment generating function of X_i, having a normal power distribution with parameter γ, does not exist, and therefore the results of Theorem 3.1 and its proof are not standard. For a proof of this theorem we refer to Albers et al. (2006).

Theorem 3.1. *Let $X_1, \ldots, X_n$ be i.i.d. r.v.'s with a normal power distribution with parameter γ. Then for each $\varepsilon > 0$*

$$\limsup_{n\to\infty} n^{-\min(1,\, 2/(1+\gamma))} \log P(|\overline{X}| > \varepsilon) < 0,$$

$$\limsup_{n\to\infty} n^{-\min(1,\, 1/(1+\gamma))} \log P(|S^2 - 1| > \varepsilon) < 0,$$

and

$$\limsup_{n\to\infty} n^{-\min(1,\, 2/(1+\gamma))} \log P(|\hat{\gamma} - \gamma| > \varepsilon) < 0.$$

3.4. In-Control Behavior

We assume that the observations $X_1, \ldots, X_n, X_{n+1}$ are i.i.d. r.v.'s each with a d.f. F, given by $F(x) = K_\gamma((x-\mu)/\sigma)$ with K_γ belonging to the normal power family. If μ, σ, and γ are known and $FAR = p$, then $UCL = \mu + \overline{K}_\gamma^{-1}(p)\sigma$. As a rule μ, σ, and γ are unknown, and we estimate them by $\overline{X}$, S, and $\hat{\gamma}$. This leads to the observed FAR, given by

$$P_n = P_n(\overline{X}, S, \hat{\gamma}) = P\left(X_{n+1} > \overline{X} + \overline{K}_{\hat{\gamma}}^{-1}(p)S\right) = \overline{K}_\gamma\left(\frac{\overline{X} - \mu}{\sigma} + \overline{K}_{\hat{\gamma}}^{-1}(p)\frac{S}{\sigma}\right).$$

3.4.1. Bias

It was already mentioned that a seemingly innocent extension of the normal family to a larger parametric family in fact causes great and often insuperable complications. The normal power family, being a natural extension in the context of control charts, offers a solution, but a lot of technicalities are still involved. We will not present all the details here, but give a sketch of the main ideas to get approximately unbiased control charts.

We write $c_u(\hat{\gamma})$ for a correction term giving (almost) unbiasedness. This leads to $UCL = \overline{X} + \{\overline{K}_{\hat{\gamma}}^{-1}(p) + c_u(\hat{\gamma})\}S$, and the observed FAR is given by

$$P_n = \overline{K}_\gamma\left(\frac{\overline{X}-\mu}{\sigma} + \{\overline{K}_{\hat{\gamma}}^{-1}(p) + c_u(\hat{\gamma})\}\frac{S}{\sigma}\right) = \overline{K}_\gamma\left(\overline{K}_\gamma^{-1}(p) + V + c_u(\hat{\gamma})\frac{S}{\sigma}\right),$$

where

$$V = \frac{\overline{X}-\mu}{\sigma} + \overline{K}_{\hat{\gamma}}^{-1}(p)\frac{S}{\sigma} - \overline{K}_\gamma^{-1}(p). \tag{3.2}$$

For the estimators $\overline{X}$, S, and $\hat{\gamma}$ we restrict attention to neighborhoods of μ, σ, and γ. The error involved with this is presented in Theorem 3.1. Letting

$$A_n(\varepsilon) = \left\{\left|\frac{\overline{X}-\mu}{\sigma}\right| > \varepsilon, \left|\left(\frac{S}{\sigma}\right)^2 - 1\right| > \varepsilon, |\hat{\gamma}-\gamma| > \varepsilon\right\},$$

we have by Theorem 3.1 $P(A_n(\varepsilon)) \leq \exp\{-\eta n^{\min(1,\,1/(1+\gamma))}\}$ for some $\eta > 0$, and hence for each $\varepsilon > 0$ we have $P(A_n(\varepsilon)) = o(n^{-1})$ as $n \to \infty$. By Taylor expansion of $Eg(P_n)$ and careful evaluation of EV and EV^2, the suitable correction term is obtained. The following theorem presents the result for $g(p) = p$. In that case the correction term is given by

$$c_u(\hat{\gamma}) = -B1_n(\hat{\gamma}) - \frac{1}{2}B2_n(\hat{\gamma})\frac{k'_{\hat{\gamma}}}{k_{\hat{\gamma}}}(\overline{K}_{\hat{\gamma}}^{-1}(p)), \tag{3.3}$$

where $k_\gamma = K'_\gamma$, the density of Z_γ, and where $B1_n(\gamma)$ and $B2_n(\gamma)$ are the first-order terms of EV and EV^2. For explicit formulas of $B1_n(\gamma)$ and $B2_n(\gamma)$, a theorem on general functions g, the proof of the theorem, and more details, we refer to Albers et al. (2004). The theorem shows that indeed the correction does what it should do: give unbiasedness up to order $o(n^{-1})$.

Theorem 3.2. *Let $X_1, \ldots, X_n, X_{n+1}$ be i.i.d. r.v.'s with $(X_i - \mu)/\sigma$ having a normal power distribution with parameter γ. Then we have*

$$EP_n = p + o(n^{-1}) \quad \text{as } n \to \infty.$$

3.4.2. Exceedance Probability

As explained while discussing the normal control chart, we may restrict ourselves without loss of generality to $g(p) = p$ when dealing with exceedance probabilities. Writing $c_e(\hat{\gamma})$ for the correction term involved in this approach, we consider $UCL = \overline{X} + \{\overline{K}_{\hat{\gamma}}^{-1}(p) + c_e(\hat{\gamma})\}S$. The correction term should be chosen in such a way that for suitable (small) values of $\varepsilon \geq 0$ and $\alpha > 0$ we get

$$P\left(\frac{P_n - p}{p} > \varepsilon\right) = P\left(\frac{\overline{K}_\gamma(\overline{K}_\gamma^{-1}(p) + V + c_e(\hat{\gamma})\frac{S}{\sigma}) - p}{p} > \varepsilon\right) \leq \alpha$$

with V given by (3.2). The following theorem gives the required (limiting) correction result.

Theorem 3.3. *Let $X_1, \ldots, X_n, X_{n+1}$ be i.i.d. r.v.'s with $(X_i - \mu)/\sigma$ having a normal power distribution with parameter γ. Define*

$$c_e(\gamma) = \sqrt{B2_n(\gamma, \varepsilon)}u_\alpha + \overline{K}_\gamma^{-1}(p(1+\varepsilon)) - \overline{K}_\gamma^{-1}(p),$$

where $B2_n(\gamma, \varepsilon)$ is obtained from $B2_n(\gamma)$ in (A.9) of Albers et al. (2004) by replacing (twice) $\overline{K}_\gamma^{-1}(p)$ with $\overline{K}_\gamma^{-1}(p(1+\varepsilon))$ and replacing (twice) $\frac{\partial \overline{K}^{-1}_{h^{-1}(\gamma^*)}(p)}{\partial \gamma^*}$ with $\frac{\partial \overline{K}^{-1}_{h^{-1}(\gamma^*)}(p(1+\varepsilon))}{\partial \gamma^*}$. Then

$$\lim_{n\to\infty} P\left(\frac{P_n - p}{p} > \varepsilon\right) = \alpha.$$

Sketch of Proof. Let

$$V_\varepsilon = \frac{\overline{X} - \mu}{\sigma} + \overline{K}_{\hat\gamma}^{-1}(p(1+\varepsilon))\frac{S}{\sigma} - \overline{K}_\gamma^{-1}(p(1+\varepsilon)).$$

We get

$$\begin{aligned} &\frac{\overline{K}_\gamma\left(\overline{K}_\gamma^{-1}(p) + V + c_e(\hat\gamma)\frac{S}{\sigma}\right) - p}{p} > \varepsilon \\ &\iff \overline{K}_\gamma^{-1}(p) + V + c_e(\hat\gamma)\frac{S}{\sigma} < \overline{K}_\gamma^{-1}(p(1+\varepsilon)) \\ &\iff V_\varepsilon + \sqrt{B2_n(\hat\gamma, \varepsilon)}u_\alpha \frac{S}{\sigma} < 0. \end{aligned}$$

Since for the normal power family $V_\varepsilon/\sqrt{B2_n(\gamma, \varepsilon)}$ is asymptotically standard normal and since $B2_n(\hat\gamma, \varepsilon)/B2_n(\gamma, \varepsilon)$ converges in probability to 1, we have

$$\begin{aligned} \lim_{n\to\infty} P\left(\frac{P_n - p}{p} > \varepsilon\right) &= \lim_{n\to\infty} P\left(V_\varepsilon + \sqrt{B2_n(\hat\gamma, \varepsilon)}u_\alpha\left(\frac{S}{\sigma} - 1\right) + \sqrt{B2_n(\hat\gamma, \varepsilon)}u_\alpha < 0\right) \\ &= \lim_{n\to\infty} P(-V_\varepsilon/\sqrt{B2_n(\gamma, \varepsilon)} > u_\alpha) = \alpha. \end{aligned}$$

□

3.5. Out-of-Control Behavior

Under the out-of-control case X_{n+1} is shifted to the right in the sense that it is distributed as $\mu + d\sigma + \sigma Z_\gamma$. Let $p_1 = \overline{K}_\gamma(\overline{K}_\gamma^{-1}(p) - d)$ be the out-of-control rate when the parameters μ, σ, and γ are known. The expectation of the random out-of-control rate when applying the corrected control chart with $UCL = \overline{X} + \{\overline{K}_{\hat\gamma}^{-1}(p) + c_e(\hat\gamma)\}S$ can be approximated in the following way (here E_d denotes the out-of-control expectation and E refers to the in-control expectation, that is, with $d = 0$):

$$\begin{aligned} E_d P_n &= E\overline{K}_\gamma\left(\frac{\overline{X} - \mu}{\sigma} + \{\overline{K}_{\hat\gamma}^{-1}(p) + c_e(\hat\gamma)\}\frac{S}{\sigma} - d\right) \approx \overline{K}_\gamma\left(\overline{K}_\gamma^{-1}(p) + c_e(\gamma) - d\right) \\ &\approx p_1 - c_e(\gamma)k_\gamma\left(\overline{K}_\gamma^{-1}(p) - d\right). \end{aligned}$$

Straightforward calculation shows that (for $p_1 < \frac{1}{2}$)

$$\frac{k_\gamma\big(\overline{K}_\gamma^{-1}(p) - d\big)}{p_1} = \frac{k_\gamma\big(\overline{K}_\gamma^{-1}(p) - d\big)}{\overline{K}_\gamma\big(\overline{K}_\gamma^{-1}(p) - d\big)} = \frac{u_{p_1}^{-\gamma}}{(1+\gamma)c(\gamma)}\frac{\varphi(u_{p_1})}{\overline{\Phi}(u_{p_1})} \approx \frac{4(1+u_{p_1})}{5(1+\gamma)c(\gamma)u_{p_1}^{\gamma}}.$$

Hence, we get

$$\frac{E_d P_n - p_1}{p_1} \approx -c_e(\gamma)\frac{4(1+u_{p_1})}{5(1+\gamma)c(\gamma)u_{p_1}^{\gamma}}.$$

The same holds in the bias case, replacing $c_e(\gamma)$ by $c_u(\gamma)$. Just as for the normal control chart, we may conclude that the correction terms do not disturb the behavior of the control charts in the out-of-control situation.

4. NONPARAMETRIC CONTROL CHARTS

The *ME* can be avoided completely by using a nonparametric control chart. The idea is as follows. Suppose that F is known. Then a control chart with $FAR = p$ is easily obtained by taking $UCL = F^{-1}(p)$. The nonparametric control chart is obtained by estimating $F(x)$ by the empirical d.f. $F_n(x) = n^{-1}\sum_{i=1}^{n} 1(X_i \le x)$ with $1(A) = 1$ if A holds and 0 otherwise. The corresponding quantile function $F_n^{-1}(t) = \inf\{x | F_n(x) \ge t\}$ leads to $UCL = \overline{F}_n^{-1}(p) = F_n^{-1}(1-p) = X_{(n-[np])}$. For some closely related charts, see Ion et al. (2000) and Willemain and Runger (1996); for a recent overview of nonparametric charts in general, see, e.g., Chakraborti et al. (2001).

4.1. In-Control Behavior

Consider the in-control situation, that is, $X_1, \ldots, X_n, X_{n+1}$ are i.i.d. r.v.'s each with (continuous) d.f. F. The uncorrected nonparametric control chart has $ME = 0$, but its *SE* is very large. Take, e.g., $p = 0.001$ and $n = 500$; then $r = 0$ and the random *FAR* $P_{100} = \overline{F}(X_{(500)})$, and thus $EP_{500} = 1/501$, which is about twice as much as it should be, even though we have 500 Phase I observations. As for the normal and parametric control chart, we discuss both the bias and the exceedance probability approach.

4.1.1. Bias

To reduce the bias we can apply a randomization procedure as follows. Let $U_{(1)} \le \cdots \le U_{(n)}$ be the order statistics of the random sample $U_1, \ldots, U_n$ from a uniform distribution on (0, 1) and define $U_{(0)} = 0$ and $U_{(n+1)} = 1$. For an increasing g, define the integer r with $0 \le r = r(p) \le n$ by

$$Eg(U_{(r)}) \le g(p) < Eg(U_{(r+1)}). \tag{4.1}$$

Let V be an r.v. independent of $X_1, \ldots, X_{n+1}$ taking as its values 0 and 1. Replace the control chart by

$$X_{n+1} > VX_{(n-r)} + (1-V)X_{(n-r+1)} \quad \text{with} \quad P(V=1) = \frac{g(p) - Eg(U_{(r)})}{Eg(U_{(r+1)}) - Eg(U_{(r)})}, \tag{4.2}$$

where in the case $r = 0$ we define $X_{(n+1)} = \infty$.

In particular, for $g(p) = p$ we get $r = [p(n+1)]$, and the nonparametric control chart reads as

$$X_{n+1} > VX_{(n-[p(n+1)])} + (1-V)X_{(n-[p(n+1)]+1)} \quad \text{with} \quad P(V=1) = p(n+1) - [p(n+1)].$$

Similarly, for a decreasing g, define $0 \le r = r(p) \le n$ by

$$Eg(U_{(r)}) \ge g(p) > Eg(U_{(r+1)}). \tag{4.3}$$

The control chart is again given by (4.2). In particular, for $g(p) = \frac{1}{p}$ we get $r = [np] + 1$, and provided that $r \ge 2$ (that is $np \ge 1$), the nonparametric control chart reads as

$$X_{n+1} > VX_{(n-[np]-1)} + (1-V)X_{(n-[np])} \quad \text{with} \quad P(V=1) = \frac{([np]+1)(np-[np])}{np}.$$

When $r = 1$ and $g(p) = \frac{1}{p}$, the nonparametric control chart gives an out-of-control signal if $X_{n+1} > X_{(n-1)}$, and hence $P_n = \overline{F}(X_{(n-1)})$, implying $E\frac{1}{P_n} = E\frac{1}{U_{(2)}} = n < \frac{1}{p}$.

Theorem 4.1. *Let $X_1, \ldots, X_n, X_{n+1}$ be i.i.d. r.v.'s each with (continuous) d.f. F. Let g be an increasing or a decreasing function and let r be defined by* (4.1) *and* (4.3), *respectively. Assume that $|Eg(U_{(r+1)})| < \infty$ and $|Eg(U_{(r)})| < \infty$. The control chart given by* (4.2) *satisfies*

$$Eg(P_n) = g(p).$$

Proof. Note that P_n is now defined as the probability of a false alarm, given $X_1, \ldots, X_n$ and V, that is, $P_n = V\overline{F}(X_{(n-r)}) + (1-V)\overline{F}(X_{(n-r+1)})$. Since $\overline{F}(X_{(n-r)})$ and $\overline{F}(X_{(n-r+1)})$ are distributed as $U_{(r+1)}$ and $U_{(r)}$, respectively, we get

$$\begin{aligned} Eg(P_n) &= P(V=1)Eg(U_{(r+1)}) + P(V=0)Eg(U_{(r)}) \\ &= Eg(U_{(r)}) + P(V=1)\{Eg(U_{(r+1)}) - Eg(U_{(r)})\} = g(p). \end{aligned}$$ □

From a practical point of view the nonparametric control chart is still questionable for $r = 0$, because it implies that with positive probability we will never get an out-of-control signal! Therefore a modification of the nonparametric control in case $r = 0$ is presented in Albers et al. (2006). We do not discuss this modification here.

4.1.2. Exceedance Probability

As before, we can restrict ourselves without loss of generality to $g(p) = p$. To obtain

$$P\left(\frac{P_n - p}{p} > \varepsilon\right) \le \alpha$$

for the uncorrected nonparametric control chart at some reasonable values of ε and α, we need really huge sample sizes. For instance, taking $p = 0.001$, $\varepsilon = 0.1$, and $\alpha = 0.2$, we need $n = 88,021$. To find suitable corrections, we consider $UCL = VX_{(n-[np]+k-1)} + (1-V)X_{(n-[np]+k)}$ for some $k \ge 0$. Let $B(n, \tilde{p}, y)$ denote the cumulative binomial probability $P(Y \le y)$ with Y $bin(n, \tilde{p})$. Then the following theorem gives the right correction.

Theorem 4.2. *Let* $k \geq 0$ *be such that* $B(n, p(1+\varepsilon), [np]-k) \leq \alpha < B(n, p(1+\varepsilon), [np]-k+1)$, *and let*

$$P(V=1) = \frac{\alpha - B(n, p(1+\varepsilon), [np]-k)}{B(n, p(1+\varepsilon), [np]-k+1) - B(n, p(1+\varepsilon), [np]-k)}.$$

Then

$$P\left(\frac{P_n - p}{p} > \varepsilon\right) = \alpha.$$

Proof. We have

$$\begin{aligned}
P\left(\frac{P_n - p}{p} > \varepsilon\right) &= P(P_n > p(1+\varepsilon)) \\
&= P(V=1)P(\overline{F}(X_{(n-[np]+k-1)}) > p(1+\varepsilon)) \\
&\quad + P(V=0)P(\overline{F}(X_{(n-[np]+k)}) > p(1+\varepsilon)) \\
&= P(V=1)P(U_{([np]-k+2)} > p(1+\varepsilon)) \\
&\quad + P(V=0)P(U_{([np]-k+1)} > p(1+\varepsilon)) \\
&= P(V=1)B(n, p(1+\varepsilon), [np]-k+1) \\
&\quad + P(V=0)B(n, p(1+\varepsilon), [np]-k) \\
&= \alpha.
\end{aligned}$$

□

When $[np] = 0$ and $\lim_{n\to\infty} np(1+\varepsilon) < |\log \alpha|$, then we get $\lim_{n\to\infty} B(n, p(1+\varepsilon), [np]) = \lim_{n\to\infty}(1-p(1+\varepsilon))^n > \alpha$, and hence $k = 1$, implying that with positive probability we will never get an out-of-control signal. Hence, we should have a sufficiently large sample size to avoid such effects. On the other hand, much smaller sample sizes are needed than without correction.

4.2. Out-of-Control Behavior

The new observation X_{n+1} has in the out-of-control situation d.f. $F(x-d)$ with $d > 0$, as we restrict attention to *UCL*s. Typically $p_1 = \overline{F}(\overline{F}^{-1}(p) - d)$ may still be small, but not extremely so, like p. We compare the uncorrected chart where $UCL = \overline{F}_n^{-1}(p) = X_{(n-[np])}$ with a corrected one of the form $UCL = VX_{(n-[np]+k-1)} + (1-V)X_{(n-[np]+k)}$ for some $k \geq 0$. The following theorem gives the result.

Theorem 4.3. *Replacement of* $UCL = X_{(n-[np])}$ *by* $UCL = VX_{(n-[np]+k-1)} + (1-V)X_{(n-[np]+k)}$ *for some* $k \geq 0$ *results in a relative change in* $E_d P_n$ *approximately equal to*

$$-\frac{\{k - P(V=1)\}}{p_1}\frac{f(\overline{F}^{-1}(q) - d)}{f(\overline{F}^{-1}(q))}$$

in which $f = F'$, $q = ([np]+1)/(n+1)$, *provided that* $[np]$ *is not too small. For* $[np] = 0$ *and* $k = 1$ *the reduction of* $E_d P_n$ *equals* $P(V=1)$.

Proof. If X_{n+1} has d.f. $F(x-d)$, it follows that P_n with the uncorrected control limit $UCL = X_{(n-[np])}$ is distributed as $\overline{F}(\overline{F}^{-1}(U_{([np]+1)}) - d)$, and thus E_dP_n can be approximated by $\overline{F}(\overline{F}^{-1}(EU_{([np]+1)}) - d) = \overline{F}(\overline{F}^{-1}(q) - d)$. The change in E_dP_n caused by replacing $X_{(n-[np])}$ by $X_{(n-[np]+k)}$ approximately equals $-kf(\overline{F}^{-1}(q) - d)/f(\overline{F}^{-1}(q))$. Therefore, the change in E_dP_n when taking $UCL = VX_{(n-[np]+k-1)} + (1-V)X_{(n-[np]+k)}$ instead of $X_{(n-[np])}$ equals $-(k-1)P(V=1)$ $f(\overline{F}^{-1}(q) - d)/f(\overline{F}^{-1}(q)) - kP(V=0)f(\overline{F}^{-1}(q) - d)/f(\overline{F}^{-1}(q))$, and the first result of the theorem immediately follows. When $[np] = 0$ and $k = 1$, we have $\overline{F}(VX_{(n-[np]+k-1)} + (1-V)X_{(n-[np]+k)}) = \overline{F}(VX_{(n)} + (1-V)X_{(n+1)}) = V\overline{F}(X_{(n)})$, and thus E_dP_n is reduced by a factor $P(V=1)$. □

Examples show that a considerable price has to be paid in terms of out-of-control performance, unless n or p are sufficiently large. For more details we refer to Albers and Kallenberg (2004b).

5. COMBINED CONTROL CHART

All three types of charts discussed so far have their own merits, if they are used individually; however, all three also have disadvantages if the proper conditions for the specific chart are not opportune. For instance, when normality holds, we should not use the nonparametric chart, etc. Therefore, we introduce a combined chart by choosing between the three available charts. Since the form of the distribution in the tails is the key issue, the choice between the three charts is based on the tail behavior, as expressed by the data. Hence, we take the rescaled maximum $(X_{(n)} - \overline{X})/S$ as starting point. We restrict ourselves here to the bias situation with $g(p) = p$.

We consider the following combined control chart. When

$$\overline{\Phi}^{-1}\left(\frac{-0.7 + 0.5\log n}{n}\right) \le \frac{X_{(n)} - \overline{X}}{S} \le \overline{\Phi}^{-1}\left(\frac{5}{n\sqrt{n}}\right) \tag{5.1}$$

the normal chart is chosen, that is, we take as the UCL

$$UCL_N = \overline{X} + \left(u_p + \frac{u_p(u_p^2 + 3)}{4n}\right)S.$$

The idea is to stay as long as possible at the normal chart. Under standard normality we have $P(X_{(n)} < \overline{\Phi}^{-1}((-0.7 + 0.5\log n)/n)) \approx 2/\sqrt{n}$ and $P(X_{(n)} > \overline{\Phi}^{-1}(5/(n\sqrt{n}))) \approx 5/\sqrt{n}$. Distributions with heavier tails than the normal one give problems with the in-control behavior, leading for common distributions to EP_n being 4 or even 12 times as large as it should be; see Table 1 in Albers et al. (2004). Distributions with thinner tails are conservative in the in-control case with in consequence, a loss in the out-of-control case. Because errors in the in-control case are more serious (the control chart is then invalid) than those in the out-of-control case, and since a positive ME as large as p can easily occur, while a negative ME is at most $-p$, we take the selection rule unbalanced. The particular choice of the boundaries is partly by theoretical arguments (see Theorem 5.1) but also based on our simulation experience. For a more extensive discussion we refer to Albers et al. (2006).

When (5.1) does not hold and

$$\overline{K}_{\hat{\gamma}}^{-1}\left(\frac{-0.2+0.5\log n}{n}\right) \le \frac{X_{(n)}-\overline{X}}{S} \le \overline{K}_{\hat{\gamma}}^{-1}\left(\frac{3}{n\sqrt{n}}\right), \tag{5.2}$$

the parametric chart is chosen with *UCL*

$$UCL_P = \overline{X} + \left\{\overline{K}_{\hat{\gamma}}^{-1}(p) + c_u(\hat{\gamma})\right\}S$$

with $c_u(\hat{\gamma})$ given by (3.3). When both (5.1) and (5.2) are violated, the nonparametric chart is chosen with *UCL*

$$UCL_{NP} = VX_{(n-[p(n+1)])} + (1-V)X_{(n-[p(n+1)]+1)} \quad \text{with} \quad P(V=1) = p(n+1) - [p(n+1)].$$

The next theorem shows that the combined chart behaves asymptotically as well as each of the individual charts on their own domain, both with respect to the in-control case as for the out-of-control case.

Theorem 5.1.

(i) *Let* $X_1, \ldots, X_n, X_{n+1}$ *be i.i.d. r.v.'s with* $X_i \sim N(\mu, \sigma^2)$ *for* $i = 1, \ldots, n$ *and* $X_{n+1} \sim N(\mu + d\sigma, \sigma^2)$. *Then for* $d = 0$ (*in control*) *as well as for* $d > 0$ (*out of control*), *we have*

$$|E_d P_n^c - E_d P_n^N| \le \frac{e^{0.7}+5}{\sqrt{n}}(1+o(1)) \quad \text{as } n \to \infty,$$

where P_n^c *is the observed FAR of the combined control chart and* P_n^N *the one of the normal control chart.*

(ii) *Let* $X_1, \ldots, X_n, X_{n+1}$ *be i.i.d. r.v.'s with* X_i *distributed as* $\mu + \sigma Z_\gamma$ *for* $i = 1, \ldots, n$ *and* X_{n+1} *distributed as* $X_1 + d\sigma$, *where* Z_γ *has a normal power distribution with* $\gamma \ne 0$. *Then for* $d = 0$ (*in control*) *as well as for* $d > 0$ (*out of control*), *we have*

$$|E_d P_n^c - E_d P_n^P| \le \frac{e^{0.2}+3}{\sqrt{n}}(1+o(1)) \quad \text{as } n \to \infty,$$

where P_n^c *is the observed FAR of the combined control chart and* P_n^P *the one of the parametric control chart.*

(iii) *Let* $X_1, \ldots, X_n, X_{n+1}$ *be i.i.d. r.v.'s with* X_i *having d.f.* F *for* $i = 1, \ldots, n$ *and* X_{n+1} *distributed as* $X_1 + d$. *Let* $EX_1 = \mu$, $var(X_1) = \sigma^2$, *and let* γ *be defined as the limit of the estimator* $\hat{\gamma}$ *under* F, *that is, by*

$$\gamma = \frac{\log\left(\frac{F^{-1}(0.95)-\mu}{F^{-1}(0.75)-\mu}\right)}{\log\left(\frac{\Phi^{-1}(0.95)}{\Phi^{-1}(0.75)}\right)} - 1.$$

Then, for each $\varepsilon_i, \eta_i, \zeta_i > 0$, $i = 1, \ldots 4$, *with* $\zeta_3, \zeta_4 < 1 + \gamma$, *we have for sufficiently large* n

$$\left|E_d P_n^c - E_d P_n^{NP}\right| \le \min\{m_1, m_2\} + \min\{m_3, m_4\},$$

where P_n^c *is the observed FAR of the combined control chart and* P_n^{NP} *the one of the nonparametric control chart and where*

$$m_1 = F\big(\mu + \sigma\big(\sqrt{1+\varepsilon_1} + \zeta_1\big)\sqrt{2\log n}\big)^n + P\bigg(\bigg|\frac{\overline{X}-\mu}{\sigma}\bigg| > \eta_1\bigg) + P\bigg(\bigg|\frac{S^2}{\sigma^2} - 1\bigg| > \varepsilon_1\bigg),$$

$$m_2 = 1 - F\big(\mu + \sigma\big(\sqrt{1-\varepsilon_2} - \zeta_2\big)\sqrt{2\log n}\big)^n + P\bigg(\bigg|\frac{\overline{X}-\mu}{\sigma}\bigg| > \eta_2\bigg) + P\bigg(\bigg|\frac{S^2}{\sigma^2} - 1\bigg| > \varepsilon_2\bigg),$$

$$m_3 = F\big(\mu + \sigma\big(\sqrt{\log n}\big)^{1+\gamma+2\zeta_3}\big)^n + P\bigg(\bigg|\frac{\overline{X}-\mu}{\sigma}\bigg| > \eta_3\bigg) + P\bigg(\bigg|\frac{S^2}{\sigma^2} - 1\bigg| > \varepsilon_3\bigg) + P(|\hat{\gamma} - \gamma| > \zeta_3),$$

$$m_4 = 1 - F\big(\mu + \sigma\big(\sqrt{\log n}\big)^{1+\gamma-2\zeta_4}\big)^n + P\bigg(\bigg|\frac{\overline{X}-\mu}{\sigma}\bigg| > \eta_4\bigg) + P\bigg(\bigg|\frac{S^2}{\sigma^2} - 1\bigg| > \varepsilon_4\bigg) + P(|\hat{\gamma} - \gamma| > \zeta_4).$$

Theorem 5.1 only makes sense if F differs from the normal family in the sense that for some $\varepsilon > 0$

$$\lim_{n\to\infty}\big[F\big(\mu + \sigma(1+\varepsilon)\sqrt{2\log n}\big)\big]^n = 0$$

(heavier tail than the normal distribution) or

$$\lim_{n\to\infty}\big[F\big(\mu + \sigma(1-\varepsilon)\sqrt{2\log n}\big)\big]^n = 1$$

(thinner tail than the normal distribution) and F is outside the normal power family in the sense that for some $\varepsilon > 0$

$$\lim_{n\to\infty}\big[F\big(\mu + \sigma(\sqrt{\log n})^{1+\gamma+\varepsilon}\big)\big]^n = 0$$

(heavier tail than the normal power family) or

$$\lim_{n\to\infty}\big[F\big(\mu + \sigma(\sqrt{\log n})^{1+\gamma-\varepsilon}\big)\big]^n = 1$$

(lighter tail than the normal power family).

Proof. It is not hard to see that

$$\big|E_d P_n^c - E_d P_n^N\big| \le P\bigg(\frac{X_{(n)} - \overline{X}}{S} \notin \bigg[\overline{\Phi}^{-1}\bigg(\frac{-0.7 + 0.5\log n}{n}\bigg), \overline{\Phi}^{-1}\bigg(\frac{5}{n\sqrt{n}}\bigg)\bigg]\bigg).$$

A careful analysis, using large deviation theory, leads to

$$P\bigg(\frac{X_{(n)} - \overline{X}}{S} < \overline{\Phi}^{-1}\bigg(\frac{-0.7 + 0.5\log n}{n}\bigg)\bigg) = \bigg(1 - \frac{-0.7 + 0.5\log n}{n}\bigg)^n (1 + o(1)) = \frac{e^{0.7}}{\sqrt{n}}(1 + o(1)) \text{ as } n \to \infty,$$

and

$$P\left(\frac{X_{(n)} - \overline{X}}{S} > \overline{\Phi}^{-1}\left(\frac{5}{n\sqrt{n}}\right)\right) = \frac{5}{\sqrt{n}}(1 + o(1)) \quad \text{as } n \to \infty,$$

which completes the proof of (i). The proofs of (ii) and (iii) are along the same line of argument, but in particular the proof of (ii) is technically much more complicated. For details of the proof and for more general statements of the theorem we refer to Albers et al. (2006). □

6. GROUPED OBSERVATIONS

As shown in the previous section, the combined chart has very nice properties in the sense that it behaves as the appropriate chart according to the underlying distribution. When the nonparametric chart is chosen, even though this is the best thing to do, a lot of Phase I observations are needed to have a good performance; see also Section 4. In fact, in such a case we cannot improve much when considering an individual chart. As noted in the introduction, a more fundamental solution is to use a (small) group of observations. The essential point is that we may postpone the decision until somewhat more observations have arrived. When the process goes out of control, it is sometimes hard to see it on the basis of one observation, but if two or more observations show different behavior, it is easier to recognize it. In this section we discuss several charts for grouped observations. In fact, two types of comparisons play a role. In the first place, for each fixed value of the group size m, various monitoring statistics can be compared. Second, each given type of statistic can also be compared for varying m. Even the normal case is not quite trivial in this respect and still leads to some interesting insights. The point is, of course, that we are not dealing with a single given out-of-control situation, implying that the optimal choice of m will vary according to the alternative considered. We do not focus here on the estimation part of the problem, but estimation is nevertheless present in the background, since the UCLs of the monitoring statistics should be estimated in a nonparametric way, and the possibility and consequences of such an estimation procedure should be taken into account.

So we start by considering only Phase II observations with a known (but not necessarily normal) underlying distribution. That is, we have a (small) group of observations $X_{n+1}, \ldots, X_{n+m}$ (with $m = 1, \ldots, 5$, thus including the individual chart as well), which are either in control, that is, they are distributed as X_1, with d.f. F, say, or they are out of control and are distributed as $X_1 + d$ with $d > 0$. A chart is defined by a statistic $w(X_{n+1}, \ldots, X_{n+m})$ and a $UCL(w, m)$, and an alarm is produced when

$$w(X_{n+1}, \ldots, X_{n+m}) > UCL(w, m).$$

To compare the charts for different values of m in a fair way, we match the ARLs in the in-control situation. Hence, writing $F_{w,m}$ for the d.f. of $w(X_{n+1}, \ldots, X_{n+m})$ in the in-control case, we have

$$UCL(w, m) = \overline{F}_{w,m}^{-1}(mp). \tag{6.1}$$

The performance of several statistics $w(X_{n+1}, \ldots, X_{n+m})$ (and several values of m) are investigated in Albers and Kallenberg (2006a) by their ARL under the out-of-control case: the smaller the ARL, the better the chart. Here we restrict attention to two of them, the obvious first choice (at least under normality) taking the AVE and the MIN of $X_{n+1}, \ldots, X_{n+m}$.

6.1. AVE

The AVE chart is based on

$$w(X_{n+1}, \ldots, X_{n+m}) = m^{1/2}\overline{X}^{(m)} \quad \text{with } \overline{X}^{(m)} = m^{-1}\sum_{i=1}^{m} X_{n+i}.$$

When normality holds this clearly is the optimal choice, but also in a nonparametric context it is a potential candidate. When F is known and F_m^* is the d.f. of the convolution $X_1 + \cdots + X_m$, then we get (see equation 6.1)

$$UCL = \overline{F}_{w,m}^{-1}(mp) = m^{-1/2}\overline{F}_m^{*\,-1}(mp). \tag{6.2}$$

Let us discuss some results on the estimation step for this chart in the nonparametric case. Suppose that we have Phase I observations $X_1, \ldots, X_n$. For the (uncorrected) individual chart ($m = 1$) we take $UCL = \overline{F}_n^{-1}(p)$, where F_n is the empirical d.f. of $X_1, \ldots, X_n$. Similarly, the d.f. of the convolution F_m^* is estimated nonparametrically by the empirical d.f. of the convolution, defined by

$$F_{mn}^*(x) = \frac{1}{\binom{n}{m}} \sum_{1 \le i_1 < \cdots < i_m \le n}\cdots\sum 1(X_{i_1} + \cdots + X_{i_m} \le x).$$

This leads, according to (6.2), to

$$UCL = m^{-1/2}\overline{F}_{mn}^{*\,-1}(mp).$$

Consider the exceedance probability criterion. Then we are looking for a corrected version of the form $UCL = m^{-1/2}\overline{F}_{mn}^{*-1}(mq)$, say, with $q = q(\varepsilon, \alpha)$ such that for suitable (small) values of $\varepsilon \ge 0$ and $\alpha > 0$ we get

$$P\left(\frac{P_n - mp}{mp} > \varepsilon\right) \le \alpha,$$

where P_n is the observed FAR, given by

$$P_n = P\big(m^{1/2}\overline{X}^{(m)} > m^{-1/2}\overline{F}_{mn}^{*\,-1}(mq)\big) = \overline{F}_m^*\big(\overline{F}_{mn}^{*\,-1}(mq)\big).$$

It can be shown (see Lemma 1 in Albers and Kallenberg, 2006b) that

$$P(P_n > mp(1+\varepsilon)) = P\left(\overline{F}_{mn}^*\big(\overline{F}_m^{*\,-1}(mp(1+\varepsilon))\big) \le \frac{[\binom{n}{m}mq]}{\binom{n}{m}}\right).$$

The question is whether taking a group of size m is helpful in the estimation part in the sense that the range of p and n for which we get a useful asymptotic expression is larger than in the individual case. When relying on asymptotic normality we therefore have to consider the limiting behavior of $\overline{F}^*_{mn}\big(\overline{F}^{*\,-1}_m(mp_n(1+\varepsilon_n))\big)$ or, more generally, $\overline{F}^*_{mn}(t_n)$. On the one hand, the number of terms in the empirical d.f. of the convolution is much larger than for the empirical d.f. of $X_1, \ldots, X_n$. On the other hand, the terms are dependent. More terms are in general favorable for asymptotic normality, but dependence has a negative influence. The following theorem gives the asymptotic normality.

Theorem 6.1. *Define*

$$\{s_n(t)\}^2 = P(X_1 + X_2 + \cdots + X_m > t, X_1 + \widetilde{X}_2 + \cdots + \widetilde{X}_m > t) - \{\overline{F}^*_m(t)\}^2,$$

where $X_1, X_2, \ldots, X_m, \widetilde{X}_2, \ldots, \widetilde{X}_m$ are i.i.d. r.v.'s with d.f. F. Further define

$$\gamma_{0,n} = n^{-1/2} E\left|\frac{\overline{F}^*_{m-1}(t_n - X_1) - \overline{F}^*_m(t_n)}{s_n(t_n)}\right|^3,$$

$$\gamma_{3,r,n} = \frac{4(m-1)}{n^{1/2}(n-1)} \frac{[\overline{F}^*_m(t_n)\{1-\overline{F}^*_m(t_n)\}^r + \{1-\overline{F}^*_m(t_n)\}\{\overline{F}^*_m(t_n)\}^r]^{1/r}}{s_n(t_n)} \quad \textit{for } r \geq 1.$$

Then there exists a constant $C \in \mathbb{R}$, such that for $\frac{3}{2} \leq r < 2$,

$$\sup_{x\in\mathbb{R}}\left|P\left(\frac{\sqrt{n}}{m}\frac{\overline{F}^*_{mn}(t_n) - \overline{F}^*_m(t_n)}{s_n(t_n)} \leq x\right) - \Phi(x)\right|$$
$$\leq C\left(\gamma_{0,n} + \frac{1}{2-r} n^{13/6}\gamma_{0,n}^{1/3}\gamma_{3,r,n}^r + n^{4/3}\gamma_{0,n}^{2/3}\gamma_{3,3/2,n}\right).$$

The estimate remains true for $r = 2$ if $1/(2-r)$ is replaced by $\log n$.

The proof is based on the Berry-Esseen bound given in Theorem 2.1(a), (c) of Friedrich (1989); see Albers and Kallenberg (2005b). Application of Theorem 6.1 yields for $m = 2$ and ε_n bounded when $F = \Phi$, the standard normal distribution, that asymptotic normality of $\overline{F}^*_{mn}\big(\overline{F}^{*\,-1}_m(mp_n(1+\varepsilon_n))\big)$ holds if

$$\lim_{n\to\infty} np_n|\log p_n|^{1/2} = \infty \quad \text{or} \quad p_n = \frac{a_n}{n\sqrt{\log n}} \quad \text{with } a_n \to \infty,$$

while for $F(x) = 1 - \exp(-x)$, the standard exponential distribution, we get

$$\lim_{n\to\infty} \frac{np_n}{|\log p_n|} = \infty \quad \text{or} \quad p_n = \frac{a_n \log n}{n} \quad \text{with } a_n \to \infty.$$

Compared to $m = 1$, where asymptotic normality is obtained when $\lim_{n\to\infty} np_n = \infty$, a relaxation in the sense of a slightly larger range of p_n's for which asymptotic normality holds is possible ($F = \Phi$) as well as a restriction to a smaller ranges of admissible p_n's ($F(x) = 1 - \exp(-x)$), depending on the d.f. of the observations. When $m = 1$ and $\lim_{n\to\infty} np_n < \infty$, we get convergence to a Poisson distribution. This is not true for $m > 1$; see Albers and Kallenberg (2005b). The conclusion therefore is that in the estimation step we do not get a helpful progress when taking groups and applying *AVE*.

6.2. MIN

The statistic involved here is the smallest of $X_{n+1}, \dots, X_{n+m}$, that is,

$$w(X_{n+1}, \dots, X_{n+m}) = \min(X_{n+1}, \dots, X_{n+m}).$$

When using *MIN* we take advantage of the effect that in a group the observations intensify each other. That is, if m observations are pretty large and not necessarily extremely large, this is already enough evidence to give an alarm. In contrast to when taking the maximum, here really the group is used; see also Albers and Kallenberg (2006a). Because under the in-control case

$$\overline{F}_{MIN,m}(y) = P(\min(X_{n+1}, \dots, X_{n+m}) > y) = \{\overline{F}(y)\}^m,$$

we get as the *UCL* (see equation 6.1)

$$UCL = \overline{F}^{-1}_{MIN,m}(mp) = \overline{F}^{-1}(\{mp\}^{1/m}).$$

As concerns the estimation step, it is now easily seen that for asymptotic normality it is only needed that $\lim_{n\to\infty} np_n^{1/m} = \infty$, and indeed, when using *MIN* we benefit from dealing with much less extreme quantiles, which facilitates the estimation step substantially. While getting asymptotic exceedance probability equal to α for the *AVE* chart requires a lot of intricate conditions (see Theorem 4 in Albers and Kallenberg, 2005b), for *MIN* this is much easier, as is seen in the following theorem.

Theorem 6.2. *Let* p_n *satisfy* $\lim_{n\to\infty} p_n = 0, \lim_{n\to\infty} np_n^{1/m} = \infty$, *and suppose that* $\varepsilon_n \geq 0$ *is bounded. Let* P_n *be the observed FAR for the corrected minimum control chart with* $UCL = \overline{F}_n^{-1}(\{mq_n\}^{1/m})$, *where*

$$q_n = p_n(1+\varepsilon_n) - \frac{mu_\alpha p_n(1+\varepsilon_n)}{\sqrt{n\{mp_n(1+\varepsilon_n)\}^{1/m}}}(1+o(1)) \quad \text{as } n \to \infty,$$

then

$$\lim_{n\to\infty} P\left(\frac{P_n - mp_n}{mp_n} > \varepsilon_n\right) = \alpha$$

Proof. A signal is given when

$$\min(X_{n+1}, \dots, X_{n+m}) > \overline{F}_n^{-1}(\{mq_n\}^{1/m}),$$

and hence

$$P_n = \left\{\overline{F}\left(\overline{F}_n^{-1}(\{mq_n\}^{1/m})\right)\right\}^m.$$

This implies (see also the proof of Theorem 4.2)

$$\begin{aligned} P\left(\frac{P_n - mp_n}{mp_n} > \varepsilon_n\right) &= P(P_n > mp_n(1+\varepsilon_n)) \\ &= P\left(\overline{F}\left(\overline{F}_n^{-1}(\{mq_n\}^{1/m})\right) > \{mp_n(1+\varepsilon_n)\}^{1/m}\right) \\ &= B(n, \{mp_n(1+\varepsilon_n)\}^{1/m}, [n\{mq_n\}^{1/m}]). \end{aligned}$$

The proof is completed by using the asymptotic normality of the binomial distribution. □

6.3. Comparison of *AVE* and *MIN* under the Out-of-Control Case

Clearly, from the estimation point of view *MIN* is far more attractive than *AVE*. However, we should also compare their out-of-control behavior. Therefore we consider the *ARL* of both procedures under the out-of-control case. We restrict ourselves here to the situation where F is known, thus ignoring the estimation effects. They have been considered before (when the process is in control), and they are less important under the out-of-control case. The *FAR* of *MIN* during the out-of-control case is given by

$$P\big(\min(X_{n+1}, \ldots, X_{n+m}) + d > \overline{F}^{-1}(\{mp\}^{1/m})\big) = \big\{\overline{F}(\overline{F}^{-1}(\{mp\}^{1/m}) - d)\big\}^m,$$

and thus

$$ARL(MIN, m, d) = \frac{m}{\{\overline{F}(\overline{F}^{-1}((mp)^{1/m}) - d)\}^m}.$$

The most favorable distribution for *AVE* is the normal distribution. When $F = \Phi$ we get $UCL = u_{mp}$ and

$$ARL(AVE, m, d) = \frac{m}{\overline{\Phi}(u_{mp} - m^{1/2}d)}.$$

Figures 1 and 2 give an impression of the *ARL*s for different shifts. On the horizontal axis the *ARL* of the individual chart is presented, while on the vertical axis the difference with the individual chart is given, that is, in Figure 1 $ARL(1, d) - ARL(AVE, m, d)$ against $ARL(1, d)$, and in Figure 2 $ARL(1, d) - ARL(MIN, m, d)$ against $ARL(1, d)$, where $ARL(1, d) = ARL(AVE, 1, d) = ARL(MIN, 1, d)$. In the figures $ARL(1, d)$ is shortly denoted as *IND*, $ARL(AVE, m, d)$ as $AVE(m)$, and $ARL(MIN, m, d)$ as $MIN(m)$.

Both for *AVE* and *MIN* a substantial gain can be obtained when using larger values of m, in particular for smaller shifts and hence larger *ARL*s. For shifts with $d \geq 1$, corresponding to $ARL(1, d) \leq 55$, the differences between $m = 3, 4, 5$ are rather small. Further, we see that even for normally distributed observations *MIN* actually performs quite well, in particular if we compare it with the individual chart. For example, at $d = 1$ we get $ARL(1, d) = 54.6$; it is improved with 26.7 by taking *MIN* with $m = 3$, yielding $ARL = 27.9$; the further improvement when using *AVE* with $m = 3$ is much less: 8.5, giving $ARL = 19.4$.

As a second distribution we consider a skew distribution, the gamma distribution with parameters 4 and 1 having density $\frac{1}{6}x^3e^{-x}$. Its coefficient of skewness equals 1. In Figure 3 the difference of the *ARL*s of *AVE* and *MIN* are plotted against the *ARL* of *AVE*.

It is seen that *MIN* is somewhat better than *AVE*. Both of them are much better than the individual chart. For instance, the *ARL* of the individual chart at $d = 1$ equals 213.2, and the *ARL* of the *MIN* chart at $d = 1$ equals 79.6, 41.1, 26.2, 19.3 for $m = 2, 3, 4, 5$, respectively, while the *AVE* chart at $d = 1$ gives 87.1, 47.8, 31.4, 23.3 for $m = 2, 3, 4, 5$, respectively.

From these and other distributions we have investigated (see Albers and Kallenberg, 2006a), together with the results on the estimation step, we conclude that the chart based on a group of $m = 2, \ldots, 5$ in general performs better than the

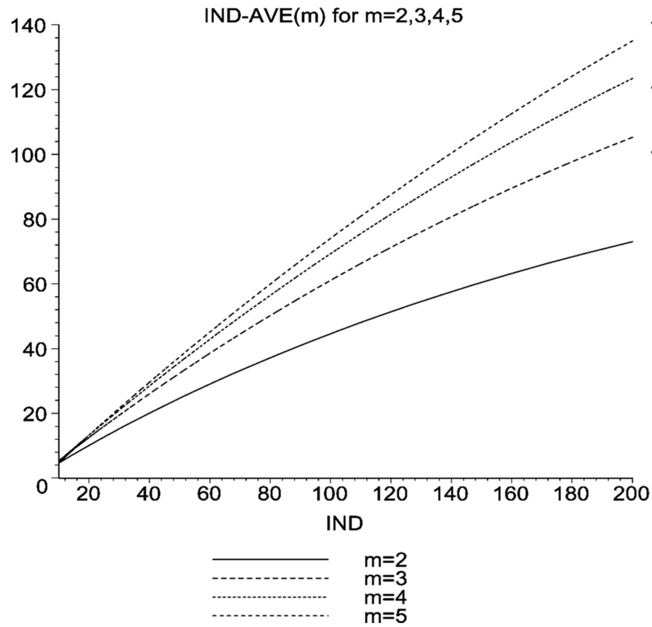

Figure 1. *AVE* chart under normality.

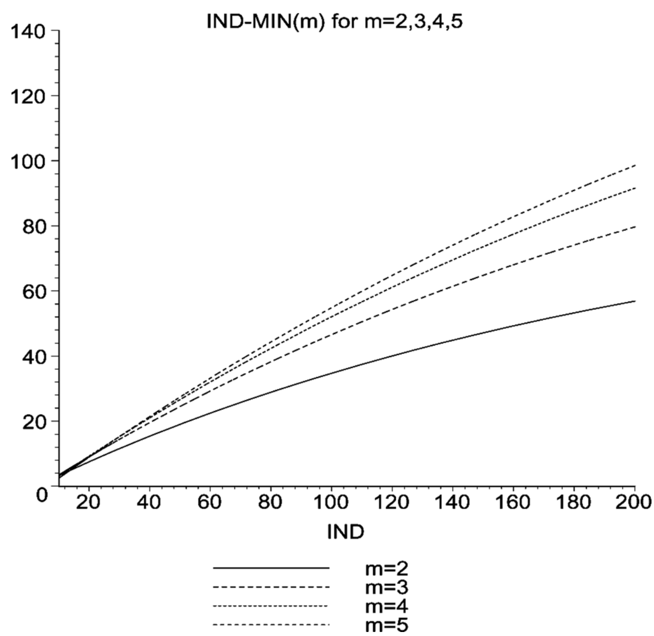

Figure 2. *MIN* chart under normality.

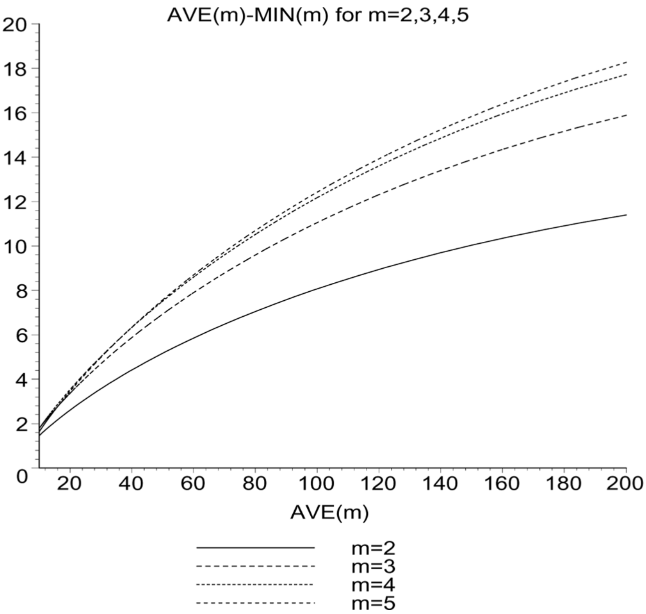

Figure 3. Difference between the *ARL*s of *AVE* and *MIN*.

individual chart, and that accurate nonparametric estimation for the *MIN* chart is quite straightforward for moderate values of n, but that nonparametric estimation for the *AVE* chart gives no improvement compared to the individual chart, and hence no solution for moderate n and current values of p. Therefore, when the nonparametric chart is the most appropriate one, the *MIN* chart is recommended.

REFERENCES

Albers, W. and Kallenberg, W. C. M. (2004a). Are Estimated Control Charts in Control? *Statistics* 38: 67–79.

Albers, W. and Kallenberg, W. C. M. (2004b). Empirical Nonparametric Control Charts: Estimation Effects and Corrections, *Journal of Applied Statistics* 31: 345–360.

Albers, W. and Kallenberg, W. C. M. (2004c). Estimation in Shewhart Control Charts: Effects and Corrections, *Metrika* 59: 207–234.

Albers, W. and Kallenberg, W. C. M. (2005a). New Corrections for Old Control Charts, *Quality Engineering* 17: 467–473.

Albers, W. and Kallenberg, W. C. M. (2005b). Tail Behavior of the Empirical Distribution Function of Convolutions, *Mathematical Methods of Statistics* 14: 133–162.

Albers, W. and Kallenberg, W. C. M. (2006a). Alternative Shewhart-Type Charts for Grouped Observations, *Metron*, in press.

Albers, W. and Kallenberg, W. C. M. (2006b). Self Adapting Control Charts, *Statistica Neerlandica* 60: 292–308.

Albers, W., Kallenberg, W. C. M., and Nurdiati, S. (2004). Parametric Control Charts, *Journal of Statistical Planning and Inference* 124: 159–184.

Albers, W., Kallenberg, W. C. M., and Nurdiati, S. (2005). Exceedance Probabilities for Parametric Control Charts, *Statistics* 39: 429–443.

Albers, W., Kallenberg, W. C. M., and Nurdiati, S. (2006). Data Driven Choice of Control Charts, *Journal of Statistical Planning and Inference* 136: 909–941.

Barndorff-Nielsen, O. E. (1996). Normal Inverse Gaussian Distributions and Stochastic Volatility Modelling, *Scandinavian Journal of Statistics* 24: 1–13.

Chakraborti, S. (2000). Run Length, Average Run Length and False Alarm Rate of Shewhart $\overline{X}$ Chart: Exact Derivations by Conditioning, *Communications in Statistics – Simulation & Computation* 29: 61–81.

Chakraborti, S., van der Laan, P., and Bakir, S. T. (2001). Nonparametric Control Charts: An Overview and Some Results, *Journal of Quality Technology* 33: 304–315.

Chan, L. K., Hapuarachchi, K. P., and Macpherson, B. D. (1988). Robustness of $\overline{X}$ and R Charts, *IEEE Transactions on Reliability* 37: 117–123.

Chen, G. (1997). The Mean and Standard Deviation of the Run Length Distribution of $\overline{X}$ Charts when Control Limits are Estimated, *Statistica Sinica* 7: 789–798.

Friedrich, K. O. (1989). A Berry-Esseen Bound for Functions of Independent Random Variables, *Annals of Statistics* 17: 170–183.

Ghosh, B. K., Reynolds, M. R., Jr., and Hui, Y. V. (1981). Shewhart $\overline{X}$-Charts with Estimated Process Variance, *Communications in Statistics – Theory & Methods* 10: 1797–1822.

Ion, R. A., Does, R. J. M. M., and Klaassen, C. A. J. (2000). A Comparison of Shewhart Control Charts Based on Normality, Nonparametrics, and Extreme-Value Theory, report 00–8, University of Amsterdam.

Nedumaran, G. and Pignatiello, J. J., Jr. (2001). On Estimating $\overline{X}$ Control Chart Limits, *Journal of Quality Technology* 33: 206–212.

Pappanastos, E. A. and Adams, B. M. (1996). Alternative Designs of the Hodges-Lehmann Control Chart, *Journal of Quality Technology* 28: 213–223.

Quesenberry, C. P. (1991). SPC Q Charts for Start-Up Processes and Short or Long Runs, *Journal of Quality Technology* 23: 213–224.

Quesenberry, C. P. (1993). The Effect of Sample Size on Estimated Limits for $\overline{X}$ and X Control Charts, *Journal of Quality Technology* 25: 237–247.

Ramberg, J. S. and Schmeisser, B. W. (1972). An Approximate Method for Generating Symmetric Random Variables, *Communications of Association for Computing Machinery* 15: 987–990.

Ramberg, J. S. and Schmeisser, B. W. (1974). An Approximate Method for Generating Symmetric Random Variables, *Communications of Association for Computing Machinery* 17: 78–82.

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

Roes, C. B., Does, R. J. M. M., and Schurink, Y. (1993). Shewhart-Type Control Charts for Individual Observations, *Journal of Quality Technology* 25: 188–198.

Willemain, T. R. and Runger, G. C. (1996). Designing Control Charts Using an Empirical Reference Distribution, *Journal of Quality Technology* 28: 31–38.

Woodall, W. H. and Montgomery, D. C. (1999). Research Issues and Ideas in Statistical Process Control, *Journal of Quality Technology* 31: 376–386.

Yang, C. H. and Hillier, F. S. (1970). Mean and Variance Control Limits Based on a Small Number of Subgroups, *Journal of Quality Technology* 2: 9–16.