Faculty of Mathematical Sciences

University of Twente

University for Technical and Social Sciences

P.O. Box 217
7500 AE Enschede
The Netherlands
Phone: +31-53-4893400
Fax: +31-53-4893114
Email: memo@math.utwente.nl

Memorandum No. 1559

Estimation in Shewhart control charts

W. Albers and W.C.M. Kallenberg

December 2000

ISSN 0169-2690

Estimation in Shewhart control charts

Willem Albers and Wilbert C.M. Kallenberg
Faculty of Mathematical Sciences
University of Twente
P.O. Box 217, 7500 AE Enschede
The Netherlands

Abstract The influence of the estimation of parameters in Shewhart control charts is investigated. It is shown by simulation and asymptotics that (very) large sample sizes are needed to accurately determine control charts if estimators are plugged in. Correction terms are developed to get accurate control limits for common sample sizes in the in-control situation. Simulation and theory show that the new corrections work very well. The performance of the corrected control charts in the out-of-control situation is studied as well. It turns out that the correction terms do not disturb the behavior of the control charts in the out-of-control situation. On the contrary, for moderate sample sizes the corrected control charts remain powerful and therefore, the recommendation to take at least 300 observations can be reduced to 40 observations when corrected control charts are applied.

Keywords and phrases: statistical process control, Phase II control limits, second order unbiasedness, out-of-control.

2000 Mathematics Subject Classification: 62F12, 62P30

1 Introduction

The basic Shewhart $\overline{X}$-chart for monitoring the mean consists of upper and lower control limits. An out-of-control signal is given when an observation falls beyond the control limits. Assuming normality of the observations, the control limits depend on the mean and standard error of the normal distribution in the in-control situation. As a rule these parameters are unknown and hence to apply the control chart one has to estimate the parameters.

Estimation is based on the observations obtained in the so called Phase I. We suppose that the observations in Phase I belong to the in-control situation. The monitoring phase is called Phase II.

Woodall and Montgomery (1999) describe the question of the effect of estimation error in control charts as follows: "In most evaluations and comparisons of control chart performance in Phase II, it is assumed that the in-control values of the parameters are known. In practice, however, the parameters must be estimated in Phase I. The effects of this estimation on control chart performance have been studied, but only for relatively few types of charts (see, e.g., Ghosh, Reynolds and Hui (1981); Quesenberry (1993); and Chen (1997)). Much more research is needed in this area recognizing that the Phase II control limits are, in fact, random variables. Research shows that more data than has been recommended is needed to accurately determine control chart limits."

In Quesenberry (1993) and Chen (1997) simulations and numerical calculations are performed for the mean and standard deviation of the run length distribution, when the mean and standard error of the normal distribution are estimated. The simulation results confirm the conclusion of Quesenberry (1993) that the classical empirical rules for choosing the number of observations to estimate the mean and standard error of the normal distribution are inadequate and should be taken much larger. Is is recommended to take at least 300 observations. For a further discussion on this topic, see also Roes (1995), in particular Section 2.2.2.

In the literature we did not find any suggestion how to *correct* the control limits in order to have accurate control limits for commonly used sample sizes. As far as we know, the present paper is the first contribution to find such correction terms.

It should be noted that the introduction of Q-charts in Quesenberry (1991) does not solve the problem if we consider, for instance, the mean of the run length distribution or other quantities based on the run length. The reason is that, due to dependence involved by the estimators, the run length distribution is no longer a geometric distribution, see Quesenberry (1993) for an extensive discussion on that point.

The aim of this paper is threefold:

1. to show that (indeed) a great many data are needed to get accurate control charts limits, when estimators are simply plugged in;
2. to develop correction terms in order that also for common sample sizes accurate control charts are obtained;
3. to investigate the consequences for the out-of-control situation of using the corrected control limits.

Nowadays short production runs are more and more in demand and hence there may be not enough data to accurately estimate process parameters, all the more as many data are needed when estimators are simply plugged in. This calls for the search for simple, but efficient correction terms.

Although the first point mentioned above can be established by Monte Carlo simulation, the exploration of correction terms requires analytic insight in the structure of the problem. Asymptotic methods can reveal the main quantities and their relations with respect to each other. It turns out that the error implied by estimating the standard error is larger than the one induced by estimating the mean, especially if the probability of incorrectly signaling out-of-control is small.

The idea now is to devise approximations, which are simple enough to make corrections possible and, on the other hand, are still sufficiently accurate. Fortunately, the correction terms, obtained in this way, are easy to apply, even when the derivation of these terms requires the more complex approach of second order asymptotics. The corrected control limits do their job very well, giving already for moderate sample sizes accurate results.

In situations where the correction leads to a more stringent control limit, the obvious consequence is that in an out-of-control case the control limit is exceeded less often. Quantification of the out-of-control performance is provided by asymptotic methods. It is shown by simulation results that the approximations describe the out-of-control behavior very well.

It turns out that the out-of-control behavior is not disturbed at all by the more stringent control limits. The loss due to estimation is not very great. Therefore, the recommendation of taking at least 300 observations can be reduced to 40 observations when applying corrected control charts. This implies that for today's practice, where often not many data can be obtained, the corrected control charts provide a solution for the problem that in fact no accurate control charts were available.

The paper is organized as follows. In Section 2 our set-up is given and it is shown by simulation and asymptotic theory that very large samples are needed to get accurate control limits in case no correction is made. The correction terms are derived in Section 3, where also the performance of the corrected control limits is exhibited. The out-of-control behavior is treated in Section 4. Each section is closed by summarizing the conclusions of that section.

2 The effect of estimation error

Let $X_1, \ldots, X_n, X_{n+1}$ be independent and identically distributed random variables (r.v.'s), each with a $N(\mu, \sigma^2)$-distribution. 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 Sections 2 and 3 we consider for Phase II the in-control situation, that is X_{n+1} has the same distribution as $X_1, \ldots, X_n$.

For convenience we consider a control chart with only an upper limit. The more standard case of upper- and lower control limit is treated in a similar way. Also, generalizations to a set-up with X_i replaced by a group of observations, for instance $X_{i1}, \ldots, X_{i5}$, is fairly straightforward.

If μ and σ are known and p is the probability of incorrectly concluding that the process is out-of-control, then the upper control limit (UCL) equals

$$\mu + u_p \sigma \text{ with } u_p = \Phi^{-1}(1-p),$$

where Φ denotes the distribution function of the standard normal distribution. The density of the standard normal distribution is denoted by φ. However, as a rule μ and σ are unknown and they have to be estimated on the basis of $X_1, \ldots, X_n$. We consider as estimator of μ the sample mean $\overline{X} = n^{-1} \sum_{i=1}^n X_i$ and denote this estimator by $\hat{\mu}$. As estimator of σ we sometimes consider

$$S = \sqrt{S^2} \text{ with } S^2 = \frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n-1},$$

but more often we take

$$\hat{\sigma} = \frac{S}{c_4(n)}, \tag{2.1}$$

where $c_4(n)$ is such that $\hat{\sigma}$ is an unbiased estimator of σ, implying

$$c_4(n) = \frac{\sqrt{2}\Gamma(n/2)}{\sqrt{n-1}\Gamma((n-1)/2)}. \tag{2.2}$$

Note that in view of Basu's theorem any location-invariant estimator is independent of $\hat{\mu}$ (cf. Lehmann (1986) page 191 Example 1). A general location-invariant estimator of σ is denoted by σ^* and particular examples are $\hat{\sigma}$ and S. Plugging in the estimators $\hat{\mu}$ and σ^* in the UCL leads to the estimated UCL

$$\hat{\mu} + u_p \sigma^*.$$

The interest is in the probability of an incorrect signal that the process should be out-of-control. This probability does depend on the estimators $\hat{\mu}$ and σ^* and therefore it is a r.v. It is given by, writing $\overline{\Phi}(x) = 1 - \Phi(x)$,

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

Several aspects of the closeness of P_n to p can be investigated. The most obvious quantity to consider is its expectation EP_n. This expectation is then compared with p. Also functions of P_n are of interest. The average run length (ARL) is given by $1/p$ and hence the estimated average run length equals $1/P_n$. Therefore, we may compare $E(1/P_n)$ with $1/p$. Another family of functions of interest are the probabilities that the run length is at most some specified value k, cf. also Remark 2.1. Typical values of interest for k are fractions of the ARL, i.e. $k = [\gamma/p]$ with, for instance, $\gamma = \frac{1}{10}, \frac{1}{4}, \frac{1}{2}$ or 1. This probability is given by $1 - (1-p)^k$ and is estimated by $R_{n,k} = 1 - (1 - P_n)^k$. The expectation $ER_{n,k}$ is compared with $1 - (1-p)^k$.

More generally, we consider a function $g(p)$, estimate it by $g(P_n)$ and compare $Eg(P_n)$ with $g(p)$. In particular, the previous functions

$$\begin{aligned} g(p) &= p, \\ g(p) &= 1/p, \\ g(p) &= 1 - (1-p)^k \end{aligned} \tag{2.3}$$

are of interest. Other functions g can be treated in a similar way, as for instance the standard deviation of the run length, corresponding to $g(p) = \sqrt{1-p}/p$.

A criterion for closeness of $Eg(P_n)$ to $g(p)$ is that the relative error should be at most 10%, in formula

$$\left| \frac{Eg(P_n) - g(p)}{g(p)} \right| \leq 0.1. \tag{2.4}$$

We are looking for the smallest n, for which (2.4) holds. (Of course, other values than 0.1 can be chosen if desired.)

A simulation study is made to show the performance of $Eg(P_n)$ for the functions, given in (2.3). The estimator of σ applied in the simulation study is $\hat{\sigma}$, given by (2.1) and (2.2). The number of repetitions in the simulation study equals 100,000. We apply criterion (2.4) in case of $g(p) = p$ and $g(p) = 1/p$, while for $g(p) = 1-(1-p)^k$ we require an absolute error of at most 0.01, since this seems to be more appropriate when dealing with probabilities not close to 0. The results are given for $p = 0.001$ in Table 1 and for $p = 0.01$ in Table 2.

Table 1 Simulation results for $p = 0.001$.

	$EP_n \times 10^3$	$E(1/P_n)$	$ER_{n,1000}$	$ER_{n,500}$	$ER_{n,250}$	$ER_{n,100}$
$g(p)$	1.000	1000	0.6323	0.3936	0.2213	0.0952
tolerated interval	(0.900, 1.100)	(900, 1100)	(0.6223, 0.6423)	(0.3836, 0.4036)	(0.2113, 0.2313)	(0.0852, 0.1052)
$n = 25$	2.6701	8020	0.5967	0.4601	0.3266	0.1822
$n = 50$	1.7291	2328	0.6095	0.4407	0.2876	0.1429
$n = 75$	1.4503	1705	0.6158	0.4298	0.2688	0.1271
$n = 100$	1.3260	1472	0.6194	0.4228	0.2581	0.1191
$n = 150$	1.2145	1285	0.6242	0.4159	0.2476	0.1115
$n = 200$	1.1584	1200	0.6266	0.4114	0.2415	0.1074
$n = 250$	1.1252	1155	0.6279	0.4083	0.2376	0.1050
$n = 300$	1.1042	1128	0.6287	0.4062	0.2351	0.1034
$n = 350$	1.0878	1109	0.6290	0.4043	0.2330	0.1021
$n = 400$	1.0771	1095	0.6294	0.4031	0.2316	0.1013
$n = 450$	1.0694	1082	0.6301	0.4024	0.2307	0.1007
$n = 500$	1.0624	1075	0.6301	0.4015	0.2297	0.1001

It is seen from the simulations that EP_n is in the tolerated interval for $n \geq 312$, that $E(1/P_n)$ is in the tolerated interval for $n \geq 377$, while $ER_{n,k}$ satisfies the requirements for $n \geq 124, 401, 419, 243$, when $k = 1000, 500, 250, 100$, respectively.

Typically it is recommended to have at least 20-25 samples of size 4-5 each to base the estimators on (cf. Woodall and Montgomery (1999) page 379). The idea is that when taking the mean of 4-5 single observations as a "combined" observation, normality is a reasonable assumption, which may be more disputable for the single observations. If we indeed associate a sample of size 4-5 with one observation in our set-up, the recommended number of observations equals 20-25. The simulations clearly show that this number is far too small to get an accurate estimate. Even if we should consider the observations as single observations in the recommendation, leading to n between 80 and 125, this number of observations is still too small to get accurate results.

The simulation results clearly confirm point 1 in the introduction: *very many data are needed to get accurate control charts limits, when estimators are simply plugged in.* This conclusion agrees with the results of e.g. Quesenberry (1993) and Chen (1997).

The situation becomes somewhat better if we consider $p = 0.01$, as is shown in Table 2, but still rather large sample sizes are needed. Moreover, in control charts $p = 0.001$ is far more often applied than $p = 0.01$.

Table 2 Simulation results for $p = 0.01$.

	$EP_n \times 10^2$	$E(1/P_n)$	$ER_{n,100}$	$ER_{n,50}$	$ER_{n,25}$	$ER_{n,10}$
$g(p)$	1.000	100	0.6340	0.3950	0.2222	0.0956
tolerated interval	(0.900, 1.100)	(90,110)	(0.6240, 0.6440)	(0.3850, 0.4050)	(0.2122, 0.2322)	(0.0856, 0.1056)
$n = 10$	2.3963	1878	0.5924	0.4575	0.3240	0.1785
$n = 25$	1.5061	212	0.6119	0.4339	0.2755	0.1322
$n = 50$	1.2402	140	0.6210	0.4182	0.2513	0.1141
$n = 75$	1.1602	123	0.6260	0.4124	0.2427	0.1082
$n = 100$	1.1186	117	0.6277	0.4084	0.2378	0.1050
$n = 125$	1.0959	113	0.6293	0.4064	0.2350	0.1033
$n = 150$	1.0786	111	0.6298	0.4043	0.2328	0.1019
$n = 175$	1.0663	109	0.6305	0.4030	0.2312	0.1009
$n = 200$	1.0579	108	0.6307	0.4020	0.2301	0.1003
$n = 225$	1.0513	107	0.6310	0.4012	0.2292	0.0998
$n = 250$	1.0467	106	0.6314	0.4008	0.2286	0.0994

It is seen from the simulations that EP_n is in the tolerated interval for $n \geq 118$, that $E(1/P_n)$ is in the tolerated interval for $n \geq 163$, while $ER_{n,k}$ satisfies the requirements for $n \geq 59, 139, 164, 95$, when $k = 100, 50, 25, 10$, respectively.

Remark 2.1 It follows from Tables 1 and 2 that P_n is positively biased: $EP_n > p$. This agrees with the idea that we have to pay for estimating parameters: the expected probability of an incorrect signal becomes larger. At first sight one may think that if P_n tends to be larger than p, this will imply that $1/P_n$ is smaller than $1/p$, cf. e.g. Quesenberry (1993, pages 241 and 242). However, it turns out that also $1/P_n$ is positively biased: $E(1/P_n) > 1/p$.

One reason for it is that (very) small values of P_n imply (very) large values of $1/P_n$: even if the probabilities of getting such small values of P_n are not so large, the high values of $1/P_n$ cause a high expectation. Especially for small n this phenomenon is rather strongly present, see, for instance, Table 5 on page 245 of Quesenberry (1993).

Since $E(1/P_n)$ is strongly determined by the occurrence of extremely long runs, which are not very relevant in practice, Roes (1995, page 34) remarks that $E(1/P_n)$ does not adequately summarize the run length properties of the chart, cf. also Quesenberry (1993, page 242).

To avoid this problem of "outliers" one may apply the strategy, often successfully applied in robust statistics, to replace the average by the median. The median of a geometric distribution with parameter p is given by the function

$$g(p) = \frac{-\log 2}{\log(1-p)}.$$

However, for small p, this function behaves as $(\log 2)/p$ and hence, the same problem arises with the median run length.

The introduction of the criterion $ER_{n,k}$ for several values of k is sometimes motivated by giving a more sensible performance measure than the ARL, see e.g. Does and Schriever (1992), Roes (1995) pages 102, 103, Del Castillo and Montgomery (1995) and Quesenberry (1995). Note however, that the practical relevance of $ER_{n,k}$ with k as large as 1000 is disputable too. What we want is protection against small run lengths, when we are in-control. The profit obtained by estimation of the parameters for $ER_{n,1000}$ (see Table 1) is therefore of far less importance than the disadvantage for $ER_{n,k}$ with $k = 100, 250, 500$.

By the same type of argument as above it follows that the standard deviation of the run length distribution exceeds the mean of the run length distribution, cf. Quesenberry (1993, page 242).

Some further comments on the positive bias of $EP_n, E(1/P_n), ER_{n,k}$ for $k = 100, 250, 500$ and the negative bias of $ER_{n,k}$ for $k = 1000$ are given in Remark 2.4. □

In order to develop correction terms such that accurate control charts are obtained for smaller sample sizes, we need to have more analytic insight in the problem. Moreover, to understand *why* so much larger sample sizes are needed here than recommended in earlier days, more insight is needed too. Such insight can be provided by asymptotic methods. In many statistical problems it has turned out that the more transparent approximations show the important feature of the problem without throwing away the accuracy. As a rule, numerical work (alone) cannot give the insight needed to derive appropriate correction terms.

We start with a first type of asymptotics, for a great part based on asymptotic normality of the estimators. For some criteria the first type approximations can not be easily applied. Therefore, a second asymptotic approach is considered as well.

The standard asymptotical method is with respect to the number of observations n tending to infinity. However, a commonly used value for u_p is 3 and, for instance u_p^4 then equals 81, which may be of the same order of magnitude as n for several situations considered in the paper. Therefore, we follow a more delicate way, taking into account not only n, but also u_p in our asymptotical approach. For simplification of notation we write u instead of u_p.

(i) *First type asymptotics.*

We start with a theorem, presenting the limiting distribution of P_n. We put the following condition on the estimator σ^*.

Condition A The estimator σ^* satisfies

$$\left(\frac{\sigma^*}{\sigma} - 1\right)\sqrt{2(n-1)} \xrightarrow{D} N(0,1).$$

It is easily seen that condition A holds for $\sigma^* = S$ and $\sigma^* = \hat{\sigma}$.

Define

$$z(u) = \frac{u\overline{\Phi}(u)}{\varphi(u)};$$

then z is very close to 1 for large values of u. For instance, for $u > 0$, it holds that

$$\frac{u^2}{u^2+1} < z(u) < 1.$$

Theorem 2.1 *Suppose that condition A holds, that $u \geq 1$ and that $u = o(n^{1/2})$ as $n \to \infty$; then*

$$P_n \approx pY_n \textit{ with } Y_n \sim \text{lognormal}\left(0, \left\{\frac{u}{z(u)}\right\}^2 \left\{\frac{1}{n} + \frac{u^2}{2(n-1)}\right\}\right)$$

in the sense that

$$\frac{\log\left(\dfrac{P_n}{p}\right)}{\dfrac{u}{z(u)}\sqrt{\dfrac{1}{n} + \dfrac{u^2}{2(n-1)}}} \xrightarrow{D} N(0,1)$$

as $n \to \infty$, uniformly for all sequences $u = u(n)$ satisfying $u(n) \geq 1$ and $\lim_{n\to\infty} u(n)n^{-1/2} = 0$.

Proof Let

$$\Delta(u) = \frac{\hat{\mu} - \mu}{\sigma} + u\left(\frac{\sigma^*}{\sigma} - 1\right).$$

In view of the definition of $\hat{\mu}$ and condition A we have

$$\frac{\Delta(u)}{\sqrt{\dfrac{1}{n} + \dfrac{u^2}{2(n-1)}}} \xrightarrow{D} N(0,1). \tag{2.5}$$

Direct calculation gives

$$\frac{P_n}{p} = \frac{u}{u + \Delta(u)} \frac{z(u + \Delta(u))}{z(u)} \exp\{-u\Delta(u) - \tfrac{1}{2}\Delta^2(u)\}.$$

On the set $B = \{u \geq 1, |\Delta(u)| \leq \frac{1}{2}\}$ we have, for some constant $c_1 > 0$,

$$\left|\log\left(\frac{u}{u+\Delta(u)}\frac{z(u+\Delta(u))}{z(u)}\right) - \Delta(u)\left\{-\frac{1}{u} + \frac{z'(u)}{z(u)}\right\}\right| \leq c_1\Delta^2(u) \tag{2.6}$$

and hence, noting that

$$-u-\frac{1}{u}+\frac{z'(u)}{z(u)}=-\frac{u}{z(u)}$$

and writing

$$R(u)=\log\left(\frac{u}{u+\Delta(u)}\frac{z(u+\Delta(u))}{z(u)}\right)-\Delta(u)\left\{-\frac{1}{u}+\frac{z'(u)}{z(u)}\right\}-\tfrac{1}{2}\Delta^2(u),$$

we get

$$\log\left(\frac{P_n}{p}\right)=-\frac{u}{z(u)}\Delta(u)+R(u)1_B(u)+R(u)1_{\overline{B}}(u), \tag{2.7}$$

where $\overline{B}$ denotes the complement of B. In view of (2.5) and (2.6) we obtain

$$\frac{|R(u)1_B(u)|}{\frac{u}{z(u)}\sqrt{\frac{1}{n}+\frac{u^2}{2(n-1)}}}\leq\frac{\sqrt{\frac{1}{n}+\frac{u^2}{2(n-1)}}}{\frac{u}{z(u)}}(c_1+\tfrac{1}{2})\left(\frac{\Delta(u)}{\sqrt{\frac{1}{n}+\frac{u^2}{2(n-1)}}}\right)^2\xrightarrow{D}0. \tag{2.8}$$

Since $u=o(n^{1/2})$ and hence, by (2.5), $P(|\Delta(u)|>\frac{1}{2})\to 0$, it follows that

$$\frac{|R(u)1_{\overline{B}}(u)|}{\frac{u}{z(u)}\sqrt{\frac{1}{n}+\frac{u^2}{2(n-1)}}}\xrightarrow{D}0. \tag{2.9}$$

Combination of (2.5), (2.7), (2.8) and (2.9) gives the result. □

Remark 2.2 For somewhat larger values of u, it is seen that the estimation of σ is more important than that of μ, in the sense that the contribution to the asymptotic variance due to estimating σ is in case of $u=3$ a factor $u^2/2=4.5$ larger than the contribution due to estimation of μ. □

We are interested in approximations of $Eg(P_n)$, in particular when g is one of the functions given in (2.3). Theorem 2.1 suggests the following first type approximation:

$$Eg(P_n)\approx Eg(pY_n) \text{ with } Y_n\sim \text{ lognormal }\left(0,\left\{\frac{u}{z(u)}\right\}^2\left\{\frac{1}{n}+\frac{u^2}{2(n-1)}\right\}\right).$$

Taking $g(x)=x$, this approximation gives

$$EP_n\approx p\exp\left\{\frac{1}{2}\left(\frac{u}{z(u)}\right)^2\left(\frac{1}{n}+\frac{u^2}{2(n-1)}\right)\right\}. \tag{2.10}$$

For $g(x) = 1/x$, this approximation gives

$$E\frac{1}{P_n} \approx \frac{1}{p}\exp\left\{\frac{1}{2}\left(\frac{u}{z(u)}\right)^2\left(\frac{1}{n}+\frac{u^2}{2(n-1)}\right)\right\}. \tag{2.11}$$

Remark 2.3 The paradox of getting positive bias for P_n and $1/P_n$ is explained in Remark 2.1 and Remark 2.4 below. Note that EP_n/p and $E(1/P_n)/(1/p)$ have the same first type approximation, which is greater than 1. Hence, the first type approximation gives a positive bias for P_n and $1/P_n$. In this sense the approximation $P_n \approx pY_n$ with Y_n a lognormal distribution with $\mu = 0$ does what it should do. □

Finally, consider $g(p) = 1-(1-p)^k$. In this case the approximation leads to

$$\begin{aligned} E\left\{1-(1-P_n)^k\right\} &\approx E\left\{1-(1-pY_n)^k\right\} \\ &= \sum_{j=1}^{k}\binom{k}{j}(-p)^{j+1}\exp\left\{\frac{1}{2}j^2\left(\frac{u}{z(u)}\right)^2\left(\frac{1}{n}+\frac{u^2}{2(n-1)}\right)\right\}. \end{aligned} \tag{2.12}$$

We compare the approximation results, given in (2.10) and (2.11) with the simulation results. Since the approximation in (2.12) is rather complicated, we do not further consider this case in our first type approximation. In the second type approximation it will be treated again.

Table 3 Comparison of first type approximations and simulation results for $p = 0.001$.

	$EP_n \times 10^3$			$E(1/P_n)$		
	simulation	appr.(2.10)	difference	simulation	appr.(2.11)	difference
$n = 25$	2.670	3.875	1.205	8020	3875	4145
$n = 50$	1.729	1.946	0.217	2328	1946	382
$n = 75$	1.450	1.555	0.105	1705	1555	150
$n = 100$	1.326	1.391	0.065	1472	1391	81
$n = 150$	1.215	1.245	0.030	1285	1245	40
$n = 200$	1.158	1.179	0.021	1200	1179	21
$n = 250$	1.125	1.140	0.015	1155	1140	15
$n = 300$	1.104	1.116	0.012	1128	1116	12
$n = 350$	1.088	1.098	0.010	1109	1098	11
$n = 400$	1.077	1.085	0.008	1095	1085	10
$n = 450$	1.069	1.076	0.007	1082	1076	6
$n = 500$	1.062	1.068	0.006	1075	1068	7

According to the approximations, EP_n and $E(1/P_n)$ are in the tolerated interval for $n \geq 345$. Note that in the simulations we had that EP_n is in the tolerated interval for $n \geq 312$, that $E(1/P_n)$ is in the tolerated interval for $n \geq 377$.

It is seen from Table 3 that the first type approximation shows very well what is going on. However, a more handsome approximation of $ER_{n,k}$ should be obtained. Therefore, we take a slightly different and more direct approach pointed to expectations rather than (limiting) distributions.

(ii) *Second type asymptotics.*

Restricting attention to expectations, it should be remarked that the $(2k-1)^{th}$ and $(2k)^{th}$ moments of $\Delta(u)$ are of the same order of magnitude. Therefore, in terms of expectations it seems more natural to consider not only terms of order $\Delta(u)$, but also of order $\Delta^2(u)$.

We are interested in approximations of $Eg(P_n)$, in particular when g is one of the functions, given in (2.3). As

$$P_n = \overline{\Phi}\left(\frac{\hat{\mu}-\mu}{\sigma} + u\frac{\sigma^*}{\sigma}\right) = \overline{\Phi}(u+\Delta(u)),$$

we introduce the notation

$$h(u) = g(\overline{\Phi}(u))$$

and investigate $Eg(P_n) = Eh(u+\Delta(u))$. As our second type approximation we take the two-step Taylor expansion

$$Eh(u+\Delta(u)) \approx h(u) + h'(u)E\Delta(u) + \tfrac{1}{2}h''(u)E\Delta^2(u). \tag{2.13}$$

Direct calculation gives the following derivatives of h for the functions g, presented in (2.3).

Table 4 Derivatives of h for the functions g from (2.3).

$g(p)$	$h(u)$	$h'(u)$	$h''(u)$
p	$\overline{\Phi}(u)$	$-\varphi(u)$	$u\varphi(u)$
$\frac{1}{p}$	$\frac{1}{\overline{\Phi}(u)}$	$\frac{\varphi(u)}{\overline{\Phi}^2(u)}$	$\frac{2\varphi^2(u)}{\overline{\Phi}^3(u)} - \frac{u\varphi(u)}{\overline{\Phi}^2(u)}$
$1-(1-p)^k$	$1-\Phi^k(u)$	$-k\varphi(u)\Phi^{k-1}(u)$	$-k(k-1)\varphi^2(u)\Phi^{k-2}(u)$ $+ku\varphi(u)\Phi^{k-1}(u)$

Remark 2.4 The positive bias of EP_n and $ER_{n,k}$ for $k = 100, 250$ and 500 are in line with the idea that we have to pay for estimating the parameters. Indeed, with respect to these criteria we have a higher probability of an incorrect signal (EP_n) or a higher probability of getting a smaller run length ($ER_{n,k}$ for $k = 100, 250$ and 500). For $E(1/P_n)$ and $ER_{n,k}$ with $k = 1000$ estimation looks

profitable. In Remark 2.1 there is already given an explanation for the positive bias of $E(1/P_n)$. However, this explanation (some very high values of $1/P_n$ cause a high expectation) can not be used for $ER_{n,k}$, since $R_{n,k}$ is bounded by 1.

There is another argument that explains the bias. If h is a convex function, then Jensen's inequality gives $Eh(X) > h(EX)$. Consider

$$Eg(P_n) = Eh(u + \Delta(u)).$$

For $p = 0.001$ we have $u = 3.09$ and further, $\Delta(u)$ converges to 0, implying that $u + \Delta(u)$ is with high probability in a neighborhood of 3.

If $g(p) = p$ or $g(p) = 1/p$, we have $h''(u) > 0$ for $u > 0$ (see Table 4) and hence in both cases the function h is convex for $u > 0$. Therefore, Jensen's inequality strongly suggests $Eh(u + \Delta(u)) > h(u + E\Delta(u)) = h(u)$, implying $EP_n > p$ and $E(1/P_n) > 1/p$.

If $g(p) = 1 - (1 - p)^k$, we have $h''(u) > 0$ for $u > 2.375$ and $k = 100$, for $u > 2.689$ and $k = 250$ and for $u > 2.908$ and $k = 500$ (see Table 4) and hence the function h is convex for these combinations of u- and k-values. Therefore, Jensen's inequality suggests $ER_{n;k} = Eh(u+\Delta(u)) > h(u+E\Delta(u)) = h(u) = 1-(1-p)^k$ for $k = 100, 250, 500$. If $g(p) = 1 - (1 - p)^{1000}$, we have $h''(u) < 0$ for $u < 3.115$ and hence the function h is concave for $u < 3.115$. Therefore, Jensen's inequality suggests $ER_{n;1000} = Eh(u + \Delta(u)) < h(u + E\Delta(u)) = h(u) = 1 - (1 - p)^{1000}$.

This explains the positive bias of $EP_n, E(1/P_n)$ and $ER_{n,k}$ for $k = 100, 250$ and 500 and the negative bias of $ER_{n;1000}$ as seen in Table 1. □

The following theorem gives an upper bound for the (relative) error of the approximation. We put the following condition on the estimator σ^*.

Condition B1 The estimator σ^* satisfies for $u \geq 1$ with $u = O(n^{1/4})$ as $n \to \infty$,

$$E\Delta^3(u) = O(u^3 n^{-2}), E\Delta^4(u) = O(u^4 n^{-2}), E|\Delta(u)|^4 \exp\{u|\Delta(u)| + \tfrac{1}{2}\Delta^2(u)\} = O(u^4 n^{-2}),$$

$$E|\Delta(u)|^9 \exp\{u|\Delta(u)| + \tfrac{1}{2}\Delta^2(u)\} = O(u^9 n^{-9/2}).$$

Condition B2 The estimator σ^* satisfies for $u \geq 1$ with $u = O(n^{1/4})$ as $n \to \infty$,

$$E\Delta^3(u) = O(u^3 n^{-2}), E\Delta^4(u) = O(u^4 n^{-2}).$$

It can be shown by rather straightforward calculation that conditions B1 and B2 hold for $\sigma^* = S$ and $\sigma^* = \hat{\sigma}$. We omit the details.

Theorem 2.2 *Suppose that $u \geq 1$ and that $u = O(n^{1/4})$ as $n \to \infty$. Assume that h is 4 times differentiable.*
(i) Suppose that condition B1 holds and that for some constants $c_2 > 0$ and $c_3 > 0$

$$\left|\frac{h'''(u)}{h(u)}\right| \leq c_2 u^3 \text{ and } \left|\frac{h^{iv}(u + y)}{h(u)}\right| \leq c_3(u^4 + |y|^5)\exp\{u|y| + \tfrac{1}{2}y^2\} \qquad (2.14)$$

for all $u \geq 1$ and all $y \in \mathbb{R}$, then

$$\left|\frac{Eh(u+\Delta(u)) - h(u)}{h(u)} - \frac{h'(u)E\Delta(u) + \frac{1}{2}h''(u)E\Delta^2(u)}{h(u)}\right| = O(u^8 n^{-2}).$$

(ii) Suppose that condition B2 holds and that h''' and h^{iv} are bounded, then

$$\left|Eh(u+\Delta(u)) - h(u) - \{h'(u)E\Delta(u) + \tfrac{1}{2}h''(u)E\Delta^2(u)\}\right| = O(u^4 n^{-2}).$$

Proof (i) By Taylor expansion we get, for some $0 \leq \eta \leq 1$,

$$Eh(u+\Delta(u)) =$$

$$h(u)+h'(u)E\Delta(u)+\tfrac{1}{2}h''(u)E\Delta^2(u)+\tfrac{1}{6}Eh'''(u)\Delta^3(u)+\tfrac{1}{24}Eh^{iv}(u+\eta\Delta(u))\Delta^4(u)$$

and hence, by (2.14) and condition B1, we have

$$\left|\frac{Eh(u+\Delta(u)) - h(u)}{h(u)} - \frac{h'(u)E\Delta(u) + \frac{1}{2}h''(u)E\Delta^2(u)}{h(u)}\right|$$

$$\leq \tfrac{1}{6}c_2 u^3|E\Delta^3(u)| + \tfrac{1}{24}c_3 E[(u^4 + |\Delta(u)|^5)|\Delta(u)|^4 \exp\{u|\Delta(u)| + \tfrac{1}{2}\Delta^2(u)\}]$$

$$= O(u^8 n^{-2}).$$

(ii) By Taylor expansion we get, for some $0 \leq \eta \leq 1$,

$$Eh(u+\Delta(u)) =$$

$$h(u)+h'(u)E\Delta(u)+\tfrac{1}{2}h''(u)E\Delta^2(u)+\tfrac{1}{6}Eh'''(u)\Delta^3(u)+\tfrac{1}{24}Eh^{iv}(u+\eta\Delta(u))\Delta^4(u)$$

and hence, since h''' and h^{iv} are bounded and condition B2 holds,

$$|Eh(u+\Delta(u)) - h(u) - \{h'(u)E\Delta(u) + \tfrac{1}{2}h''(u)E\Delta^2(u)\}|$$

$$\leq c_4\{|E\Delta^3(u)| + E\Delta^4(u)\} = O(u^4 n^{-2}). \qquad \square$$

Using

$$\overline{\Phi}(u) = \frac{\varphi(u)}{u}(1 + o(1)) \text{ as } u \to \infty,$$

it is not hard to show that (2.14) is satisfied for $h(u) = g(\overline{\Phi}(u))$ with $g(p) = p$ and $g(p) = 1/p$, and that h''' and h^{iv} are bounded for $g(p) = 1 - (1-p)^k$. Hence the relative error in the approximation (2.13) is $O(u^8 n^{-2})$ for $g(p) = p$ and $g(p) = 1/p$. As argued before, for $g(p) = 1 - (1-p)^k$ we consider the absolute error, which equals $O(u^4 n^{-2})$.

Straightforward calculation gives

$$E\Delta(u) = uE\left(\frac{\sigma^*}{\sigma} - 1\right) \text{ and } E\Delta^2(u) = \frac{1}{n} + u^2 E\left(\frac{\sigma^*}{\sigma} - 1\right)^2.$$

In particular, for $\sigma^* = \hat{\sigma}$ we get

$$E\Delta(u) = 0 \text{ and } E\Delta^2(u) = \frac{1}{n} + u^2\left\{\frac{\Gamma\left(\frac{n+1}{2}\right)\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)^2} - 1\right\} = a(u,n), \text{ say,} \quad (2.15)$$

and for $\sigma^* = S$ we get

$$E\Delta(u) = u\{c_4(n) - 1\} \text{ and } E\Delta^2(u) = \frac{1}{n} - 2u^2\{c_4(n) - 1\}, \quad (2.16)$$

with $c_4(n)$ as in (2.2). Inserting (2.15) and the derivatives from Table 4 in (2.13) leads to the following approximations for the estimator $\sigma^* = \hat{\sigma}$:

$$\begin{aligned} EP_n &\approx p + \tfrac{1}{2}a(u,n)u\varphi(u) \\ E(1/P_n) &\approx 1/p + \tfrac{1}{2}a(u,n)\left\{\frac{2\varphi^2(u)}{(1-p)^3} - \frac{u\varphi(u)}{(1-p)^2}\right\} \\ ER_{nk} &\approx 1-(1-p)^k + \tfrac{1}{2}a(u,n)\{-k(k-1)\varphi^2(u)(1-p)^{k-2} + ku\varphi(u)(1-p)^{k-1}\}. \end{aligned} \quad (2.17)$$

Remark 2.5 The paradox of getting positive bias for P_n and $1/P_n$ is explained in Remark 2.1 and Remark 2.4. The second type approximations of $EP_n - p$ and $E(1/P_n) - (1/p)$ are both positive and hence, the second type approximations give positive bias for P_n and $1/P_n$, as they should do. □

We compare the approximation results, given in (2.17) with the simulation results. Also the case $g(p) = 1 - (1-p)^k$ is covered now.

Table 5 Comparison of second type approximations and simulation results for $p = 0.001$.

	$EP_n \times 10^3$			$E(1/P_n)$		
	simulation	appr.	difference	simulation	appr.	difference
$n = 25$	2.670	2.254	0.416	8020	2478	5542
$n = 50$	1.729	1.614	0.115	2328	1724	604
$n = 75$	1.450	1.406	0.044	1705	1479	226
$n = 100$	1.326	1.304	0.022	1472	1358	114
$n = 150$	1.215	1.202	0.013	1285	1238	47
$n = 200$	1.158	1.151	0.007	1200	1178	22
$n = 250$	1.125	1.121	0.004	1155	1142	13
$n = 300$	1.104	1.100	0.004	1128	1118	10
$n = 350$	1.088	1.086	0.002	1109	1102	7
$n = 400$	1.077	1.075	0.002	1095	1089	6
$n = 450$	1.069	1.067	0.002	1082	1079	3
$n = 500$	1.062	1.060	0.002	1075	1071	4

	$ER_{n,1000}$			$ER_{n,500}$		
	simulation	appr.	difference	simulation	appr.	difference
$n = 25$	0.5967	0.5910	0.0057	0.4601	0.5670	0.1069
$n = 50$	0.6095	0.6121	0.0026	0.4407	0.4785	0.0378
$n = 75$	0.6158	0.6189	0.0031	0.4298	0.4498	0.0200
$n = 100$	0.6194	0.6223	0.0029	0.4228	0.4356	0.0128
$n = 150$	0.6242	0.6257	0.0015	0.4159	0.4215	0.0056
$n = 200$	0.6266	0.6273	0.0007	0.4114	0.4145	0.0031
$n = 250$	0.6279	0.6283	0.0004	0.4083	0.4103	0.0020
$n = 300$	0.6287	0.6290	0.0003	0.4062	0.4075	0.0013
$n = 350$	0.6290	0.6295	0.0005	0.4043	0.4055	0.0012
$n = 400$	0.6294	0.6298	0.0004	0.4031	0.4040	0.0009
$n = 450$	0.6301	0.6301	0.0000	0.4024	0.4029	0.0005
$n = 500$	0.6301	0.6303	0.0002	0.4015	0.4019	0.0004

	$ER_{n,250}$			$ER_{n,100}$		
	simulation	appr.	difference	simulation	appr.	difference
$n = 25$	0.3266	0.3993	0.0727	0.1822	0.1965	0.0143
$n = 50$	0.2876	0.3084	0.0208	0.1429	0.1448	0.0019
$n = 75$	0.2688	0.2790	0.0102	0.1271	0.1280	0.0009
$n = 100$	0.2581	0.2644	0.0063	0.1191	0.1197	0.0006
$n = 150$	0.2476	0.2499	0.0023	0.1115	0.1115	0.0000
$n = 200$	0.2415	0.2427	0.0012	0.1074	0.1074	0.0000
$n = 250$	0.2376	0.2384	0.0008	0.1050	0.1050	0.0000
$n = 300$	0.2351	0.2356	0.0005	0.1034	0.1033	0.0001
$n = 350$	0.2330	0.2335	0.0005	0.1021	0.1022	0.0001
$n = 400$	0.2316	0.2320	0.0004	0.1013	0.1013	0.0000
$n = 450$	0.2307	0.2308	0.0001	0.1007	0.1006	0.0001
$n = 500$	0.2297	0.2298	0.0001	0.1001	0.1001	0.0000

It is seen from Table 5 that the second type approximation is very accurate, also for $ER_{n,k}$.

The next table gives the sample sizes needed to get the required accuracy, given by the relative error (2.4) for $g(p) = p$ and $g(p) = 1/p$ and by the absolute error 0.01 for $ER_{n,k}$.

Table 6 Comparison of the first and second type approximation of the needed sample size with simulation results.

	1^{st} type appr.	2^{nd} type appr.	simulation
EP_n	345	302	312
$E(1/P_n)$	345	356	377
$ER_{n,1000}$	-	101	124
$ER_{n,500}$	-	417	401
$ER_{n,250}$	-	428	419
$ER_{n,100}$	-	244	243

Remark 2.6 A further simplification can be obtained by using

$$\frac{\Gamma\left(\frac{n+1}{2}\right)\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)^2} \approx \frac{1}{2n}, \tag{2.18}$$

leading, for $\sigma^* = \hat{\sigma}$, to

$$EP_n \approx p + \frac{1}{2}\frac{u^2+2}{2n}u\varphi(u)$$

$$E(1/P_n) \approx 1/p + \frac{1}{2}\frac{u^2+2}{2n}\left\{\frac{2\varphi^2(u)}{(1-p)^3} - \frac{u\varphi(u)}{(1-p)^2}\right\}$$

$$ER_{n,k} \approx 1-(1-p)^k + \frac{1}{2}\frac{u^2+2}{2n}\left\{-k(k-1)\varphi^2(u)(1-p)^{k-2} + ku\varphi(u)(1-p)^{k-1}\right\}. \qquad \square$$

Remark 2.7 In a similar way, using (2.16), approximations of $EP_n, E(1/P_n)$ and $ER_{n,k}$ can be evaluated when $\sigma^* = S$. □

We summarize the conclusions of this section.

1. Very many data are needed to get accurate control charts limits, when estimators are simply plugged in.
2. The first type approximation, using a lognormal distribution, gives a very good and easy approximation for the distribution of P_n, for EP_n and for $E(1/P_n)$, but is not easily applicable to $ER_{n,k}$.
3. The more direct second type approximation of $Eg(P_n)$, based on a two-step Taylor expansion, works very well and can be applied easily.

3 The correction terms

Simulations show very clearly that huge sample sizes are needed to get accurate control charts limits when estimators are simply plugged in. In order to get for commonly used sample sizes an accurate UCL we apply a correction term. The idea is as follows. Starting with an UCL which has for known parameters μ and σ a probability p of incorrectly concluding that the process is out-of-control, we arrive at a value unequal to p. Therefore, we change the starting value p to q, say, in such a way that, when estimating μ and σ, we end up with $Eg(P_n) = g(p)$ (or at least close to it). In other words, we do not use u_p, but u_q for an appropriate value of q. Instead of u_q we write $u_p + c$ with c being the correction term, which by the way depends on g. Indeed, when $g(p) = p$, simulation shows that $EP_n > p$ and hence we need a larger UCL: $c > 0$. When $g(p) = 1/p$, simulation shows that $E(1/P_n) > 1/p$ and hence, now P_n should be larger in order to get $E(1/P_n) = 1/p$, that is, we need a smaller ULC: $c < 0$.

When, for instance, looking at EP_n, which is larger than p, one might argue that the corrected UCL could be obtained by simply making it a little bit larger: $\hat{\mu} + u_p\sigma^* + \tilde{c}$. However, the smaller σ, the smaller such $\tilde{c}$ should be. Hence, in that case we could replace it by $c\sigma$. As σ is unknown, we should estimate σ by σ^*, thus arriving at the same form of the corrected UCL as described before: $\hat{\mu} + (u_p + c)\sigma^*$.

Remark 3.1 The replacement of the estimator S in the control limit by $\hat{\sigma}$ is done to get an unbiased estimator of σ. However, in fact the problem is not to get an unbiased estimator of σ, but the aim is to get an unbiased estimator of $g(p)$. Since $g(p)$ is a nonlinear function of σ, it is not enough to replace S by $\hat{\sigma}$. Apart from the correction factor $c_4(n)$, a further correction factor $\breve{c}$, say, is needed. Instead of writing $u_p\breve{c}(S/c_4(n)) = u_p\breve{c}\hat{\sigma}$, we may also write $(u_p + c)\hat{\sigma}$, thus obtaining the form mentioned before. Note that it in fact does not matter whether we start with S or $\hat{\sigma}$. After correction we get the same UCL. □

To calculate an appropriate correction term c we need insight in the way $Eg(P_n)$ changes, when a correction term c is added to u_p. The asymptotics of the previous section give an answer to this question.

We start with the first type approximation, given in (2.10). Replacing u by $u + c$ (also in $p = \overline{\Phi}(u)$) and then equating EP_n to p leads to the following equation

$$\overline{\Phi}(u+c)\exp\left\{\frac{1}{2}\left(\frac{u+c}{z(u+c)}\right)^2\left(\frac{1}{n}+\frac{(u+c)^2}{2(n-1)}\right)\right\} = p = \overline{\Phi}(u).$$

Using

$$\overline{\Phi}(u) = \varphi(u)\frac{z(u)}{u},$$

taking logarithms, expanding with respect to c and deleting terms of order c^2 and of order cn^{-1}, we arrive at

$$-uc - \frac{c}{u} + c\frac{z'(u)}{z(u)} + \frac{1}{2}\left(\frac{u}{z(u)}\right)^2\left(\frac{1}{n}+\frac{u^2}{2n}\right) = 0$$

and hence

$$c = \frac{u(u^2+2)}{4n}\frac{1}{z^2(u)\left(1+\dfrac{1}{u^2}-\dfrac{z'(u)}{uz(u)}\right)}. \tag{3.1}$$

Noting that

$$z(u) = 1 - u^{-2} + O(u^{-4}) \text{ and } z'(u) = O(u^{-3}) \text{ as } u \to \infty,$$

it follows that

$$z^2(u)\left(1+\frac{1}{u^2}-\frac{z'(u)}{uz(u)}\right) = 1 + O(u^{-4}) \text{ as } u \to \infty. \tag{3.2}$$

Similarly, in view of (2.11), the correction term based on the first type approximation in order to get $E(1/P_n)$ close to $1/p$ equals

$$c = -\frac{u(u^2+2)}{4n}\frac{1}{z^2(u)\left(1+\dfrac{1}{u^2}-\dfrac{z'(u)}{uz(u)}\right)}, \tag{3.3}$$

which is exactly the opposite of the correction term for $g(p) = p$. Note that, as argued before, indeed the correction term for $g(p) = p$ turns out to be positive and the correction term for $g(p) = 1/p$ is negative.

Next we consider the second type approximation, cf. (2.13),

$$Eh(u+c+\Delta(u+c)) \approx h(u+c) + h'(u+c)E\Delta(u+c) + \tfrac{1}{2}h''(u+c)E\Delta^2(u+c).$$

Ignoring terms of order c^2 and lower order terms like $cE\Delta(u)$ etc., that is replacing $h(u+c)$ by $h(u)+ch'(u)$ and $h'(u+c)E\Delta(u+c)+\frac{1}{2}h''(u+c)E\Delta^2(u+c)$ by its leading term $h'(u)E\Delta(u)+\frac{1}{2}h''(u)E\Delta^2(u)$, the correction term c is given by

$$ch'(u) + h'(u)E\Delta(u) + \tfrac{1}{2}h''(u)E\Delta^2(u) = 0.$$

Hence, we get

$$c = -E\Delta(u) - \frac{1}{2}\frac{h''(u)}{h'(u)}E\Delta^2(u). \tag{3.4}$$

Taking $\sigma^* = \hat{\sigma}$ and applying the further simplification given in (2.18), this reduces to

$$c = -\frac{(u^2+2)}{4n}\frac{h''(u)}{h'(u)}. \tag{3.5}$$

In particular, we get

$$\begin{aligned} c &= \frac{u(u^2+2)}{4n} && \text{if } g(p) = p \\ c &= \frac{u^2+2}{4n}\left\{u - \frac{2\varphi(u)}{p}\right\} && \text{if } g(p) = \frac{1}{p} \\ c &= \frac{u^2+2}{4n}\left\{u - \frac{(k-1)\varphi(u)}{1-p}\right\} && \text{if } g(p) = 1-(1-p)^k. \end{aligned} \tag{3.6}$$

Since for large u

$$2\frac{\varphi(u)}{p} = 2\frac{\varphi(u)}{\overline{\Phi}(u)} \approx 2u,$$

the correction term for $g(p) = 1/p$ is opposite to the correction term for $g(p) = p$; the latter is positive, the first one is negative, as they should be. Moreover, both correction terms are close to the first type correction terms, given in (3.1) and (3.3), cf. also (3.2).

The correction term for $g(p) = 1-(1-p)^k$ is smaller than the one for $g(p) = p$ and larger than the one for $g(p) = 1/p$ (as long as $k < 2/p - 1$, which will always be the case).

For $g(p) = p$ it is even possible to get an exact correction term if the distribution of σ^* is manageable. We consider $\sigma^* = \hat{\sigma}$. Then we have

$$EP_n = P(X_{n+1} > \hat{\mu} + (u+c)\hat{\sigma}).$$

Since

$$\frac{X_{n+1}-\hat{\mu}}{\hat{\sigma}} = \frac{X_{n+1}-\hat{\mu}}{S/c_4(n)}$$

has the same distribution as $\sqrt{1+\frac{1}{n}}c_4(n)T_{n-1}$, where T_{n-1} has a Student-distribution with $n-1$ degrees of freedom, the exact correction term for $g(p)=p$ is given by

$$c = \sqrt{1+\frac{1}{n}}c_4(n)t_{n-1;p} - u \text{ with } P(T_{n-1} > t_{n-1;p}) = p. \tag{3.7}$$

Remark 3.2 If we consider $\sigma^* = S$, the exact correction term equals $c = \sqrt{1+\frac{1}{n}}t_{n-1;p} - u$ and (of course) the same control chart is obtained, cf. also Remark 3.1. □

Remark 3.3 The control chart with the exact correction, given in (3.7), can already be found on page 1806 of Ghosh, Reynolds and Van Hui (1981), where it is remarked that "the problem with this approach is that the run length distribution and ARL are still unknown". It corresponds also to the so called Q-chart, presented by Quesenberry (1991), when applying (7) on page 215 of that paper with $r = n+1$. Roes, Does and Schurink (1993) present exact corrections for control charts with several other estimators as well. □

The following table gives the various correction terms.

Table 7 Correction terms for $Eg(P_n)$ with $g(p) = p, g(p) = 1/p$ and $g(p) = 1-(1-p)^k$.

	EP_n			$E(1/P_n)$	
	exact	1st type	2nd type	1st type	2nd type
$n=10$	1.2931	0.9722	0.8923	-0.9722	-1.0521
$n=20$	0.5296	0.4861	0.4461	-0.4861	-0.5261
$n=30$	0.3325	0.3241	0.2974	-0.3241	-0.3507
$n=40$	0.2423	0.2431	0.2231	-0.2431	-0.2630
$n=50$	0.1906	0.1944	0.1785	-0.1944	-0.2104
$n=60$	0.1570	0.1620	0.1487	-0.1620	-0.1754
$n=70$	0.1335	0.1389	0.1275	-0.1389	-0.1503
$n=80$	0.1162	0.1215	0.1115	-0.1215	-0.1315
$n=90$	0.1028	0.1080	0.0991	-0.1080	-0.1169
$n=100$	0.0922	0.0972	0.0892	-0.0972	-0.1052

	$ER_{n;1000}$	$ER_{n;500}$	$ER_{n;250}$	$ER_{n;100}$
	2nd type	2nd type	2nd type	2nd type
$n = 10$	-0.0799	0.4067	0.6499	0.7959
$n = 20$	-0.0400	0.2033	0.3250	0.3980
$n = 30$	-0.0266	0.1356	0.2166	0.2653
$n = 40$	-0.0200	0.1017	0.1625	0.1990
$n = 50$	-0.0160	0.0813	0.1300	0.1592
$n = 60$	-0.0133	0.0678	0.1083	0.1327
$n = 70$	-0.0114	0.0581	0.0928	0.1137
$n = 80$	-0.0100	0.0508	0.0812	0.0995
$n = 90$	-0.0089	0.0452	0.0722	0.0884
$n = 100$	-0.0080	0.0407	0.0650	0.0796

Remark 3.4 The correction terms are mostly positive leading to a larger UCL. Exceptions are $g(p) = 1/p$ and $g(p) = 1-(1-p)^k$ with $k = 1000$. For $g(p) = 1/p$ this is due to the fact that in that case a substantial positive bias occurs, when estimators of the parameters are plugged in, although the function g is decreasing. Explanations of this phenomenon are given in Remark 2.1 and Remark 2.4.

The bias of $ER_{n,k}$ is slightly negative, while $g(p) = 1-(1-p)^k$ is increasing. A reason for it is given in Remark 2.4. As a consequence of the negative bias, the corresponding correction term is slightly negative. □

The effect of the correction terms is seen by simulation of $Eg(P_n)$ with the corrected UCL. The following table gives the results.

Table 8 Simulation results of $Eg(P_n)$ after correction.

	$EP_n \times 10^3$		$E(1/P_n)$	
$g(p)$	1.000		1000	
	1st type	2nd type	1st type	2nd type
$n = 10$	1.6042	1.7710	337	279
$n = 20$	1.0882	1.1958	993	848
$n = 30$	1.0189	1.0899	1029	917
$n = 40$	0.9994	1.0491	1041	959
$n = 50$	0.9887	1.0407	1025	961
$n = 60$	0.9848	1.0271	1044	988
$n = 70$	0.9834	1.0242	1030	988
$n = 80$	0.9849	1.0143	1027	990
$n = 90$	0.9854	1.0119	1023	993
$n = 100$	0.9852	1.0093	1020	993

	$ER_{n,1000}$	$ER_{n,500}$	$ER_{n,250}$	$ER_{n,100}$
$g(p)$	0.6323	0.3936	0.2213	0.0952
$n = 10$	0.6103	0.3221	0.1843	0.0959
$n = 20$	0.6156	0.3522	0.2016	0.0950
$n = 30$	0.6187	0.3658	0.2082	0.0945
$n = 40$	0.6224	0.3728	0.2128	0.0945
$n = 50$	0.6225	0.3805	0.2147	0.0947
$n = 60$	0.6237	0.3825	0.2161	0.0948
$n = 70$	0.6259	0.3854	0.2174	0.0952
$n = 80$	0.6260	0.3863	0.2185	0.0954
$n = 90$	0.6269	0.3883	0.2193	0.0954
$n = 100$	0.6279	0.3889	0.2194	0.0952

Table 8 shows that the correction terms work very well. It was shown in Section 2 that simply plugging in the estimators in the control limits requires very large sample sizes to get accurate results. With the correction terms we already have very accurate UCL's for common sample sizes like 30. To be more precise, the next table shows for which n the same criterion is met as in Section 2, a relative error equal to 0.1 for $g(p) = p$ or $g(p) = 1/p$ and an absolute error equal to 0.01 for $g(p) = 1 - (1-p)^k$.

Table 9 Sample sizes needed to get a relative error 0.1 ($g(p) = p$ or $g(p) = 1/p$) and an absolute error 0.01 ($g(p) = 1 - (1-p)^k$).

	1^{st} type	2^{nd} type
EP_n	21	29
$E(1/P_n)$	15	25
$ER_{n,1000}$	-	43
$ER_{n,500}$	-	65
$ER_{n,250}$	-	36
$ER_{n,100}$	-	10

Comparison with Table 6 shows that the correction terms indeed do their job: the very large sample sizes needed in Section 2 are reduced to common sample sizes, usually available in practice.

Applying $\hat{\sigma}$ as estimator of σ, we therefore propose the following UCL's for the corrected control charts

$$\hat{\mu} + (u + c)\hat{\sigma}$$

with for the first type approximation c given by (3.1) if $g(p) = p$ and by (3.3) if $g(p) = 1/p$, and for the second type approximation c given by (3.6).

We summarize the conclusions of this section.

1. The corrected control limits work very well, giving accurate results for common sample sizes.
2. The correction terms are easy to apply.
3. The first and second type correction terms perform both very well, with the first type correction term even slightly better for $n \leq 50$. However, the second type correction term can be applied as well for correcting $ER_{n,k}$.
4. The asymptotics are needed to derive the correction terms. It seems impossible to "guess" the right form of the correction terms by just doing simulations. In fact, $c_4(n)$ is an example of a widely used correction, which, however, does not work.

4 The out-of-control situation

It is seen in Section 2 that simply plugging in estimators of μ and σ in the UCL leads to accurate control charts for the in-control situation only for very large sample sizes. The correction terms of Section 3 solve this problem. The question remains what the influence of this correction is for moderate sample sizes on the out-of-control case. If the corrections are so large that we almost never have a signal when the system is out-of-control, then the control chart would become useless. Fortunately, as we will show in this section, the corrections do not at all disturb the out-of-control behavior drastically. On the contrary, the gain in taking very large sample sizes is marginal and will in most cases not be in balance with the costs of the extra observations.

In general, one may think that estimation of parameters brings some extra costs, when compared to the situation of known parameters. Indeed, as a rule this is right, in the sense that we have to pay for the estimation in the out-of-control situation. This is guessed from Table 7, where, except for $E(1/P_n)$ and $ER_{n,1000}$, the correction terms are positive and hence lead to a higher UCL, implying that it is more difficult to get a signal, also in the out-of-control case.

The function $g(p) = 1/p$ and $g(p) = 1-(1-p)^{1000}$ are exceptions, as explained in Remark 2.1 and Remark 2.4. For those cases the correction terms are negative, since according to these criteria estimation is "profitable": in the in-control case the expectation of the estimated ARL is larger than the ARL with known parameters and $ER_{n,1000}$ is smaller than $1-(1-p)^{1000}$. For some authors, cf. Roes (1995) page 34 and Does and Schriever (1992), this is a reason to dispute the criterion $E(1/P_n)$. The same argument holds w.r.t. $ER_{n,1000}$.

What we want to do in this section is firstly to see how in the out-of-control case the performance of the corrected control charts changes when the number of observations grows and secondly, how much difference there is compared to control charts with known parameters. In this way we can also compare the strategy of simply plugging in estimators, and hence having the cost of taking very large sample sizes, with the strategy of applying the corrected control charts for moderate sample sizes.

On the other hand, due to the enormous flexibility in production processes (shorter product life cycles, product diversity, products tailored to specific customer requirements etc.), it is not possible in a lot of practical situations today to collect many observations. Therefore, in many cases we cannot avoid correction terms and it is of great interest to see how the performance of control charts changes with the number of observations in the region of moderate sample sizes.

Our set-up in the out-of-control case is as follows. Let $X_1, \ldots, X_n$ be independent and identically distributed r.v.'s, each with a $N(\mu, \sigma^2)$-distribution and let X_{n+1} have a $N(\mu_1, \sigma_1^2)$-distribution. The r.v.'s $X_1, \ldots, X_n$ are the observations on which the estimators of μ and σ are based, while X_{n+1} is the out-of-control variable. As we restrict attention to UCL's it is assumed that $\mu_1 > \mu$.

Denoting the (corrected) UCL by

$$\hat{\mu} + (u + c)\sigma^*,$$

the probability of a signal that the process is out-of-control is given by

$$P_n = P(X_{n+1} > \hat{\mu} + (u + c)\sigma^*) = \overline{\Phi}\left(\frac{\hat{\mu} - \mu_1}{\sigma_1} + (u + c)\frac{\sigma^*}{\sigma_1}\right)$$

$$= \overline{\Phi}\left(\frac{\mu - \mu_1}{\sigma_1} + u\frac{\sigma}{\sigma_1} + \frac{\sigma}{\sigma_1}\left\{c + \frac{\hat{\mu} - \mu}{\sigma} + (u + c)\left(\frac{\sigma^*}{\sigma} - 1\right)\right\}\right)$$

$$= \overline{\Phi}\left(u_1 + \frac{\sigma}{\sigma_1}\{c + \Delta(u + c)\}\right),$$

where

$$u_1 = \frac{\mu - \mu_1}{\sigma_1} + u\frac{\sigma}{\sigma_1}.$$

Writing

$$p_1 = \overline{\Phi}(u_1),$$

by a similar argument as in (the proof of) Theorem 2.1 we arrive at the first type approximation

$$P_n \approx p_1 W_n \text{ with } W_n \sim \text{ lognormal}\left(-c\frac{\sigma}{\sigma_1}\frac{u_1}{z(u_1)}, \left\{\frac{u_1}{z(u_1)}\right\}^2\left(\frac{\sigma}{\sigma_1}\right)^2\left\{\frac{1}{n} + \frac{(u + c)^2}{2(n - 1)}\right\}\right).$$

Hence, we get as first type approximation in the out-of-control situation

$$\begin{aligned} EP_n &\approx p_1 \exp\left[-c\frac{\sigma}{\sigma_1}\frac{u_1}{z(u_1)} + \frac{1}{2}\left\{\frac{u_1}{z(u_1)}\right\}^2 \left(\frac{\sigma}{\sigma_1}\right)^2 \left\{\frac{1}{n} + \frac{(u+c)^2}{2(n-1)}\right\}\right], \\ E\frac{1}{P_n} &\approx \frac{1}{p_1} \exp\left[c\frac{\sigma}{\sigma_1}\frac{u_1}{z(u_1)} + \frac{1}{2}\left\{\frac{u_1}{z(u_1)}\right\}^2 \left(\frac{\sigma}{\sigma_1}\right)^2 \left\{\frac{1}{n} + \frac{(u+c)^2}{2(n-1)}\right\}\right]. \end{aligned} \tag{4.1}$$

If the parameters μ and σ are known, we get p_1 and $(1/p_1)$, respectively. Note that c is relatively small compared to u (see Table 7). Ignoring c in the second part of the exponential terms, that is replacing $(u+c)^2$ by u^2, the second part gives the positive bias in EP_n and $E(1/P_n)$ when estimators are simply plugged in. The (main) influence due to the correction term c is in the factors

$$\exp\left\{-c\frac{\sigma}{\sigma_1}\frac{u_1}{z(u_1)}\right\} \text{ and } \exp\left\{c\frac{\sigma}{\sigma_1}\frac{u_1}{z(u_1)}\right\}, \tag{4.2}$$

respectively. Since the correction term for EP_n equals minus the one for $E(1/P_n)$, see (3.1) and (3.3), both factors in (4.2) are equal and they are smaller than 1. It turns out (see Table 10) that for $E(1/P_n)$ and in most cases also for EP_n, the factors in (4.2) dominate the factors coming from the second part of the exponential term in (4.1). Hence, the exponential terms in (4.1) result in a factor smaller than 1, thus showing for EP_n the penalty we have to pay for correct estimation of the parameters μ and σ and for $E(1/P_n)$ the "gain" obtained by this estimation.

The following table shows the results in the out-of-control situation. Since the control charts are designed for the mean, we consider only a change in the mean and keep the variance equal to the variance in the out-of-control case. However, the control charts considered here can also detect changes in variance. The influence of changing the variance is seen in (4.1).

Table 10 First type approximation (4.1) of EP_n and $E(1/P_n)$ with $\mu_1 = \mu + a\sigma$ and $\sigma_1 = \sigma$.

	$a = 0.5$		$a = 1$		$a = 2$	
	$p_1 = 0.0048$	$1/p_1 = 208.5$	$p_1 = 0.0183$	$1/p_1 = 54.6$	$p_1 = 0.138$	$1/p_1 = 7.3$
n	EP_n	$E(1/P_n)$	EP_n	$E(1/P_n)$	EP_n	$E(1/P_n)$
10	0.0208	54.0	0.0359	14.4	0.107	2.4
20	0.0060	133.2	0.0178	33.0	0.104	4.5
30	0.0050	163.4	0.0167	40.6	0.111	5.3
40	0.0048	177.3	0.0167	44.4	0.116	5.8
50	0.0047	185.1	0.0168	46.6	0.119	6.1
60	0.0047	189.9	0.0169	48.1	0.122	6.3
70	0.0047	193.1	0.0171	49.1	0.124	6.4
80	0.0047	195.4	0.0172	49.8	0.125	6.5
90	0.0047	197.2	0.0173	50.4	0.127	6.6
100	0.0047	198.5	0.0173	50.8	0.128	6.7
150	0.0047	202.3	0.0176	52.2	0.131	6.9
200	0.0047	204.0	0.0178	52.8	0.132	7.0
300	0.0047	205.6	0.0179	53.4	0.134	7.1

It is seen from Table 10 that with moderate sample sizes already very good behavior is achieved in the out-of-control situation. Therefore, the goal of the correction terms is reached: the in-control behavior is regulated and the loss in the out-of-control case is small for moderate sample sizes, even when compared to control charts with known parameters.

It is seen that for $n \geq 50$, say, the behavior changes only slowly with n and in most practical situations the improvement will be outweighed by the costs of taking more observations. Even for such small values of n as 20, already very reasonable results are obtained.

Although, when simply plugging in the estimators of the parameters μ and σ, very many observations are needed to get accurate control charts, we can conclude from Table 10 that control charts with correction terms can be based on moderate sample sizes, as the latter charts are accurate for the in-control situation and powerful in the out-of-control case. Hence, in cases where no large sample sizes are available, which in today's practice occurs more and more, the corrected control charts provide a good solution, which was not present before.

The second type approximation is based on the following expansion

$$Eg(P_n) = Eh\left(u_1 + \frac{\sigma}{\sigma_1}\{c + \Delta(u+c)\}\right)$$

$$\approx Eh(u_1) + h'(u_1)\frac{\sigma}{\sigma_1}\{c + E\Delta(u+c)\} + \frac{1}{2}h''(u_1)\left(\frac{\sigma}{\sigma_1}\right)^2 E\Delta^2(u+c),$$

leading for $\sigma^* = \hat{\sigma}$ to

$$\begin{aligned} EP_n &\approx p_1 - c\frac{\sigma}{\sigma_1}\varphi(u_1) + \frac{1}{2}\left(\frac{\sigma}{\sigma_1}\right)^2 a(u+c,n)u_1\varphi(u_1) \\ E\frac{1}{P_n} &\approx \frac{1}{p_1} + c\frac{\sigma}{\sigma_1}\frac{\varphi(u_1)}{\overline{\Phi}^2(u_1)} + \frac{1}{2}\left(\frac{\sigma}{\sigma_1}\right)^2 a(u+c,n)\left(\frac{2\varphi^2(u_1)}{\overline{\Phi}^3(u_1)} - \frac{u_1\varphi(u_1)}{\overline{\Phi}^2(u_1)}\right). \end{aligned} \tag{4.3}$$

Note that c is relatively small compared to u (see Table 7). Ignoring c in the last terms, that is replacing $a(u+c,n)$ by $a(u,n)$, these last terms give the bias in EP_n and $E(1/P_n)$ when estimators are simply plugged in. The (main) influence due to the correction term c is in the terms

$$-c\frac{\sigma}{\sigma_1}\varphi(u_1) \text{ and } c\frac{\sigma}{\sigma_1}\frac{\varphi(u_1)}{\overline{\Phi}^2(u_1)}. \tag{4.4}$$

Since the correction term for EP_n equals minus the one for $E(1/P_n)$, see (3.1) and (3.3), both terms in (4.4) are negative. It turns out (see Table 11) that for $E(1/P_n)$ and in most cases also for EP_n the terms in (4.4) dominate the last terms on the right-hand side of (4.3), thus showing for EP_n the penalty we have to pay for correct estimation of the parameters μ and σ and for $E(1/P_n)$ the "gain" obtained by estimation.

Table 11 Second type approximation (4.3) of EP_n and $E(1/P_n)$ with $\mu_1 = \mu + a\sigma$ and $\sigma_1 = \sigma$.

	$a = 0.5$		$a = 1$		$a = 2$	
	$p_1 = 0.0048$	$1/p_1 = 208.5$	$p_1 = 0.0183$	$1/p_1 = 54.6$	$p_1 = 0.138$	$1/p_1 = 7.3$
n	EP_n	$E(1/P_n)$	EP_n	$E(1/P_n)$	EP_n	$E(1/P_n)$
10	0.0105	-100.4	0.0254	-22.8	0.062	-0.8
20	0.0055	109.5	0.0163	26.6	0.086	3.9
30	0.0049	155.8	0.0159	38.6	0.100	5.2
40	0.0047	174.1	0.0161	43.6	0.109	5.8
50	0.0046	183.5	0.0164	46.3	0.114	6.1
60	0.0046	189.1	0.0166	47.9	0.118	6.3
70	0.0046	192.7	0.0168	49.1	0.120	6.5
80	0.0046	195.3	0.0169	49.9	0.123	6.6
90	0.0046	197.1	0.0171	50.5	0.124	6.6
100	0.0047	198.5	0.0172	50.9	0.126	6.7
150	0.0047	202.4	0.0175	52.3	0.130	6.9
200	0.0047	204.2	0.0177	52.9	0.132	7.0
300	0.0047	205.8	0.0179	53.5	0.134	7.1

The same comments as given below Table 10 apply to the results of Table 11. (Clearly, the approximations for such a small sample size as 10 should not be taken seriously. In view of the conditions in Theorem 2.2 this is not surprising. Note also that the conditions in Theorem 2.2 are slightly stronger than in Theorem 2.1: indeed, the aberrance for $n = 10$ in Table 11 exceeds the corresponding one in Table 10.)

Of course, the conclusions based on Table 10 and 11 are only valid if the approximations work well. To see whether the approximations are accurate enough a simulation study is performed with the following results.

Table 12 Simulation results for EP_n and $E(1/P_n)$ with first type correction terms, given by (3.1) and (3.3), when $\mu_1 = \mu + a\sigma$ and $\sigma_1 = \sigma$.

	$a = 0.5$		$a = 1$		$a = 2$	
	$p_1 = 0.0048$	$1/p_1 = 208.5$	$p_1 = 0.0183$	$1/p_1 = 54.6$	$p_1 = 0.138$	$1/p_1 = 7.3$
n	EP_n	$E(1/P_n)$	EP_n	$E(1/P_n)$	EP_n	$E(1/P_n)$
10	0.0051	82.7	0.0140	20.0	0.073	3.2
20	0.0043	172.9	0.0141	39.1	0.090	5.1
30	0.0043	190.2	0.0147	45.2	0.102	5.8
40	0.0044	195.2	0.0153	48.4	0.109	6.2
50	0.0044	200.5	0.0158	49.5	0.113	6.4
60	0.0044	200.8	0.0161	50.6	0.119	6.5
70	0.0045	202.4	0.0164	51.6	0.120	6.6
80	0.0045	203.7	0.0166	51.7	0.122	6.7
90	0.0045	204.5	0.0167	52.1	0.123	6.8
100	0.0045	204.6	0.0168	52.3	0.124	6.8
150	0.0046	206.1	0.0173	52.9	0.129	7.0
200	0.0047	207.0	0.0175	53.6	0.131	7.0
300	0.0047	207.7	0.0178	54.0	0.133	7.1

Table 13 Simulation results for EP_n and $E(1/P_n)$ with second type correction terms, given by (3.6), when $\mu_1 = \mu + a\sigma$ and $\sigma_1 = \sigma$.

	$a = 0.5$		$a = 1$		$a = 2$	
	$p_1 = 0.0048$	$1/p_1 = 208.5$	$p_1 = 0.0183$	$1/p_1 = 54.6$	$p_1 = 0.138$	$1/p_1 = 7.3$
n	EP_n	$E(1/P_n)$	EP_n	$E(1/P_n)$	EP_n	$E(1/P_n)$
10	0.0058	58.2	0.0155	15.0	0.079	2.8
20	0.0046	140.7	0.0151	34.8	0.096	4.7
30	0.0045	170.0	0.0155	42.0	0.106	5.5
40	0.0045	184.1	0.0160	45.1	0.112	6.0
50	0.0046	189.1	0.0163	47.5	0.117	6.2
60	0.0046	193.2	0.0166	49.1	0.119	6.4
70	0.0046	195.3	0.0168	49.9	0.122	6.5
80	0.0046	198.2	0.0169	50.3	0.123	6.6
90	0.0046	198.0	0.0171	50.8	0.125	6.7
100	0.0046	200.0	0.0171	51.3	0.126	6.7
150	0.0047	204.0	0.0177	52.6	0.130	6.9
200	0.0047	204.4	0.0176	53.1	0.132	7.0
300	0.0047	205.8	0.0179	53.5	0.134	7.1

Although for extremely small n the second type approximation is clearly unreliable, the simulations show that for moderate sample sizes both approximations work very well and hence approximations like (4.1) and (4.3) are very useful. Therefore, the conclusions based on Tables 10 and 11 are confirmed by the simulation results.

It is seen from the simulations in Tables 12 and 13 that, in terms of EP_n and $E(1/P_n)$, the out-of-control behavior of the corrected control charts with the second type of correction is slightly better than the one with the first type of correction. This is simply due to the better correction (for $n \leq 50$) obtained by the first type of correction when we have the in-control situation, cf. Table 8, leading to larger correction terms (and hence larger UCL's) for EP_n and smaller ones (implying smaller UCL's) for $E(1/P_n)$, cf. Table 7.

Simulations w.r.t. $ER_{n,k}$ (not presented here) confirm the findings discussed following some lines after Table 10 that the corrected control charts behave very well with only a small loss in the out-of-control case for moderate sample sizes.

We summarize the conclusions of this section.

1. Inserting the correction terms of Section 3 does not disturb the behavior of the control charts in the out-of-control situation. On the contrary, for moderate sample sizes the corrected control charts work very well.
2. When no large sample sizes are available, no accurate control charts were possible; the corrected control charts provide a solution for this problem.
3. The gain in taking (very) large sample sizes is marginal and will usually be outweighed by the costs of these observations.
4. The recommendation to take at least 300 observations to get accurate and powerful control charts can be reduced to 40 observations when corrected control charts are applied.
5. The behavior in the out-of-control situation can be described very well by simple approximations as in (4.1) and (4.3).

Acknowledgement The authors would like to thank Sri Nurdiati for doing the Monte Carlo studies.

References

Chen, G. (1997). The mean and standard deviation of the run length distribution of $\overline{X}$ charts when control limits are estimated. *Statist. Sinica* **7** 789-798.

Del Castillo, E. and Montgomery, D.C. (1995). A Kalman filtering process control scheme with an application in semiconductor short run manufacturing. *Quality and Reliability Engineering International* **11** 101-105.

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

Ghosh, B.K., Reynolds, M.R.Jr. and Hui, Y.V. (1981). Shewhart $\overline{X}$-charts with estimated process variance. *Comm. Statist. Theory Methods* **10** 1797-1822.

Lehmann, E.L. (1986). *Testing Statistical Hypotheses* 2^{nd} ed. Wiley, New York.

Quesenberry, C.P. (1991). SPC Q charts for start-up processes and short or long runs. *J. Quality Technol.* **23** 213-224.

Quesenberry, C.P. (1993). The effect of sample size on estimated limits for $\overline{X}$ and X control charts. *J. Quality Technol.* **25** 237-247.

Quesenberry, C.P. (1995). On properties of Q charts for variables. *J. Quality Technol.* **27** 184-203.

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

Roes, C.B., Does, R.J.M.M. and Schurink, Y. (1993). Shewhart-type control charts for individual observations. *J. Quality Technol.* **25** 188-198.

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