CONTROL OF CONTROL CHARTS

The research presented in this thesis was carried out at the group of Statistics and Probability, Department of Applied Mathematics, Faculty of Electrical Engineering, Mathematics and Computer Science, Twente University, Enschede, The Netherlands.

The research was made possible by the Technology Foundation STW (TWI. 5422), applied science division of NWO and technology program of the Ministry of Economic Affairs, The Netherlands.

Cover designed by Anggoro Santoso Edy Widayat

Printed by Wohrmann Print Service, The Netherlands

ISBN 90-365-2128-9

CONTROL OF CONTROL CHARTS

DISSERTATION

to obtain
the doctor's degree at the University of Twente,
on the authority of the rector magnificus,
prof. dr. W.H.M. Zijm,
on account of the decision of the graduation committee,
to be publicly defended
on Friday 21st January 2005 at 15.00

by

Sri Nurdiati
born on 26 November 1960
in Malang, East Java, Indonesia

Dit proefschrift is goedgekeurd door de promotor,
prof. dr. W. Albers

en de assistent-promotor,
dr. W. C. M. Kallenberg

O, ye who believe!
If ye will aid,
(the cause of) God,
He will aid you,
and plant your feet firmly.
(Q.S. 47:7)

Verily, with every difficulty
there is relief.
Therefore, when thou art
free (from thine immediate task),
still labour hard.
And to thy Lord
turn (all) thy attention.
(Q.S. 94:6-8)

Contents

Chapter 1

Introduction

Maintaining a production process in good condition is an important issue for many practitioners. For that reason it is necessary to monitor such a process so that as soon as it starts to act differently from the required conditions a signal will be produced. With the occurrence of this signal, the production process will be stopped for some adjustment before it is restarted again.

This monitoring task is often based on a so called Shewhart $\overline{X}$-chart which consists of a center line, an upper control limit (A1) and a lower control limit (B1), see Figure 1.1. The function of the Shewhart $\overline{X}$-chart is to monitor the mean of a process, which is expected to be around the center line. The area between the two control limits is considered as a permissible area for the observations. As long as the observations scatter randomly inside this area, the process is considered to be in-control. Otherwise, if the observations form a pattern or there is an observation which falls outside this permissible area, then there is an indication of an out-of-control situation. In the present research, we focus on an out-of-control case that is indicated by an observation which exceeds the permissible area.

Since the effect of either stopping the running process or letting the process run in the out-of-control situation can be very substantial from the economic point of view, practitioners need to be very careful to decide when the process really has to be stopped. In other words, it is very crucial to choose the most suitable control limits to be used in practice. If the control limits are too tight (e.g. A2 and B2 in Figure 1.1) there will too often be a false out-of-control signal, and if these are too loose (e.g. A3 and B3 in Figure 1.1), most likely there will be no out-of-control signal at all while the process is gone out-of-control. Therefore, the objective of this research is to determine the most suitable control limits.

In Section 1.1, the problem which concerns the determination of control limits is discussed. There, some results which are related to this topic in the literature are

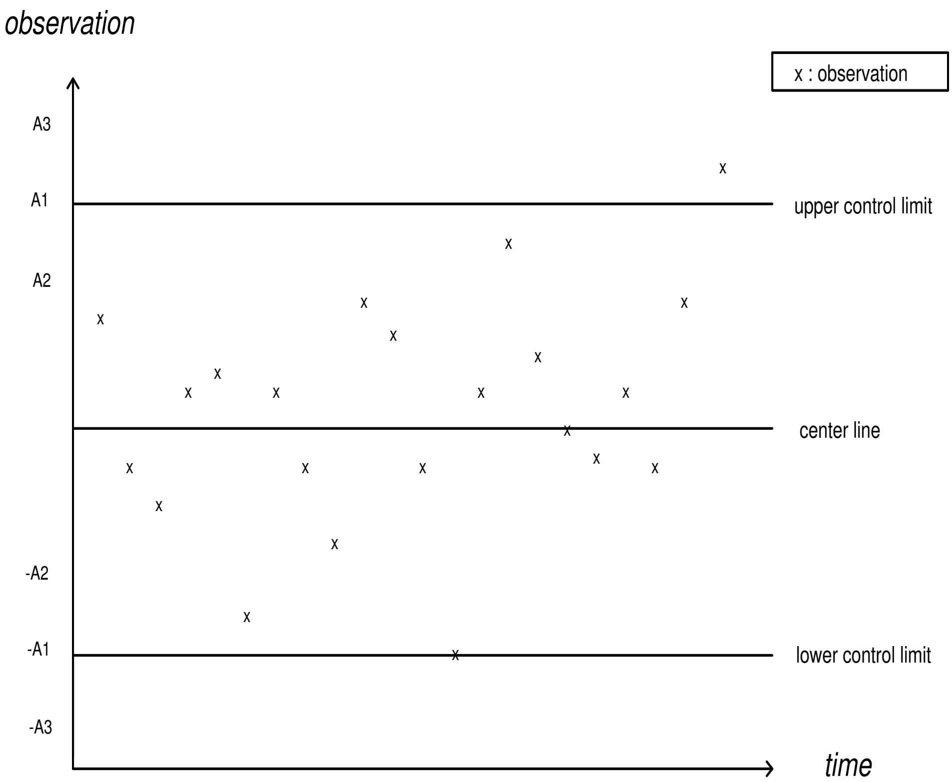

Figure 1.1. Shewhart $\overline{X}$-chart.

included. Criteria for setting the control limits are given in the following section. The contents of the thesis will be summarized in the final section.

1.1 The problem of setting the control limits

The Shewhart $\overline{X}$-chart is widely used because of its simplicity. It assumes normality of the underlying distribution of the observations. Hence, the control limits depend on the mean and standard error of the normal distribution in the in-control situation. In practice, applying this control chart to monitor the mean of a process may lead to two types of problems. The first concerns the typically unknown parameters involved in the distribution, while the second concerns the validity of the assumption itself.

In the case that the assumption is true that the observations are normally distributed but with unknown parameters, one has to estimate the parameters based on observations from the past, which are also known as Phase I observations. Since the estimation usually involves very small p, which is the probability of an in-control process producing an out-of-control signal, it may result in large relative errors, unless (very) large sample sizes are applied. In the present research, the so called stochastic error as an implication of the estimation process will be discussed. It turns out that the error caused 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.

Previously, the effect of the estimation step has been discussed a.o. by Woodall and Montgomery (1999). They mention that more data than are usually recommended are needed to accurately determine control chart limits. This is based on earlier research conducted e.g. by Quesenberry (1993) and Chen (1997) who concluded that the classical empirical rules for choosing the number of observations to estimate the mean and standard error of the normal distribution were inadequate and should be taken much larger. It was recommended to take at least 300 observations. For a further discussion on this topic, see also Roes (1995). For more recent references about the problems concerning estimation in control charts see Chakraborti (2000), Chakraborti et al. (2001) and Nedumaran and Pignatiello (2001).

In this thesis we first consider a normal approach by adopting the Shewhart $\overline{X}$-chart to investigate the effect of the estimation error. The results show that although the observations are normally distributed, we still need too large sample sizes to construct accurate control limits. Hence, we confirm the conclusions offered in the previous studies that basically very many data are needed to construct accurate control limits. Since such large sample sizes are usually not available in practice, we need to find a way to correct the control limits in order to have accurate limits for commonly used sample sizes. 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 introduced by the estimators, the run length distribution is no longer a geometric distribution. See Quesenberry (1993) for an extensive discussion on this point. It turns out that by adding suitable correction terms to the control limits, we are able to reduce substantially the sample sizes needed to construct accurate control limits.

In practice, the distribution of the observations can be different from the normal family. Hence, the normality assumption made for the observations often fails, which causes a model error. In such as a case, the probability of incorrectly producing an out-of-control signal may be seriously in error, see e.g. Chan et al.(1988) and Pappanastos and Adams (1996). Basically, the problem is that the normality assumption may be fair for the central part of the distribution, but produces large relative errors in the tail. And, in view of the small values of p

typically used, the tails are what we are dealing with. Therefore, a larger model is needed, providing more flexibility to describe the behavior in the far tail. For that purpose we propose to use a parametric as well as a nonparametric approach.

In the parametric approach we assume that the distribution of the observations is from a larger class of parametric models containing the normal family as a submodel. Several suitable models have been considered for this purpose, such as random or deterministic mixtures. However, for the following reasons we choose a so called normal power family as a supposed model. Firstly, using the normal power family gives more flexibility and hence improvement over simply using normality and ignoring the facts that in practice normality often fails. If control limits based on normality are applied, this should implicitly mean that the distribution is approximately normal, and in that case the normal power family is certainly appropriate, since normality is a submodel of the normal power family. Secondly, by the extension to the normal power family many other commonly used distributions show up, which are covered sufficiently well by members of the normal power family. Thirdly, the normal power family is not so large that accurate estimation is only possible with huge sample sizes as in the nonparametric approach.

Based on the assumption that the true distribution of the observations is from the normal power family, we can construct a so called parametric chart. Applying the parametric chart in a real life situation is able to reduce the model errors of many distributions considerably. On the other hand, as more parameters are needed to be estimated, larger stochastic errors may be expected. Therefore, the next step to be taken is to control these stochastic errors by developing correction terms for the control limits. A second order asymptotic approach is applied to derive the correction terms, since the first order asymptotics does not produce sufficiently accurate results, due to the very small values involved in the quantiles. With the addition of the correction terms in the control limits, the corrected parametric chart performs much better since it reduces the stochastic errors substantially.

The parametric approach based on the normal power family offers a satisfactory solution over a broad range of underlying distributions. However, situations do occur where even such a larger model is inadequate since the true distribution may still not be in it and the resulting model error may remain unacceptably large. In such a situation nothing remains but to consider a nonparametric approach. In principle these form ideal solutions, since the model error is simply not present here, and the stochastic error becomes arbitrarily small as the sample size increases. But on the other hand, for common sample sizes encountered in practice, this is not a lot of help. If we want to estimate e.g. the 0.999-quantile, it is clear that with 100 observations we will not get anywhere with a nonparametric estimate. The problem is left that huge sample sizes are required for the estimation step. Otherwise, the resulting stochastic error is so large that the control limits are very unstable, a disadvantage which seems to outweigh the advantage of avoiding the model error from the parametric case.

In this thesis we analyze under what conditions the nonparametric approach actually becomes feasible as compared to the parametric approach. In particular, corrected versions of control limits are suggested for which a possible change point is reached at sample sizes which are markedly less huge (but still larger than the customary range). These corrections serve to control the behavior during the in-control situation (markedly wrong outcomes of the estimates only occur sufficiently rarely). The price for this protection clearly will be some loss of detection power during the out-of-control situation. A change point comes in view as soon as this loss can be made sufficiently small.

We will restrict attention to the obvious choice based on the empirical quantile function, as this will already provide a clear picture of what can be expected in general. Some previous work on closely related topics can be found in Willemain and Runger (1996) and in Ion et al. (2000). For a recent overview of nonparametric charts in general, see e.g. Chakraborti et al. (2001). Incidentally, as these authors point out, several of the procedures that have been proposed are in fact not truly nonparametric or distribution-free. They are based on a nonparametric estimator, like the Hodges-Lehmann estimator, rather than on $\overline{X}$, but the actual in-control run length distribution involved does depend on the underlying distribution of the observations.

Each of the three approaches discussed above has advantages as well as disadvantages with respect to another. The Shewhart $\overline{X}$-chart, also known as the normal chart, is preferable as long as the true distribution is (very) close to normality. However, for distributions farther away from normality, but still close to the larger parametric family, the best choice is to adopt the parametric chart. Of course, there also are distributions so far outside the larger parametric family that the parametric chart is not satisfactory either and the nonparametric chart should be applied. To make a choice we need to find a way such that the data can select the most suitable control chart for themselves.

A first idea might be to execute a (standard) goodness of fit test to investigate normality. If normality is not rejected, use the normal chart. If normality is rejected, apply a goodness of fit test for the normal power family. Again when this is not rejected, apply the parametric chart. Otherwise, use the nonparametric chart (if this makes sense). Although this way of thinking looks attractive, it has a serious drawback. Standard goodness of fit tests are looking at the majority of the data, and as such concentrate on the middle of the distribution, while here we are not interested in this middle part, but in the (extreme) tail. Therefore, standard goodness of fit tests are not appropriate for the situation at hand. For the same reason, less formal methods like “a good look at the data” or “an inspection of a histogram” are completely insufficient to judge the possible normality in the far tail.

The choice between the three control charts can be seen as a model selection problem. This area in statistics is nowadays in the center of interest and therefore it looks promising to apply such methods not merely for the three charts mentioned here, but for a whole range of (nested) models. Unfortunately, two problems arise. First of all, it is far from easy to develop control charts in each of these models. Secondly, again the common selection rules are intended for the bulk of the data and not for the extreme tail. The motivation to switch from the normal control chart to the parametric control chart or even to the nonparametric control chart is provided by the model error. As this model error is the discriminating quantity in deciding which control chart to use and since moreover the data should tell us what the appropriate model is, it is natural to base the decision between the three control charts on the estimated model error. The idea to let the data tell what control chart to use is discussed in more detail in this thesis under the so called combined approach.

1.2 Criteria for setting the control limits

To study the problem of setting control limits as described in Section 1.1, first of all a performance characteristic for the given control limits has to be selected, such as the out-of-control signal probability (p), the average run length ($ARL = 1/p$), or any well-behaved function g of p ($g(p)$). Due to the estimation, this performance characteristic becomes random. Hence, p is replaced by a stochastic counterpart P, the $ARL = 1/p$ likewise by $1/P$ and generally $g(p)$ by $g(P)$. Subsequently, for each characteristic the absolute error (e.g. $|P - p|$) or the relative error (e.g. $(P/p-1)$) can be studied with respect to aspects such as expectation, semi-variance and exceedance probability, each of which in its own way helps to characterize the behavior of the estimated control limits.

In this thesis we apply both bias and exceedance probability criteria for setting the control limits. A bias criterion aims at reducing the discrepancy between the actual result and the true value of the performance characteristic, while the exceedance probability criterion is applied for controlling the probability of having a relative error larger than a given value. More precise descriptions of the two criteria are given in the following subsections.

1.2.1 Reducing bias

The term "bias", or sometimes systematic error , is used to express the difference between the measurement result and its unknown "true value". In this study we try to reduce the bias between the performance characteristic and its counterpart for the case of known parameters as much as possible. Since in the estimated case the performance characteristic is a random variable, we need a measurement for the random variable before we can compare it with another value. One way to measure the random variable is by using its expectation which shows the average performance of the random variable over a long series of applications. For example $E(P)$ can be compared with p, $E(1/P)$ can be compared with the fixed $ARL = 1/p$, or $E(g(P))$ with $g(p)$.

Furthermore, the comparison can be done e.g. by investigating either the relative error or the absolute error between the two values. In order to reduce the bias, we have to restrict the two errors to be at most a small number. For example, it can be investigated when the stochastic counterpart P of p satisfies $|EP - p|/p \leq 0.1$, or satisfies $|EP - p| \leq 0.01$, etc. In our case, for $p = 0.001$ the relative error of at most 10% is satisfied when we have more than 300 observations, while for $p = 0.01$ we need to have more than 100 observations (see Section 2.2).

Of course, instead of using the expectation, we can also use either the semi-variance or the exceedance probability to check the performance of the control charts. The latter will be described in the following subsection.

1.2.2 Controlling exceedance probability

In the bias criterion we only used the expectation of a random variable to investigate the performance characteristics of the control chart. Using expectation, however, only captures part of the picture. Since the variability of P around its expected value is rather large, it also seems worthwhile to control the exceedance probabilities involved. Hence, rather than only worrying about $|EP - p|/p > 0.1$, we now try to figure out how large $P((P - p)/p > 0.1)$ is. This probability indicates how likely it is that the Phase I observations produce an outcome of P larger than 0.0011 when $p = 0.001$. Once such an outcome has occurred, we are so to say stuck with it during the subsequent application of the chart. The fact that EP may be well behaved only means that the present large outcome is expected to be balanced by a small one during a future application of the newly estimated chart. Hence bias correction is certainly useful with respect to the long-term behavior of the chart in a series of separate applications. But exceedance probabilities supply information about how likely it is that errors beyond a certain level occur in a single given application. From a practical point of view, this may even be the more interesting topic to discuss.

1.3 Outline of the thesis

In this thesis we propose four different approaches to determine the control limit. The difference is mainly due to the different assumptions made for the underlying distribution of the observations. The following approaches have been considered:

- Normal approach, assuming that the true distribution of the observations comes from a normal family (Chapter 2).

- Parametric approach, assuming that the true distribution is from a broader parametric class than the normal family. In this case we assume the true distribution is from a so called normal power family (Chapters 3 and 4).

- Nonparametric approach, applying a distribution-free method. Hence, no assumption is needed except the continuity of the true distribution (Chapter 5).

- Combined approach, as the name suggests, offering to combine the normal, parametric and nonparametric approaches and let the data, based on certain criteria, select the most suitable control limit for themselves (Chapter 6).

The results presented in Chapter 2 do confirm that although the normality assumption for the observations is fulfilled, too large sample sizes are needed to get accurate control limits when estimators of the parameters are simply plugged in. However, the samples sizes needed to construct accurate control limits can be reduced substantially if we add suitable correction terms to the control limits. The discussions throughout Chapters 3-5 point out that with the addition of appropriate correction terms, the performance of the other control limits is also highly improved. The derivation of the control limits and their corresponding correction terms is based on asymptotic as well as on approximation theory. To check the value of the results, extensive simulation studies are performed. The results of these simulations indeed confirm those from the asymptotic theory. Some examples are given to provide better illustrations of the material presented there.

From the discussion presented in Chapters 2-5 we conclude that applying the corrected versions of the normal, parametric and nonparametric charts to monitor the production process in practice offers considerable advantages. However, applying these charts individually may cause some disadvantages due to the specific characteristics of the observations. Therefore, we recommend a combined chart to be used in practice. The derivation of the combined chart, as a product of the combined approach, is discussed in Chapter 6. Finally, Chapter 7 provides a guideline to implement the material presented throughout Chapters 2-6 in real life situations.

Chapter 2

Normal Approach

This chapter discusses a problem which occurs when applying a standard control chart and proposes a way to deal with it. Supposing that the normality assumption for the observations is fulfilled, too large sample sizes are needed to get accurate control limits when estimators of the parameters are simply plugged in.

To fix the problem, correction terms are developed and applied to the control limits. The determination of these correction terms is based on so called bias and exceedance probability criteria. Using the corrected normal charts much smaller sample sizes are needed to get accurate control limits. The new charts are also shown to behave quite well in an out-of-control situation.

2.1 Introduction

Assuming normality of the observations in the in-control situation, the control limits of the standard chart depend on the mean and standard deviation of the normal distribution. For this reason, the standard chart is also known as a normal chart, and the latter will be used quite frequently in this thesis. With respect to the mean and standard deviation of the distribution, these parameters are most likely unknown and therefore, to apply the control limits, one has to estimate the parameters.

Estimation is based on the observations obtained in Phase I that we suppose to belong to the in-control situation. The monitoring phase is based on the observations from Phase II. Hence, an estimation error will have occurred. 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 effect 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 deviation of the normal distribution are inadequate and should be taken much larger. The studies recommend to take at least 300 observations to get accurate limits. Further discussion on this topic is given by Roes (1995), in particular in Section 2.2.2.

In line with these studies, we perform some simulations and an asymptotics study in the present chapter to show that (indeed) a great many data are needed to get accurate control charts limits, when estimators are simply plugged in. Since nowadays short production runs are more and more in demand, there may be not enough data to accurately estimate the process parameters.

Due to estimation the probability P of incorrectly concluding that the process is out-of-control is no longer deterministic, but a random variable. In this thesis with "accurate control charts" in the bias case we mean for example that P is close to a prescribed false alarm rate p, in the sense that EP is close to p, or in a more general way $Eg(P)$ is close to $g(p)$. More details on functions g of special interest are given in Section 2.2.2. This can be interpreted as meaning that in the long run the user always gets what he wants.

In the literature there is only a limited number of suggestions how to correct the control limits. Moreover, they are restricted to $g(p) = p$ and do not deal with exceedance probability. Hillier (1969) introduced a method to set control limits with an exact correction term based on a small number of subgroups. He estimated σ with the average range. Yang and Hillier (1970) proposed a similar correction term using average subgroup variance and sample variance to estimate σ^2. Exact correction terms using other estimators were also introduced by Ghosh, Reynolds and Van Hui (1981) and by Roes, Does and Schurink (1993). Other than that, we could not find any suggestion to have accurate control limits for commonly used sample sizes. This calls for the search for simple, but efficient correction terms to be applied in the control limits.

Hence, our goal is to develop correction terms in order that also for common sample sizes accurate control charts can be obtained. For this, we adopt so called bias and exceedance probability criteria. Based on these criteria we develop correction terms for the control limits. The correction term based on the bias criterion has order of $1/n$, while the correction term developed under the exceedance probability criterion has order of $1/\sqrt{n}$. From this we can see that the bias removal approach requires a moderate correction, while controlling the exceedance probability requires a larger correction for an initial control limit. The different types of protection give different effects in the out-of-control performance. Weak protection which is associated with the moderate correction has a low price, whereas strong protection which is associated with the larger correction, has a moderate price. Hence, by adding these correction terms to the initial control limit, we get two different types of control charts. Practitioners can choose either one of these which suit their needs the best.

As far as we know, what is presented in this chapter is the first contribution to finding 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 in the estimators, the run length distribution is no longer a geometric distribution, see Quesenberry (1993) for an extensive discussion on that point.

Fortunately, the correction terms, obtained by an asymptotic method, 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 accurate results already for moderate sample sizes.

Having developed the correction terms, we also want to investigate the consequences for the out-of-control situation of using the corrected control limits. 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 much by the more stringent control limits. The loss due to estimation is not very large.

The chapter is organized as follows. In Section 2.2 our set-up for the normal chart 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 2.3, based on the bias and exceedance probability criteria, where also the performance of the corrected control limits is exhibited. The out-of-control behavior is treated in Section 2.4. Finally, a concluding remark is given to close the chapter.

2.2 The characteristics of the normal chart

This section discusses the characteristics of the normal chart. First, we present the initial set up for the control limit of the chart. Next, using a simulation study and asymptotic theory we show that very large samples are needed to get accurate control limits when estimators are simply plugged in. We use bias and exceedance probability criteria to investigate the effect of the estimation error.

2.2.1 The initial control limit of the normal chart

Let $X_1, \cdots, X_n, X_{n+1}$ be independent and identically distributed (i.i.d.) distributed random variables (r.v.'s), each with a $N(\mu, \sigma^2)$-distribution. The r.v.'s $X_1, \cdots, X_n$ 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 2.2 and 2.3 we consider the in-control situation for Phase II, that is we assume that X_{n+1} has the same distribution as $X_1, \cdots, X_n$.

We mainly consider a control chart with an upper limit. The more standard case of upper- and lower control limit can be treated in a similar way and will be given only as a remark. Generalization to a set-up with X_i replaced by a group of observations, for instance $X_{i1}, \cdots, X_{i5}$, is fairly straightforward and will be excluded here.

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) of the normal chart equals

$$\mu + u_p\, \sigma$$

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, very often μ and σ are unknown and they have to be estimated on the basis of $X_1, \cdots, X_n$. We consider as estimator of μ the sample mean $\overline{X} = n^{-1} \sum_{i=1}^{n} X_i$ and denote this estimator by $\widehat{\mu}$. As estimator of σ we sometimes consider

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

but more often we take

$$\widehat{\sigma} = \frac{S}{c_4(n)}$$

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

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

Note that in view of Basu's theorem any location-invariant estimator is independent of $\widehat{\mu}$ (cf. Lehmann (1986) page 191 Example 1). A general location-invariant estimator of σ is denoted by σ^* and particular examples are $\widehat{\sigma}$ and $\breve{\sigma}$.

Plugging in the estimators $\widehat{\mu}$ and σ^* in the UCL leads to

$$\widehat{UCL} = \widehat{\mu} + u_p\, \sigma^*. \tag{2.1}$$

Our interest is to observe the probability of an incorrect signal that the process should be out-of-control. This probability depends on the estimators $\widehat{\mu}$ and σ^* and therefore it is a r.v. It is given by

$$P_n = P_n(\widehat{\mu}, \sigma^*) = P(X_{n+1} > \widehat{\mu} + u_p\, \sigma^*) = \overline{\Phi}(\frac{\widehat{\mu} - \mu}{\sigma} + u_p \frac{\sigma^*}{\sigma}), \tag{2.2}$$

where $\overline{\Phi}(x) = 1 - \Phi(x)$.

2.2.2 The effect of the estimation error for the bias

In order to learn more about the performance characteristics of the control chart, several aspects of the closeness of P_n to p can be investigated. The first choice is based on the bias of P_n with respect to p. Also functions of P_n are of interest. For example, 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 $E(1/P_n)$ to $1/p$.

Another family of functions of interest are the probabilities that the run length is at most some specified value k. 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 then compared with $1 - (1 - p)^k$. Typical values of interest for k are fractions of the ARL, i.e. $k = [\delta/p]$ with, for instance, $\delta = \frac{1}{10}, \frac{1}{4}, \frac{1}{2}$ or 1. With these given values of δ, $k = 100, 250, 500$ or 1000, respectively. These values are commonly used since they are feasible from the practical point of view.

More generally, focusing on the bias 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

$$g(p) = p,\; g(p) = \frac{1}{p},\; g(p) = 1 - (1 - p)^k \tag{2.3}$$

are of interest. Other functions g can be treated in a similar way. For instance, the standard deviation of the run length, corresponding to $g(p) = \sqrt{1 - p}/p$. As a criterion for closeness of $Eg(P_n)$ to $g(p)$ we restrict the relative error to be at most 10%, in formula :

$$\left| \frac{Eg(P_n) - g(p)}{g(p)} \right| \leq 10\%. \tag{2.4}$$

We are looking for the smallest n for which (2.4) holds. Of course, other values than 10% can be chosen if desired. This criterion is suitable for $g(p) = p$ and

$g(p) = 1/p$, while for $g(p) = 1 - (1-p)^k$ an absolute error of at most 1% seems to be more appropriate since it deals with probabilities not close to zero.

A simulation study using $\widehat{\sigma}$ as the estimator of σ, is carried out to see the performance of $Eg(P_n)$ for the functions given in (2.3). The number of repetitions in the simulation study equals 100,000. The results are given for $p = 0.001$ in Table 2.1 and for $p = 0.01$ in Table 2.2.

Table 2.1 Simulation results for $p = 0.001$.

	Relative Error (%)		Absolute Error (%) for $g(p) = 1 - (1-p)^k$			
n	$g(p) = p$	$g(p) = 1/p$	$k = 1000$	$k = 500$	$k = 250$	$k = 100$
25	167.0	702.0	3.56	6.65	10.53	8.70
50	72.9	132.8	2.28	4.71	6.63	4.77
75	45.0	70.5	1.65	3.62	4.75	3.19
100	32.6	47.2	1.29	2.92	3.68	2.39
150	21.5	28.5	0.81	2.23	2.63	1.63
200	15.8	20.0	0.57	1.78	2.02	1.22
250	12.5	15.5	0.44	1.47	1.63	0.98
300	10.4	12.8	0.36	1.26	1.38	0.82
350	8.8	10.9	0.33	1.07	1.17	0.69
400	7.7	9.5	0.29	0.95	1.03	0.61
450	6.9	8.2	0.22	0.88	0.94	0.55
500	6.2	7.5	0.22	0.79	0.84	0.49

From Table 2.1 we can see that the relative error of EP_n is larger than 70% if $n = 50$, and it is still larger than 30% if n is increased to 100. The simulations give that the relative error of EP_n satisfies (2.4) for $n \geq 312$. In line with this, the relative error of $E(1/P_n)$ is larger than 40% if $n = 100$. This error is smaller than 10% only for $n \geq 377$. Moreover, $ER_{n,k}$ satisfies the requirements of having an absolute error smaller than 1% for $n \geq 124, 381, 419, 243$, when $k = 1000, 500, 250, 100$ respectively.

Woodall and Montgomery (1999) recommend to have at least 20-25 samples of size 4-5 each to base the estimators on. 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 associate a sample of size 4-5 with one observation in our set-up, the recommended number of observations equals 20-25; which is far too small to get an accurate estimate according to the simulation result. Even if we would 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 evidently confirm that 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).

Allowing p to be a little bit larger, with a value 0.01 instead of 0.001, the situation becomes slightly better, as shown in Table 2.2, although rather large sample sizes are still needed. Moreover, in control charts $p = 0.001$ is far more often applied than $p = 0.01$.

Table 2.2 Simulation results for $p = 0.01$.

	Relative Error (%)		Absolute Error (%) for $g(p) = 1 - (1-p)^k$			
n	$g(p) = p$	$g(p) = 1/p$	$k = 100$	$k = 50$	$k = 25$	$k = 10$
10	139.6	1778.0	4.16	6.25	10.18	8.29
25	50.6	112.0	2.21	3.89	5.33	3.66
50	24.0	40.0	1.30	2.32	2.91	1.85
75	16.0	23.0	0.80	1.74	2.05	1.26
100	11.9	17.0	0.63	1.34	1.56	0.94
125	9.6	13.0	0.47	1.14	1.28	0.77
150	7.9	11.0	0.42	0.93	1.06	0.63
175	6.6	9.0	0.35	0.80	0.90	0.53
200	5.8	8.0	0.33	0.70	0.79	0.47
225	5.1	7.0	0.30	0.62	0.70	0.42
250	4.7	6.0	0.26	0.58	0.64	0.38

It is seen, cf. Table 2.2 , that EP_n has a relative error smaller than 10% for $n \geq 118$, and that $E(1/P_n)$ satisfies (2.4) for $n \geq 163$, while $ER_{n,k}$ satisfies the requirements to have an absolute error smaller than 1% for $n \geq 59, 139, 164, 95$, when $k = 100, 50, 25, 10$ respectively.

Remark 2.2.1 The simulation results show that EP_n is always larger than p, which means that P_n is positively biased. In this case, as a consequence 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, that is $E(1/P_n) > 1/p$. 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 to replace the average by the median. This strategy is often successfully applied in robust statistics. The median of a geometric distribution with parameter p is given by the function

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

However, for a 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. Noting that we want to protect a process against small run lengths, when we are in the in-control situation, the benefit obtained by estimation of the parameters for $ER_{n,1000}$ is therefore of far less importance than the disadvantage that resulted for $ER_{n,k}$, with $k = 100, 250, 500$.

The positive bias of EP_n and $ER_{n,k}$ for $k = 100, 250$ and 500 is 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 is profitable. An explanation for the positive bias of $E(1/P_n)$ was given above. However, this explanation (some very high values of $1/P_n$ cause high expectation) can not be used for $ER_{n,k}$, since $R_{n,k}$ is bounded by 1. Therefore, to explain the positive bias of $ER_{n,k}$ we use Jensen's inequality later in this section. □

In order to develop correction terms such that accurate control charts are obtained for smaller sample sizes, as well as to understand the reason that more observations are needed here than were recommended in the earlier days, we need to have more analytic insight in the problem. Such insight can be provided by asymptotic methods. More transparent approximations which are used in many statistical problems are capable of showing 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, which is for a great part based on asymptotic normality of the estimators. However since, for some criteria, the first type approximations can not be easily applied, a second asymptotic approach is considered as well.

Although n plays a very important role in the standard asymptotic method, since it always deals with a very large number, other variables involved in the model may also influence the final result significantly, being large as well. Therefore their effects should not be ignored. For example, a commonly used value for u_p is 3, hence u_p^4 equals 81, which may be of the same order of magnitude as n for several situations considered here. Hence, we follow a more delicate way, taking into account not only n, but also u_p in the asymptotical approach. For simplification of notation, u is used instead of u_p in the rest of the present chapter.

(i) *First type asymptotics*

We start with a theorem which presents the limiting distribution of P_n, and for that we put the following condition on the estimator σ^* from Section 2.2.1.

Condition 1 *The estimator σ^* satisfies*

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

This condition holds for $\sigma^* = \check{\sigma}$ as well as for $\sigma^* = \widehat{\sigma}$. Next, we define

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

so that 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 2.2.1 *Suppose that Condition 1 holds, that $u \geq 1$ and that $u = o(n^{1/2})$ as $n \to \infty$, then*

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

in the sense that

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

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

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

In view of the definition of $\widehat{\mu}$ and Condition 1 we have

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

Direct calculation gives

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

On the set $B = \left\{u \geq 1, |\Delta(u)| \leq \frac{1}{2}\right\}$ 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

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

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

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

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 2.2.2 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 the estimating σ is in the case of $u = 3$ a factor $u^2/2 = 4.5$ larger than the contribution due to estimation of μ. □

To approximate $Eg(P_n)$, in particular when g is one of the functions given in (2.3), Theorem 2.2.1 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 $g(p) = p$, 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 $g(p) = \dfrac{1}{p}$, 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 2.2.3 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, cf. Remark 2.2.1 and Remark 2.2.4 below. □

Finally, for $g(p) = 1-(1-p)^k$ the first type approximation leads to

$$E\left\{1-(1-P_n)^k\right\} \approx E\left\{1-(1-p\,Y_n)^k\right\}$$

$$= \sum_{j=1}^{k}\binom{k}{j}(-1)^{j+1}p^j\exp\left\{\frac{1}{2}j^2\left(\frac{u}{z(u)}\right)^2\left(\frac{1}{n}+\frac{u^2}{2(n-1)}\right)\right\}. \tag{2.12}$$

The first type approximations of EP_n and $E(1/P_n)$ are easily applied, but the first type approximation of $ER_{n,k}$ is more complicated. Therefore, in the following subsection we take a slightly different and more direct approach focusing on 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 $\Delta^2(u)$.

Here, we are going to find an approximation of $Eg(P_n)$, in particular when g is one of the functions given in (2.3). Since according to (2.2)

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

we use the notation $h(u) = g(\overline{\Phi}(u))$ to investigate the second type approximation of $Eg(P_n) = Eh(u+\Delta(u))$ using the two-step Taylor expansion as follows:

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

The following table presents the derivatives of h for g given in (2.3).

Table 2.3 Derivatives of h for the functions g given in (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 2.2.4 In Remark 2.2.1 an argument has been given to justify the occurrence of the positive bias of both EP_n and $E(1/P_n)$. The argument could not be applied on $ER_{n,k}$. In addition, the positive bias of EP_n, $E(1/P_n)$, and $ER_{n,k}$ for $k = 100$, 250, 500, and the negative bias of $ER_{n,1000}$ can be explained using Jensen's inequality as follows. 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 $u + \Delta(u)$ gives high probability to a neighborhood of 3. If $g(p) = p$ or $g(p) = 1/p$, then from Table 2.3 we have $h''(u) > 0$ for $u > 0$, and hence in both cases the function h is convex for $u > 0$. Therefore, Jensen's inequality strongly suggests that $Eh(u + \Delta(u)) > h(u + E\Delta(u)) = h(u)$, implying $EP_n > p$ and $E(1/P_n) > 1/p$.

If $g(p) = 1 - (1 - p)^k$, where $k = 100,\ 250, \text{and } 500$, we can see from Table 2.3 that $h''(u) > 0$ for $u > 2.375$, $u > 2.689$, and $u > 2.908$, respectively. Hence, the function h is convex for these combinations of $u-$ and $k-$values. Therefore, Jensen's inequality suggests that $ER_{n,k} = Eh(u+\Delta(u)) > h(u+E\Delta(u)) = h(u) = 1 - (1 - \overline{\Phi}(u))^k = 1 - (1 - p)^k$, for $k = 100,\ 250, \text{and } 500$.

For $g(p) = 1-(1-p)^{1000}$, 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 2.1. □

The following theorem gives an upper bound for the (relative) error of the approximation. We impose Conditions 2 and 3 on the estimator σ^*. It can be shown that the two conditions hold for $\sigma^* = \check{\sigma}$ and $\sigma^* = \widehat{\sigma}$.

Condition 2 *The estimator σ^* satisfies*

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

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

$$E\,|\Delta(u)|^9 \exp\left\{u\,|\Delta(u)| + \frac{1}{2}\Delta^2(u)\right\} = O(u^9 n^{-9/2}),$$

$$for\ u \geq 1\ with\ u = O(n^{1/4})\ as\ n \to \infty.$$

Condition 3 *The estimator σ^* satisfies,*

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

$$for\ u \geq 1\quad with\quad u = O(n^{1/4})\ as\ n \to \infty.$$

Theorem 2.2.2 *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 1 holds and that for some constants $c_2 > 0$ and $c_3 > 0$

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

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 2 holds and that h''' and h^{iv} are bounded, then

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

Proof.

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

$$\begin{aligned} Eh(u+\Delta(u)) &= h(u) + h'(u)E\Delta(u) + \frac{1}{2}h''(u)E\Delta^2(u) \\ &+ \frac{1}{6}h'''(u)E\Delta^3(u) + \frac{1}{24}Eh^{iv}(u+\eta\Delta(u))\Delta^4(u), \end{aligned}$$

and hence, by (2.14) and Condition 2, we have

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

$$\leq \frac{1}{6}c_2u^3\left|E\Delta^3(u)\right| + \frac{1}{24}c_3E\left\{\left(u^4+|\Delta(u)|^5\right)|\Delta(u)|^4\exp\left\{u\,|\Delta(u)|+\frac{1}{2}\Delta^2(u)\right\}\right\}$$
$$= O\left(u^8n^{-2}\right).$$

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

$$\begin{aligned} Eh(u+\Delta(u)) \quad &= h(u) + h'(u)E\Delta(u) + \frac{1}{2}h''(u)E\Delta^2(u) \\ &+\frac{1}{6}h'''(u)E\Delta^3(u) + \frac{1}{24}Eh^{iv}(u+\eta\Delta(u))\Delta^4(u), \end{aligned}$$

and hence, since h''' and h^{iv} are bounded and Condition 3 holds,

$$\left|Eh(u+\Delta(u)) - h(u) - \left\{h'(u)E\Delta(u) + \frac{1}{2}h''(u)E\Delta^2(u)\right\}\right|$$
$$\leq c_4\left\{\left|E\Delta^3(u)\right| + E\Delta^4(u)\right\} = O\left(u^4n^{-2}\right).$$

■

Using
$$\overline{\Phi}(u) = \frac{\varphi(u)}{u}\left(1+o(1)\right) \text{ as } u\to\infty,$$
it can be shown 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 $O(u^8n^{-2})$ 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 is $O(u^4n^{-2})$.

Further calculation gives

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

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

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

Writing $E\Delta^2(u) = a(u,n)$ in (2.15), we get for $\sigma^* = \check{\sigma}$

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

where $c_4(n)$ is as given in Section 2.2.1. Inserting the derivatives of h, given in Table 2.3 and (2.15) in (2.13) leads to the following approximations for the estimator $\sigma^* = \widehat{\sigma}$:

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

From (2.17) we can see that the second type approximations also give positive bias for P_n and $1/P_n$ as they should do.

A further simplification can be obtained by realizing that

$$\frac{\Gamma(\frac{n+1}{2})\Gamma(\frac{n-1}{2})}{\Gamma(\frac{n}{2})^2} \approx 1 + \frac{1}{2n}. \qquad (2.18)$$

Using this approximation we get for $\sigma^* = \widehat{\sigma}$

$$\begin{aligned}
EP_n &\approx p + \frac{1}{2}\frac{u^2+2}{2n}u\varphi(u) \\
E(\frac{1}{P_n}) &\approx \frac{1}{p} + \frac{1}{2}\frac{u^2+2}{2n}\left\{\frac{2\varphi^2(u)}{p^3} - \frac{u\varphi(u)}{p^2}\right\} \\
ER_{n,k} &\approx 1-(1-p)^k + \frac{1}{2}\frac{u^2+2}{2n} \\
&\quad \left\{-k(k-1)\varphi^2(u)(1-p)^{k-2} + ku\varphi(u)(1-p)^{k-1}\right\}. \qquad (2.19)
\end{aligned}$$

Remark 2.2.5 In a similar way, using (2.16), approximations of EP_n, $E(1/P_n)$ and $ER_{n,k}$ can be evaluated when $\sigma^* = \check{\sigma}$. □

The following tables compare the approximation results with the simulation results for the first two functions given in (2.3). From these tables we can see that both the first type approximation and the second type approximation represent EP_n and $E(1/P_n)$ very well.

According to the first type approximation and the second type approximation, EP_n has already a relative error smaller than 10% for $n \geq 345$ and $n \geq 302$, respectively. This seems in line with the result of the simulation which gives $n \geq 312$, cf. Table 2.4. Similarly, we get for $g(p) = 1/p$ that according to the first type approximation and the second type approximation, $E(1/P_n)$ has a relative error smaller than 10% for $n \geq 345$ and $n \geq 356$, respectively. This agrees very well with the result of the simulation. The table shows that both type approximations are very accurate.

Table 2.4 The comparison of the relative error of $Eg(P_n)$ for $g(p) = p$ and $g(p) = 1/p$ with $p = 0.001$.

$g(p)$	p			$1/p$		
n	simulation	1^{st}-app	2^{nd}-app	simulation	1^{st}-app	2^{nd}-app
25	167.0	287.5	125.4	702.0	287.5	147.8
50	72.9	94.6	61.4	132.8	94.6	72.4
75	45.0	55.5	40.6	70.5	55.5	47.9
100	32.6	39.1	30.4	47.2	39.1	35.8
150	21.5	24.5	20.2	28.5	24.5	23.8
200	15.8	17.9	15.1	20.0	17.9	17.8
250	12.5	14.0	12.1	15.5	14.0	14.2
300	10.4	11.6	10.0	12.8	11.6	11.8
350	8.8	9.8	8.6	10.9	9.8	10.2
400	7.7	8.5	7.5	9.5	8.5	8.9
450	6.9	7.6	6.7	8.2	7.6	7.9
500	6.2	6.8	6.0	7.5	6.8	7.1

Table 2.5 The comparison of the absolute error of $ER_{n,k}$ for $p = 0.001$.

	$k = 1000$		$k = 500$		$k = 250$		$k = 100$	
n	simul.	appr.	simul.	appr.	simul.	appr.	simul.	appr.
25	3.56	4.13	6.65	17.34	10.53	17.80	8.70	10.13
50	2.28	2.02	4.71	8.49	6.63	8.71	4.77	4.96
75	1.65	1.34	3.62	5.62	4.75	5.77	3.19	3.28
100	1.29	1.00	2.92	4.20	3.68	4.31	2.39	2.45
150	0.81	0.66	2.23	2.79	2.63	2.86	1.63	1.63
200	0.57	0.50	1.78	2.09	2.02	2.14	1.22	1.22
250	0.44	0.40	1.47	1.67	1.63	1.71	0.98	0.98
300	0.36	0.33	1.26	1.39	1.38	1.43	0.82	0.81
350	0.33	0.28	1.07	1.19	1.17	1.22	0.69	0.70
400	0.29	0.25	0.95	1.04	1.03	1.07	0.61	0.61
450	0.22	0.22	0.88	0.93	0.94	0.95	0.55	0.54
500	0.22	0.22	0.79	0.83	0.84	0.85	0.49	0.49

Since the first type approximation for the third function is rather cumbersome, it is not carried out further. However, the comparison of this function using the second type approximation is presented in Table 2.5. It is seen from Table 2.5 that also for $ER_{n,k}$ the second type approximation is very accurate, in the sense that the results of this approximation agree very well with the results of the simulation.

Table 2.6 compares the simulation results with the approximation results in terms of the sample sizes needed to get the required accuracy, given by (2.4) for $g(p) = p$ and $g(p) = 1/p$ and by an absolute error smaller than 1% for $g(p) = 1 - (1 - p)^k$. This table shows that the two types approximations are highly accurate for representing the expectation of the performance characteristics P_n, $1/P_n$ and $R_{n,k}$, with $k = 100,\ 250,\ 500,\ 1000$.

Table 2.6 Comparison of the first and second type approximation of the needed sample size with simulation results.

	1^{st} type approx	2^{nd} type approx	simulation
EP_n	345	302	312
$E(1/P_n)$	345	356	377
$ER_{n,1000}$	-	101	124
$ER_{n,500}$	-	417	381
$ER_{n,250}$	-	428	419
$ER_{n,100}$	-	244	243

2.2.3 The effect of the estimation error for the exceedance probability

So far, we have only used the expectation as a criterion to investigate the performance characteristics P_n, $1/P_n$ and $R_{n,k}$ (with $k = 100,\ 250,\ 500,\ 1000$) of the control chart. This criterion, however, only captures part of the picture and is merely useful for the long term behavior of the chart over a series of separate applications. In the case that we also want to control the probability that errors beyond a certain level occur in a single application, it is more interesting to investigate the following exceedance probability criterion:

$$P\left(\frac{g(P_n) - g(p)}{g(p)} > \epsilon\right), \tag{2.20}$$

where g is an increasing function, and

$$P\left(\frac{g(P_n) - g(p)}{g(p)} < -\epsilon\right), \tag{2.21}$$

if g is a decreasing function, and where in both cases ϵ is a small non-negative number.

The exceedance probability criterion can be considered as more ambitious than the bias criterion, since controlling each separate case is harder than controlling the average long term behavior. An illustration of this fact is given in what follows, where g is tacitly assumed to be positive . Using P_n in (2.2) with $g(p)$ is an increasing function, the exceedance probability in (2.20) can be written as:

$$P\left(\frac{\widehat{\mu} - \mu}{\sigma} + u\frac{\sigma^*}{\sigma} < \overline{\Phi}^{-1}\left\{g^{-1}(g(p)(1+\varepsilon))\right\}\right), \tag{2.22}$$

whereas, if $g(p)$ is a decreasing function, (2.21) can be written as:

$$P\left(\frac{\widehat{\mu}-\mu}{\sigma}+u\frac{\sigma^*}{\sigma}<\overline{\Phi}^{-1}\left\{g^{-1}(g(p)(1-\varepsilon))\right\}\right). \tag{2.23}$$

Using $\sigma^* = \breve{\sigma}$, expressions (2.22) and (2.23), subsequently can be written in the form

$$P\left(\frac{Z}{n^{1/2}}+u\frac{S}{\sigma}<b\right), \tag{2.24}$$

where the r.v. Z has a standard normal distribution function and is independent from S. In this case, of course, b depends on g and ϵ. For example, referring to the functions given in (2.3), b equals $\overline{\Phi}^{-1}(p(1+\varepsilon))$, $\overline{\Phi}^{-1}(p/(1-\varepsilon)$, and $\overline{\Phi}^{-1}\left(1-\left[1-\left\{1-(1-p)^k\right\}(1+\varepsilon)\right]^{1/k}\right)$ for $g(p)=p$, $g(p)=1/p$ and $g(p)=1-(1-p)^k$, respectively.

Modifying the expression (2.24) a little bit further, we get

$$P\left(-\frac{Z}{n^{1/2}}-u\frac{S}{\sigma}>-b\right)=P\left(\frac{Z}{n^{1/2}}+b>u\frac{S}{\sigma}\right)=P\left(\frac{Z+bn^{1/2}}{S/\sigma}>un^{1/2}\right)$$
$$=\overline{G}_{n-1,bn^{1/2}}(un^{1/2})=1-G_{n-1,bn^{1/2}}(un^{1/2}), \tag{2.25}$$

where $G_{n-1,\delta}$ stands for the distribution function of the noncentral t-distribution with $n-1$ degrees of freedom and noncentrality parameter δ (cf. Ghosh, Reynolds and Hui (1981) for earlier use of the non-central t in this connection). Of course, in case we use $\sigma^* = \widehat{\sigma}$, we need to change u to $u/(c_4(n))$.

Using (2.25) and (2.20) we can evaluate the effect of the estimation error using the exceedance probability criterion. The following equation which is the realized value of the exceedance probability shows how large the effect is of the estimation error if we apply no correction term on the control limits.

$$\tilde{\alpha}=\overline{G}_{n-1,n^{1/2}b}(n^{1/2}u), \tag{2.26}$$

To see the effect under the three different functions given in (2.3) we use $\epsilon = 0.1$ and $p = 0.001$ and present the realized value of the exceedance probability in the following table.

Recall that in this table, b not only depends on ϵ, but also on g. For example, if we use $\epsilon = 0.1$ and $p = 0.001$ we get for $g(p) = p$, $b = \overline{\Phi}^{-1}(p(1+\varepsilon)) = 3.062$, and for $g(p) = 1/p$, $b = \overline{\Phi}^{-1}(p/(1-\varepsilon) = 3.059$. For $g(p) = 1-(1-p)^k$, $b = \overline{\Phi}^{-1}\left(1-\left[1-\left\{1-(1-p)^k\right\}(1+\varepsilon)\right]^{1/k}\right)$, which for $k = 100,\ 250,\ 500$ and 1000, respectively equals 3.123, 3.125, 3.130 and 3.141.

Table 2.7 The realized value of the exceedance probability as in (2.26) with $\epsilon = 0.1$ and $p = 0.001$

$g(p)$	p	$1/p$	$1-(1-p)^k$			
			$k = 100$	$k = 250$	$k = 500$	$k = 1000$
$n = 25$	0.5104	0.5080	0.5092	0.5071	0.5029	0.4914
$n = 50$	0.4903	0.4868	0.4885	0.4856	0.4797	0.4633
$n = 75$	0.4784	0.4741	0.4762	0.4725	0.4653	0.4452
$n = 100$	0.4695	0.4645	0.4669	0.4627	0.4544	0.4313
$n = 150$	0.4558	0.4498	0.4527	0.4475	0.4374	0.4093
$n = 200$	0.4451	0.4381	0.4416	0.4356	0.4239	0.3917
$n = 250$	0.4361	0.4283	0.4321	0.4255	0.4125	0.3768
$n = 500$	0.4028	0.3920	0.3973	0.3880	0.3702	0.3218
$n = 1000$	0.3590	0.3442	0.3514	0.3389	0.3150	0.2521
$n = 2000$	0.3015	0.2822	0.2916	0.2753	0.2449	0.1701
$n = 5000$	0.2029	0.1788	0.1903	0.1704	0.1354	0.0646

Comparing Table 2.7 with Table 2.1, we can see that the exceedance probability criterion imposes a stronger requirement than does the bias criterion. As shown in Table 2.1, for example, with $g(p) = p$ where $p = 0.001$, the bias criterion already satisfies the requirement $|EP_n - p|/p < 0.1$, for $n \geq 345$, whereas with much larger n, say $n = 500$ the exceedance probability criterion produces $P((P_n - p)/p > 0.1) = 0.4$, as shown in Table 2.7. The result shows that the new criterion needs much larger sample sizes to reduce the probability to reasonable levels like 0.1 or 0.2.

Also for the case $g(p) = 1/p$, already for $n \geq 500$, the first criterion gives a relative error smaller than 0.1, while the second criterion results in a quite high value of 0.39. The same situation holds for $g(p) = 1 - (1-p)^k$, where $k = 100,\ 250,\ 500,$ and 1000. Also in this case, the exceedance probability criterion needs much larger sample sizes to reduce the probability to reasonable levels like 0.1 or 0.2.

To end the present section, we observe that the estimation error leads to an inaccurate control limit. This can be fixed by either enlarging the sample sizes or by developing a correction term for the control limits. The second option will be the main topic to be discussed in the following section.

2.3 Corrected normal chart

It has been shown in the previous section that for observations having a normal distribution with unknown parameters μ and σ, we have to use a large number of observations to get an accurate control limit. To reduce the sample size needed to get the accurate control limit, we modify the initial control limit given in (2.1) by adding a correction term. Hence, the corrected control limit will be somewhat larger or smaller than the initial control limit. To derive the correction term we apply bias and exceedance probability criteria as follows.

The first criterion is based on the bias of EP_n with respect to p, or in a more general form $Eg(P_n)$ with respect to $g(p)$: require e.g. that the relative error is at most 10%, that is $|Eg(P_n) - g(p)|/g(p) \leq 0.1$. Hence, for $p = 0.001$ the expected false alarm rate in the estimated chart should stay between 0.0009 and 0.0011. Similarly, $E(1/P_n)$ can be required to lie between 900 and 1100. The exceedance probability criterion, on the other hand, intends to restrict $P((P_n - p)/p) > \epsilon)$, where ϵ is a small non-negative number, to be smaller than a specified small positive number (α). This probability indicates how likely it is that Phase I observations produce an outcome of, e.g. if $\epsilon = 0.1$, P_n larger than 0.0011 when $p = 0.001$. As in the bias criterion, the exceedance probability criterion can also be extended to a more general function $g(p)$, hence producing $P((g(P_n) - g(p))/g(p) > \epsilon) < \alpha$ if g is an increasing function, and $P((g(P_n) - g(p))/g(p) < -\epsilon) < \alpha$ if g is a decreasing function.

The following subsections present the derivation of the correction terms using each criterion separately.

2.3.1 Bias criterion

It has been demonstrated in Section 2.2.2., by simulation as well as by asymptotic theory, that huge sample sizes are needed to get accurate control chart limits when estimators are simply plugged in. In order to get an accurate control limit for commonly used sample sizes 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 $E(P_n) = p$, or at least close to it (of course, we may generalize p to $g(p)$ with its stochastic counterpart $g(P_n)$). In other words, we do not use u, but use u_q for an appropriate value of q. Instead of u_q, we write $u + c$ with c being the correction term, which depends on g. Using this scenario, the corrected UCL becomes

$$\widehat{UCL}_c = \widehat{\mu} + (u + c)\sigma^*, \tag{2.27}$$

and the corresponding P_n, cf. (2.2), becomes

$$P_n = P(X_{n+1} > \widehat{\mu} + (u + c)\sigma^*) = \overline{\Phi}\left(\frac{\widehat{\mu} - \mu}{\sigma} + (u + c)\frac{\sigma^*}{\sigma}\right). \tag{2.28}$$

Indeed, when $g(p) = p$, simulation for the case $c = 0$ shows that $EP_n > p$ and hence, we need a larger UCL or $c > 0$. When $g(p) = 1/p$, simulation for the case

$c = 0$ shows that $E(1/P_n) > 1/p$ and hence P_n should be larger in order to get $E(1/P_n) = 1/p$, that is, we need a smaller ULC or $c < 0$. When, for instance, looking at EP_n, which is larger than p, one might argue that the corrected UCL could be obtained by simply adding $\tilde{c}$ to it. Hence, the new UCL equals $\widehat{\mu} + u\sigma^* + \tilde{c}$. However, the smaller σ, the smaller such $\tilde{c}$ should be. Accordingly, 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 in (2.27).

Remark 2.3.1 The replacement of the estimator $\breve{\sigma}$ in the control limit by $\widehat{\sigma}$ is meant to get an unbiased estimator of σ. However, in fact the problem is not to get an unbiased estimator of σ, but to have an unbiased estimator of $g(p)$. Since $g(p)$ is a nonlinear function of σ, it is not enough to replace S by $\widehat{\sigma}$. Apart from the correction factor $c_4(n)$, a further correction factor $\breve{c}$, say, is needed. Instead of writing $u\,\breve{c}\,(S/c_4(n)) = u\,\breve{c}\,\widehat{\sigma}$, we may also write $(u + c)\widehat{\sigma}$, thus obtaining the form mentioned before. Notice that it does not matter whether we start with $\breve{\sigma}$ or with $\widehat{\sigma}$. After the correction we arrive at 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. The asymptotic methods developed in the previous section give an answer to this question.

We start with the first type approximation given in (2.10). Replacing u by $u + c$ 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

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

and hence

$$c = \frac{u(u^2+2)}{4n\,z(u)}. \tag{2.29}$$

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

$$c = -\frac{u(u^2+2)}{4n\,z(u)}, \tag{2.30}$$

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 whereas the correction term for $g(p) = 1/p$ is negative.

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

$$\begin{aligned} Eh(u+c+\Delta(u+c)) \quad &\approx \quad h(u+c)+h'(u+c)E\Delta(u+c) \\ &+\frac{1}{2}h''(u+c)E\Delta^2(u+c). \end{aligned}$$

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

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

Hence, we get

$$c=-E\Delta(u)-\frac{1}{2}\frac{h''(u)}{h'(u)}E\Delta^2(u). \tag{2.31}$$

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

$$c=-\frac{(u^2+2)}{4n}\frac{h''(u)}{h'(u)}. \tag{2.32}$$

In particular, we get

$$\begin{aligned} c= &\quad \frac{u(u^2+2)}{4n} && \text{if } g(p)=p \\ c= &\quad \frac{u^2+2}{4n}\left\{u-\frac{2\varphi(u)}{p}\right\} && \text{if } g(p)=\frac{1}{p} \\ c= &\quad \frac{u^2+2}{4n}\left\{u-\frac{(k-1)\varphi(u)}{1-p}\right\} && \text{if } g(p)=1-(1-p)^k. \end{aligned} \tag{2.33}$$

Since, for large u,

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

the correction term for $g(p)=1/p$ approximately equals minus the correction term of $g(p)$; 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 (2.29) and (2.30), because $z(u)=1+O(u^{-2})$ as $u\to\infty$.

Since we always take $k<(2/p)-1$, the correction term for $g(p)=1-(1-p)^k$ is smaller than that of $g(p)=p$ and larger than that of $g(p)=1/p$.

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

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

Since

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

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

$$c = \sqrt{1 + \frac{1}{n}} c_4(n) t_{n-1;p} - u, \quad \text{with } P(T_{n-1} > t_{n-1;p}) = p. \tag{2.34}$$

Remark 2.3.2 If $\sigma^* = \check{\sigma}$, 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. □

Remark 2.3.3 The control chart with the exact correction, given in (2.34), can already be found on page 1806 of Ghosh, Reynolds and Van Hui (1981). It corresponds also to the so called Q-chart, presented by Quesenberry (1991), when applying (7) on page 215 of that paper with $r = n + 1$. The idea of using an exact correction term for the control limits was introduced for the first time by Hillier (1969). This correction term is based on a small number of subgroups with the average range as an estimator of σ. Yang and Hillier (1970) proposed a similar correction term using average subgroup variance and sample variance as estimators of σ^2. Roes, Does and Schurink (1993) present exact corrections for control charts with several other estimators as well. □

Remark 2.3.4 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 Subsection 2.2.2.

The bias of $ER_{n,k}$ is slightly negative, while $g(p) = 1 - (1 - p)^k$ is increasing. A reason for it has been given in Remark 2.2.4. As a consequence of the negative bias, the corresponding correction term is slightly negative. □

The following table provides the various correction terms for the functions g given in (2.3).

Table 2.8 Correction terms for $Eg(P_n)$ with $g(p) = p$, $g(p) = 1/p$ and $g(p) = 1-(1-p)^k$ and $p = 0.001$.

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

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

Table 2.9 Simulation results of $Eg(P_n)$ after correction.

	Relative Error (%)				Absolute Error (%)			
n	first type		second type		second type			
	EP_n	$E(1/P_n)$	EP_n	$E(1/P_n)$	$ER_{n,1000}$	$ER_{n,500}$	$ER_{n,250}$	$ER_{n,100}$
10	60.4	66.3	77.1	72.1	2.2	7.2	3.7	0.07
20	8.8	0.7	19.6	15.2	1.7	4.1	2.0	0.02
30	1.9	2.9	9.0	8.3	1.4	2.8	1.3	0.07
40	0.6	4.1	4.9	4.1	1.0	2.1	0.9	0.07
50	1.1	2.5	4.1	3.9	1.0	1.3	0.7	0.05
60	1.5	4.4	2.7	1.2	0.9	1.1	0.5	0.04
70	1.7	3.0	2.4	1.2	0.6	0.8	0.4	0.00
80	1.5	2.7	1.4	1.0	0.6	0.7	0.3	0.02
90	1.5	2.3	1.2	0.7	0.5	0.5	0.2	0.02
100	1.5	2.0	0.9	0.7	0.4	0.5	0.2	0.00

The effect of the correction terms is checked by simulation of $Eg(P_n)$ with the corrected UCL. Table 2.9 gives the results. This table shows that the correction terms work very well. It was shown in Subsection 2.2.2 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 Subsection 2.2.2, 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 2.10 The Required Sample Sizes

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

Comparison with Table 2.6 shows that the correction terms indeed do their job: the very large sample sizes needed in Subsection 2.2.2 are reduced to common sample sizes, usually available in practice. For $g(p) = p$ and $g(p) = 1/p$, the required sample sizes are reduced from more than 300 to less than 30, while for $g(p) = 1 - (1 - p)^k$, with $k = 1000,\ 500,\ 250,$ and 100, respectively, the sample sizes are reduced from more than 100 to less than 45, from more than 400 to only 65, and from more than 400 to less than 40, and from more than 200 to only 10.

Applying $\widehat{\sigma}$ as estimator of σ, we therefore propose the following UCL's for the bias corrected control charts $\widehat{\mu} + (u + c)\widehat{\sigma}$:

- with for the first type approximation : if $g(p) = p$ then c is given by (2.29), and if $g(p) = 1/p$ then c is given by (2.30)
- with for the second type approximation : if $g(p)$ is any of the three functions mentioned in (2.3) then c is given by (2.33).
- with exact correction term c given by (2.34) if $g(p) = p$.

To show how these correction terms for controlling the bias work, we present the application of the various control limits in a real data example.

Example 2.3.1

Data on charge weights of an insecticide dispenser are taken from Table 7.1 on page 130 of Burr (1979). The first 25 data out of a total of 50 are supposed to be a preliminary run. The data are as follows.

472	462	458	476	462	463	464	461	450	479	453	467	458
431	454	476	476	456	498	448	453	470	474	461	467	
444	455	455	439	445	456	475	454	452	467	456	466	452
440	451	464	462	470	459	466	476	473	466	480	454	

Normal $Q-Q$ plots of the first 25 data and of all data show that normality of the data is reasonable assumption (p-values of the Anderson-Darling normality test are 0.47 and 0.67, respectively). For the first 25 observations we get $\bar{x} = 463.56$ and $s = 13.026$, while $c_4(25) = 0.9896$. Taking $p = 0.001$, the UCL without correction based on the first 25 observations gives: 504.23. Hence, none of the next 25 observations gives a signal that the process is out-of-control.

Applying the UCL without correction results in the following expected probability of an incorrect signal (here $\bar{X}_{25}$ and S_{25} are the sample mean and sample standard deviation based on the first 25 observations)

$$P\left(X_{26} > \bar{X}_{25} + \frac{u_{0.001}\,S_{25}}{c_4(25)}\right) = P\left(T_{24} > \frac{u_{0.001}}{c_4(25)\sqrt{1+\frac{1}{25}}}\right) = 0.0023$$

being more than twice as much as it should be. So, indeed a large error is introduced, when dealing with the UCL without correction.

The exact correction term, given by (2.34), when $g(p) = p$ equals 0.4086. Therefore, the UCL in this case equals 509.61.

Similarly, we can calculate the other UCL's with the correction terms for the configurations as given in Table 2.8. This leads to the following UCL's.

p			$1/p$	
exact	1st type	2nd type	1st type	2nd type
509.61	509.35	508.93	499.12	498.69

$1-(1-p)^{1000}$	$1-(1-p)^{500}$	$1-(1-p)^{250}$	$1-(1-p)^{100}$
1st type	2nd type	1st type	2nd type
503.81	506.38	507.66	508.42

When all 50 observations are used for setting the control limits, we get $\bar{x} = 461.32$ and $s = 12.218$, while $c_4(50) = 0.9949$. Taking $p = 0.001$, the UCL without correction based on all 50 observations gives: 499.27. The remaining UCL's based on all observations are as follows

p			$1/p$	
exact	1st type	2nd type	1st type	2nd type
501.61	501.66	501.46	496.88	496.69

$1-(1-p)^{1000}$	$1-(1-p)^{500}$	$1-(1-p)^{250}$	$1-(1-p)^{100}$
1st type	2nd type	1st type	2nd type
499.07	500.27	500.87	501.22

This shows how UCL's are changing when more observations become available. □

2.3.2 Exceedance probability criterion

In the previous subsection we have discussed the benefits of developing a correction term based on the bias criterion. Using the developed correction term, an accurate control limit can be obtained using much smaller sample sizes of about 40 as compared to 300 without the correction term. However, the bias criterion, which is focused on the expectation of the distribution, only captures part of the picture and is merely useful for controlling the long term behavior of the chart over a series of applications. If we also want to control each application individually, then the exceedance probability criterion provides an answer to this problem.

As mentioned before, the exceedance probability criterion intends to restrict the probability of the relative error between $g(P_n)$ and $g(p)$, as given in (2.20) and (2.21) to be smaller than α, where α is a small positive number. For the case $g(p) = p$, the exceedance probability can be written as

$$P\left(\frac{P_n - p}{p} > \epsilon\right) \leq \alpha. \tag{2.35}$$

Substituting P_n of (2.28) in (2.20) with g tacitly assumed to be positive, we get the following expression

$$P\left(\frac{\widehat{\mu} - \mu}{\sigma} + (u + c)\frac{\sigma^*}{\sigma} < \overline{\Phi}^{-1}\left(g^{-1}(g(p)(1 + \varepsilon))\right)\right). \tag{2.36}$$

Recall that this equation only holds for an increasing function $g(p)$. For a decreasing function $g(p)$, our starting point is (2.21) and we end up with (2.36) with $(1 + \epsilon)$ replaced by $(1 - \epsilon)$.

As in the no-correction case, cf. (2.22) and (2.23), (2.36) can be written as

$$P\left(\frac{Z}{n^{1/2}} + a\frac{S}{\sigma} < b\right), \tag{2.37}$$

where $a = u + c$ for $\sigma^* = \check{\sigma}$, $a = (u + c)/c_4(n)$ for $\sigma^* = \widehat{\sigma}$ and where the r.v. Z has a standard normal distribution function and is independent from S. As usual, b depends on g and ϵ. For example, referring to the functions given in (2.3), b equals $\overline{\Phi}^{-1}(p(1 + \varepsilon))$, $\overline{\Phi}^{-1}(p/(1 - \varepsilon)$, and $\overline{\Phi}^{-1}\left(1 - \left[1 - \left\{1 - (1 - p)^k\right\}(1 + \varepsilon)\right]^{1/k}\right)$ for $g(p) = p$, $g(p) = 1/p$ and $g(p) = 1 - (1 - p)^k$, respectively.

Modifying the expression (2.37) a little bit further, we get

$$P\left(-\frac{Z}{n^{1/2}} - a\frac{S}{\sigma} > -b\right) = P\left(\frac{Z}{n^{1/2}} + b > a\frac{S}{\sigma}\right) = P\left(\frac{Z + bn^{1/2}}{S/\sigma} > an^{1/2}\right)$$

$$= \overline{G}_{n-1,bn^{1/2}}(an^{1/2}) = 1 - G_{n-1,bn^{1/2}}(an^{1/2}), \tag{2.38}$$

where as before, $G_{n-1,\delta}$ stands for the distribution function of the noncentral t-distribution with $n-1$ degrees of freedom and noncentrality parameter δ, and $\overline{G}_{n-1,\delta} = 1 - G_{n-1,\delta}$.

Consequently, an outcome α will result for $a = n^{-1/2}\overline{G}^{-1}_{n-1,n^{\frac{1}{2}}b}(\alpha)$. For $\sigma^* = \breve{\sigma}$, we have $a = u + c$ and hence the correction c which precisely produces equality in (2.36), is given by

$$c = n^{-1/2}\overline{G}^{-1}_{n-1,bn^{1/2}}(\alpha) - u, \tag{2.39}$$

For $\sigma^* = \widehat{\sigma}$, the factor $n^{-1/2}$ in (2.39) becomes $n^{-1/2}\,c_4(n)$, as $a = (u+c)/c_4(n)$ in this case.

Hence for any combination of n, p, ε and α, the exact correction c can now readily be evaluated using (2.39).

Recall that Table 2.7 shows how bad the situation is if we apply no correction term in the control limit. Therefore, the exact correction term given in (2.39) will be very useful to reduce the effect of the estimation error presented in the table. In the following table we provide some examples of the exact correction terms to show that the resulting corrections are indeed quite (to very) large.

Table 2.11 The exact correction terms as in (2.39) with $\epsilon = 0.1$, $\alpha = 0.1$ and $p = 0.001$

n	$g(p) = p$	$g(p) = 1/p$	$g(p) = 1-(1-p)^k$			
			$k = 100$	$k = 250$	$k = 500$	$k = 1000$
25	0.757	0.753	0.832	0.835	0.840	0.854
50	0.482	0.479	0.552	0.555	0.560	0.573
75	0.374	0.371	0.442	0.445	0.450	0.463
100	0.314	0.310	0.381	0.383	0.388	0.401
150	0.245	0.242	0.310	0.313	0.318	0.330
200	0.205	0.202	0.270	0.273	0.278	0.289
250	0.179	0.175	0.243	0.246	0.251	0.262
500	0.115	0.112	0.178	0.181	0.186	0.197
1000	0.072	0.068	0.134	0.137	0.141	0.153
2000	0.042	0.038	0.104	0.106	0.111	0.122
5000	0.015	0.012	0.077	0.080	0.084	0.096

Although the correction in (2.39) is exact, however, it does not show how matters depend on n, p, ε, α and g. In practice, such qualitative information would be quite valuable to allow sensible behavior. Hence in the following paragraphs we shall present a simple and transparent approximation which does reveal such relationships.

Firstly we present the approximation to (2.37). Let Z be a $N(0,1)$ random variable, independent of S^2 from Subsection 2.2.1. Moreover, let r be a fixed positive constant and define

$$F_n(x) = P((r^2+1)^{-1/2}\{rZ + (2n)^{1/2}(S/\sigma - 1)\} \leq x), \tag{2.40}$$

then we have the following result.

Lemma 2.3.1 *The distribution function from (2.40) satisfies*

$$F_n(x) = \Phi(x) + \frac{2^{1/2}\varphi(x)}{4\{(r^2+1)n\}^{1/2}}\left\{1 - \frac{x^2-1}{3(r^2+1)}\right\} + O(n^{-1}). \tag{2.41}$$

Hence in particular, we have that $F_n(x) = \Phi(x) + O(n^{-1/2})$.

Proof. It is well-known that $(n-1)S^2/\sigma^2$ is χ^2_{n-1}-distributed and as such can be represented by $\sum_{i=1}^{n-1} Z_i^2$, where $Z_1, \cdots, Z_{n-1}$ are independent $N(0,1)$ random variables. This fact not only implies asymptotics normality for $(n-1)S^2/\sigma^2$ by virtue of the central limit theorem, but also allows further refinements in the form of Edgeworth expansions (see e.g. Bickel and Doksum (2001), p. 318).

Let $\tilde{F}_n(x) = P(\{(n-1)/2\}^{1/2}(S^2/\sigma^2 - 1) \leq x)$, then we not only have that $\tilde{F}_n(x) = \Phi(x) + O(n^{-1/2})$, but for example also

$$\tilde{F}_n(x) = \Phi(x) - \frac{2^{1/2}\varphi(x)}{3n^{1/2}}(x^2-1) + O(n^{-1}). \tag{2.42}$$

Next, consider S instead of S^2. Let $F_n^*(x) = P(\{2(n-1)\}^{1/2}(S/\sigma - 1) \leq x)$, then

$$F_n^*(x) = P(S^2/\sigma^2 - 1 \leq 2\{2(n-1)\}^{-1/2}x + \{2(n-1)\}^{-1}x^2) = \tilde{F}_n(x + (8n)^{-1/2}x^2). \tag{2.43}$$

Combination of (2.42) and (2.43) readily gives that

$$F_n^*(x) = \Phi(x) + \frac{2^{1/2}\varphi(x)}{4n^{1/2}}\left\{1 - \frac{1}{3}(x^2-1)\right\} + O(n^{-1}). \tag{2.44}$$

As Z and S are independent, it is immediate that

$$P(\{rZ+\{2(n-1)\}^{1/2}(S/\sigma-1)\}\leq x(r^2+1)^{1/2})=\int F_n^*(x(r^2+1)^{1/2}-rz)d\Phi(z). \tag{2.45}$$

Straightforward calculation shows that substitution of (2.44) into (2.45) indeed produces the right-hand side of (2.41). It remains to note that the difference caused by using $(2n)^{1/2}$ in $F_n(x)$ (see (2.40)) and $\{2(n-1)\}^{1/2}$ in (2.45), is of order n^{-1} as well. ■

Corollary 2.3.1 *Let $\xi_{n,\alpha}$ be the lower α-quantile of F_n, then $\xi_{n,\alpha}$ satisfies*

$$\xi_{n,\alpha}=-u_\alpha-\frac{2^{1/2}}{4\{(r^2+1)n\}^{1/2}}\left\{1-\frac{u_\alpha^2-1}{3(r^2+1)}\right\}+O(n^{-1}). \tag{2.46}$$

Hence in particular, we have that $\xi_{n,\alpha}=-u_\alpha+O(n^{-1/2})$.

Proof. From (2.41) it follows that $F_n(x)=\Phi(x+2^{-3/2}\{(r^2+1)n\}^{-1/2}\{1-(x^2-1)/(3(r^2+1))\})+O(n^{-1})$ and thus that $\alpha=F_n(\xi_{n,\alpha})=\Phi(\xi_{n,\alpha}+2^{-3/2}\{(r^2+1)n\}^{-1/2}\{1-(\xi_{n,\alpha}^2-1)/(3(r^2+1))\})+O(n^{-1})$. This in its turn implies (2.46). ■

Next, the approximations for the correction term are derived to reveal the relationship between c and n, p, ϵ, α and g.

Lemma 2.3.2 *For the correction term c from (2.39) we have that, for $\sigma^*=\breve{\sigma}$,*

$$c=b-u+u_\alpha\left(\frac{b^2+2}{2n}\right)^{1/2}+\frac{b}{(b^2+2)n}\left\{\left(\frac{5}{12}b^2+1\right)u_\alpha^2+\left(\frac{1}{3}b^2+\frac{1}{2}\right)\right\}+O(n^{-3/2}). \tag{2.47}$$

Recall that b depends on g, p and ϵ, in which for $g(p)=p$, $b=\overline{\Phi}^{-1}(p(1+\varepsilon))$, for $g(p)=1/p$, $b=\overline{\Phi}^{-1}(p/(1-\varepsilon)$, and for $g(p)=1-(1-p)^k$, $b=\overline{\Phi}^{-1}\left(1-\left[1-\left\{1-(1-p)^k\right\}(1+\varepsilon)\right]^{1/k}\right)$. For $\sigma^*=\widehat{\sigma}$, the term $b^2/3+1/2$ in (2.47) should be replaced by $b^2/12$. Hence in particular we have for either choice of σ^*

$$c=b-u+u_\alpha\left(\frac{b^2+2}{2n}\right)^{1/2}+O(n^{-1}). \tag{2.48}$$

Proof. We need to evaluate $\tilde{P} = P(n^{-1/2}Z + aS/\sigma < b)$, where $a = u + c$ for $\sigma^* = \breve{\sigma}$, $a = (u + c)/c_4(n)$ for $\sigma^* = \widehat{\sigma}$. After some rearrangement we obtain

$$\tilde{P} = P((2^{1/2}/a)Z + (2n)^{1/2}(S/\sigma - 1) < (2n)^{1/2}(b/a - 1)), \tag{2.49}$$

and thus we apply Lemma 2.3.1 with $r = 2^{1/2}/a$, which leads to

$$\tilde{P} = F_n((b - a)\{(2n)/(a^2 + 2)\}^{1/2}). \tag{2.50}$$

The desired result $\tilde{P} = \alpha$ is achieved if c (and thus a) is chosen such that $(b - a)\{(2n)/(a^2 + 2)\}^{1/2} = \xi_{n,\alpha}$. As $\xi_{n,\alpha} = -u_\alpha + O(n^{-1/2})$, it follows that $a = b + u_\alpha\{(a^2 + 2)/(2n)\}^{1/2} + O(n^{-1})$, and thus

$$a = b + u_\alpha \left(\frac{b^2 + 2}{2n}\right)^{1/2} + O(n^{-1}). \tag{2.51}$$

Note that (2.51) implies that $\{(a^2+2)/(2n)\}^{1/2} = \{(b^2+2)/(2n)\}^{1/2} + bu_\alpha/(2n) + O(n^{-3/2})$. Combination of this fact with (2.46) and (2.51) produces (2.47) after straightforward computation. Finally, if we go from $\sigma^* = \breve{\sigma}$ to $\sigma^* = \widehat{\sigma}$, then $a = u+c$ changes into $a = (u+c)/c_4(n)$, and thus c becomes $c_4(n)a - u$. As $c_4(n) = 1 - (4n)^{-1} + O(n^{-3/2})$, this modification results to $O(n^{-3/2})$ in the subtraction of $-b/(4n)$ from the expression in (2.47). ■

From (2.48), for each of the two choices of σ^* we get the following approximation for the correction term,

$$c \approx b - u + u_\alpha \left(\frac{b^2 + 2}{2n}\right)^{1/2}. \tag{2.52}$$

By taking into account not only that n becomes large, but also that ϵ (usually) will be small and that p is small, using the following corollary, further simplification can be achieved, leading for example for the cases $g(p) = p$ and $g(p) = 1/p$, to

$$c \approx -\frac{\varepsilon}{u} + u_\alpha \left(\frac{u^2 + 2}{2n}\right)^{1/2}. \tag{2.53}$$

Corollary 2.3.2 *For the case $g(p) = p$, the correction term c from (2.39) moreover satisfies*

$$c = -\varepsilon\frac{p}{\varphi(u)} + u_\alpha \left(\frac{u^2 + 2}{2n}\right)^{1/2} + R_1 = -\frac{\varepsilon}{u} + u_\alpha \left(\frac{u^2 + 2}{2n}\right)^{1/2} + R_2, \tag{2.54}$$

where $|R_1| \le C_1(\varepsilon,p)n^{-1} + C_2(p)\varepsilon^2$ *and* $|R_2| \le C_1(\varepsilon,p)n^{-1} + C_2(p)\varepsilon^2 + C_3u^{-6}$*, in which the* $C_i, i = 1,2,3$*, are constants depending as indicated on* p *and/or* ε*.*

Proof. Since for $g(p) = p$, $b = u_{p(1+\varepsilon)}$, it is immediate that $b - u = -\varepsilon p/\varphi(u) + O(\varepsilon^2)$. Moreover, changing b into u in the term of order $n^{-1/2}$ merely adds a term of order $\varepsilon n^{-1/2} = O(n^{-1} + \varepsilon^2)$ to the remainder. Finally, as $\overline{\Phi}(x)/\varphi(x) = x^{-1} + O(x^{-3})$ for x large, it follows that the second result in (2.54) requires an additional term $O(\varepsilon u^{-3})$, which is covered by adding u^{-6} to the remainder. ■

The results given in (2.53) and (2.54) only hold for $g(p) = p$. Of course, they can be extended to a more general function, such as a positive function g given in (2.3). In a more general form, expansion with respect to ϵ (cf. Corollary 2.3.2) now leads to

$$b - u \approx \overline{\Phi}^{-1}(p + \varepsilon g(p)/g'(p)) - u \approx -\varepsilon g(p)/\{\varphi(u)g'(p)\} \approx (-\varepsilon/u)h(p), \quad (2.55)$$

with $h(p) = g(p)/\{pg'(p)\}$ and the convention that $+\varepsilon$ is replaced by $-\varepsilon$ for decreasing g. Note that this more general form for the first term in (2.52) boils down to $-\varepsilon/u$ for both $g(p) = p$ and $g(p) = 1/p$, while for the third choice we obtain $h(p) = (1-p)\{(1-p)^{-k}-1\}/(kp) \approx (e^{kp}-1)/kp) \approx 1+kp/2$ for k such that kp is small. Hence we may conclude that, no matter which of these three choices of g is favored by the practitioner, the actual correction to be used is fortunately (almost) the same.

Remark 2.3.5 For the two-sided case with $g(p) = p$, (2.53) translates into

$$c2 \approx -\frac{\varepsilon}{u_{p/2}} + u_\alpha \left(\frac{u_{p/2}^2}{2n}\right)^{1/2}, \quad (2.56)$$

which result, just as in the one-sided case, holds true for the ARL as well.

The derivation of (2.56) is as follows. Here P_n from (2.28) is replaced by $P(X_{n+1} > \widehat{\mu} + (u_{p/2}+c)\sigma^*) + P(X_{n+1} < \widehat{\mu} - (u_{p/2}+c)\sigma^*) = \overline{\Phi}((\widehat{\mu}-\mu)/\sigma + (u_{p/2}+c)\sigma^*/\sigma) + \overline{\Phi}(-(\widehat{\mu}-\mu)/\sigma + (u_{p/2}+c)\sigma^*/\sigma)$. Note that an explicit exact approach (cf.(2.36), (2.37), (2.38)) is no longer possible in this two-sided case. Hence we move on directly to the approximations. To avoid repetition we shall not go into much detail. Just note that now $P_n - p \approx -2\varphi(u_{p/2})\{u_{p/2}(\sigma^*/\sigma - 1) + c\}$, which shows that (2.35) will approximately hold for this case if $P(u_{p/2}(\sigma^*/\sigma - 1) < -\varepsilon p/(2\varphi(u_{p/2})) - c) \approx \alpha$. This in its turn implies in view of (2.44) that

$$c \approx -\varepsilon \frac{p}{2\varphi(u_{p/2})} + u_\alpha \left(\frac{u_{p/2}^2}{2n}\right)^{1/2}, \quad (2.57)$$

cf. (2.54). The first term in (2.57) can be further simplified into $-\varepsilon/u_{p/2}$; it is also straightforward to check that the case of general g is covered here as well by multiplying this first term by $h(p)$ from (2.55). (Note: $h(p)$, and not $h(p/2)$!) □

To demonstrate the quality of the approximations, we once more use the exact method to evaluate the realized exceedance probabilities. In (2.26) the uncorrected case was considered, and here we collect in Table 2.12

$$\tilde{\alpha} = \overline{G}_{n-1,n^{1/2}b}(n^{1/2}(u+c)), \tag{2.58}$$

where c is given by (2.52), by its further refinement (2.47), or for $g(p) = p$ by its further simplification (2.53), respectively.

Table 2.12 Values of $\tilde{\alpha}$ from (2.58) for $g(p) = p$ with $p = 0.001, \varepsilon = 0.1, \alpha = 0.1$ and $\alpha = 0.2$, and c from (2.52), (2.47) and (2.53).

α	0.1			0.2		
c	(2.52)	(2.47)	(2.53)	(2.52)	(2.47)	(2.53)
$n = 25$	0.1548	0.1132	0.1546	0.2514	0.2120	0.2519
$n = 50$	0.1369	0.1065	0.1373	0.2358	0.2058	0.2362
$n = 75$	0.1294	0.1043	0.1302	0.2279	0.2038	0.2301
$n = 100$	0.1252	0.1032	0.1264	0.2239	0.2029	0.2267
$n = 150$	0.1202	0.1022	0.1221	0.2193	0.2019	0.2231
$n = 200$	0.1174	0.1016	0.1198	0.2166	0.2014	0.2213
$n = 250$	0.1154	0.1013	0.1184	0.2148	0.2011	0.2202
$n = 500$	0.1107	0.1006	0.1156	0.2104	0.2006	0.2189
$n = 1000$	0.1075	0.1003	0.1152	0.2072	0.2003	0.2202
$n = 2000$	0.1053	0.1002	0.1170	0.2051	0.2001	0.2244
$n = 5000$	0.1074	0.1001	0.1235	0.2032	0.2001	0.2354

From Table 2.12 it is clear that the approximations behave as expected: c from (2.52) converges quite well and seems sufficiently accurate for practical purposes. Moreover, the refinement indeed is very accurate and converging rapidly, while (2.53) is reasonable but involves an error which does not vanish as n increases (cf. (2.54)). As concerns the uncorrected exceedance probability $\tilde{\alpha}$ from (2.26), it is easily shown (see Corollary 2.3.3) that

$$\tilde{\alpha} = \Phi\left((b-u)\left(\frac{u^2+2}{2n}\right)^{-1/2}\right). \tag{2.59}$$

Corollary 2.3.3 *For the realized exceedance probability $\tilde{\alpha}$ from (2.26) we have that*

$$\tilde{\alpha} = F_n\left((b-u)\left(\frac{u^2+2}{2n}\right)^{-1/2}\right), \tag{2.60}$$

which through (2.41) provides an expansion to $O(n^{-1})$ for $\tilde{\alpha}$. In particular, we obtain that

$$\tilde{\alpha} = \Phi\left((b-u)\left(\frac{u^2+2}{2n}\right)^{-1/2}\right) + O(n^{-1/2}). \tag{2.61}$$

Proof. If no correction is applied, obviously $c = 0$, and thus $a = u$. Insertion of this value into the expression for the realized exceedance probability $\tilde{P}$ from (2.50), then already produces (2.60). ■

Next, we exploit (2.52) and (2.53) to study the behavior of the correction in relation to n, p, ε, α and g.

1) role of n

It was demonstrated in Subsection 2.3.1 that a term $u(u^2+2)/(4n)$ is dominant in correcting the bias of P_n. From (2.53) we see that here a term $u_\alpha\{(u^2+2)/(2n)\}^{1/2}$ occurs. Hence it indeed turns out that controlling the variability calls for stronger intervention: it requires a term of order $n^{-1/2}$ rather than of order n^{-1}. Consequently, much larger sample sizes are needed before this type of correction becomes negligible and thus superfluous. To be more specific consider $g(p) = p$ or $g(p) = 1/p$ and note that 'no correction' means $c = 0$. Hence (2.52) and (2.53) immediately supply approximations to the minimal sample size for which (2.35) is met without correction:

$$n \approx \frac{1}{2}u_\alpha^2(b^2+2)/(u-b)^2 \approx \frac{1}{2}u_\alpha^2 u^2(u^2+2)/\varepsilon^2. \tag{2.62}$$

For example, for $p = 0.001$, $\varepsilon = 0.1$ and $\alpha = 0.25$, we find $n \approx 3200$ and $n \approx 2500$, respectively. Using the exact approach from the previous section, we find for these values of n that the realized value of α equals 0.254 and 0.28, respectively. Hence (2.52) is quite accurate, but even the more simple (2.53) suffices to get a clear picture of what is going on in practice. Returning to c itself, we get for the above p and ε that (2.52) reduces to

$$c \approx 2.38u_\alpha n^{-1/2} - 0.028, \tag{2.63}$$

(for (2.53) replace the last number in (2.63) by –0.032), which for $\alpha = 0.1$ and $\alpha = 0.2$ boils down to $c = 3.06n^{-1/2} - 0.028$ and $c = 2.01n^{-1/2} - 0.028$, respectively. It is easily checked (see Table 2.11) that these simple expressions nicely explain the corresponding exact values.

2) role of p

Things go (much) less well than intuitively anticipated because we are dealing with small quantiles. Indeed, if we for example go from $p = 0.001$, to $p = 0.01$, matters improve as u decreases. Again consider $g(p) = p$ or $g(p) = 1/p$. The n from (2.62) is reduced by a factor 0.4, leading in the example above to $n \approx 1300$ and $n \approx 900$, respectively (with corresponding realized values of α of 0.255 and 0.29, respectively). Likewise, (2.63) becomes $c = 1.90u_\alpha n^{-1/2} - 0.036$ (replace the last number by -0.043 for (2.53) rather than (2.52)). Hence c is reduced in two respects: the coefficient of the positive term decreases, whereas the magnitude of the negative term increases. Of course, this is in view of (2.53) hardly surprising, since $(u^2+2)^{\frac{1}{2}} = u(1+2/u^2)^{1/2} \approx u+1/u$ behaves like u, while $b-u$ behaves like $-\varepsilon/u$.

3) role of ε

Clearly, as ε decreases, it becomes harder to satisfy (2.35) and both c in (2.52) and (2.53) and n in (2.62) will grow. In the limiting case $\varepsilon = 0$ a correction can only be avoided by taking $\alpha = 1/2$, which is a simple consequence of the fact that P_n is asymptotically normal with mean p. The role of ε of reversing to some extent the need for correction (cf. its negative sign in (2.53)) is, as we saw above, weakened as p decreases. Taking once more $g(p) = p$ or $g(p) = 1/p$, $p = 0.001$ and $\alpha = 0.1$, we obtain for small ε through (2.53) that $c \approx 3.06n^{-1/2} - 0.32\varepsilon$. This shows that the effect of the second term is rather limited, unless n is really large. Of course, matters change when ε is not small. E.g. for $p = 0.001$ we then get from (2.52) that $c \approx 2.38u_\alpha n^{-1/2} - \{3.09 - \overline{\Phi}^{-1}(0.001(1+\varepsilon))\}$. For e.g. $\varepsilon = 9$, the last term equals 0.764, which implies that $c \approx 0$ for $u_\alpha n^{-1/2} \approx 0.321$, i.e. $n \approx 9.75u_\alpha^2$. For $\alpha = 0.05$, this produces $n = 26$. Hence the combination of a small p, a small n and a small α indeed leads to trouble: without correction, P_n can be off by a huge factor (e.g. 10, as in this example). Incidentally, the approximation again works quite well: from (2.39) we get for the values under consideration $c = 0.04$, which is indeed rather negligible compared to $u = 3.09$.

4) role of α

This is pretty straightforward: in both (2.52) and (2.53), the first term is linear in u_α. Similar remarks can be made and similar examples can be given as in *1) – 3)* above.

5) role of g

Recall that b depends on g, and since c depends on b, c also depends on g. Fortunately, for g given in (2.3), the resulted correction terms are almost the same cf. (2.55).

Summarizing the exposition above, it illustrates that without correction, and with

p small and n not large, one should not have great expectations about α and ε! Approximation (2.52) and (2.53) make transparent how bad the actual situation is, and what needs to be done to achieve control again.

Having developed and evaluated the correction terms using the exceedance probability criterion, we therefore propose the following UCL's for the corrected control chart $\widehat{\mu} + (u + c)\sigma^*$ with either $\widehat{\sigma}$ or $\breve{\sigma}$ as estimator of σ:

- for any $g(p)$ in (2.3), an exact correction c is given in (2.39)
- for any $g(p)$ in (2.3), an approximation of c is given by (2.52) with a further refinement given in (2.47)
- for $g(p) = p$ and $g(p) = 1/p$, an approximation of c is given by (2.53), and for $g(p) = 1-(1-p)^k$, the first term in (2.53) is replaced by $(-\epsilon/u)(1+(kp)/2)$ provided that kp is small. This approximation however, requires ϵ to be small. Moreover, it involves an error which does not vanish as n increases. Hence, it is less recommended than the others.

To illustrate the results obtained, we once again consider data from Example 3.1. For $\varepsilon = 0.1$ and $\alpha = 0.1$, we obtain from the exact corrections 0.76 for $n = 25$ and 0.48 for $n = 50$, respectively. These results lead to replacement of the uncorrected values 504.2 and 499.3 by 514.3 and 505.4, respectively. Clearly, these values are indeed considerably larger than the corresponding values 509.6 and 501.6 used in bias reduction, which reflects the fact that controlling exceedance probabilities has a larger impact. Also note that, apart from being larger in an absolute sense, the present corrections also converge more slowly than those from the bias case. To be specific, for $n = 25$ the ratio of the increases over the uncorrected value in the two approaches equals (509.6-504.2)/(514.3-504.2) = 0.535, while after doubling the sample size, i.e. for $n = 50$, this ratio has gone down to (501.6-499.3)/(505.4-499.3) = 0.378. Now $(0.535/0.378)^2 = 2.01$, which nicely reflects the fact that the rates of convergence involved are $n^{-1/2}$ and n^{-1}, respectively (cf. the remarks to this effect under *1) role of* n above).

2.4 Out-of-control case

It has been shown in Section 2.3 that merely using the UCL in (2.1) without any attempt to correct it, we can get accurate control charts for the in-control situation only for very large sample sizes. To solve this problem we have developed correction terms based on the bias and exceedance probability criteria. We have also demonstrated in the previous section that with the addition of these correction terms, the corrected UCL leads to accurate charts for much smaller sample sizes.

Although the corrected version of the UCL thus has a positive effect on the in-control behavior, we still have to verify to what extent such corrections affect the out-of-control behavior. If the corrections are so large that we almost never detect a signal when the system is out-of-control, then the control chart would become of no use. Fortunately, as we will show in this section, the corrections do not at all disturb the out-of-control performance drastically.

The objective of this section is to see how in the out-of-control case the performance of the corrected control charts changes when the number of observations grows. Moreover, we want to compare them with control charts with known parameters. In addition, we also want to investigate the 'loss' in the out-of-control situation due to correcting for estimation in the in-control case. The notion 'loss' is of course somewhat virtual, as the case of known parameters to which it refers, hardly ever occurs and as such forms no real alternative option. In this way, we 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. However, knowing that very often in many practical situations today it is impossible to collect many observations due to the enormous flexibility in production processes (shorter product life cycles, product diversity, products tailored to specific customer requirements, etc.), in many cases we cannot avoid to apply the correction terms and it is of great interest to see how the performance of control charts change with the number of observations in the region of moderate sample sizes.

The out-of-control situation is described as follows. Let $X_1, \cdots, X_n$ be i.i.d. distributed r.v.'s, each having a $N(\mu, \sigma^2)$-distribution, and let X_{n+1} be a r.v. with a $N(\mu_1, \sigma_1^2)$-distribution, where $\mu_1 = \mu + d\sigma$ for some $d > 0$. If the parameters were known, the false alarm rate in the out-of-control state would be $p_1 = \overline{\Phi}(u - d)$. Since typically $p_1 >> p$, the estimation effects around p_1 in terms of relative error, are (much) smaller than those around p. Therefore, we expect for the out-of-control behavior small effects due to the moderate bias correction, and moderate effects due to the larger corrections associated with the exceedance probability.

In the following paragraphs, we will discuss the effects in detail. We start with the bias corrected control charts.

2.4.1 The effect of removing the bias

Using the corrected UCL given in (2.27), the probability to detect an out-of-control signal is given by

$$
\begin{aligned}
P_n &= P\left(X_{n+1} > \widehat{\mu} + (u+c)\sigma^*\right) = \overline{\Phi}\left(\frac{\widehat{\mu} - \mu_1}{\sigma_1} + (u+c)\frac{\sigma^*}{\sigma_1}\right) \\
&= \overline{\Phi}\left(\frac{\mu - \mu_1}{\sigma_1} + u\frac{\sigma}{\sigma_1} + \frac{\sigma}{\sigma_1}\left\{c + \frac{\widehat{\mu} - \mu}{\sigma} + (u+c)\left(\frac{\sigma^*}{\sigma} - 1\right)\right\}\right) \\
&= \overline{\Phi}\left(u_1 + \frac{\sigma}{\sigma_1}\left\{c + \Delta(u+c)\right\}\right),
\end{aligned}
$$

where

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

Using a similar strategy as has been developed in Section 2.2.2, and writing $p_1 = \overline{\Phi}(u_1)$, we arrive at the following first type approximation $P_n \approx p_1 W_n$ with

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

From this, in the out-of-control situation we get a first type approximation for $g(p) = p$ as follows:

$$EP_n \approx p_1 \exp\left[-c\frac{\sigma}{\sigma_1}\frac{u_1}{z(u_1)} + \frac{1}{2}\left\{\frac{u_1}{z(u_1)}\right\}^2 \left(\frac{\sigma}{\sigma_1}\right)^2 \left\{\frac{1}{n} + \frac{(u+c)^2}{2(n-1)}\right\}\right], \quad (2.64)$$

and for $g(p) = 1/p$:

$$E\left(\frac{1}{P_n}\right) \approx \frac{1}{p_1} \exp\left[c\frac{\sigma}{\sigma_1}\frac{u_1}{z(u_1)} + \frac{1}{2}\left\{\frac{u_1}{z(u_1)}\right\}^2 \left(\frac{\sigma}{\sigma_1}\right)^2 \left\{\frac{1}{n} + \frac{(u+c)^2}{2(n-1)}\right\}\right]. \quad (2.65)$$

As in Section 2.2.2, we do not give a first type approximation for the third function here. Instead we will present this function using the second type approximation.

If the parameters μ and σ are known, we get p_1 and $(1/p_1)$, respectively. In general, c is relatively small compared to u (see Table 2.8). 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

$$\exp\left\{-c\frac{\sigma}{\sigma_1}\frac{u_1}{z(u_1)}\right\} \text{ and } \exp\left\{c\frac{\sigma}{\sigma_1}\frac{u_1}{z(u_1)}\right\}, \quad (2.66)$$

Since the correction term for EP_n equals minus the one for $E(1/P_n)$, see (2.29) and (2.30), both factors in (2.66) are equal and they are smaller than 1. It turns out (see Table 2.13 and Table 2.14) that for $E(1/p_n)$ and in most cases also for EP_n, the factors in (2.66) dominate the factors coming from the second part of the exponential term in (2.64) and (2.65), respectively for EP_n and $E(1/P_n)$. Hence, the exponential terms in (2.64) 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 $E(1/P_n)$ cf. (2.65), the 'gain' obtain by the estimation.

The second type approximation is based on the following expansion:

$$\begin{aligned}
Eg(P_n) &= Eh\left(u_1 + \frac{\sigma}{\sigma_1}\{c + \Delta(u+c)\}\right) \\
&\approx h(u_1) + h'(u_1)\frac{\sigma}{\sigma_1}\{c + E\Delta(u+c)\} + \frac{1}{2}h''(u_1)\left(\frac{\sigma}{\sigma_1}\right)^2 E\Delta^2(u+c).
\end{aligned}$$

In view of Table 2.3 and (2.15) with $\sigma^* = \widehat{\sigma}$ we get the second type approximation for the three functions given in (2.3) as follows:

$$\begin{aligned}
EP_n &\approx p_1 - c\frac{\sigma}{\sigma_1}\varphi(u_1) + \frac{1}{2}\left(\frac{\sigma}{\sigma_1}\right)^2 a(u+c,n)u_1\varphi(u_1) \\
E(\frac{1}{P_n}) &\approx \frac{1}{p_1} + c\frac{\sigma}{\sigma_1}\frac{\varphi(u_1)}{{p_1}^2} + \\
&\quad \frac{1}{2}\left(\frac{\sigma}{\sigma_1}\right)^2 a(u+c,n)\left(\frac{2\varphi^2(u_1)}{{p_1}^3} - \frac{u_1\varphi(u_1)}{{p_1}^2}\right) \\
E(1-(1-P_n)^k) &\approx 1-(1-p_1)^k - c\frac{\sigma}{\sigma_1}k\varphi(u_1)(1-p_1)^{k-1} + \\
&\quad \frac{1}{2}\left(\frac{\sigma}{\sigma_1}\right)^2 a(u+c,n)\{-k(k-1)\varphi^2(u_1)(1-p_1)^{k-2} \\
&\quad + k\ u_1\ \varphi(u_1)(1-p_1)^{k-1}\}
\end{aligned} \tag{2.67}$$

respectively for $g(p) = p$, $1/p$, and $1-(1-p)^k$.

Remark 2.4.1 Let $\sigma_1 = \sigma$ and $\mu_1 = \mu + d\sigma$. Noting that c is relatively small compared to u we have $a(u+c,n) \approx a(u,n) \approx (u^2+2)/(2n)$ cf. (2.15) and (2.18). For $g(p) = p$, the relative error can be approximated as follows:

$$\begin{aligned}
\frac{EP_n - p_1}{p_1} &\approx -c\frac{\varphi(u-d)}{p_1} + \frac{1}{2}\frac{(u-d)\varphi(u-d)}{p_1}\left(\frac{u^2+2}{2n}\right) \\
&\approx \frac{\varphi(u-d)}{p_1}\frac{u^2+2}{4n}\{-u+u-d\} \\
&\approx -d\left(\frac{u^2+2}{5n}\right)\{1+u-d\},
\end{aligned}$$

where in the last line we have used that for $0 \leq x \leq 3.5$ the function $\kappa(x) = \varphi(x)/\overline{\Phi}(x)$ can be approximated adequately by $4(1+x)/5$. □

Having developed the approximations under the out-of-control situation, we investigate the performance of the chart using various sample sizes. The behavior of the chart using the two types approximation is presented in Table 2.13 for $g(p) = p$

and in Table 2.15 for $g(p) = 1/p$. 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 in-control case. However, the control charts considered here can also detect changes in variance. The influence of changing the variance is seen in (2.64), (2.65) and (2.67).

In addition, we carry out some simulations to check the accuracy of these approximations. The simulation results are given in Table 2.14 and Table 2.16, respectively. Furthermore, in Table 2.17 we present the result of the second type approximation together with the simulation result for $g(p) = 1 - (1-p)^k$.

Table 2.13 First type (2.64) and second type (2.67) approximation of EP_n with $\mu_1 = \mu + d\sigma$ and $\sigma_1 = \sigma$.

d	0.5		1		2	
p_1	0.0048	0.0048	0.0183	0.0183	0.138	0.138
approx	$1^{st}type$	$2^{nd}type$	$1^{st}type$	$2^{nd}type$	$1^{st}type$	$2^{nd}type$
$n = 10$	0.0208	0.0105	0.0359	0.0254	0.107	0.062
$n = 20$	0.0060	0.0055	0.0178	0.0163	0.104	0.086
$n = 30$	0.0050	0.0049	0.0167	0.0159	0.111	0.100
$n = 40$	0.0048	0.0047	0.0167	0.0161	0.116	0.109
$n = 50$	0.0047	0.0046	0.0168	0.0164	0.119	0.114
$n = 60$	0.0047	0.0046	0.0169	0.0166	0.122	0.118
$n = 70$	0.0047	0.0046	0.0171	0.0168	0.124	0.120
$n = 80$	0.0047	0.0046	0.0172	0.0169	0.125	0.123
$n = 90$	0.0047	0.0046	0.0173	0.0171	0.127	0.124
$n = 100$	0.0047	0.0047	0.0173	0.0172	0.128	0.126
$n = 150$	0.0047	0.0047	0.0176	0.0175	0.131	0.130
$n = 200$	0.0047	0.0047	0.0178	0.0177	0.132	0.132
$n = 300$	0.0047	0.0047	0.0179	0.0179	0.134	0.134

Table 2.13 shows that the results from the two types of approximation are very close to each other. The results show that inserting the correction terms for $g(p) = p$ does not disturb the behavior of the control charts in the out-of-control situation. Already for moderate sample sizes the corrected control charts work very well, in the sense that a very good behavior is achieved in the out-of-control situation. This fact shows that the objective of the correction terms is reached: the in-control behavior is regulated and the loss in the out-of-control case is relatively small for moderate sample sizes.

From this table we can say that, for $n \geq 50$, the behavior changes very slowly with n, and in most practical situations the improvement will be outweighed by the costs of taking more observations. Even for small values of n such as 20, already very reasonable results are obtained. For that reason, the corrected control charts based on moderate sample sizes are suitable for practitioners, since the corrected charts are accurate for the in-control situation and powerful in the out-of-control case.

To check whether these approximations are valid, we perform a simulation study. The results of the simulation, which are given in Table 2.14, show that the two approximations can be considered as very accurate. Although for exceptionally small sample sizes the approximations seem unreliable, for moderate sample sizes both approximations work very well. Hence, Table 2.14 confirms the conclusion based on Table 2.13.

Table 2.14 Simulation results for EP_n with first and second type correction terms, given by (2.29) and (2.33), respectively, when $\mu_1 = \mu + d\sigma$ and $\sigma_1 = \sigma$.

d	0.5		1		2	
p_1	0.0048	0.0048	0.0183	0.0183	0.138	0.138
approx	$1^{st}type$	$2^{nd}type$	$1^{st}type$	$2^{nd}type$	$1^{st}type$	$2^{nd}type$
$n = 10$	0.0051	0.0058	0.0140	0.0155	0.073	0.079
$n = 20$	0.0043	0.0046	0.0141	0.0151	0.090	0.096
$n = 30$	0.0043	0.0045	0.0147	0.0155	0.102	0.106
$n = 40$	0.0044	0.0045	0.0153	0.0160	0.109	0.112
$n = 50$	0.0044	0.0046	0.0158	0.0163	0.113	0.117
$n = 60$	0.0044	0.0046	0.0161	0.0166	0.119	0.119
$n = 70$	0.0045	0.0046	0.0164	0.0168	0.120	0.122
$n = 80$	0.0045	0.0046	0.0166	0.0169	0.122	0.123
$n = 90$	0.0045	0.0046	0.0167	0.0171	0.123	0.125
$n = 100$	0.0045	0.0046	0.0168	0.0171	0.124	0.126
$n = 150$	0.0046	0.0047	0.0173	0.0177	0.129	0.130
$n = 200$	0.0047	0.0047	0.0175	0.0176	0.131	0.132
$n = 300$	0.0047	0.0047	0.0178	0.0179	0.133	0.134

Table 2.15 presents the results of the two approximations for the case $g(p) = 1/p$. Fortunately, just as in Table 2.13, the results from the two types of approximation are very close to each other. These approximations show that inserting the correction terms for $g(p) = 1/p$ does not disturb the behavior of the control charts in the out-of-control situation. In this case, the approximations for a small sample size such as 10 should not be taken seriously. As a matter of fact, this is consistent with the conditions in Theorem 2.2.2. Recall that the conditions in Theorem 2.2.2 are slightly stronger than those in Theorem 2.2.1: indeed, with the second type approximation the aberrance for $n = 10$ exceeds the corresponding one with the first type approximation.

Subsequently, the results are validated by the simulation results given in Table 2.16. The results of the simulation show that the two types approximation are very accurate, except for extremely small sample sizes, such as 10. Therefore, the conclusion based on Table 2.15 is confirmed by the result in Table 2.16.

Table 2.15 First type (2.65) and second type (2.67) approximation of $E(1/P_n)$ with $\mu_1 = \mu + d\sigma$ and $\sigma_1 = \sigma$.

d	0.5		1		2	
$1/p_1$	208.5	208.5	54.6	54.6	7.3	7.3
approx	$1^{st}type$	$2^{nd}type$	$1^{st}type$	$2^{nd}type$	$1^{st}type$	$2^{nd}type$
$n = 10$	54.0	-100.4	14.4	-22.8	2.4	-0.8
$n = 20$	133.2	109.5	33.0	26.6	4.5	3.9
$n = 30$	163.4	155.8	40.6	38.6	5.3	5.2
$n = 40$	177.3	174.1	44.4	43.6	5.8	5.8
$n = 50$	185.1	183.5	46.6	46.3	6.1	6.1
$n = 60$	189.9	189.1	48.1	47.9	6.3	6.3
$n = 70$	193.1	192.7	49.1	49.1	6.4	6.5
$n = 80$	195.4	195.3	49.8	49.9	6.5	6.6
$n = 90$	197.2	197.1	50.4	50.5	6.6	6.6
$n = 100$	198.5	198.5	50.8	50.9	6.7	6.7
$n = 150$	202.3	202.4	52.2	52.3	6.9	6.9
$n = 200$	204.0	204.2	52.8	52.9	7.0	7.0
$n = 300$	205.6	205.8	53.4	53.5	7.1	7.1

Table 2.16 Simulation results for $E(1/P_n)$ with first and second type correction terms, given by (2.30) and (2.33), respectively, when $\mu_1 = \mu + d\sigma$ and $\sigma_1 = \sigma$.

d	0.5		1		2	
$1/p_1$	208.5	208.5	54.6	54.6	7.3	7.3
approx	$1^{st}type$	$2^{nd}type$	$1^{st}type$	$2^{nd}type$	$1^{st}type$	$2^{nd}type$
$n = 10$	82.7	58.2	20.0	15.0	3.2	2.8
$n = 20$	172.9	140.7	39.1	34.8	5.1	4.7
$n = 30$	190.2	170.0	45.2	42.0	5.8	5.5
$n = 40$	195.2	184.1	48.4	45.1	6.2	6.0
$n = 50$	200.5	189.1	49.5	47.5	6.4	6.2
$n = 60$	200.8	193.2	50.6	49.1	6.5	6.4
$n = 70$	202.4	195.3	51.6	49.9	6.6	6.5
$n = 80$	203.7	198.2	51.7	50.3	6.7	6.6
$n = 90$	204.5	198.0	52.1	50.8	6.8	6.7
$n = 100$	204.6	200.0	52.3	51.3	6.8	6.7
$n = 150$	206.1	204.0	52.9	52.6	7.0	6.9
$n = 200$	207.0	204.4	53.6	53.1	7.0	7.0
$n = 300$	207.7	205.8	54.0	53.5	7.1	7.1

Table 2.17 Simulation results and second type of approximation of $E(1-(1-P_n)^k)$ with $\mu_1 = \mu + d\sigma$ and $\sigma_1 = \sigma$.

d	0.5		1		2	
$1-(1-p_1)^{1000}$	0.99	0.99	1.00	1.00	1.00	1.00
	simul	$2^{nd}app$	*simul*	$2^{nd}app$	*simul*	$2^{nd}app$
$n = 10$	0.83	0.60	0.94	1.00	1.00	1.00
$n = 20$	0.87	0.80	0.97	1.00	1.00	1.00
$n = 30$	0.90	0.87	0.99	1.00	1.00	1.00
$n = 40$	0.91	0.90	0.99	1.00	1.00	1.00
$n = 50$	0.93	0.92	1.00	1.00	1.00	1.00
$n = 60$	0.94	0.93	1.00	1.00	1.00	1.00
$n = 70$	0.94	0.94	1.00	1.00	1.00	1.00
$n = 80$	0.95	0.95	1.00	1.00	1.00	1.00
$n = 90$	0.95	0.95	1.00	1.00	1.00	1.00
$n = 100$	0.96	0.96	1.00	1.00	1.00	1.00
$n = 150$	0.97	0.97	1.00	1.00	1.00	1.00
$n = 200$	0.97	0.97	1.00	1.00	1.00	1.00
$n = 300$	0.98	0.98	1.00	1.00	1.00	1.00

d	0.5		1		2	
$1-(1-p_1)^{500}$	0.91	0.91	1.00	1.00	1.00	1.00
	simul	$2^{nd}app$	*simul*	$2^{nd}app$	*simul*	$2^{nd}app$
$n = 10$	0.57	-0.46	0.75	0.94	0.96	1.00
$n = 20$	0.66	0.31	0.88	0.99	1.00	1.00
$n = 30$	0.71	0.53	0.93	1.00	1.00	1.00
$n = 40$	0.75	0.63	0.95	1.00	1.00	1.00
$n = 50$	0.77	0.69	0.97	1.00	1.00	1.00
$n = 60$	0.79	0.73	0.98	1.00	1.00	1.00
$n = 70$	0.80	0.75	0.98	1.00	1.00	1.00
$n = 80$	0.81	0.77	0.99	1.00	1.00	1.00
$n = 90$	0.82	0.79	0.99	1.00	1.00	1.00
$n = 100$	0.83	0.80	0.99	1.00	1.00	1.00
$n = 150$	0.85	0.84	1.00	1.00	1.00	1.00
$n = 200$	0.87	0.86	1.00	1.00	1.00	1.00
$n = 300$	0.88	0.87	1.00	1.00	1.00	1.00

d	0.5		1		2	
$1-(1-p_1)^{250}$	0.70	0.70	0.99	0.99	1.00	1.00
	simul	$2^{nd}app$	*simul*	$2^{nd}app$	*simul*	$2^{nd}app$
$n = 10$	0.39	-0.41	0.58	0.45	0.88	1.00
$n = 20$	0.47	0.19	0.74	0.76	0.98	1.00
$n = 30$	0.52	0.37	0.81	0.85	1.00	1.00
$n = 40$	0.55	0.45	0.86	0.89	1.00	1.00
$n = 50$	0.57	0.50	0.88	0.91	1.00	1.00
$n = 60$	0.59	0.54	0.90	0.92	1.00	1.00
$n = 70$	0.60	0.56	0.91	0.93	1.00	1.00
$n = 80$	0.61	0.58	0.93	0.94	1.00	1.00
$n = 90$	0.62	0.59	0.94	0.95	1.00	1.00
$n = 100$	0.63	0.60	0.94	0.95	1.00	1.00
$n = 150$	0.65	0.64	0.96	0.96	1.00	1.00
$n = 200$	0.66	0.65	0.97	0.97	1.00	1.00
$n = 300$	0.67	0.67	0.98	0.98	1.00	1.00

d	0.5		1		2	
$1-(1-p_1)^{100}$	0.38	0.38	0.84	0.84	1.00	1.00
	simul	$2^{nd}app$	*simul*	$2^{nd}app$	*simul*	$2^{nd}app$
$n = 10$	0.24	0.19	0.41	-0.58	0.76	1.00
$n = 20$	0.28	0.23	0.53	0.23	0.93	1.00
$n = 30$	0.30	0.27	0.60	0.45	0.97	1.00
$n = 40$	0.31	0.30	0.65	0.55	0.99	1.00
$n = 50$	0.33	0.31	0.68	0.62	0.99	1.00
$n = 60$	0.34	0.32	0.70	0.65	1.00	1.00
$n = 70$	0.34	0.33	0.72	0.68	1.00	1.00
$n = 80$	0.35	0.34	0.73	0.70	1.00	1.00
$n = 90$	0.35	0.34	0.75	0.72	1.00	1.00
$n = 100$	0.35	0.35	0.75	0.73	1.00	1.00
$n = 150$	0.36	0.36	0.78	0.77	1.00	1.00
$n = 200$	0.37	0.36	0.79	0.79	1.00	1.00
$n = 300$	0.37	0.37	0.81	0.81	1.00	1.00

It is seen from the simulations in Tables 2.14 and Table 2.16 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 2.9, 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 2.8.

In line with the results presented in Table 2.13 and Table 2.15, using the same values of d, Table 2.17 shows that inserting the correction term for $g(p) = 1-(1-p)^k$

does not disturb the behavior of the control charts in the out-of-control situation. The second type of approximation derived for this function works quite well, except for an exceptionally small sample size, such as 10, where the approximation produces very unreliable results. The simulation results present a similar picture, hence confirms that this approximation is very accurate. However, the values of d seems to be very high for this function, since all the results are close to one; for reasons of uniformity in presentation we have taken the same d's here as well.

2.4.2 The effect of controlling the exceedance probability

In the previous subsection it is demonstrated that for bias correction the 'loss' in the out-of-control situation is negligible. To investigate the effect of controlling the exceedance probability in the out-of-control situation, we consider the following setup for the out-of-control situation: X_{n+1} comes from $N(\mu + d\sigma, \sigma_1^2)$ with $\sigma_1^2 = \sigma^2$, for some $d > 0$. In the case of known parameters, the false alarm rate would be $p_1 = \overline{\Phi}(u - d)$ and the question is how our corrected random false alarm rate P_n given in (2.28) behaves in comparison to this p_1. It may be nice that the in-control behavior has been corrected, but to what extent do such corrections affect the out-of-control performance? It has been discussed before that for the bias correction the loss is negligible. The explanation for this nice result lies in the fact that the estimation effects around p_1 are much smaller (in terms of relative error) than those around p. The latter arose, from the fact that p is usually extremely small and thus large relative errors will result. However, p_1 typically is not small, at least not extremely so.

As the same effect will play a role here as well, we expect the following overall picture: the bias removal approach requires moderate corrections with small effects under out-of-control, while the controlling of exceedance probabilities considered here requires large corrections with moderate out-of-control effects. Hence the practitioner can buy weak or strong protection at a low or moderate price, respectively.

Explanation on how this general description can be made specific is given as follows. In the out-of-control situation, P_n in (2.28) transforms into

$$P_n = \overline{\Phi}\left(\frac{\widehat{\mu} - \mu}{\sigma} + (u + c)\frac{\sigma^*}{\sigma} - d\right). \tag{2.68}$$

Moreover, arguing as in Subsection 2.3.2., we obtain, for both $\sigma^* = \widehat{\sigma}$ and $\sigma^* = \breve{\sigma}$,

$$g(P_n) \approx g\left(\overline{\Phi}(u - d + c + \tau Z)\right), \tag{2.69}$$

with Z distributed as $N(0, 1)$ and $\tau^2 = \dfrac{u^2 + 2}{2n}$.

As mentioned earlier, the behavior of $g(P_n)$ can be characterized in many ways, one of them is by using the following expectation:

$$Eg(P_n) \approx g\left(\overline{\Phi}(u-d+c)\right) \approx g(p_1) - cg'(p_1)\varphi(u-d), \tag{2.70}$$

where the last step only make sense for c still reasonably small.

Since the estimation effects during out of control are less important, we here choose to simply look at the relative error. If this is the case, from (2.70) we get the relative error

$$\frac{Eg(P_n) - g(p_1)}{g(p_1)} \approx -c\frac{g'(p_1)}{g(p_1)}\varphi(u-d), \tag{2.71}$$

where c as given in Subsection 2.3.2.

Therefore, the relative errors of the three functions given in (2.3), respectively equal to $-c(1/p_1)\varphi(u-d)$, $c(1/p_1)\varphi(u-d)$ and $(-c(k(1-p_1)^{k-1})/(1-(1-(1-p_1)^k))\varphi(u-d)$. Notice that the relative error of $g(p) = 1/p$ is minus the relative error of $g(p) = p$.

Focusing on $g(p) = p$, we get for the relative error a simple approximation:

$$\frac{EP_n - p_1}{p_1} = -c\kappa(u-d) \approx -4c\{1 + (u-d)\}/5, \tag{2.72}$$

where as mentioned in Remark 2.4.1, $\kappa(x) = \varphi(x)/\overline{\Phi}(x)$ and for $0 \le x \le 3.5$ this function can be approximated adequately by $4(1+x)/5$.

Using c in (2.53), (2.72) can be written as

$$\frac{EP_n - p_1}{p_1} \approx \frac{-4}{5}\left\{-\frac{\epsilon}{u} + u_\alpha\left(\frac{u^2+2}{2n}\right)^{1/2}\right\}\{1 + (u-d)\}. \tag{2.73}$$

From (2.73) we can easily see the role of n, ϵ, α and d in the out-of-control situation. As an illustration of the impact of the correction on the out-of-control situation, suppose we consider the following two examples. The first is the out-of-control situation with $d = 1.44$, where we get $p_1 = 0.05$, and the second is with $d = 2.25$, for which we get $p_1 = 0.20$. Moreover, for $p = 0.001$, $n = 100$, $\alpha = 0.2$ and $\epsilon = 0.1$, from (2.53) we obtain $c = 0.17$. Hence, the average out-of-control RL $(1/p_1)$ is increased from 20 to 27 in the first example, and from 5 to 6.3 in the second example. These examples show how to find a proper balance between adequate protection against estimation effects during in-control, and the price to be paid for it during out-of-control.

Comparing the relative error of $g(p) = p$ based on the bias criterion (cf. Remark 2.4.1) to that of the exceedance probability criterion (cf. (2.73)) we observe that the first one only depends on n and d, while the second one also depends on ϵ and α beside n and d. Using the same values of n and d, the latter gives a significantly larger relative error than the first one. This can also be observed from the fact that c based on the exceedance probability is much larger than that of the bias criterion.

This comparison shows that the correction term developed under the exceedance probability criterion gives a larger effect in the out-of-control situation than does the bias criterion. The larger correction in the exceedance probability criterion can be associated with a strong protection which requires a moderate price as compared to a weak protection with a low price in the bias case.

To end this section, we summarize the conclusion on the out-of-control behavior as follows :

- Generally speaking, inserting the correction terms does not disturb the behavior of the control charts in the out-of-control situation. Already for moderate sample sizes the corrected control charts work very well in this situation. Therefore, when no large sample sizes are available, no accurate control charts were possible; the corrected control charts provide a solution for this problem.

- For the bias case, the behavior of the first two functions mentioned in (2.3) in the out-of-control situation can be described very well by simple approximations as in (2.64)and (2.65), respectively, while the behavior of all the three functions in the out-of-control situation is also described very well by the second type approximation given in (2.67). For the exceedance probability case, the behavior of the all three functions in the out-of-control situation can be approximated by (2.70).

- For the bias case, the relative error of $g(p) = p$ is given in Remark 2.4.1. For the exceedance probability case, the relative error is given in (2.71). For $g(p) = p$ the relative error can be approximated further as given in (2.72).

- The bias removal approach requires moderate corrections with small effects under out-of-control (with relative error of order $1/n$), while controlling of exceedance probabilities requires large corrections with moderate out-of-control effects with relative error of order $1/\sqrt{n}(+\epsilon$ term). Hence, the practitioner can buy weak or strong protection at a low or moderate price, respectively.

2.5 Concluding remarks

This section provides the summary of the present chapter by outlining the conclusion obtained in each section.

In this chapter we have shown that estimation error may lead to inaccurate control limits when we simply plug in the estimators. asymptotic theory shows that we need huge sample sizes to get accurate control limits if parameter estimators are just plugged in. The validity of the asymptotic theory is confirmed by a simulation study. There are at least two ways to get an accurate control limit for practical use, either by enlarging the sample sizes or by developing a correction term for the control limits.

Since nowadays short production runs are more and more in demand, the first choice seems to be often out-of-order. Moreover, more data means higher operational costs which most practitioners will want to avoid. For this reason, we have developed (using asymptotic theory) correction terms for the control limits, based on bias and exceedance probability criteria. It turns out that these correction terms do their job very well. The correction terms depend on the involved requirement concerning the bias or the exceedance probability criteria. Applying these correction terms, we are able to obtain accurate control limits (in the sense that the corresponding requirement is fulfilled) for much smaller sample sizes. The results of the simulation study confirm the performance of these approximation results.

Having successfully reduced the bias and controlled the exceedance probability in the in-control situation, we also have investigated the performance of the corrected control limits under out-of-control situations. Fortunately, these corrected control limits still perform quite well in the out-of-control case. From the results of the approximation as well as the simulation study, we can say that the bias removal approach requires moderate corrections with small effects under the out-of-control situation. Controlling the exceedance probability on the other hand, requires large corrections with moderate out-of-control effects. Hence, the practitioners can buy weak or strong protection at a low or moderate price, respectively.

Chapter 3

Parametric Approach: Reducing the bias

Traditional control charts rely on the normality assumption for the observations. In the case that the assumption fails, both parametric and nonparametric approaches offer alternative solutions. However, since a fully nonparametric approach requires much larger sample sizes than practitioners can usually afford, the parametric approach can be considered as an alternative.

This chapter discusses a parametric control chart suitable for observations from a broader class of distributions. The parametric control chart adopts a so called normal power family as a model. With the addition of a suitable correction term in its control limit, developed using a bias criterion, the parametric chart performs very well not only in the considered parametric family, but also for common distributions outside the parametric family, thus covering a broad class of distributions.

3.1 Introduction

In Chapter 2, one of the problems concerning the underlying distribution has been discussed. Even when the assumption was fulfilled according to which the observations were drawn from a normal distribution with unknown parameters, it was shown that estimating these parameters could result in serious errors if the estimators were just plugged in without making any other attempt to adjust the control limit of the chart. This confirms what had been pointed out by several authors such as Nedumaran and Pignatiello(2001), Chakraborti(2000), Woodall and Montgomery (1999), Chen (1997), Quesenberry(1993), and Ghosh et al.(1981).

The present chapter discusses the problem which is caused by imposing the normality assumption itself on the distributions. In the case that the assumption fails, i.e. the observations, standardized for location and scale, have another distribution than the standard normal one, substantial model errors may occur. These model errors reflect the difference between the assumed model and the true model of the observations. The greater the difference between the two models, the larger the model errors that are produced. In this case, if the normal family is used as the supposed model, then all observations from outside this family are likely to have significant model errors.

To reduce these model errors substantially, it is necessary to do something about the normality assumption, either by considering a wider parametric class of models or by making no distributional assumptions at all. The last approach, which is also known as a nonparametric approach, requires much larger sample sizes. This nonparametric approach will be treated in Chapter 5. The first approach of considering a broader parametric class than the normal family is the main topic to be discussed here and in the next chapter. These chapters are concerned with developing an approach to handle the model error. Handling this model error, however, causes a serious stochastic error as a result of parameter estimation. To reduce this stochastic error, in the present chapter we also develop a correction term based on a bias criterion. How to deal with this stochastic error using an exceedance probability criterion will be discussed in the following chapter.

As described before, let X be a random variable to be monitored in a production process and let p be the false alarm rate. Suppose, for a moment, that the in-control process is modeled by assuming that X comes from a standard normal distribution. Then, as soon as $X > u_p$, the out-of-control signal is produced. However, the probability of incorrectly producing an out-of-control signal may be seriously in error when the distributional form of the observations differs from normality, see e.g. Chan et al. (1988), Pappanastos and Adams(1996), table 7 on page 222. Basically, the problem is that the normal approximation may be fair for the central part of the distribution, but produces large relative errors in the tails. And, in view of the small values of p typically used, the tails are what we are dealing with. Therefore, a larger model is needed, providing more flexibility to describe the behavior in the far tail.

For that purpose we consider the so called normal power family, containing the normal distributions as submodel. One may object that this larger parametric family again does not cover all distributions and that a nonparametric approach should be taken. As typically the upper p-quantile should be estimated, it is clear that with, say, 100 observations in Phase I for $p = 0.001$ such a quantile cannot be estimated nonparametrically. The upper p-quantile of the distribution is the point above which lies $100p$ percent of the mass. Hence, some assumptions should be made about the relation between the behavior of X in the far tail (where we should estimate an appropriate quantile) and the behavior of X somewhat more to the middle, where we have observations and can do the estimation. A similar type of argument is applied e.g. in de Haan and Sinha (1999), Hall and Weissman (1997) and Dekkers and de Haan (1989). The normal power family is on the one hand sufficiently rich, but on the other hand not that large that estimation is impossible or too inaccurate.

Using the normal power family gives more flexibility and hence an improvement over simply using normality and ignoring the well-known fact that, in practice, normality often fails which causes big errors in the false alarm rate. It is not our claim that the normal power family always is the "right" model. But firstly, if control limits based on normality are applied, this should implicitly mean that the distribution is approximately normal, and in that case the normal power family is certainly appropriate, since normality is a submodel of the normal power family. Secondly, by the extension to the normal power family many other commonly used distributions show up, which are covered sufficiently well by members of the normal power family. Thirdly, the normal power family is not so large that accurate estimation is only possible with huge sample sizes as in the nonparametric approach.

In addition to the normal power family, actually there are many other possibilities to extend the family of normal distributions, like random or deterministic mixtures. However, it turns out that several standard extensions of the family of normal distributions lead to very substantially more complicated control limits.

Although the parametric approach can reduce the model errors of many distributions considerably, more parameters need to be estimated in this new approach. Hence, a larger stochastic error may be expected due to the parameter estimation. Therefore, the next step to be taken is to control these stochastic errors using a bias criterion. A second order asymptotic approach is applied to derive the correction terms of the control limits, since first order asymptotics do not produce sufficiently accurate results, due to the very small values involved in the quantiles. With the addition of the correction term in the new control limit, the corrected parametric chart performs very well, not only in the considered parametric family, but also for common distributions outside the parametric family, thus covering a broad class of distributions.

The present chapter is organized as follows. In Section 3.2 an exposition of the problem is given with a discussion on the model error due to misspecification and the stochastic error implied by estimation. Next, some suitable models as candidates for the extended model are discussed. Subsequently, the characteristics of the parametric control chart will be presented, followed by a discussion on the model errors for the suitable models in Section 3.5. In this section some examples will be given to show advantages of using this parametric control chart. Further use of a correction term in the control limit to reduce the stochastic error is explained in Section 3.6. Using the bias criterion, the derivation and the impact of the correction terms will also be discussed. To see how well the asymptotic results come true for finite sample sizes, a simulation study is carried out and presented in Section 3.7. The results of the simulation show that the corrected parametric chart performs very well for observations drawn from a broader class than the normal power family when the process is in-control. Section 3.8 discusses the main questions stated in Section 3.2 including the performance of the corrected parametric chart in the out-of-control situation which turns out to be reasonably well. The last section provides the summary of the conclusions obtained throughout the chapter.

3.2 The main questions

Let $X_1, \ldots, X_n, X_{n+1}$ be i.i.d. random variables (r.v.'s) with distribution function F, that is, we consider the in-control situation. The out-of-control behavior, where X_{n+1} comes from a shifted distribution function $F(x-\Delta)$ is discussed later in this chapter. The r.v.'s $X_1, \ldots, X_n$ are the data from Phase I on which the estimators of the unknown parameters are based and X_{n+1} is the monitoring characteristic. The monitoring r.v. may be based on m observations, but here we consider the situation $m = 1$ of individual measurements to avoid additional complications, thus facilitating the explanation of the basic arguments.

The true, but unknown distribution function F is modeled by a parametric family $\{G_\theta : \theta \in \Theta\}$. Apart from this *general (parametric) model*, we consider also the *restricted model*, where the parameter space is restricted to a subset $\theta \in \Theta_0$; neither this model nor the parametric model is necessarily true. For any distribution function H we write $\overline{H} = 1 - H$ and H^{-1} and $\overline{H}^{-1}$ for the respective inverse functions.

If F equals G_θ and θ is known, the control limit equals $\overline{G}_\theta^{-1}(p)$. Often F is unknown and two problems arise: (i) F may be (slightly) outside the parametric family $\{G_\theta : \theta \in \Theta\}$ (or $\{G_\theta : \theta \in \Theta_0\}$ if the supposed model is the restricted model) and (ii) θ is unknown. This leads to two kinds of errors, the model error and the stochastic error. The nature of the model error and how to handle it will be the focus of the present chapter. In addition, the stochastic error as a result of estimating the parameters and how to deal with it using the bias criterion will also be discussed in this chapter.

Given a true model, we speak of the *restricted model error (RME)* and *restricted stochastic error (RSE)* if the supposed model is the restricted model and we simply speak of the *model error (ME)* and *stochastic error (SE)* if the supposed model is the general model.

Let us first consider the situation that the supposed model is the restricted model of normality with the upper control limit given in (2.1). Recall that due to estimation of the parameters the false alarm rate is no longer a number, but a r.v. which we denote by P_n. In this case, the discrepancy between P_n and p can be written as

$$P_n - p = P\left(X_{n+1} > \widehat{\mu} + \sigma^* u_p\right) - p = \overline{F}\left(\widehat{\mu} + \sigma^* u_p\right) - p = RME + RSE \quad (3.1)$$

with the restricted model error RME given by

$$RME = \overline{F}\left(\mu + \sigma u_p\right) - p = \overline{F}\left(\mu + \sigma u_p\right) - \overline{F}\left(\overline{F}^{-1}(p)\right) \quad (3.2)$$

and the restricted stochastic error RSE defined by

$$RSE = \overline{F}\left(\widehat{\mu} + \sigma^* u_p\right) - \overline{F}\left(\mu + \sigma u_p\right), \quad (3.3)$$

where $\mu = \int x dF$ and $\sigma = \sqrt{\int (x-\mu)^2 dF}$ in (3.2) and (3.3).

In general, if $\{G_\theta : \theta \in \Theta\}$ is the supposed model, the discrepancy between P_n and p can be written as

$$P_n - p = P(X_{n+1} > \overline{G}_{\widehat{\theta}}^{-1}(p)) - p = \overline{F}(\overline{G}_{\widehat{\theta}}^{-1}(p)) - p = ME + SE \quad (3.4)$$

with the model error ME given by

$$ME = \overline{F}(\overline{G}_{\theta}^{-1}(p)) - p = \overline{F}(\overline{G}_{\theta}^{-1}(p)) - \overline{F}(\overline{F}^{-1}(p)) \quad (3.5)$$

for some suitably chosen point $\theta \in \Theta$ (or $\theta \in \Theta_0$ if the supposed model is the restricted model), and the stochastic error SE defined by

$$SE = \overline{F}(\overline{G}_{\widehat{\theta}}^{-1}(p)) - \overline{F}(\overline{G}_{\theta}^{-1}(p)). \quad (3.6)$$

(On how θ is “suitably chosen” we come back below.)

The larger the parametric family, the smaller in general ME, but the more difficult the estimation problem (more parameters involved) and hence the larger SE. We want to control both ME and SE. To establish this, we consider three cases:

1. We are in the restricted model. A typical example is that our observations are normally distributed.
2. We are in the general parametric model. This model contains the restricted model. A typical example is an extension of the normal family with a few extra parameters. The general parametric model is denoted by $\{G_\theta : \theta \in \Theta\}$. Here $\theta = (\theta_1, \ldots, \theta_k)$, say, while in the restricted model only $\theta_1, \ldots, \theta_l$ vary, for some $0 \leq l < k$, and $\theta_{l+1}, \ldots, \theta_k$ are fixed. Without loss of generality,

we let these fixed values be 0. Hence the restricted model is presented by $\{G_\theta : \theta = (\theta_1, \ldots, \theta_l, 0, \ldots 0) \in \Theta\}$. In other words each $\theta \in \Theta_0$ is of the form $\theta = (\theta_1, \ldots, \theta_l, 0, \ldots 0)$.

3. We are outside the general parametric model.

Of course, the model error ME depends on the true model and the supposed model.

The "suitably chosen point $\theta \in \Theta$" can be characterized as follows. The parameter $\theta_i, i = 1, \ldots, k$, is assumed to be a functional of F, given by $\theta_i = \int \zeta_i dF$ for some function ζ_i. Tacitly it is assumed that the parameterization is such that $\theta_i = \int \zeta_i dG_\theta$. In this way the parameter has a meaning also outside the restricted or general model and this meaning coincides with the concept of the parameter within these models. For instance, the functional of F may correspond to the 0.75-quantile of F. In the restricted and general model the 0.75-quantile then leads to the suitable parameter values.

In view of the notation of the restricted model it is implied that $\int \zeta_i dF = 0$ for $i = l+1, \ldots, k$, when F belongs to the restricted model. Several functionals may be associated with the parameters $\theta_i, i = 1, \ldots, k$. The choice is determined by the estimation of the parameter as is exemplified in the following example.

Example 3.2.1

Consider the random mixture model $X = \mu + \sigma\{(1-W)Z_0 + WZ_1\}$, where W is independent of Z_0 and Z_1, $P(W=1) = 1 - P(W=0) = \gamma$ and where Z_0 and Z_1 have distribution functions K_0 and K_1, respectively. The parameter θ is denoted by (μ, σ, γ). Let $K_0 = \Phi$ and let K_1 be the distribution function of the standardized Student-distribution with 6 degrees of freedom. The parameters μ and σ are estimated by the usual location and scale estimators $\overline{X} = n^{-1} \sum X_i$ and $S = \sqrt{S^2}$ with $S^2 = (n-1)^{-1} \sum (X_i - \overline{X})^2$. The related functionals are of course $\mu = \int x dF(x)$ and $\sigma = \sqrt{\int (x-\mu)^2 dF(x)}$.

To estimate γ we may use the moment estimator based on the fourth moment. As the fourth moment of Z_1 equals 6, this leads to the estimator

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

which corresponds to the functional

$$\frac{\int \left(\frac{x-\mu}{\sigma} \right)^4 dF(x) - 3}{3}.$$

Writing $X_{m:n}$ for the m^{th} order statistic of $X_1, \ldots, X_n$, another estimator is obtained by equating a sample 90%-quantile to the 90%-quantile of the distribution, giving, with $m = [0.9n]$, where $[x]$ denotes the entier of x,

$$\frac{\frac{m}{n} - \Phi\left(\frac{X_{m:n}-\overline{X}}{S}\right)}{K_1\left(\frac{X_{m:n}-\overline{X}}{S}\right) - \Phi\left(\frac{X_{m:n}-\overline{X}}{S}\right)},$$

with corresponding functional

$$\frac{0.9 - \Phi\left(\frac{F^{-1}(0.9)-\mu}{\sigma}\right)}{K_1\left(\frac{F^{-1}(0.9)-\mu}{\sigma}\right) - \Phi\left(\frac{F^{-1}(0.9)-\mu}{\sigma}\right)}.$$

The idea is to relate the (unknown) true distribution F to a member of the general parametric model by fitting several characteristics of F to the nominated member of the family. Hopefully, this results in similar (extreme) quantiles of F and the chosen member of the family as well. The more characteristics are fitted, the better presumably the approximation in the far tail and hence the smaller the ME, but, on the other hand, the more difficult the estimation of those parameters, that is the larger the SE.

Moreover, not only the number of parameters is important, but also *which* characteristics are used. Since the estimator determines the suitably chosen point in the parameter space, the model error for distributions outside the supposed model depends on the estimator. In other words, one estimator will give a better fit w.r.t. the required extreme p-quantile in the parametric model than another. On the other hand, a lower ME may involve a larger SE. For instance, the perfect fit in terms of ME is obtained by taking $\theta = \theta(F)$, such that $\overline{G_\theta}^{-1}(p) = \overline{F}^{-1}(p)$. However, such $\theta(F)$ cannot be estimated with, say, 100 observations. In fact, with this estimator we would not use the parametric family at all: it boils down to the nonparametric approach. A similar reasoning holds when using less extreme quantiles. For example, using the preceding quantile-approach to estimate γ, with $n = 100$ observations there is no difference between the estimators using the 0.99-quantile or the 0.999-quantile. This shows that for some characteristics the SE will be very large.

Furthermore, additional functionals should be such that really different characteristics are involved.

Concerning the model error and the stochastic error, the following questions can be considered as interesting topics to discuss.

1. When the *restricted* model is true, that is $F = G_\theta$ with $\theta = (\theta_1, \ldots, \theta_l, 0, \ldots, 0)$, both the RME and ME are equal to 0. Hence, we only have to

control SE. As a rule, the SE will be larger than the RSE. How large is this difference? Can these RSE and SE be reduced by (simple) corrections? How large are the corrected RSE and SE when the restricted model holds?

2. When the general model is true, the ME equals 0, but as a rule the RME does not. How bad can the RME be? How does this balance with the larger SE? Can the RSE and SE be reduced by (simple) corrections? How large are the corrected RSE and SE when the observations are from the general model?

3. When we are outside the general model, the ME's can in principle be very large if we are far away from the general model. In Chapter 3 and 4 we concentrate on the parametric approach and hence we consider only situations outside the general model where the ME is not too big. In Chapter 5 the nonparametric approach will be considered and then also departures farther away from the general model are considered. But, as already observed in the introduction, the nonparametric approach only makes sense if some flexibility is allowed w.r.t. n and/or p: for e.g. $n = 100$ and $p = 0.001$ as given quantities, nothing can be done. How large are the uncorrected and corrected SE and the uncorrected and corrected RSE, when we are outside the general model and ME is not too big? How is the total error when applying the corrected control limit with as supposed model the general one and how does this compare with the total error when applying the corrected control limit with as supposed model the restricted model?

4. What is the impact on the out-of-control behavior of the chart? Does the process stop at a reasonable time when it has gone out-of-control?

Having discussed the main questions to deal with in this chapter, we present some suitable models as candidates for the general model in the next section.

3.3 Suitable models

The following special classes of models are candidates for a parametric model; they will also be used as important examples in the simulation study for checking and illustration of the theory developed in this thesis. Often, the restricted model will be the normal family, but adaptation to another choice, as for instance the family of exponential distributions, can be easily made.

We will take location and scale parameters μ and σ. In the restricted model these are the only parameters. In terms of $\theta = (\theta_1, \ldots, \theta_l, \theta_{l+1}, \ldots, \theta_k)$ this means that $l = 2$, $\theta_1 = \mu$ and $\theta_2 = \sigma$. The additional parameters $\theta_3, \ldots, \theta_k$ are denoted by γ. In this thesis we will restrict mainly to $k = 3$, that is, γ is a one-dimensional parameter. The distribution function G_θ is written as $G_\theta(x) = K_\gamma((x - \mu)/\sigma)$.

Hence, the control limit of the general model has the following form:

$$\widehat{\mu} + \sigma^* \overline{K}_{\widehat{\gamma}}^{-1}(p).$$

The models are defined in such a way that varying tail behavior can be described. Especially, heavier tails than those of the normal distribution are of interest. In terms of high upper quantiles this means larger values than the normal upper quantiles.

3.3.1 Random mixture

Suppose $K(\frac{x-\mu}{\sigma})$, with K fixed, represents a distribution function in a location-scale family, then the idea of the random mixture model is to replace K by K_γ which is a linear mixture of some (fixed) distribution functions. This model utilizes an additional parameter γ which is used as a weighting factor in the mixture. Here, K_γ is defined to be $K_\gamma = (1-\gamma)K_0 + \gamma K_1$, where K_0 , K_1 are fixed distribution functions having expectation 0 and variance 1. In the case that $\gamma = 0$, the model gives back the restricted model $K = K_0$, which is very often set equal to Φ.

Under this model, the random variable X can be written as

$$X = \mu + \sigma \left\{(1-W)Z_0 + WZ_1\right\},$$

where Z_0 and Z_1 have distribution functions K_0 and K_1 respectively, and W is independent of Z_0 and Z_1, with $P(W=1) = 1 - P(W=0) = \gamma$.

3.3.2 Deterministic mixture

Studying an out-of-control signal means that we have to focus on the quantiles of the distributions involved, e.g. in the family of normal distributions the p-quantile is given by $\mu + \sigma K^{-1}(p)$, with $K = \Phi$.

Like in the random mixture model, the idea behind the deterministic mixture model is to replace K^{-1} by $K_\gamma^{-1} = c(\gamma)\left\{(1-\gamma)K_0^{-1} + \gamma K_1^{-1}\right\}$, where $c(\gamma)$ is a normalizing factor to get the variance of K_γ to be equal to 1, while K_0 and K_1 are fixed distribution functions having expectation 0 and variance 1.

In this deterministic mixture model, the random variable X is written as

$$X = \mu + \sigma c(\gamma)\left\{(1-\gamma)Z_0 + \gamma Z_1\right\},$$

where $Z_0 = K_0^{-1}(U)$, $Z_1 = K_1^{-1}(U)$ and U is a random variable with a uniform distribution on $(0,1)$. The random variables Z_0 and Z_1 have distribution functions K_0 and K_1 respectively, and they are strongly dependent; in fact these r.v.'s are comonotone. The normalizing constant $c(\gamma)$ is given by

$$c(\gamma) = \left\{(1-\gamma)^2 + \gamma^2 + 2\gamma(1-\gamma)\rho\right\}^{-\frac{1}{2}}, \quad \text{where} \quad \rho = \int_0^1 K_0^{-1}(t)K_1^{-1}(t)dt.$$

Because of the technical difficulties with the random and deterministic mixture models, in this thesis we mainly deal with the following model.

3.3.3 Normal power family

This class of models is derived from the normal family by replacing the $(1-p)$-quantile of the standard normal distribution $\bar{\Phi}^{-1}(p)$ by $\bar{K}_\gamma^{-1}(p)$, where the latter depends on the value of γ. This is essentially done by replacing u_p by $u_p^{1+\gamma}$. For $\gamma > 0$ the effect will be a larger quantile, corresponding to a heavier tail than the normal one. For $\gamma < 0$ we get a smaller value according to a thinner tail. The new quantile $\bar{K}_\gamma^{-1}(p)$ is defined for $\gamma > -1$ by $K_\gamma^{-1}(t) = c(\gamma)\left|\Phi^{-1}(t)\right|^{1+\gamma} sign(\Phi^{-1}(t))$. As in the deterministic mixture, $c(\gamma)$ is a normalizing constant given by $c(\gamma) = \left\{E|Z|^{2(1+\gamma)}\right\}^{-1/2} = \pi^{1/4}2^{-(1+\gamma)/2}\Gamma(\gamma+\frac{3}{2})^{-1/2}$, where Z is a random variable having a standard normal distribution. Under this model the random variable X is given by

$$X = \mu + \sigma Z_\gamma, \text{ where } Z_\gamma = c(\gamma)|Z|^{1+\gamma} sign(Z) \text{ for } \gamma > -1.$$

We mention two further well known models. Also in these cases analytic evaluation of the estimators of the additional parameter(s) up to the required precision is difficult.

3.3.4 Tukey's λ-family

Tukey's λ distributions can be used to approximate a number of common distributions, such as a normal distribution or a logistic distribution. Using this family as a model, the r.v. X is given by

$$X = \mu + \sigma c(\lambda)\left\{U^\lambda - (1-U)^\lambda\right\},$$

where U has a uniform distribution on $(0,1)$ and $c(\lambda)$ is a normalizing constant such that $\frac{X-\mu}{\sigma}$ has variance 1, that is

$$c(\lambda) = \left[\frac{2}{2\lambda+1}\left\{1 - \frac{\lambda}{2}B(\lambda,\lambda)\right\}\right]^{-1/2},$$

in which B is the beta-function. For $\lambda = 0$, we define X in a continuous way, leading to the logistic distribution for $\frac{X-\mu}{\sigma}$. The choice $\lambda = 0.14$ 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) page 118. Introducing the parameter γ, given by $\gamma = 0.14 - \lambda$, we refer to K_γ as the distribution function corresponding to the r.v.

$$X = \mu + \sigma c(0.14 - \gamma)\left\{U^{0.14-\gamma} - (1-U)^{0.14-\gamma}\right\}.$$

3.3.5 Orthonormal family

The orthonormal family model is derived as follows. Based on the uniform distribution, a family of density functions with respect to the Lebesgue measure on $(0,1)$ is defined by

$$f(y,\varsigma) = c^*(\varsigma) exp\left\{\sum_{j=1}^{k} \varsigma_j \pi_j(y)\right\}, \text{ with } \varsigma = (\varsigma_1, \cdots, \varsigma_k),$$

where π_j is the j^{th} Legendre polynomial on $(0,1)$. Here, $c^*(\varsigma)$ serves as a normalizing constant such that f is a valid density function over its domain $(0,1)$.

Suppose Y is a random variable with density function $f(y,\varsigma)$. Now, let the expectation and standard deviation of $\Phi^{-1}(Y)$ be $E(\varsigma)$ and $c(\varsigma)^{-1}$ respectively. Using these quantities, a random variable X is given by

$$X = \mu + \sigma c(\varsigma)\left\{\Phi^{-1}(Y) - E(\varsigma)\right\}.$$

Differently from the previously stated models, the orthonormal model explicitly offers the possibility for more than one additional parameter beyond μ and σ. Of course it is still possible to get parametric models other than those mentioned above. For example we may consider to mix more than just two K_i's both in the random and deterministic models. As alternatives, we may also use a generalized version of Tukey's λ-family, as discussed by Hušková (1988), or the normal inverse Gaussian distributions as given by Barndorff-Nielsen (1997).

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 correction we need to evaluate (first and second) moments of $\widehat{\mu} + \sigma^* \overline{K}_{\widehat{\gamma}}^{-1}(p) - (\mu + \sigma \overline{K}_{\gamma}^{-1}(p))$ up to high precision. This implies that either $\overline{K}_{\gamma}^{-1}(p)$ should be analytically tractable as function of γ, or we should have a very precise approximation of $\overline{K}_{\gamma}^{-1}(p)$ by a simple function of γ. A particular problem arises in the well known random and deterministic mixture models, because negative values of γ are meaningless there. Hence, $\widehat{\gamma}$ is restricted to nonnegative values, which can often only be achieved by adding a suitable indicator function to the definition of the estimator. Due to the required precision this causes great (technical) problems, aggravated by the fact that $\widehat{\gamma}$ (and also the indicator function) is tied up with $\widehat{\mu}$ and σ^*. Apart from that, the truncation of negative values also introduces an artificial bias near $\gamma = 0$, which is also rather unattractive.

Among those models mentioned above, only the normal power family seems to be tractable. As explained above, other models involve considerable technical difficulties. For that reason, the normal power model is selected to be the extended model, and serves as the main topic to be discussed in the following section.

3.4 Parametric control chart

The idea of the parametric approach starts with a normal family as a restricted model and moves farther away from it by introducing at least one new parameter. The new class of models chosen should be wide enough to be able to cover as many distributions as possible, but not so wide that parameter estimation is no longer possible. In the rest of this chapter, the normal power family which has parameters (μ, σ, γ) is chosen as an extension of the restricted model and used as a general model for the following reasons. Having $u_p^{1+\gamma}$ in the control limit, the normal power family offers more flexibility to get distributions with heavier or lighter tails than those from the normal family. Hence, it can be considered as an improvement over simply using normality as an assumption, and ignoring the well-known fact that in practice normality often fails, which causes big errors in the false alarm rate.

Of course, we do not think that the unknown F should belong to the normal power family. But firstly, if control limits based on normality are applied, this should implicitly mean that the distribution is approximately normal. In that case, the normal power family is certainly appropriate, since normality is a submodel of the normal power family. Secondly, when one assumes normality, as many people do, it is not reasonable to have a distribution very far away from the normal family. Hence, a normal power family which can be considered in the neighborhood of normality should be sufficient. Thirdly, by the extension to the normal power family many other commonly used distributions show up, which are covered sufficiently well by members of the normal power family. Fourthly, the normal power family is not so large that accurate estimation is only possible with huge sample sizes as in the nonparametric approach.

In principle, better adaptation is possible with the normal power family than assuming only normality. Using the normal power family as a general model, a new control chart, to be called the *parametric control chart*, is constructed and applied to the Phase II observations. With the new assumption that the observations are generated from the normal power family, the random variable X is given by $X = \mu + \sigma Z_\gamma$ in which for $\gamma > -1$, Z_γ is defined as $Z_\gamma = c(\gamma)\,|Z|^{1+\gamma}\, sign\,(Z)$, where $c(\gamma) = \left\{E\,|Z|^{2(1+\gamma)}\right\}^{-1/2}$ is a normalizing constant with Z having a standard normal distribution.

While previously K_γ represents the distribution function of the general model, in what follows let K_γ denote the distribution function of Z_γ. Then the upper p-quantile $(0 < p < 1/2)$ of K_γ is given by

$$\overline{K}_\gamma^{-1}(p) = c\,(\gamma)\,u_p^{1+\gamma} = \pi^{1/4} 2^{-(1+\gamma)/2} \Gamma\left(\gamma + \frac{3}{2}\right)^{-1/2} u_p^{1+\gamma}. \tag{3.7}$$

Under this general (parametric) model, the estimated control limit in (2.1) is replaced by $\widehat{UCL} = \widehat{\mu} + \overline{K}_{\widehat{\gamma}}^{-1}(p)\sigma^*$ or, equivalently, by

$$\widehat{UCL} = \widehat{\mu} + c(\widehat{\gamma}) u_p^{1+\widehat{\gamma}} \sigma^*. \tag{3.8}$$

In what follows we use the usual location and scale estimators $\bar{X}$ and S as $\widehat{\mu}$ and σ^* (or $\check{\sigma}$, see Subsection 2.2.1), respectively. With regard to the additional parameter γ, there are several ways to estimate it. One way, which is based on the consideration that the whole distribution is being modeled, is to construct $\widehat{\gamma}$ using a moment estimator which involves all observations. In this case, since the first three moments of Z_γ equal 0, 1, and 0 respectively, the fourth moment seems to be a good starting point to construct the first estimator as follows.

The fourth moment is given by

$$\gamma_1^* = h_1(\gamma) = E{Z_\gamma}^4 = \frac{\sqrt{\pi}\,\Gamma(2\gamma + 5/2)}{\Gamma(\gamma + 3/2)^2}. \tag{3.9}$$

We estimate γ_1^* by

$$\widehat{\gamma}_1^* = n^{-1} \sum_{i=1}^{n} \left(\frac{X_i - \overline{X}}{S}\right)^4. \tag{3.10}$$

As a consequence, our first estimator $\widehat{\gamma}_1$ of γ is given by

$$\widehat{\gamma}_1 = h_1^{-1}(\widehat{\gamma}_1^*). \tag{3.11}$$

Define for a distribution function F the corresponding "suitably chosen point" by $\gamma_1(F) = h_1^{-1}(\mu_4(F))$ with $\mu_4(F) = E_{F_0} X^4$, where F_0 is the standardized F, that is the distribution function of $(X - \mu(F))/\sigma(F)$, when X has distribution function F, expectation $\mu(F)$ and standard deviation $\sigma(F)$. Obviously, when F is the distribution function of $X = \mu + \sigma Z_\gamma$, then $\mu_4(F) = \gamma_1^*$ and hence $\gamma_1(F) = \gamma$.

Concerning the second estimator of γ, instead of using every single observation in the production process, the estimator can be based merely on a small part of the distribution being modeled. In line with the objective to determine a false alarm rate, this small part should represent the intended far tail of the distribution. Since the normal power family is put forward as making the link between the (very) far tail and the behavior somewhat more to the middle, it is natural to base the estimator of the additional parameter on some order statistics in the ordinary tail. In this case we hope that the behavior in the ordinary tail has some relevance for the information in the far tail, where we need to go. For this purpose, we only rely on the largest k order statistics. The same type of reasoning is applied in extreme-value theory, see e.g. Dekkers and de Haan (1989) and Hall and Weissman (1997).

Moreover, since estimators based on only one quantile may lead to a problem since the corresponding functional $\overline{K}_\gamma^{-1}(q) = c(\gamma)(\overline{\Phi}^{-1}(q))^{1+\gamma}$ is not monotone in γ, for example for $q = 0.1$, the estimator should be based on at least 2 quantiles. For that we use two statistics in the ordinary tail as follows. Let $X_{1:n}, \ldots, X_{n:n}$ denote the order statistics of $X_1, \ldots, X_n$, and let $[x]$ denote the largest integer $\leq x$ then, for some q and r satisfying $0 < q < r < \frac{1}{2}$, define

$$\widehat{\gamma}_2^* = \frac{X_{[n+1-qn]:n} - \overline{X}}{X_{[n+1-rn]:n} - \overline{X}}. \tag{3.12}$$

This estimator is used to estimate γ_2^*, where

$$\gamma_2^* = \frac{\overline{K}_\gamma^{-1}(q)}{\overline{K}_\gamma^{-1}(r)} = (\frac{u_q}{u_r})^{1+\gamma} = h_2(\gamma), \tag{3.13}$$

and hence γ is estimated using the following equation:

$$\widehat{\gamma}_2 = h_2^{-1}(\widehat{\gamma}_2^*) = \frac{log(\widehat{\gamma}_2^*)}{log(\frac{u_q}{u_r})} - 1. \tag{3.14}$$

Our recommendation on the estimator of γ is to use the 95% and 75% sample quantiles as follows:

$$\widehat{\gamma} = 1.1218 \log\left(\frac{X_{[0.95n+1]:n} - \overline{X}}{X_{[0.75n+1]:n} - \overline{X}}\right) - 1, \tag{3.15}$$

noting that $\frac{1}{\log(u_{0.05}/u_{0.25})} \approx 1.1218$.

Let F be the true distribution function, corresponding to the r.v. X and let F_0 be the distribution function of $(X - \mu)/\sigma$. Then the suitably chosen point γ equals

$$\gamma_2(F) = \frac{\log\left(\frac{\overline{F}^{-1}(0.05)-\int x dF(x)}{\overline{F}^{-1}(0.25)-\int x dF(x)}\right)}{\log\left(\frac{u_{0.05}}{u_{0.25}}\right)} - 1 = \frac{\log\left(\frac{\overline{F_0}^{-1}(0.05)}{\overline{F_0}^{-1}(0.25)}\right)}{\log\left(\frac{u_{0.05}}{u_{0.25}}\right)} - 1 \tag{3.16}$$

and indeed this gives γ if X belongs to the normal power family with parameter γ. Obviously, the model error for distributions in the normal power family is reduced to 0. But also a reduction for distributions outside the normal power family is achieved. For instance, when the true distribution is the Student distribution with 6 degrees of freedom, the model error is reduced from 0.00356 to 0.00208. More detailed explanation on the model error and the stochastic error for the suitable models will be given in the next section.

3.5 $(R)ME$ and $(R)SE$ for the suitable models

The advantage of using the parametric chart is that it tremendously improves the model errors caused by imposing the normality assumption on the underlying distribution. As mentioned in Section 3.2, the model error describes the difference between the true model and the supposed model of the observations. In this case, the model error is denoted by RME if the supposed model is the restricted model of normality, and by ME if the supposed model is the general model of the normal power family.

Example 3.5.1

Suppose that the true distribution comes from the normal power family, that is $F(x) = K_\gamma\left(\frac{x-\mu}{\sigma}\right)$. Then

$$RME = \overline{K}_\gamma\left(u_p\right) - p = \overline{\Phi}\left(\left(\frac{u_p}{c(\gamma)}\right)^{\frac{1}{1+\gamma}}\right) - p.$$

For $p = 0.001$ and $\gamma = -0.25$ we get $c(-0.25) = 1.0783$ and hence $RME = -0.00098$. For $p = 0.001$ and $\gamma = 0.5$ we get $c(0.5) = 0.5(\sqrt[4]{2\pi})$ and hence $RME = 0.00558$, while for $p = 0.001$ and $\gamma = 1$ we have $c(1) = 3^{-0.5}$ and thus $RME = 0.00935$. Let the true distribution be the Student distribution with 6 degrees of freedom. Taking $p = 0.001$ we get $RME = 0.00356$.

Remark 3.5.1 From the calculations of RME it is seen that RME can be negative, implying a lower false alarm rate. From the point of view of the false alarm rate this looks nice. However, this will certainly have harmful consequences for the out-of-control behavior. For instance, when $\gamma = -0.25$ in the normal power family, $RME = -0.00098$. To illustrate what this might imply for the out-of-control behavior, we consider the simpler situation of a control chart for a standard normal distribution with $p = 0.001$ and $p = 0.001 - 0.00098 = 0.00002$. Then the expected run length to detect a shift 2 equals 7.3 for $p = 0.001$ and not less than 57.0 for $p = 0.00002$. Hence, both positive and negative model errors should be reduced. □

From Example 3.5.1 it can be seen that the restricted model error can be quite large. Often RME is several factors larger than the prescribed p. Therefore, if normality fails, there is a need for a larger model, thus reducing the model error.

In the following subsections, subsequently we present $R(ME)$ and $R(SE)$ of the suitable models to further comprehend the performance of the parametric chart.

3.5.1 RME and ME for the suitable models

Let the true model be the standard normal, then it is clear that $RME = 0$, while if the true model is the normal power family, the RME becomes:

$$\overline{K}_\gamma(u_p) - \overline{\Phi}(u_p) = \overline{\Phi}(\{\frac{u_p}{c(\gamma)}\}^{\frac{1}{1+\gamma}}) - p.$$

Using this expression, RME of the normal power family with various values of γ can be calculated. For example, given $p = 0.001$, if $\gamma = 0.25, 0.5, 0.75$ and 1, then $RME = 0.00266, 0.00558, 0.00786$ and 0.00935 respectively. For $\gamma < 0$, if $\gamma = -0.5, -0.25$, then $RME = -0.00100, -0.00098$ respectively. Regarding these negative model errors, one should note the following. Although the result looks nice in terms of implying a lower false alarm rate, a negative model error may cause a serious problem in the out-of-control situation, since it weakens the capability of detecting the shift of the process. The result shows that applying the normal chart to the observations having normal power distributions causes quite large model errors. The resulting RME varies from $2.7p$, $5.6p$ to almost $10p$.

For distributions outside the normal power family, the use of the normal family as the restricted model also causes some large RME's. For example, if the true model is the standardized Student distribution with 6 degrees of freedom (T), then $RME = 0.00356$, while if the true model is a Random Mixture (RM), then $RME = \gamma\{P(Z_1 > u_p) - p\}$. In the latter case, when Z_1 is distributed as T, $RME = 0.00356\gamma$. For example, if $\gamma = 0.5$, we get $RME = 0.00178$.

Now, suppose that the true model of the observations is the Deterministic Mixture (DM) with Z_1 the standardized Student distribution with 6 degrees of freedom, then $RME = t - p$ where t is the value satisfying the following equation:

$$c(\gamma)\{(1-\gamma)\overline{\Phi}^{-1}(t) + \gamma\overline{T_6}^{-1}(t)\} = u_p.$$

Given $p = 0.001$, t depends only on the value of γ. In this case, $\gamma = 0.5$ leads to $t = 0.00292$, and hence $RME = 0.00192$.

If the true model is from Tukey's λ-family, then

$$RME = P(c(\lambda)\{U^\lambda - (1-U)^\lambda\} > u_p) - \overline{\Phi}(u_p),$$

which depends on the value of λ. In this case, if λ is set to be $-0.1, 0$ and 0.14, respectively, then the resulting RME will be $0.00471, 0.00267$ and -0.00015, respectively.

Finally, if the true model is from the orthonormal family, the model error can be written as

$$RME = P(c(\varsigma)\{\Phi^{-1}(Y) - E(\varsigma)\} > u_p) - \overline{\Phi}(u_p).$$

As stated in the previous section, Y has density function $f(y,\varsigma)$. By selecting $\varsigma = (-0.1, -0.1, 0.1)$, we get $E(\varsigma) = -0.0787$, $c(\varsigma) = 1.0696$ and $RME = 0.00113$.

In the case that we use the normal power family as the supposed model, then it is obvious that if the true model is from the normal power family, the model errors will vanish right away producing $ME = 0$.

For distributions outside the normal power family, the use of the normal power family as the general model causes smaller ME as compared to RME. For example, if the true model is the standardized Student distribution with 6 degrees of freedom (T), then the model error is reduced from 0.00356 to -0.00009, and to 0.00208, respectively if we use $\widehat{\gamma}_1$ and $\widehat{\gamma}_2$ as estimators in the normal power family.

Now suppose the true model is Random Mixture (RM) with $Z_0 \sim N(0,1)$, $Z_1 \sim T_6$ and weighted parameter 0.5, then (with $\gamma_i = \gamma_i(F)$) $ME = 0.5P(Z_0 > \overline{K}_{\gamma_i}^{-1}(p)) + 0.5P(Z_1 > \overline{K}_{\gamma_i}^{-1}(p)) - p$, which equals -0.00011 and 0.00116, respectively if we use $\widehat{\gamma}_1$ and $\widehat{\gamma}_2$.

If the true model of the observations is the Deterministic Mixture (DM), with Z_0 is the standardized normal distribution and Z_1 is the standardized Student distribution with 6 degrees of freedom, and weighted parameter 0.5, then $ME = t - p$ where t is the value satisfying the following equation:

$$c(0.5)\{0.5\overline{\Phi}^{-1}(t) + 0.5\overline{T_6}^{-1}(t)\} = \overline{K}_{\gamma_i}^{-1}(p),$$

where $\gamma_i = \gamma_i(F)$.

Given $p = 0.001$, the model error is reduced from 0.00192 to 0.00016 and to 0.00128, respectively for γ_1 and γ_2.

For observations having Tukey's λ-family as the true model with λ set to be $-0.1, 0$ and 0.14, respectively, RME will be reduced from $0.00471, 0.00267$ and -0.00015, respectively to ME equal to 0.00000, 0.00024 and -0.00010, respectively for γ_1 and to ME equal to 0.00225, 0.00133 and -0.00019, respectively for γ_2.

Finally, if the true model is from the orthonormal family, the model error can be written as

$$ME = P(c(\varsigma)\{\Phi^{-1}(Y) - E(\varsigma)\} > \overline{K}_{\gamma_i}^{-1}(p)) - p,$$

where $\gamma_i = \gamma_i(F)$.

As stated above, if we use $\varsigma = (-0.1, -0.1, 0.1)$, we get $E(\varsigma) = -0.0787$, $c(\varsigma) = 1.0696$ and the model error will be reduced from $RME = 0.00113$ to $ME = 0.00033$ and to -0.00031, respectively for γ_1 and γ_2.

The values of ME together with RME are presented in Table 3.1, in which 1 unit in the table represents 0.001. In addition, the table also includes the values of ME using our recommended estimator $\widehat{\gamma}$ as given in (3.15) where we use 1.1218 instead of $1/\log(\frac{u_{0.05}}{u_{0.25}})$. Indeed, ME using $\widehat{\gamma}$ equals $\widehat{\gamma}_2$, cf. (3.14) and (3.15). This table compares RME to ME for various distributions. It shows that in general the RME's are reduced tremendously if the supposed model is extended from the restricted model (normal family) to the general parametric model (normal power family). For observations from the normal power family the parametric chart is able to cancel all the model errors, while for observations from outside the normal power family, this parametric chart manages to reduce the model errors significantly.

As shown in the table, if the true model is either the standardized Student (T), Tukey with $\lambda = -0.1$ ($TU(-0.1)$) or Tukey with $\lambda = 0$, ($TU(0)$) the benefit of using the parametric chart with the second (or recommended) estimator ($\widehat{\gamma}_2$ or $\widehat{\gamma}$) is the reduction of about half of the model error, whereas for other distributions: Random Mixture (RM), Deterministic Mixture (DM), Tukey with $\lambda = 0.14$ ($TU(0.14)$), and Orthonormal (O), the reduction is about $\frac{1}{3}$. For all distributions mentioned earlier, the reduction becomes larger if the first estimator ($\widehat{\gamma}_1$) is applied to the chart.

Table 3.1 Comparison of RME with ME

F_0	RME	ME	
		$\widehat{\gamma}_1$	$\widehat{\gamma}_2$ or $\widehat{\gamma}$
Φ	0.00	0.00	0.00
$K_{-0.5}$	-1.00	0.00	0.00
$K_{-0.25}$	-0.98	0.00	0.00
$K_{0.25}$	2.66	0.00	0.00
$K_{0.5}$	5.58	0.00	0.00
$K_{0.75}$	7.86	0.00	0.00
K_1	9.35	0.00	0.00
T	3.56	-0.09	2.08
RM	1.78	-0.11	1.16
DM	1.92	0.16	1.28
$TU(-0.1)$	4.71	0.00	2.25
$TU(0)$	2.67	0.24	1.33
$TU(0.14)$	-0.15	-0.10	-0.19
O	1.13	0.33	-0.31

Table 3.1 may suggest that $\widehat{\gamma}_1$ is better than $\widehat{\gamma}_2$ (or $\widehat{\gamma}$) which is, however, not always the case for several reasons. The first reason is that the control limit based on $\widehat{\gamma_1^*}$ is computationally more complicated due to the fact that in contrast to $h_2^{-1}(\widehat{\gamma_2^*})$ (or $h_2^{-1}(\widehat{\gamma^*})$) the function $h_1^{-1}(\widehat{\gamma}_1^*)$ is not explicitly given. As these functions take a prominent place in the control limit already in the first order term, their values should be calculated rather precisely, which is easy for $h_2^{-1}(\widehat{\gamma_2^*})$ (or $h_2^{-1}(\widehat{\gamma^*})$) and difficult for $h_1^{-1}(\widehat{\gamma}_1^*)$.

The second reason is that $\widehat{\gamma}_2$ (or $\widehat{\gamma}$) is based on order statistics and comes conceptually closer to the nonparametric approach, thus being an interesting intermediate step between assuming much knowledge about the distribution (control chart based on normality) and no assumptions at all (nonparametric control chart using very extreme order statistics). For the latter we need a huge amount of data, while the estimator $\widehat{\gamma}_2$ (or $\widehat{\gamma}$) uses more common order statistics and can be applied for the sample sizes that we meet in practice.

The third reason can be observed from the following table, in which the resulting $E(SE)$ is smaller if we apply $\widehat{\gamma}_2$ (or $\widehat{\gamma}$) than if we apply $\widehat{\gamma}_1$ for observations having a heavy tailed distribution. Using all those reasons, later in this chapter we propose a recommended control limit based on the recommended γ estimator given in (3.16).

The fourth reason is that $\widehat{\gamma}_2$ is different for the lower tail and upper tail. Therefore, when the distribution is skew, $\widehat{\gamma}_2$ fits the upper tail and lower tail separately, while $\widehat{\gamma}_1$ chooses only one member of the normal power family. In particular, in case of a heavy tail on the one hand and a thin tail on the other hand, $\widehat{\gamma}_1$ will adapt to the heavy tail (since it depends on the fourth moment). As a consequence, there maybe a large negative model error at the thin tail. Using $\widehat{\gamma}_2$ this is avoided, since both tails are considered separately.

3.5.2 RSE and SE for the suitable models

Controlling model errors by extending the supposed model from the normal family to the normal power family is nice. On the other hand, however, larger stochastic errors may be expected due to the estimation of the additional parameter γ. As described earlier in this chapter, the relation between model error and stochastic error in the restricted model and in the general model are given given in (3.1) and (3.4), respectively. Hence the relation can be used to calculate $E(RSE)$ if $E(P_n)$, RME, and p are known, or to calculate $E(SE)$ if $E(P_n)$, ME, and p are known. To calculate $E(RSE)$ and $E(SE)$ in this study we use $p = 0.001$, whereas $E(P_n)$ is obtained by simulating the quantity $E(P_n)$ using 3 different sample sizes (n): 100, 250 and 500. As expected, it turns out that the larger the sample sizes, the smaller $E(RSE)$ or $E(SE)$, as shown in Table 3.2. As in the previous table, in this table 1 unit equals 0.001.

Table 3.2 Comparison of simulated $E(RSE)$ with simulated $E(SE)$

F_0	$n = 100$			$n = 250$			$n = 500$		
	RSE	SE		RSE	SE		RSE	SE	
		$\widehat{\gamma_1}$	$\widehat{\gamma_2}$		$\widehat{\gamma_1}$	$\widehat{\gamma_2}$		$\widehat{\gamma_1}$	$\widehat{\gamma_2}$
Φ	0.36	1.26	1.69	0.14	0.46	0.62	0.07	0.22	0.28
$K_{-0.5}$	0.00	1.30	2.45	0.00	0.46	0.92	0.00	0.22	0.42
$K_{-0.25}$	0.04	1.14	1.82	0.01	0.41	0.69	0.01	0.20	0.31
$K_{0.25}$	0.78	1.42	1.39	0.30	0.54	0.53	0.15	0.26	0.24
$K_{0.5}$	1.12	1.64	1.38	0.43	0.65	0.51	0.21	0.32	0.24
$K_{0.75}$	1.45	1.87	1.37	0.59	0.76	0.50	0.29	0.40	0.23
K_1	1.78	2.09	1.36	0.66	0.88	0.51	0.36	0.48	0.24
T	0.74	2.51	1.59	0.31	1.27	0.69	0.16	0.82	0.31
RM	0.53	2.02	1.50	0.21	0.99	0.61	0.11	0.62	0.27
DM	0.53	1.90	1.57	0.21	0.86	0.62	0.11	0.51	0.32
$TU(-0.1)$	0.90	2.45	1.67	0.36	1.17	0.73	0.18	0.70	0.32
$TU(0)$	0.62	1.80	1.60	0.24	0.74	0.67	0.12	0.39	0.30
$TU(0.14)$	0.36	1.21	1.67	0.13	0.43	0.60	0.06	0.20	0.28
O	0.68	1.72	1.59	0.26	0.61	0.55	0.13	0.30	0.22

Table 3.2 shows that for all observations from the normal family, the parametric chart produces $E(SE)$ about 3 times as large as the normal chart if $\widehat{\gamma_1}$ is applied, while if $\widehat{\gamma_2}$ (or $\widehat{\gamma}$) is applied the produced $E(SE)$ will be about 4 times as large.

For the second group of observations which are from the normal power family, the resulting $E(SE)$'s are quite variable, and they very much depend on the value of γ itself. For $\gamma = -0.5$ and $\gamma = -0.25$, the $E(SE)$ increases more rapidly than those of the normal chart if $\widehat{\gamma_1}$ is applied in the parametric chart. And it is getting worse if $\widehat{\gamma_2}$ (or $\widehat{\gamma}$) is applied to the chart. It is no surprise that $E(RSE)$ is much smaller for negative γ's, since the normal chart has such large negative model error that almost no variation is possible. For $\gamma > 0$, the larger the value of the γ, the larger the resulting $E(SE)$. However, the larger the value of the γ, the smaller the difference between the resulting $E(SE)$ of the normal chart and that of the parametric chart, from about twice, if $\gamma = 0.25$, to a little bit more than one, if $\gamma = 1$. This situation holds if $\widehat{\gamma_1}$ is applied. If $\widehat{\gamma_2}$ (or $\widehat{\gamma}$) is applied to the parametric chart, the produced $E(SE)$'s are slightly better for $\gamma = 0.25$ and $\gamma = 0.5$. For $\gamma = 0.75$ and $\gamma = 1$, the resulting $E(SE)$'s are even better than those of the normal chart.

For the third group of observations which come from distributions outside the normal power family, the resulting $E(SE)$'s are about 3 or 4 times as large as those of the normal chart, for the case in which the parametric chart is based on $\widehat{\gamma_1}$. If $\widehat{\gamma_2}$ (or $\widehat{\gamma}$) is used for the parametric chart, the produced $E(SE)$ are slightly better, being about 2 or 3 times larger.

Summarizing what has been discussed above, the parametric chart has an advantage over the normal chart in reducing RME substantially. This parametric chart however, also has a disadvantage due to the estimation of the additional parameter γ, which causes somewhat larger SE. Since in this case RME is (much) larger than SE, the disadvantage of the new chart is less important. Nonetheless, in order to reduce the SE of the parametric chart which can further improve the performance of the new chart, a strategy similar to that of Chapter 2 is used to develop a correction term for the chart. The next section explains the derivation of the correction terms to get a corrected parametric chart.

3.6 Corrected parametric chart

Due to estimation the (observed) false alarm rate is a random variable P, given by (3.1) or (3.4). We want P to be close to p. To compare the stochastic P and the deterministic p, several criteria can be used. The most obvious one is to consider EP and to compare it with p. More generally, we consider a function $g(p)$, estimate it by $g(P)$ and compare $Eg(P)$ with $g(p)$. In particular, we focus on the functions g given in (2.3).

Elimination of the systematic error is of course not the only criterion. Due to the variation in the possible outcomes of $g(P)$ one may require that only in relatively few cases the prescribed $g(p)$ is exceeded. Corrections (of a stronger type than the one met here) can be proposed to satisfy such an exceedance probability criterion which will be given in the following chapter. In the present chapter we restrict attention to the bias as criterion.

Both theory and simulation results presented in the previous section clearly show that for sample sizes like $n = 100$ large stochastic errors occur. Starting with the control limit $\mu+\sigma\overline{K}_{\gamma}^{-1}(p)$, which has for known parameters μ, σ and γ, probability p of incorrectly concluding that the process is out-of-control, and subsequently simply plugging in the estimators, we arrive for $Eg(P)$ at a value unequal to $g(p)$. Therefore, we change the starting value p to s, say, in such a way that, when estimating the parameters, we end up with $Eg(P) = g(p)$ (or at least close to it). In other words, we do not use $\overline{K}_{\gamma}^{-1}(p)$, but $\overline{K}_{\gamma}^{-1}(s)$ for an appropriate value of s. Instead of $\overline{K}_{\gamma}^{-1}(s)$ we write $\overline{K}_{\gamma}^{-1}(p) + c$ with c being the correction term. One might argue that the corrected control limit could be obtained by simply adding a correction term $\widetilde{c}$. However, the smaller σ, the smaller the correction term and hence we give it the form $c\sigma$, estimate σ and arrive at

$$\widehat{\mu} + \sigma^* \left(\overline{K}_{\widehat{\gamma}}^{-1}(p) + c\right).$$

Furthermore, note that the correction term depends on n, p, the function g and γ. The last parameter is estimated again and one may ask whether this estimation step in the correction term gives no renewed bias term. However, the effect here is much smaller as the correction term itself is of order n^{-1}.

The idea behind the derivation of the correction term is as follows. With a given correction term c, firstly we give an expression for the bias (as a function of c). Then we calculate the correction term c such that it eliminates the bias. In order to do these calculations we have to make many approximations, which are of two types. There are numerical approximations to complicated functions like $\overline{K}_{\gamma}^{-1}(p)$ in order to do the analysis. Secondly, we apply asymptotics with respect to n, ignoring $o(n^{-1})$-terms. The derivation is still complicated, but fortunately leads to a proposal that can be applied straightforwardly. Later in this section, we propose a recommended control limit for the parametric chart.

The correction term, which is to be determined later on, is denoted by $c_u = c_u(\widehat{\mu}, \breve{\sigma}, \widehat{\gamma})$. This correction term is inserted in the parametric chart to get the following corrected control limit:

$$\widehat{UCL}_c = \widehat{\mu} + \breve{\sigma}\{\overline{K}_{\widehat{\gamma}}^{-1}(p) + c_u\}. \tag{3.17}$$

Using the control limit given in (3.17), the false alarm rate becomes :

$$\begin{aligned} P(X_{n+1} > \widehat{\mu} + \breve{\sigma}\{\overline{K}_{\widehat{\gamma}}^{-1}(p) + c_u\}) &= P(\frac{X_{n+1} - \mu}{\sigma} > \overline{K}_{\gamma}^{-1}(p) + V + \frac{\breve{\sigma}}{\sigma} c_u) \\ &= \overline{F_0}(\overline{K}_{\gamma}^{-1}(p) + V + \frac{\breve{\sigma}}{\sigma} c_u), \end{aligned}$$

where F_0 is the distribution function of $\sigma^{-1}(X_{n+1} - \mu)$ and where

$$V = \frac{\widehat{\mu} + \breve{\sigma}\overline{K}_{\widehat{\gamma}}^{-1}(p) - \left\{\mu + \sigma\overline{K}_{\gamma}^{-1}(p)\right\}}{\sigma} = \frac{\widehat{\mu} - \mu}{\sigma} + \frac{\breve{\sigma}}{\sigma}\overline{K}_{\widehat{\gamma}}^{-1}(p) - \overline{K}_{\gamma}^{-1}(p).$$

It is seen from the definitions of our estimators $\widehat{\mu}, \breve{\sigma}$ and $\widehat{\gamma}$ that, without loss of generality, we may assume that $\mu = 0$ and $\sigma = 1$ and that, since the process is in the in-control situation, the random variables $X_1, ..., X_n, X_{n+1}$ have distribution function F_0. As a result, the false alarm rate can be rewritten as $\overline{F_0}(\overline{K}_{\gamma}^{-1}(p) + V + \breve{\sigma}c_u)$, where

$$V = \widehat{\mu} + \breve{\sigma}\overline{K}_{\widehat{\gamma}}^{-1}(p) - \overline{K}_{\gamma}^{-1}(p). \tag{3.18}$$

For V given in (3.18), typically EV and EV^2 are of order $O(n^{-1})$ and $E|V|^k = o(n^{-1})$ for $k \geq 3$. Hence, the correction term c_u for correcting the bias is also of order $O(n^{-1})$.

Similarly as in (3.4), for some suitably chosen point γ, $g(p)$ can be expressed as follows :

$$g(\overline{F_0}(\overline{K}_\gamma^{-1}(p) + V + \breve{\sigma} c_u)) = g(p) + \left[g(\overline{F_0}(\overline{K}_\gamma^{-1}(p))) - g(p) \right]$$
$$+ \left[g(\overline{F_0}(\overline{K}_\gamma^{-1}(p) + V + \breve{\sigma} c_u)) - g(\overline{F_0}(\overline{K}_\gamma^{-1}(p))) \right]$$

In this case, the correction term is developed such that the bias of the last term is very close to 0. Using a Taylor expansion where only the first two terms are considered, since the other terms are of order smaller than $O(n^{-1})$, we get:

$$g(\overline{F_0}(\overline{K}_\gamma^{-1}(p) + V + \breve{\sigma} c_u)) - g(\overline{F_0}(\overline{K}_\gamma^{-1}(p))) \approx$$
$$-g'(\overline{F}_0(\overline{K}_\gamma^{-1}(p))) f_0(\overline{K}_\gamma^{-1}(p))(V + \breve{\sigma} c_u)$$
$$+\tfrac{1}{2} \left[g''(\overline{F}_0(\overline{K}_\gamma^{-1}(p))) f_0^2(\overline{K}_\gamma^{-1}(p)) - g'(\overline{F}_0(\overline{K}_\gamma^{-1}(p))) f_0'(\overline{K}_\gamma^{-1}(p)) \right] (V + \breve{\sigma} c_u)^2.$$

Simplification can be made by ignoring the terms of order Vc_u, $c_u(\breve{\sigma} - 1)$ and $(c_u)^k$ for $k \geq 2$. Subsequently, the right hand side of the equation is set equal to 0, also expectations are taken which leads to the first form of the correction term :

$$c_{u1} = -EV + \frac{1}{2} EV^2 \left[\frac{g''}{g'}(\overline{F}_0(\overline{K}_\gamma^{-1}(p))) f_0(\overline{K}_\gamma^{-1}(p)) - \frac{f_0'}{f_0}(\overline{K}_\gamma^{-1}(p)) \right] \tag{3.19}$$

As mentioned in the former chapter, g can be either one of the following functions : $g(p) = p$, $g(p) = \frac{1}{p}$, or $g(p) = 1 - (1-p)^k$, where $k = [\delta/p]$ with δ is a small positive number, depicting a small fraction of the average run length. As a result, $\frac{g''}{g'}(p)$ equals 0, $-\frac{2}{p}$, and $-\frac{(k-1)}{(1-p)} \approx -k \approx -\frac{\delta}{p}$, respectively. The correction term can be simplified farther by assuming that $\overline{F_0}$ is close to K_γ, in which we get :

$$\frac{g''}{g'}(\overline{F}_0(\overline{K}_\gamma^{-1}(p))) f_0(\overline{K}_\gamma^{-1}(p)) \approx \tilde{\lambda} \frac{f_0}{\overline{F}_0}(\overline{K}_\gamma^{-1}(p)), \tag{3.20}$$

with $\tilde{\lambda} = 0, -2$, and $-\delta$, respectively.

Moreover, for a wide class of F's we have

$$\frac{f_0}{\overline{F}_0} \approx -\frac{f_0'}{f_0}, \tag{3.21}$$

see Andrews (1973) for more details and examples.

The last two equations can be used to simplify the correction term given in (3.19) which lead us to the second form of the correction term as follows:

$$c_{u2} = -EV - \frac{1}{2} \lambda EV^2 \frac{f_0'}{f_0}(\overline{K}_\gamma^{-1}(p)),$$

where $\lambda = 1, -1$, and $1-\delta$, respectively.

This form, however, still consists of unknown parts which need to be estimated. The density of f_0 is supposed to be close to k_γ, which is the density of Z_γ. Moreover, EV and EV^2 depend on the unknown F_0, which can be replaced by K_γ. Hence, in both cases we need to estimate the parameter γ. Finally, we get the third form of the correction term as follows:

$$c_u = c_u(\widehat{\mu}, \check{\sigma}, \widehat{\gamma}) = -\widehat{EV} - \frac{1}{2}\lambda\widehat{EV^2}\frac{k'_{\widehat{\gamma}}}{k_{\widehat{\gamma}}}(\overline{K}_{\widehat{\gamma}}^{-1}(p)). \tag{3.22}$$

Using this correction term, the rule for providing a signal that the process may be out-of-control reads as

$$X > \widehat{\mu} + \check{\sigma}\left\{\overline{K}_{\widehat{\gamma}}^{-1}(p) - \widehat{EV} - \frac{1}{2}\lambda\widehat{EV^2}\frac{k'_{\widehat{\gamma}}}{k_{\widehat{\gamma}}}(\overline{K}_{\widehat{\gamma}}^{-1}(p))\right\}. \tag{3.23}$$

From equation (3.23) it is clear that $\widehat{EV}$, $\widehat{EV^2}$, and $\overline{K}_{\widehat{\gamma}}^{-1}(p)$ become important factors to calculate the correction term. Using either $\widehat{\gamma}_1$ or $\widehat{\gamma}_2$ as the γ estimator (see Section 3.4), V can be expressed as

$$V = \widehat{\mu} + \check{\sigma}\overline{K}_{h_i^{-1}(\widehat{\gamma}_i^*)}^{-1}(p) - \overline{K}_{h_i^{-1}(\gamma_i^*)}^{-1}(p).$$

It turns out that the derivation of $\widehat{EV}$ and $\widehat{EV^2}$ involves very complicated expressions. For that matter, we replace $\overline{K}_\gamma^{-1}(p)$, V, and V^2 by suitable quadratic approximations.

In the following subsections, consecutively we derive $\widehat{EV}$, $\widehat{EV^2}$, and $\overline{K}_{\widehat{\gamma}}^{-1}(p)$ to get the formula of the corrected control limit using $\widehat{\gamma}_1$, $\widehat{\gamma}_2$ and the recommended $\widehat{\gamma}$.

In order to do the calculations for the correction we make the following quadratic approximation in γ (with exact fit at the points $\gamma = -0.5, 0, 1$)

$$\begin{aligned}\overline{K}_\gamma^{-1}(p) \quad \approx \quad & \left(0.3849u_p^2 + 1.4927\sqrt{u_p} - 2u_p\right)\gamma^2 \\ & + \left(0.1925u_p^2 - 1.4927\sqrt{u_p} + u_p\right)\gamma + u_p.\end{aligned} \tag{3.24}$$

3.6.1 Corrected control limit using $\widehat{\gamma}_1^*$

We start with $\widehat{\gamma}_1^*$ to get an estimator of γ, where $\gamma_1^* = EX_1^4$ is the suitably chosen point in the parameter space. Regarding function h_1 given in Section 3.4, we note that, since it behaves as an exponential function, h_1^{-1} can be well approximated by

$$h_1^{-1}(\gamma_1^*) \approx 0.7363\log(\frac{\gamma_1^*}{3}). \tag{3.25}$$

Inserting (3.25) in (3.24) and applying a second order Taylor expansion gives the approximation

$$\overline{K}^{-1}_{h_1^{-1}(\widehat{\gamma}_1^*)}(p) - \overline{K}^{-1}_{h_1^{-1}(\gamma_1^*)}(p) \approx A_1(\widehat{\gamma}_1^* - \gamma_1^*) + B_1(\widehat{\gamma}_1^* - \gamma_1^*)^2,$$

with

$$\begin{aligned} A_1 &= \frac{\log\gamma_1^*}{\gamma_1^*}(0.4173u_p^2 + 1.6185\sqrt{u_p} - 2.1686u_p) \\ &\quad + \frac{1}{\gamma_1^*}(-0.3168u_p^2 - 2.8772\sqrt{u_p} + 3.1188u_p) \\ B_1 &= \frac{\log\gamma_1^*}{(\gamma_1^*)^2}(-0.2087u_p^2 - 0.8093\sqrt{u_p} + 1.0843u_p) \\ &\quad + \frac{1}{(\gamma_1^*)^2}(0.3671u_p^2 + 2.2479\sqrt{u_p} - 2.6437u_p). \end{aligned}$$

For the calculation of EV we approximate V by

$$\widehat{\mu} + (\check{\sigma} - 1)\overline{K}^{-1}_{h_1^{-1}(\gamma_1^*)}(p) + A_1(\widehat{\gamma}_1^* - \gamma_1^*) + B_1(\widehat{\gamma}_1^* - \gamma_1^*)^2 + (\check{\sigma} - 1)A_1(\widehat{\gamma}_1^* - \gamma_1^*), \tag{3.26}$$

while for EV^2 we use

$$\begin{aligned} V^2 &\approx \widehat{\mu}^2 + (\check{\sigma} - 1)^2\left(\overline{K}^{-1}_{h_1^{-1}(\gamma_1^*)}(p)\right)^2 + A_1^2(\widehat{\gamma}_1^* - \gamma_1^*)^2 + 2\widehat{\mu}(\check{\sigma} - 1)\overline{K}^{-1}_{h_1^{-1}(\gamma_1^*)}(p) \\ &\quad + 2\widehat{\mu}A_1(\widehat{\gamma}_1^* - \gamma_1^*) + 2(\check{\sigma} - 1)\overline{K}^{-1}_{h_1^{-1}(\gamma_1^*)}(p)A_1(\widehat{\gamma}_1^* - \gamma_1^*). \end{aligned} \tag{3.27}$$

The following formulas give approximations for the moments of the estimators $\widehat{\mu}, \check{\sigma}$ and $\widehat{\gamma}_1^*$ assuming that $X_1, \ldots, X_n$ are i.i.d. r.v.'s with distribution function F_0, thus having expectation 0 and variance 1. For the derivation of the correction terms in case of the normal power family, we may restrict to $F_0 = K_\gamma$ for some γ. However, we need more general formulas in Section 3.7, and therefore we only assume expectation 0 and variance 1 (and existence of further appropriate moments). Writing $\mu_i = EX_1^i$, we have

$$
\begin{aligned}
E\widehat{\mu} &= 0, \quad E\left(\check{\sigma}-1\right) \approx -\frac{\mu_4-1}{8n}, \\
E(\widehat{\gamma}_1^*-\mu_4) &\approx \frac{-2\mu_6+3\mu_4^2-5\mu_4+8\mu_3^2+6}{n} \\
E\widehat{\mu}^2 &= \frac{1}{n}, \quad E\left(\check{\sigma}-1\right)^2 \approx \frac{\mu_4-1}{4n}, \quad E\widehat{\mu}\left(\check{\sigma}-1\right) \approx \frac{\mu_3}{2n}, \\
E(\widehat{\gamma}_1^*-\mu_4)^2 &\approx \frac{\mu_8-4\mu^4\mu_6+4\mu_4^3-\mu_4^2+16\mu_3^2-8\mu_3\mu_5+16\mu_3^2\mu_4}{n}, \\
E\widehat{\mu}(\gamma_1^*-\mu_4) &\approx \frac{\mu_5-4\mu_3-2\mu_4\mu_3}{n}, \\
E(\check{\sigma}-1)(\widehat{\gamma}_1^*-\mu_4) &\approx \frac{\mu_6-\mu_4}{2n}-\frac{\mu_4(\mu_4-1)}{n}-\frac{2\mu_3^2}{n}.
\end{aligned} \tag{3.28}
$$

Hence, we get (note that by definition of the 'suitably chosen parameter' $\gamma_1^*=\mu_4$)

$$
\begin{aligned}
EV \approx\ & -\frac{\mu_4-1}{8n}\overline{K}^{-1}_{h_1^{-1}(\gamma_1^*)}(p) \\
& +A_1\frac{-2\mu_6+3\mu_4^2-5\mu_5+8\mu_3^2+6}{n} \\
& +B_1\frac{\mu_8-4\mu_4\mu_6+4\mu_4^3-\mu_4^2+16\mu_3^2-8\mu_3\mu_5+16\mu_3^2\mu_4}{n} \\
& +A_1\left\{\frac{\mu_6-\mu_4}{2n}-\frac{\mu_4(\mu_4-1)}{n}-\frac{2\mu_3^2}{n}\right\}
\end{aligned} \tag{3.29}
$$

and

$$
\begin{aligned}
EV^2 \approx\ & \frac{1}{n}+\frac{\mu_4-1}{4n}\left(\overline{K}^{-1}_{h_1^{-1}(\gamma_1^*)}(p)\right)^2 \\
& +A_1^2\frac{\mu_8-4\mu_4\mu_6+4\mu_4^3-\mu_4^2+16\mu_3^2-8\mu_3\mu_5+16\mu_3^2\mu_4}{n} \\
& +\frac{\mu_3}{n}\overline{K}^{-1}_{h_1^{-1}(\gamma_1^*)}(p)+2A_1\frac{\mu_5-4\mu_3-2\mu_4\mu_3}{n} \\
& +2\overline{K}^{-1}_{h_1^{-1}(\gamma_1^*)}(p)A_1\left\{\frac{\mu_6-\mu_4}{2n}-\frac{\mu_4(\mu_4-1)}{n}-\frac{2\mu_3^2}{n}\right\}.
\end{aligned} \tag{3.30}
$$

Returning to the normal power family, we have $\mu_i=0$ for odd i and

$$\begin{aligned}
\mu_4 &= \gamma_1^* = EZ_\gamma^4 = \frac{\sqrt{\pi}\Gamma(2\gamma+5/2)}{\Gamma(\gamma+3/2)^2}, \\
\mu_6 &= EZ_\gamma^6 = \frac{\pi\Gamma(3\gamma+7/2)}{\Gamma(\gamma+3/2)^3}, \\
\mu_8 &= EZ_\gamma^8 = \frac{\pi^{3/2}\Gamma(4\gamma+9/2)}{\Gamma(\gamma+3/2)^4}.
\end{aligned}$$

The estimated versions of EV and EV^2 are

$$\begin{aligned}
\widehat{EV} &\approx -\frac{\widehat{\gamma}_1^*-1}{8n}\overline{K}^{-1}_{h_1^{-1}(\widehat{\gamma}_1^*)}(p) + \widehat{A}_1\frac{-2\widehat{\mu}_6+3(\widehat{\gamma}_1^*)^2-5\widehat{\gamma}_1^*+6}{n} \\
&+ \widehat{B}_1\frac{\widehat{\mu}_8-4\widehat{\gamma}_1^*\widehat{\mu}_6+4(\widehat{\gamma}_1^*)^3-(\widehat{\gamma}_1^*)^2}{n} + \widehat{A}_1\left\{\frac{\widehat{\mu}_6-\widehat{\gamma}_1^*}{2n} - \frac{\widehat{\gamma}_1^*(\widehat{\gamma}_1^*-1)}{n}\right\}
\end{aligned} \tag{3.31}$$

and

$$\begin{aligned}
\widehat{EV^2} &\approx \frac{1}{n} + \frac{\widehat{\gamma}_1^*-1}{4n}\left(\overline{K}^{-1}_{h_1^{-1}(\widehat{\gamma}_1^*)}(p)\right)^2 + \widehat{A}_1^2\frac{\widehat{\mu}_8-4\widehat{\gamma}_1^*\widehat{\mu}_6+4(\widehat{\gamma}_1^*)^3-(\widehat{\gamma}_1^*)^2}{n} \\
&+ 2\overline{K}^{-1}_{h_1^{-1}(\widehat{\gamma}_1^*)}(p)\widehat{A}_1\left\{\frac{\widehat{\mu}_6-\widehat{\gamma}_1^*}{2n} - \frac{\widehat{\gamma}_1^*(\widehat{\gamma}_1^*-1)}{n}\right\},
\end{aligned} \tag{3.32}$$

where $\widehat{A}_1$ and $\widehat{B}_1$ are obtained from A_1 and B_1, respectively, by inserting $\widehat{\gamma}_1^*$ for γ_1^* and where $\widehat{\mu}_6$ and $\widehat{\mu}_8$ are obtained from μ_6 and μ_8, respectively, by inserting $h_1^{-1}(\widehat{\gamma}_1^*)$ for γ.

Finally, the control limit including the correction term using $\widehat{\gamma}_1^*$ is given by

$$\widehat{\mu} + \breve{\sigma}\left\{\overline{K}^{-1}_{h_1^{-1}(\widehat{\gamma}_1^*)}(p) - \widehat{EV} + \frac{1}{2}\lambda\widehat{EV^2}\frac{u_p^2+h_1^{-1}(\widehat{\gamma}_1^*)}{\{1+h_1^{-1}(\widehat{\gamma}_i^*)\}c\left(h_1^{-1}(\widehat{\gamma}_1^*)\right)u_p^{1+h_1^{-1}(\widehat{\gamma}_1^*)}}\right\} \tag{3.33}$$

with $\widehat{EV}$ and $\widehat{EV^2}$ replaced by the right-hand side of (3.31) and (3.32), respectively.

Remark 3.6.1 To represent the restricted model of normality, take in (3.33) $\widehat{\gamma}_1^*$ equal to $\gamma_1^* = 3$ and hence $h_1^{-1}(\widehat{\gamma}_1^*) = 0$, according to the fact that we do not need to estimate γ. Moreover, by the same reason disregard the contribution of $\overline{K}^{-1}_{h_1^{-1}(\widehat{\gamma}^*)}(p) - \overline{K}^{-1}_{h_1^{-1}(\gamma^*)}(p) \approx A_1(\widehat{\gamma}_1^*-\gamma_1^*) + B_1(\widehat{\gamma}_1^*-\gamma_1^*)^2$ in $\widehat{EV}$ and $\widehat{EV^2}$. Then (3.33) reduces to $\widehat{\mu}+\breve{\sigma}u_p+(4n)^{-1}\breve{\sigma}u_p\{1+\lambda(u_p^2+2)\}$ which is (as it should be) the correction term for the control limit based on normality given in (2.33). In this case, λ equals 1, -1 and $1-\delta$, respectively for functions g given in(2.3). Also recall that in Section 2.3 we use $\widehat{\sigma}$ instead of $\breve{\sigma}$, hence we have an extra term $1/(4n)$ here. In addition, for the case $g(p) = 1/p$ and $g(p) = 1-(1-p)^k$, we get slightly different results which involve $\varphi(u)/p$ terms (see remark following (2.33)). $\square$

3.6.2 Corrected control limit with $\widehat{\gamma}_2^*$

For the second estimator $\widehat{\gamma}_2^*$ given in (3.12), insertion of (3.14) in (3.24) and application of a second order Taylor expansion gives the approximation

$$\overline{K}^{-1}_{h_2^{-1}(\widehat{\gamma}_2^*)}(p) - \overline{K}^{-1}_{h_2^{-1}(\gamma_2^*)}(p) \approx A_2(\widehat{\gamma}_2^* - \gamma_2^*) + B_2(\widehat{\gamma}_2^* - \gamma_2^*)^2,$$

with

$$\begin{aligned} A_2 &= \frac{\log \gamma_2^*}{\gamma_2^*}\left\{\frac{0.7698u_p^2 + 2.9854\sqrt{u_p} - 4u_p}{\left(\log \frac{u_q}{u_r}\right)^2}\right\} \\ &\quad + \frac{1}{\gamma_2^*}\left\{\frac{-0.5773u_p^2 - 4.4781\sqrt{u_p} + 5u_p}{\log(\frac{u_q}{u_r})}\right\} \end{aligned}$$

and

$$\begin{aligned} B_2 &= \frac{\log \gamma_2^*}{(\gamma_2^*)^2}\left\{\frac{-0.3849u_p^2 - 1.4927\sqrt{u_p} + 2u_p}{\left(\log \frac{u_q}{u_r}\right)^2}\right\} \\ &\quad + \frac{1}{(\gamma_2^*)^2}\left\{\frac{0.3849u_p^2 + 1.4927\sqrt{u_p} - 2u_p}{\left(\log \frac{u_q}{u_r}\right)^2} + \frac{0.2887u_p^2 + 2.2391\sqrt{u_p} - 2.5u_p}{\log(\frac{u_q}{u_r})}\right\}. \end{aligned}$$

The following formulas give approximations for the moments of the estimators $\widehat{\mu}$, $\breve{\sigma}$, and $\widehat{\gamma}_2^*$, assuming that $X_1, \ldots, X_n$ are i.i.d r.v.'s with distribution function F_0, thus having expectation 0 and variance 1. We need these more general formulas in Section 3.7. For the derivation of the correction terms and the proof of Theorem 3.6.2 we only need these formulas under K_γ. In that case the error terms are

$o(n^{-1})$. Writing $x_q = F_0^{-1}(1-q)$, $q_n = 1 - \dfrac{[n+1-qn]}{n+1}$, $x_r = F_0^{-1}(1-r)$, $r_n = 1 - \dfrac{[n+1-rn]}{n+1}$ we have for $q \le r$ (if $q > r$ replace $q(1-r)$ by $r(1-q)$)

$$\begin{aligned} E\left(\widehat{\gamma}_2^* - \gamma_2^*\right) \approx\; & \frac{F_0^{-1}(1-q_n)}{F_0^{-1}(1-r_n)} - \frac{F_0^{-1}(1-q)}{F_0^{-1}(1-r)} - \frac{q(1-q)}{2nx_r}\frac{f_0'}{f_0^3}(x_q) \\ & -\frac{1}{nx_r^2}\left\{\frac{q(1-r)}{f_0(x_q)f_0(x_r)} - \frac{\int_{x_q}^{\infty} y f_0(y)\,dy}{f_0(x_q)} - \frac{\int_{x_r}^{\infty} y f_0(y)\,dy}{f_0(x_r)} + 1\right\} \\ & + \frac{x_q r(1-r)}{2nx_r^2}\frac{f_0'}{f_0^3}(x_r) + \frac{x_q}{nx_r^3}\left\{\frac{r(1-r)}{[f_0(x_r)]^2} - 2\frac{\int_{x_r}^{\infty} y f_0(y)\,dy}{f_0(x_r)} + 1\right\}, \end{aligned}$$

$$\begin{aligned} E\left(\check{\sigma} - 1\right)\left(\widehat{\gamma}_2^* - \gamma_2^*\right) \approx\; & \frac{1}{2nx_r}\left\{\frac{\int_{x_q}^{\infty}\left(y^2-1\right) f_0(y)\,dy}{f_0(x_q)} - \mu_3\right\} \\ & - \frac{x_q}{2nx_r^3}\left\{\frac{\int_{x_r}^{\infty}\left(y^2-1\right) f_0(y)\,dy}{f_0(x_r)} - \mu_3\right\}, \end{aligned}$$

$$E\widehat{\mu}\left(\widehat{\gamma}_2^* - \gamma_2^*\right) \approx \frac{1}{nx_r}\left\{\frac{\int_{x_q}^{\infty} y f_0(y)\,dy}{f_0(x_q)} - 1 - \frac{x_q}{x_r}\left[\frac{\int_{x_r}^{\infty} y f_0(y)\,dy}{f_0(x_r)} - 1\right]\right\},$$

$$E\left(\widehat{\gamma}_2^* - \gamma_2^*\right)^2 \approx \frac{1}{nx_r^2}\left\{\frac{q\,(1-q)}{\left[f_0\,(x_q)\right]^2} - 2\frac{\int_{x_q}^{\infty} y f_0\,(y)\,dy}{f_0\,(x_q)} + 1\right\}$$

$$+\frac{x_q^2}{nx_r^4}\left\{\frac{r\,(1-r)}{\left[f_0\,(x_r)\right]^2} - 2\frac{\int_{x_r}^{\infty} y f_0\,(y)\,dy}{f_0\,(x_r)} + 1\right\}$$

$$-\frac{2x_q}{nx_r^3}\left\{\frac{q\,(1-r)}{f_0\,(x_q)\,f_0\,(x_r)} - \frac{\int_{x_q}^{\infty} y f_0\,(y)\,dy}{f_0\,(x_q)} - \frac{\int_{x_r}^{\infty} y f_0\,(y)\,dy}{f_0\,(x_r)} + 1\right\}.$$

Approximating V in a similar way as in Subsection 3.6.1 and noting that in case of the normal power family we have $F_0 = K_\gamma$, $f_0 = k_\gamma$ and $f_0' = k_\gamma'$, we get

$$EV \approx -\frac{\mu_4 - 1}{8n}\overline{K}^{-1}_{h_2^{-1}(\gamma_2^*)}(p) + A_2(C_{21} + C_{22}) + B_2 C_{23}, \tag{3.34}$$

with

$$C_{21} = \frac{F_0^{-1}\,(1-q_n)}{F_0^{-1}\,(1-r_n)} - \frac{F_0^{-1}\,(1-q)}{F_0^{-1}\,(1-r)} - \frac{q\,(1-q)}{2nx_r}\frac{f_0'}{f_0^3}\,(x_q) + \frac{x_q r\,(1-r)}{2nx_r^2}\frac{f_0'}{f_0^3}\,(x_r)$$

$$-\frac{1}{nx_r^2}\left\{\frac{q\,(1-r)}{f_0\,(x_q)\,f_0\,(x_r)} - \frac{\int_{x_q}^{\infty} y f_0\,(y)\,dy}{f_0\,(x_q)} - \frac{\int_{x_r}^{\infty} y f_0\,(y)\,dy}{f_0\,(x_r)} + 1\right\}$$

$$+\frac{x_q}{nx_r^3}\left\{\frac{r\,(1-r)}{\left[f_0\,(x_r)\right]^2} - 2\frac{\int_{x_r}^{\infty} y f_0\,(y)\,dy}{f_0\,(x_r)} + 1\right\},$$

$$C_{22} = \frac{1}{2nx_r}\left\{\frac{\int_{x_q}^{\infty}\left(y^2-1\right) f_0\,(y)\,dy}{f_0\,(x_q)} - \mu_3\right\} - \frac{x_q}{2nx_r^3}\left\{\frac{\int_{x_r}^{\infty}\left(y^2-1\right) f_0\,(y)\,dy}{f_0\,(x_r)} - \mu_3\right\}$$

$$C_{23} = \frac{1}{nx_r^2}\left\{\frac{q(1-q)}{[f_0(x_q)]^2} - 2\frac{\int_{x_q}^{\infty} y f_0(y)\,dy}{f_0(x_q)} + 1\right\}$$

and

$$+\frac{x_q^2}{nx_r^4}\left\{\frac{r(1-r)}{[f_0(x_r)]^2} - 2\frac{\int_{x_r}^{\infty} y f_0(y)\,dy}{f_0(x_r)} + 1\right\}$$

$$-\frac{2x_q}{nx_r^3}\left\{\frac{q(1-r)}{f_0(x_q)\,f_0(x_r)} - \frac{\int_{x_q}^{\infty} y f_0(y)\,dy}{f_0(x_q)} - \frac{\int_{x_r}^{\infty} y f_0(y)\,dy}{f_0(x_r)} + 1\right\}.$$

By a similar approximation as in Subsection 3.6.1, for V^2 we have

$$\begin{aligned} EV^2 \approx\ & \frac{1}{n} + \frac{\mu_4 - 1}{4n}\left(\overline{K}^{-1}_{h_2^{-1}(\gamma_2^*)}(p)\right)^2 + \frac{\mu_3}{n}\overline{K}^{-1}_{h_2^{-1}(\gamma_2^*)}(p) \\ & + A_2^2 C_{23} + 2A_2\left(C_{24} + \overline{K}^{-1}_{h_2^{-1}(\gamma_2^*)}(p) C_{22}\right), \end{aligned} \tag{3.35}$$

with

$$C_{24} = \frac{1}{nx_r}\left\{\frac{\int_{x_q}^{\infty} y f_0(y)\,dy}{f_0(x_q)} - 1 - \frac{x_q}{x_r}\left[\frac{\int_{x_r}^{\infty} y f_0(y)\,dy}{f_0(x_r)} - 1\right]\right\}.$$

For the normal power family we have $\mu_i = 0$ for odd i and

$$\mu_4 = EZ_\gamma^4 = \frac{\sqrt{\pi}\,\Gamma\left(2\gamma + \frac{5}{2}\right)}{\Gamma\left(\gamma + \frac{3}{2}\right)^2}.$$

Moreover,

$$x_q = c(\gamma) u_q^{1+\gamma},\ f_0(x_q) = \frac{\varphi(u_q)}{c(\gamma)(1+\gamma)u_q^{\gamma}},\ f_0'(x_q) = -\frac{\varphi(u_q)\left\{u_q^{1-2\gamma} + \gamma u_q^{-1-2\gamma}\right\}}{\{c(\gamma)(1+\gamma)\}^2},$$

$$\int_{x_q}^{\infty} y f_0(y)dy = EZ_\gamma 1_{Z_\gamma > x_q} = c(\gamma) EZ^{1+\gamma} 1_{Z > u_q} = c(\gamma)\int_{u_q}^{\infty} x^{1+\gamma}\varphi(x)dx$$

$$\int_{x_q}^{\infty}(y^2-1)f_0(y)dy = E(Z_\gamma^2-1)1_{Z_\gamma>x_q} = E\left\{\left[c(\gamma)Z^{1+\gamma}\right]^2-1\right\}1_{Z>u_q}$$

$$= \int_{u_q}^{\infty}\left\{\left[c(\gamma)x^{1+\gamma}\right]^2-1\right\}\varphi(x)dx.$$

The estimated versions of EV and EV^2 are

$$\widehat{EV} \approx -\frac{\widehat{\mu}_4-1}{8n}\overline{K}^{-1}_{h_2^{-1}(\widehat{\gamma}_2^*)}(p)+\widehat{A}_2(\widehat{C}_{21}+\widehat{C}_{22})+\widehat{B}_2\widehat{C}_{23} \tag{3.36}$$

and

$$\begin{aligned}\widehat{EV^2} &\approx \frac{1}{n}+\frac{\widehat{\mu}_4-1}{4n}\left(\overline{K}^{-1}_{h_2^{-1}(\widehat{\gamma}_2^*)}(p)\right)^2 \\ &+\widehat{A}_2^2\widehat{C}_{23}+2\widehat{A}_2\left(\widehat{C}_{24}+\overline{K}^{-1}_{h_2^{-1}(\widehat{\gamma}_2^*)}(p)\widehat{C}_{22}\right),\end{aligned} \tag{3.37}$$

where $\widehat{A}_2$, $\widehat{B}_2$, etc. are obtained from A_2, B_2, etc. by inserting $\widehat{\gamma}_2^*$ for γ_2^* and where $\widehat{\mu}_4$ is obtained from μ_4 by inserting $h_2^{-1}(\widehat{\gamma}_2^*)$ for γ. Finally, the control limit including the correction term using $\widehat{\gamma}_2^*$ is given by

$$\widehat{\mu}+\breve{\sigma}\left\{\overline{K}^{-1}_{h_2^{-1}(\widehat{\gamma}_2^*)}(p)-\widehat{EV}+\frac{1}{2}\lambda\widehat{EV^2}\frac{u_p^2+h_2^{-1}(\widehat{\gamma}_2^*)}{\{1+h_2^{-1}(\widehat{\gamma}_2^*)\}c\left(h_2^{-1}(\widehat{\gamma}_2^*)\right)u_p^{1+h_2^{-1}(\widehat{\gamma}_2^*)}}\right\} \tag{3.38}$$

with $\widehat{EV}$ and $\widehat{EV^2}$ replaced by the right hand sides of (3.37) and (3.38), respectively.

3.6.3 Recommended control limit

For the reasons mentioned earlier, the corrected control limit given in (3.33) is computationally more complicated than (3.38). However, for practical implementation the correction term given by (3.38) is also rather complicated. Therefore, in this subsection we present a simpler approximation and completely explicit control limit which can be applied straightforwardly. Basically, this control limit is derived from (3.38) by fitting polynomials in γ and u_p to

$$A_2,\ n\left[EV-A_2\left\{\frac{F_0^{-1}(1-q_n)}{F_0^{-1}(1-r_n)}-\frac{F_0^{-1}(1-q)}{F_0^{-1}(1-r)}\right\}\right]\quad\text{and}\quad n\times\text{coefficient of }\lambda.$$

This leads to the following control limit

$$\widehat{\mu} + \breve{\sigma} \left\{ \overline{K}_{\widehat{\gamma}}^{-1}(p) - C1(\widehat{\gamma})\, C2(\widehat{\gamma}) - \frac{C3(\widehat{\gamma})}{n} + \lambda \frac{C4(\widehat{\gamma})}{n} \right\}, \tag{3.39}$$

where $\lambda = 1, -1$, or $1-\delta$ according to $g(p) = p$, $g(p) = \frac{1}{p}$, or $g(p) = 1-(1-p)^k$, $k = [\delta/p]$, respectively, and $\widehat{\gamma}$ is given in (3.15), and where moreover

$$\begin{aligned} C1(\gamma) &= -1.23 - 0.63\gamma + 0.73\gamma^2 + 0.74u_p - 0.08\gamma u_p - 0.14\gamma^2 u_p, \\ C2(\gamma) &= \left(\frac{\Phi^{-1}\left(\frac{[0.95n+1]}{n+1}\right)}{\Phi^{-1}\left(\frac{[0.75n+1]}{n+1}\right)} \right)^{1+\gamma} - 2.4387^{1+\gamma}, \\ C3(\gamma) &= -10.86 - 27.77\gamma - 22.36\gamma^2 + 4.72u_p + 9.98\gamma u_p + 7.29\gamma^2 u_p, \\ C4(\gamma) &= -87.23 - 147.89\gamma - 104.29\gamma^2 + 40.25u_p + 63.69\gamma u_p + 44.47\gamma^2 u_p. \end{aligned}$$

Remark 3.6.2 Incidentally, $\left(X_{[0.95n+1]:n} - \overline{X}\right) / \left(X_{[0.75n+1]:n} - \overline{X}\right)$ may be negative. This happens with extremely small probability and therefore it does not matter much how this is repaired; we may take e.g. the absolute value, leading to $\widehat{\gamma} = 1.1218 \log \left(\left| \frac{X_{[0.95n+1]:n} - \overline{X}}{X_{[0.75n+1]:n} - \overline{X}} \right| \right) - 1.$ □

Next, we will show that for the recommended control limit given in (3.39), $Eg(P)$ is close to $g(p)$. Hence it fulfills the objective of reducing the bias. For the estimators $\widehat{\mu}$, $\breve{\sigma}$ and $\widehat{\gamma}$ we restrict attention to neighborhoods of $\mu(=0)$, $\sigma(=1)$ and γ. The error involved by this is presented in the following theorem. For mathematical convenience the formal definition of the estimator $\widehat{\gamma}$ in the theorem is the "exact" representation with $1/\log(u_{0.05}/u_{0.25})$ instead of its numerical value up to four decimals 1.1218 (cf. (3.15)). Similarly, we use the "exact" form of $C2$ with $u_{0.05}/u_{0.25}$ instead of its numerical value up to four decimals 2.4387. The difference due to this replacement is extremely small (e.g. $\bar{K}_\gamma(\bar{K}_{\tilde{\gamma}}^{-1}(p)) - \bar{K}_\gamma(\bar{K}_\gamma^{-1}(p))$ with $\tilde{\gamma} = 1.1218 \log(u_{0.05}/u_{0.25})^{1+\gamma} - 1$ is of order 10^{-7}) and is therefore ignored.

Theorem 3.6.1 *Let $X_1, \ldots, X_n$ be i.i.d. r.v.'s with a normal power distribution with parameter γ. Then for each $\varepsilon > 0$*

$$\limsup_{n\to\infty} n^{-\min(1,\, 2/(1+\gamma))} \log P\left(\left|\overline{X}\right| > \varepsilon \right) < 0, \tag{3.40}$$

$$\limsup_{n\to\infty} n^{-\min(1,\, 1/(1+\gamma))} \log P\left(\left|S^2 - 1\right| > \varepsilon \right) < 0 \tag{3.41}$$

and

$$\limsup_{n\to\infty} n^{-\min(1,\,2/(1+\gamma))} \log P\left(|\widehat{\gamma}_2 - \gamma| > \varepsilon\right) < 0. \tag{3.42}$$

The following proof is based on large deviation theory. 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 this theorem and the proof are not quite standard.

Proof. Denote by $X_{(1)} \leq \ldots \leq X_{(n)}$ the order statistics of $X_1, \ldots, X_n$ and let $U_{(1)} \leq \ldots \leq U_{(n)}$ be the order statistics of the random sample $U_1, \ldots, U_n$ from a uniform distribution on (0,1). Let $0 < s < 1$ be fixed and let $j = j(n)$ satisfy $\lim_{n\to\infty} j/n = s$. Then, for any $\varepsilon > 0$,

$$\begin{aligned} P(X_{(j)} > K_\gamma^{-1}(s) + \varepsilon) &= P(U_{(j)} > K_\gamma(K_\gamma^{-1}(s) + \varepsilon)) \\ &= P\left(\sum_{i=1}^{n} 1_{U_i > K_\gamma(K_\gamma^{-1}(s)+\varepsilon)} \geq n - j\right) \end{aligned} \tag{3.43}$$

and by Chernoff's theorem, see e.g. Bahadur (1971), we get

$$\lim_{n\to\infty} n^{-1} \log P\left(\sum_{i=1}^{n} 1_{U_i > K_\gamma(K_\gamma^{-1}(s)+\varepsilon)} \geq n - j\right) < 0.$$

Similarly, we have

$$\lim_{n\to\infty} n^{-1} \log P(X_{(j)} < K_\gamma^{-1}(s) - \varepsilon) < 0$$

and hence

$$\lim_{n\to\infty} n^{-1} \log P\left(\left|X_{(j)} - K_\gamma^{-1}(s)\right| > \varepsilon\right) < 0. \tag{3.44}$$

If $1 + \gamma \leq 2$, application of Chernoff's theorem easily gives, for any $\varepsilon > 0$,

$$\lim_{n\to\infty} n^{-1} \log P\left(\left|\overline{X}\right| > \varepsilon\right) < 0. \tag{3.45}$$

For $1 + \gamma > 2$, the moment generating function of X_i does not exist. However, by Nagaev (1969), see also Nagaev (1979, formulas (2.31) and (2.32) on page 764), it follows that

$$\limsup_{n\to\infty} n^{-2/(1+\gamma)} \log P\left(\overline{X} > \varepsilon\right) \le \lim_{n\to\infty} n^{-2/(1+\gamma)} \log P\left(X_1 > n\varepsilon/2\right)$$

$$= \lim_{n\to\infty} n^{-2/(1+\gamma)} \log P\left(c(\gamma)\,|Z|^{1+\gamma}\, sign(Z) > n\varepsilon/2\right) = -\tfrac{1}{2}\left(\tfrac{\varepsilon/2}{c(\gamma)}\right)^{2/(1+\gamma)}. \tag{3.46}$$

By symmetry, we get

$$P\left(\overline{X} < -\varepsilon\right) = P\left(\overline{X} > \varepsilon\right)$$

and hence we obtain for $\gamma > 1$

$$\limsup_{n\to\infty} n^{-2/(1+\gamma)} \log P\left(\left|\overline{X}\right| > \varepsilon\right) \le -\frac{1}{2}\left(\frac{\varepsilon/2}{c(\gamma)}\right)^{2/(1+\gamma)}. \tag{3.47}$$

In view of (3.45) and (3.47) we arrive at

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

thus proving (3.40).

Since

$$\begin{aligned} P\left(\left|S^2-1\right| > \varepsilon\right) &\le P\left(\left|S^2-1\right| > \varepsilon, \left|\overline{X}\right| < \varepsilon\right) + P\left(\left|\overline{X}\right| > \varepsilon\right) \\ &\le P\left(\left|\sum_{i=1}^{n}\left(X_i^2-1\right)\right| > (n-1)\varepsilon - n\varepsilon^2 - 1\right) + P\left(\left|\overline{X}\right| > \varepsilon\right), \end{aligned}$$

it follows by an argument similar to that in the proof of (3.40) that (3.41) holds true.

By continuity there exist for each $\varepsilon > 0$ constants $\varepsilon_i = \varepsilon_i(\varepsilon), i = 1,2,3$ such that

$$\begin{aligned} P\left(|\widehat{\gamma}_2 - \gamma| > \varepsilon\right) \quad &\le \quad P\left(\left|X_{([0.95n+1])} - K_\gamma^{-1}(0.95)\right| > \varepsilon_1\right) \\ &\quad + P\left(\left|X_{([0.75n+1])} - K_\gamma^{-1}(0.75)\right| > \varepsilon_2\right) + P\left(\left|\overline{X}\right| > \varepsilon_3\right). \end{aligned}$$

The result now easily follows from (3.44) and (3.47). ■

Assume that $X_1, \ldots, X_n$ are i.i.d. having a normal power distribution with parameter γ and define

$$A_n(\varepsilon) = \left\{|\widehat{\mu}| \le \varepsilon, \left|\check{\sigma}^2 - 1\right| \le \varepsilon, |\widehat{\gamma}_2^* - \gamma_2^*| \le \varepsilon\right\}$$

and $A_n^c(\varepsilon)$ its complement, then Theorem 3.6.1 implies

$$P\left(A_n^c(\varepsilon)\right) \le \exp\left\{-\eta n^{\min(1,\, 1/(1+\gamma))}\right\} \quad \text{for some} \quad \eta > 0$$

and hence for each $\varepsilon > 0$ we have

$$P\left(A_n^c(\varepsilon)\right) = o(n^{-1}) \text{ as } n \to \infty.$$

Since $V = \widehat{\mu} + \check{\sigma}\overline{K}_{\widehat{\gamma}_2}^{-1}(p) - \overline{K}_{\gamma}^{-1}(p)$ (cf. (3.18)) and $\widehat{\gamma}_2 = h_2^{-1}\left(\widehat{\gamma}_2{}^*\right)$ (cf. (3.14)), Taylor expansion gives

$$\begin{aligned} EV\ 1_{A_n(\varepsilon)} &= E\widehat{\mu}1_{A_n(\varepsilon)} + E\check{\sigma}1_{A_n(\varepsilon)}\overline{K}_{\gamma}^{-1}(p) - \overline{K}_{\gamma}^{-1}(p) \\ &\quad + E\check{\sigma}\left(\widehat{\gamma}_2{}^* - \gamma_2^*\right)1_{A_n(\varepsilon)}\frac{\partial \overline{K}_{h_2^{-1}(\gamma_2^*)}^{-1}(p)}{\partial \gamma_2^*} \\ &\quad + \frac{1}{2}E\check{\sigma}\left(\widehat{\gamma}_2^* - \gamma_2^*\right)^2 1_{A_n(\varepsilon)}\frac{\partial^2 \overline{K}_{h_2^{-1}(\gamma_2^*)}^{-1}(p)}{\partial \gamma_2^{*2}} + o(n^{-1}), \end{aligned} \tag{3.48}$$

where $1_{A_n(\varepsilon)}$ is the indicator function of the set $A_n(\varepsilon)$.

Let

$$\begin{aligned} D1_n(\gamma) &= -\frac{\mu_4 - 1}{8n}\overline{K}_{\gamma}^{-1}(p) + \left\{C_{21}(\gamma) + C_{22}(\gamma)\right\}\frac{\partial \overline{K}_{h_2^{-1}(\gamma_2^*)}^{-1}(p)}{\partial \gamma_2^*} \\ &\quad + \frac{1}{2}C_{23}(\gamma)\frac{\partial^2 \overline{K}_{h_2^{-1}(\gamma_2^*)}^{-1}(p)}{\partial \gamma_2^{*2}}, \end{aligned} \tag{3.49}$$

where $C_{21}(\gamma)$, $C_{22}(\gamma)$ and $C_{23}(\gamma)$ are C_{21}, C_{22} and C_{23} applied to the normal power family with parameter γ. Then, using formulas derived in Subsection 3.6.2 we get

$$EV\ 1_{A_n(\varepsilon)} = D1_n(\gamma) + o(n^{-1}).$$

Note that the formulas of Subsection 3.6.2 hold with error term $o(n^{-1})$ when $X_1, \ldots, X_n$ are i.i.d. having a normal power distribution with parameter γ.

Similarly, we have

$$
\begin{aligned}
EV^2 1_{A_n(\varepsilon)} &= E\widehat{\mu}^2 1_{A_n(\varepsilon)} + E\left(\check{\sigma} 1_{A_n(\varepsilon)} - 1\right)^2 \left\{\overline{K}_\gamma^{-1}(p)\right\}^2 \\
&+ E\left(\widehat{\gamma}_2^* - \gamma_2^*\right)^2 1_{A_n(\varepsilon)} \left\{\frac{\partial \overline{K}_{h_2^{-1}(\gamma_2^*)}^{-1}(p)}{\partial \gamma_2^*}\right\}^2 \\
&+ 2E\widehat{\mu}\left(\check{\sigma} - 1\right) 1_{A_n(\varepsilon)} \overline{K}_\gamma^{-1}(p) + 2E\widehat{\mu}\left(\widehat{\gamma}_2^* - \gamma_2^*\right) 1_{A_n(\varepsilon)} \frac{\partial \overline{K}_{h_2^{-1}(\gamma_2^*)}^{-1}(p)}{\partial \gamma_2^*} \\
&+ 2E\left(\check{\sigma} - 1\right)\left(\widehat{\gamma}_2^* - \gamma_2^*\right) 1_{A_n(\varepsilon)} \overline{K}_\gamma^{-1}(p) \frac{\partial \overline{K}_{h_2^{-1}(\gamma_2^*)}^{-1}(p)}{\partial \gamma_2^*} + o(n^{-1})
\end{aligned} \tag{3.50}
$$

and hence

$$
EV^2 1_{A_n(\varepsilon)} = D2_n(\gamma) + o(n^{-1})
$$

with

$$
\begin{aligned}
D2_n(\gamma) &= \frac{1}{n} + \frac{\mu_4 - 1}{4n}\left\{\overline{K}_\gamma^{-1}(p)\right\}^2 + C_{23}(\gamma)\left\{\frac{\partial \overline{K}_{h_2^{-1}(\gamma_2^*)}^{-1}(p)}{\partial \gamma_2^*}\right\}^2 \\
&+ 2\left\{C_{24}(\gamma) + C_{22}(\gamma)\overline{K}_\gamma^{-1}(p)\right\} \frac{\partial \overline{K}_{h_2^{-1}(\gamma_2^*)}^{-1}(p)}{\partial \gamma_2^*}.
\end{aligned} \tag{3.51}
$$

Noting that

$$
\frac{\overline{K}_\gamma^{-1}(1-q_n)}{\overline{K}_\gamma^{-1}(1-r_n)} - \frac{\overline{K}_\gamma^{-1}(1-q)}{\overline{K}_\gamma^{-1}(1-r)} = O(n^{-1})
$$

it is immediately seen that $D1_n(\gamma) = O(n^{-1})$ and $D2_n(\gamma) = O(n^{-1})$.

Remark 3.6.3 Inserting the numerical approximations A_2 and B_2 given in Subsection 3.6.2 in $D1_n(\gamma)$ and $D2_n(\gamma)$ leads to the approximation $\widetilde{D1}_n(\gamma)$ of EV and $\widetilde{D2}_n(\gamma)$ of EV^2 given by (cf. also (3.34) and (3.35))

$$
\begin{aligned}
\widetilde{D1}_n(\gamma) &= -\frac{\mu_4-1}{8n}\overline{K}_\gamma^{-1}(p) + \{C_{21}(\gamma)+C_{22}(\gamma)\}A_2 + C_{23}(\gamma)B_2 \\
\widetilde{D2}_n(\gamma) &= \frac{1}{n} + \frac{\mu_4-1}{4n}\left\{\overline{K}_\gamma^{-1}(p)\right\}^2 + C_{23}(\gamma)A_2^2 + \\
& \quad 2\left\{C_{24}(\gamma) + C_{22}(\gamma)\overline{K}_\gamma^{-1}(p)\right\}A_2,
\end{aligned} \tag{3.52}
$$

where as before $C_{21}(\gamma)$, $C_{22}(\gamma)$, $C_{23}(\gamma)$ and $C_{24}(\gamma)$ are C_{21}, C_{22}, C_{23} and C_{24} applied to the normal power family with parameter γ. Note that A_2 and B_2 are in fact also functions of γ, see Subsection 3.6.2. □

Let

$$
c_{u1}(\widehat{\gamma}_2) = \left\{-D1_n(\widehat{\gamma}_2) + \frac{1}{2}D2_n(\widehat{\gamma}_2)\left[\frac{g''(p)}{g'(p)}k_{\widehat{\gamma}_2}\left(\overline{K}_{\widehat{\gamma}_2}^{-1}(p)\right) - \frac{k'_{\widehat{\gamma}_2}}{k_{\widehat{\gamma}_2}}\left(\overline{K}_{\widehat{\gamma}_2}^{-1}(p)\right)\right]\right\}1_{A_n(\varepsilon)}.
$$

Since $D1_n(\gamma) = O(n^{-1})$ and $D2_n(\gamma) = O(n^{-1})$, we get

$$
Ec_{u1}(\widehat{\gamma}_2) = -D1_n(\gamma) + \frac{1}{2}D2_n(\gamma)\left[\frac{g''(p)}{g'(p)}k_\gamma\left(\overline{K}_\gamma^{-1}(p)\right) - \frac{k'_\gamma}{k_\gamma}\left(\overline{K}_\gamma^{-1}(p)\right)\right] + o(n^{-1}).
$$

Let

$$
c_{u2}(\widehat{\gamma}_2) = \left\{-C1(\widehat{\gamma}_2)\,C2(\widehat{\gamma}_2) - \frac{C3(\widehat{\gamma}_2)}{n} + \lambda\frac{C4(\widehat{\gamma}_2)}{n}\right\}1_{A_n(\varepsilon)}, \text{ cf. (3.39) and (3.17),}
$$

then

$$
Ec_{u2}(\widehat{\gamma}_2) = -C1(\gamma)\,C2(\gamma) - \frac{C3(\gamma)}{n} + \lambda\frac{C4(\gamma)}{n} + o(n^{-1}).
$$

The corresponding false alarm rates are denoted by $P^{(1)}$ and $P^{(2)}$, respectively, that is

$$
P^{(j)} = \overline{K}_\gamma\left(\overline{K}_\gamma^{-1}(p) + V + \check{\sigma}c_{uj}(\widehat{\gamma}_2)\right), j = 1, 2.
$$

The correction term $c_{u2}(\widehat{\gamma})$ is on the set $A_n(\varepsilon)$ the same as our recommended one. So, apart from a set with extremely small probability, the false alarm rate of the recommended control limit equals $P^{(2)}$. Define

$$\begin{aligned} W_g(p,\gamma) &= -D1_n(\gamma) + \frac{1}{2}D2_n(\gamma)\left[\frac{g''(p)}{g'(p)}k_\gamma\left(\overline{K}_\gamma^{-1}(p)\right) - \frac{k'_\gamma}{k_\gamma}\left(\overline{K}_\gamma^{-1}(p)\right)\right] \\ &\quad - \left\{-C1(\gamma)\,C2(\gamma) - \frac{C3(\gamma)}{n} + \lambda\frac{C4(\gamma)}{n}\right\}, \end{aligned}$$

then $Ec_{u1}(\widehat{\gamma}_2) - Ec_{u2}(\widehat{\gamma}_2) = W_g(p,\gamma) + o(n^{-1})$.

Theorem 3.6.2 *Let $X_1,\ldots,X_n,X_{n+1}$ be i.i.d. r.v.'s with a normal power distribution with parameter γ. Let g be three times continuously differentiable at p. Then for sufficiently small $\varepsilon > 0$ we have*

$$E\left\{g\left(P^{(1)}\right)1_{A_n(\varepsilon)}\right\} = g(p) + o(n^{-1}) \text{ as } n \to \infty, \tag{3.53}$$

and

$$\begin{aligned} &\left|E\left\{g\left(P^{(1)}\right)1_{A_n(\varepsilon)}\right\} - E\left\{g\left(P^{(2)}\right)1_{A_n(\varepsilon)}\right\}\right| \\ &\le \left|W_g(p,\gamma)g'(p)\,k_\gamma\left(\overline{K}_\gamma^{-1}(p)\right)\right| n^{-1} + o(n^{-1}) \end{aligned} \tag{3.54}$$

as $n \to \infty$. For $g(p) = p$ we get $\left|W_g(p,\gamma)g'(p)\,k_\gamma\left(\overline{K}_\gamma^{-1}(p)\right)\right| \le 14.4p$ for $-0.5 \le \gamma \le 1, 0.001 \le p \le 0.01$ and hence a maximal relative error equal to $14.4n^{-1} + o(n^{-1})$ as $n \to \infty$.

If g is bounded, then $E\left\{g\left(P^{(j)}\right)1_{A_n^c(\varepsilon)}\right\} = o(n^{-1})$ as $n \to \infty$ for $j = 1,2$ and hence $1_{A_n(\varepsilon)}$ can be deleted in (3.53) and (3.54).

Proof. Let $X_1,\ldots,X_n,X_{n+1}$ be i.i.d. r.v.'s with a normal power distribution with parameter γ. Let g be three times continuously differentiable at p and let $\varepsilon > 0$ be sufficiently small. It is easily seen that both $Ec_{u1}(\widehat{\gamma}_2)$ and $Ec_{u2}(\widehat{\gamma}_2)$ are $O(n^{-1})$ as $n \to \infty$. By Taylor's expansion, noting that $P\left(A_n^c(\varepsilon)\right) = o(n^{-1})$, we get for $j = 1,2$

$$
\begin{aligned}
& E\left\{g\left(P^{(j)}\right) 1_{A_n(\varepsilon)}\right\} = E\left\{g\left(\overline{K}_\gamma\left(\overline{K}_\gamma^{-1}(p) + V\ 1_{A_n(\varepsilon)} + \check{\sigma} c_{uj}(\widehat{\gamma}_2)\right)\right) 1_{A_n(\varepsilon)}\right\} \\
= \ & g(p) - E\left\{V\ 1_{A_n(\varepsilon)} + \check{\sigma} c_{uj}(\widehat{\gamma}_2)\right\} g'(p)\, k_\gamma\left(\overline{K}_\gamma^{-1}(p)\right) \\
& + \frac{1}{2} E\left\{V\ 1_{A_n(\varepsilon)} + \check{\sigma} c_{uj}(\widehat{\gamma}_2)\right\}^2 \left\{g''(p)\, k_\gamma^2\left(\overline{K}_\gamma^{-1}(p)\right) - g'(p)\, k_\gamma'\left(\overline{K}_\gamma^{-1}(p)\right)\right\} \\
& + o(n^{-1}) \\
= \ & g(p) - E\left\{V\ 1_{A_n(\varepsilon)} + c_{uj}(\widehat{\gamma}_2)\right\} g'(p)\, k_\gamma\left(\overline{K}_\gamma^{-1}(p)\right) \\
& + \frac{1}{2} E V^2 1_{A_n(\varepsilon)} \left\{g''(p)\, k_\gamma^2\left(\overline{K}_\gamma^{-1}(p)\right) - g'(p)\, k_\gamma'\left(\overline{K}_\gamma^{-1}(p)\right)\right\} + o(n^{-1}) \\
= \ & g(p) - \left\{D1_n(\gamma) + Ec_{u1}(\widehat{\gamma}_2)\right\} g'(p)\, k_\gamma\left(\overline{K}_\gamma^{-1}(p)\right) \\
& + \frac{1}{2} D2_n(\gamma) \left\{g''(p)\, k_\gamma^2\left(\overline{K}_\gamma^{-1}(p)\right) - g'(p)\, k_\gamma'\left(\overline{K}_\gamma^{-1}(p)\right)\right\} \\
& + \left\{Ec_{u1}(\widehat{\gamma}) - Ec_{uj}(\widehat{\gamma}_2)\right\} g'(p)\, k_\gamma\left(\overline{K}_\gamma^{-1}(p)\right) + o(n^{-1}) \\
= \ & g(p) + \left\{Ec_{u1}(\widehat{\gamma}_2) - Ec_{uj}(\widehat{\gamma}_2)\right\} g'(p)\, k_\gamma\left(\overline{K}_\gamma^{-1}(p)\right) + o(n^{-1})
\end{aligned}
$$

as $n \to \infty$. This completes the proof of (3.53) and (3.54). Numerical calculations show that $\left|W_g(p,\gamma) g'(p)\, k_\gamma\left(\overline{K}_\gamma^{-1}(p)\right)\right| \leq 14.4p$ for $-0.5 \leq \gamma \leq 1, 0.001 \leq p \leq 0.01$ when $g(p) = p$. If g is bounded, then $E\left\{g\left(P^{(j)}\right) 1_{A_n^c(\varepsilon)}\right\} = O\left(P\left(A_n^c(\varepsilon)\right)\right) = o(n^{-1})$. ■

Having developed the correction term, in the following section we perform a simulation study and some numerical calculations to evaluate the performance of the recommended chart as compared to the normal chart. It turns out that the recommended chart given in (3.39) works very well both in reducing the model errors and the stochastic errors.

3.7 Simulation and other numerical results

A simulation study is performed to see to what kind of improvement the various steps lead: firstly, extending the restricted model of normality to the larger model of the normal power family and secondly, the application of the correction terms in the larger model. For comparison we also consider the restricted model with correction terms. In the simulation study we want to cover the restricted model of normality, the normal power family and distributions outside the normal power family (but not too far away from it).

As criterion we take EP and compare this to p. In terms of Section 3.6 this means that we take $g(p) = p$ in the simulation study. Similar results as presented here hold for the other mentioned functions cf. (2.3). In the simulations we always choose $p = 0.001$ and for n we take $100, 250$ and 500. Our main attention is focused on $n = 100$, since nowadays large samples are usually not available. The columns at the following tables with $n = 250, 500$ give an impression of the rate of convergence. The number of repetitions in the simulation study equals $100, 000$.

In the simulation study we take the following distributions (with $\mu = 0$, $\sigma = 1$):

Φ: standard normal distribution function;

K_γ: normal power distribution function with $\gamma = -0.5, -0.25, 0 (= \Phi)$, 0.25, $0.5, 0.75, 1$;

T: standardized Student distribution function with 6 degrees of freedom;

RM: random mixture: $\frac{1}{2}\Phi + \frac{1}{2}T$;

DM: deterministic mixture, given by: $DM^{-1} = c^* \left\{\Phi^{-1} + T^{-1}\right\}$ with c^* a constant to ensure unit variance;

TU: Tukey's λ-family, with $\lambda = -0.1$, $\lambda = 0$ (standardized logistic distribution), $\lambda = 0.14$ (very close to the standard normal (outside the tails!));

O: Orthonormal family with $k = 3$ and $(\varsigma_1, \varsigma_2, \varsigma_3) = (-0.1, -0.1, 0.1)$.

We also calculate numerically two approximations, denoted by $App1$ and $App2$. The first approximation of EP, $App1$, is obtained by taking into account convergence up to order $o(1)$. Since the correction terms are of order $O(n^{-1})$, this means that both for the uncorrected and the corrected control limit $App1$ equals $\overline{F}_0\left(u_p\right)$ with F_0 the distribution function of $\sigma^{-1}\left(X_{n+1} - \mu\right)$ when the supposed model is normality and $\overline{F}_0\left(\overline{K}_\gamma^{-1}(p)\right)$ with γ the suitably chosen point given by (3.16) when the supposed model is the normal power family. Hence the first approximation is related to the (restricted) model error by $App1 = p + (R)ME$, see (3.2) and (3.5).

The second approximation is obtained by expanding EP up to terms of order $O(n^{-1})$. Note that there are no $O(n^{-1/2})$-terms. In general the derivation of these formulas is as follows. We get, with V given by (3.18),

$$EP = E\left\{\overline{F}_0\left(\overline{K}_\gamma^{-1}(p) + V + \breve{\sigma}c_u\right)\right\}.$$

In the uncorrected case we have $c_u = 0$. Furthermore, when dealing with the restricted model of normality, we take $\overline{K}_\gamma^{-1}(p) = u_p$ and $V = \overline{X} + Su_p - u_p$. Typically we have that EV and EV^2 are of order $O\left(n^{-1}\right)$, and $E\left|V\right|^k = o\left(n^{-1}\right)$ for $k \geq 3$. The correction term c_u for correcting the bias typically is of order $O\left(n^{-1}\right)$. Let F_0 have density f_0. Ignoring terms of order V^k for $k \geq 3$ and terms of order $c_u\left(\breve{\sigma} - 1\right)$, c_uV and $(c_u)^k$ for $k \geq 2$, Taylor expansion gives

$$EP \approx \overline{F}_0\left(\overline{K}_\gamma^{-1}(p)\right) - f_0\left(\overline{K}_\gamma^{-1}(p)\right) E\left(V + c_u\right) - \frac{1}{2} EV^2 f_0'\left(\overline{K}_\gamma^{-1}(p)\right). \quad (3.55)$$

*App*2 is obtained from the right-hand side of (3.55) by taking suitable approximations of $E\left(V + c_u\right)$ and EV^2. The formulas are given below in the corresponding subsections.

In the following subsections, we consider four cases of the supposed model : normality without correction, normality with correction, normal power family without correction and normal power with correction. In the latter cases we only use the recommended γ estimator for the normal power family. Recall that the control limit depends on the supposed model, on the estimators of the parameters and on whether we make a correction or not. For convenience the unit in the tables equals 0.001, thus, for instance, 1.25 means 0.00125.

3.7.1 Supposed model: normality, no correction

As mentioned in the former chapter, assuming that we are in the restricted model of normality and applying no correction for using estimators, the control limit simply equals

$$\overline{X} + Su_p.$$

For the first approximation of the expected (observed) false alarm rate we get $App1 = \overline{F}_0\left(u_p\right) = p + RME$ with F_0 the distribution function of $\sigma^{-1}\left(X_{n+1} - \mu\right)$.

The second approximation of EP equals, cf. also (3.28),

$$App2 = \overline{F}_0\left(u_p\right) + \frac{\left(\mu_4 - 1\right) f_0\left(u_p\right) u_p}{8n} - \frac{f_0'\left(u_p\right)\left\{1 + \mu_3 u_p + \frac{1}{4}\left(\mu_4 - 1\right) u_p^2\right\}}{2n},$$

where μ_k is the k^{th} moment under F_0.

The simulation results, the first and second approximations are presented in Table 3.3. It is seen from Table 3.3 that the false alarm rate may be completely wrong if we are not in the restricted model, that is if normality does not hold. This confirms previous results of e.g. Chan et al. (1988) and Pappanastos and Adams (1996). However, they only give the total error. By splitting up the total error in the restricted model error and the restricted stochastic error more insight is obtained about the roles of misspecification on the one hand and estimation of the location and scale parameter on the other hand.

The restricted model error RME is the difference between $App1$ and 0.001 $(= p)$. It is illuminating that, while in the middle of the distribution there is hardly any difference between $TU(0.14)$ and the standard normal distribution, the far tails are different. The effect of this difference is clearly seen by comparing $TU(0.14)$ and Φ in Table 3.3. Often it is also stated that in the middle there is not much difference between normal and logistic distributions. Comparison of $TU(0)$ and Φ in Table 3.3 shows that for the problem at hand sloppy inspection of the data in the middle, leading to the conclusion that "normality is not so bad" may produce serious errors.

Apart from the misspecification the effect of the estimation is also not negligible. Due to the estimation the total error becomes substantially larger in case of positive RME's, which may occur in practice more often than negative RME's. In the latter situation the total error is compensated by the (positive) restricted stochastic error. This, for instance, occurs for $TU(0.14)$. However, such compensation is in an uncontrolled way and may be far too small (see $K_{-0.25}$) or may lead to overcompensation ($TU(0.14)$, $n = 100$).

Table 3.3 Simulated expected (observed) false alarm rate without correction for the estimation of the mean and variance and corresponding first and second approximation assuming normality as model. The unit in the table is 0.001.

F_0	simulation			$App1$	$App2$		
	$n = 100$	$n = 250$	$n = 500$		$n = 100$	$n = 250$	$n = 500$
Φ	1.36	1.14	1.07	1.00	1.33	1.13	1.07
$K_{-0.5}$	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$K_{-0.25}$	0.06	0.03	0.03	0.02	0.05	0.03	0.03
$K_{0.25}$	4.44	3.96	3.81	3.66	4.39	3.95	3.81
$K_{0.5}$	7.70	7.01	6.79	6.58	7.67	7.02	6.80
$K_{0.75}$	10.31	9.45	9.15	8.86	10.30	9.44	9.15
K_1	12.13	11.01	10.71	10.35	12.17	11.08	10.71
T	5.30	4.87	4.72	4.56	5.37	4.89	4.73
RM	3.31	2.99	2.89	2.78	3.34	3.00	2.89
DM	3.45	3.13	3.03	2.92	3.43	3.13	3.02
$TU(-0.1)$	6.61	6.07	5.89	5.71	6.66	6.09	5.90
$TU(0)$	4.29	3.91	3.79	3.67	4.26	3.90	3.79
$TU(0.14)$	1.21	0.98	0.91	0.85	1.19	0.98	0.91
O	2.81	2.39	2.26	2.13	2.76	2.38	2.26

Clearly, $App1$ gives an impression of EP, the expected false alarm rate, but is not very precise. $App2$ gives a very good prediction of the simulation.

3.7.2 Supposed model: normality, with correction

Assuming that we are in the restricted model of normality and applying the suitable correction term for using estimators as developed in Section 2.3.1, cf. (2.33) but note we use $\breve{\sigma}$ instead of $\widehat{\sigma}$, the control limit equals

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

The correction term is of order $O(n^{-1})$ and hence $App1$ is the same as in Table 3.3. In view of (3.55), cf. also (3.28), we get

$$\begin{aligned} App2 &= \overline{F}_0\left(u_p\right) - f_0\left(u_p\right)\frac{u_p}{4n}\left\{u_p^2 + 3 - \frac{1}{2}\left(\mu_4 - 1\right)\right\} \\ &\quad -\frac{1}{2}f_0'\left(u_p\right)\frac{1 + \mu_3 u_p + \frac{1}{4}\left(\mu_4 - 1\right)u_p^2}{n}. \end{aligned}$$

It is seen that indeed $App2 = p$ for $F_0 = \Phi$.

The simulation results, the first and second approximations are presented in Table 3.4. Note that $App1$ is the same as in Table 3.3.

It is seen from the row denoted by Φ in Table 3.4 that the correction for estimation of the parameters works very well, see also Subsection 2.3.1 for more details. Comparison of Table 3.3 and Table 3.4 shows that the correction reduces the stochastic error not only for the normal distribution, but for the other distributions as well, thus bringing the total errors closer to the restricted model error.

Nevertheless, if normality fails still the false alarm rate may be completely wrong due to the restricted model error. As a consequence of the bias correction, $App1$ and $App2$ are much closer to each other. Again, $App2$ gives a very good prediction of the simulation.

Table 3.4 Simulated expected (observed) false alarm rate with correction for the estimation of the mean and variance and corresponding first and second approximation assuming normality as model. The unit in the table is 0.001.

F_0	simulation			*App*1	*App*2		
	$n = 100$	$n = 250$	$n = 500$		$n = 100$	$n = 250$	$n = 500$
Φ	1.01	1.00	1.00	1.00	1.00	1.00	1.00
$K_{-0.5}$	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$K_{-0.25}$	0.03	0.03	0.02	0.02	0.03	0.03	0.02
$K_{0.25}$	3.68	3.67	3.66	3.66	3.65	3.66	3.66
$K_{0.5}$	6.72	6.63	6.60	6.58	6.71	6.63	6.61
$K_{0.75}$	9.27	9.03	8.94	8.86	9.28	9.03	8.95
K_1	11.10	10.67	10.51	10.35	11.17	10.68	10.51
T	4.64	4.61	4.59	4.56	4.73	4.63	4.60
RM	2.80	2.80	2.79	2.78	2.86	2.81	2.80
DM	2.89	2.91	2.93	2.92	2.90	2.91	2.92
$TU(-0.1)$	5.86	5.78	5.75	5.71	5.94	5.80	5.76
$TU(0)$	3.63	3.65	3.66	3.67	3.62	3.65	3.66
$TU(0.14)$	0.87	0.85	0.85	0.85	0.87	0.85	0.85
O	2.18	2.15	2.14	2.13	2.16	2.14	2.14

3.7.3 Supposed model: normal power family, no correction.

In the present subsection we assume that the observations are from the normal power family and that we use the recommended $\widehat{\gamma}$ based on quantiles in the ordinary tail. Then the control limit without correcting for the estimation equals

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

The first approximation of the false alarm rate is given by $App1 = \overline{F}_0\left(\overline{K}_{\gamma}^{-1}(p)\right)$ with γ the suitably chosen point given by

$$\gamma = 1.1218 \log\left(\frac{\overline{F}_0^{-1}(0.05)}{\overline{F}_0^{-1}(0.25)}\right) - 1. \tag{3.56}$$

It is related to the model error by $\overline{F}_0\left(\overline{K}_{\gamma}^{-1}(p)\right) = p + ME$, cf. (3.5). Inserting $c_u = 0$ in (3.55) the second approximation of EP equals

$$App2 = \overline{F}_0\left(\overline{K}_{\gamma}^{-1}(p)\right) - f_0\left(\overline{K}_{\gamma}^{-1}(p)\right)\widetilde{D1}_n - \frac{1}{2}\widetilde{D2}_n f_0'\left(\overline{K}_{\gamma}^{-1}(p)\right),$$

with $\widetilde{D1}_n$ and $\widetilde{D2}_n$ the approximations to EV and EV^2 given by the right hand sides of (3.34) and (3.35), respectively. The involved moments μ_i should be taken under F_0.

Table 3.5 Simulated expected (observed) false alarm rate using $\widehat{\gamma}$ as estimator without correction for the estimation of the mean, the variance and the parameter γ and corresponding first and second order approximation assuming the normal power family as model. The unit in the table is 0.001.

F_0	simulation			$App1$	$App2$		
	$n=100$	$n=250$	$n=500$		$n=100$	$n=250$	$n=500$
Φ	2.53	1.59	1.27	1.00	2.26	1.52	1.25
$K_{-0.5}$	3.41	1.92	1.42	1.00	3.15	1.89	1.43
$K_{-0.25}$	2.73	1.66	1.30	1.00	2.50	1.63	1.30
$K_{0.25}$	2.39	1.54	1.25	1.00	2.15	1.48	1.23
$K_{0.5}$	2.34	1.51	1.23	1.00	2.11	1.46	1.22
$K_{0.75}$	2.33	1.51	1.23	1.00	2.12	1.46	1.22
K_1	2.34	1.51	1.23	1.00	2.18	1.48	1.24
T	4.68	3.78	3.39	3.08	4.56	3.73	3.38
RM	3.63	2.76	2.43	2.16	3.44	2.71	2.42
DM	3.78	2.90	2.57	2.28	3.61	2.85	2.55
$TU(-0.1)$	4.92	3.97	3.58	3.25	4.79	3.93	3.56
$TU(0)$	3.96	3.00	2.63	2.33	3.75	2.94	2.61
$TU(0.14)$	2.33	1.41	1.08	0.81	2.11	1.35	1.07
O	2.18	1.23	0.94	0.69	1.82	1.15	0.91

The simulation results and the first and second approximations are presented in Table 3.5. Compared to Tables 3.3 and 3.4 it is seen from Table 3.5 that the difference between the expected (observed) false alarm rate and the required value 0.001 is seriously reduced by considering the normal power family as parametric model. Moreover, when the restricted model of normality holds, the loss is not large. The second approximation gives a very good prediction of the simulation throughout the table. This makes it rather easy to analyze and predict the behavior of the control chart under all kinds of distributions.

3.7.4 Supposed model: normal power family, with correction.

In order to improve the performance of the parametric chart by reducing the produced SE, in this subsection we apply the corrected control chart as given in (3.39). In this case, the control limit equals

$$\overline{X} + S\left\{\overline{K}_{\widehat{\gamma}}^{-1}(p) - C1(\widehat{\gamma})\,C2(\widehat{\gamma}) - \frac{C3(\widehat{\gamma})}{n} + \frac{C4(\widehat{\gamma})}{n}\right\}.$$

The correction term is of order $O(n^{-1})$ and hence $App1$ is the same as in Table 3.5. The second approximation equals, with γ the suitably chosen point given by (3.56),

$$
\begin{aligned}
App2 &= \overline{F}_0\left(\overline{K}_\gamma^{-1}(p)\right) - \frac{1}{2}\widetilde{D2}_n f_0'\left(\overline{K}_\gamma^{-1}(p)\right) \\
&\quad - f_0\left(\overline{K}_\gamma^{-1}(p)\right)\left\{\widetilde{D1}_n - C1(\gamma)\,C2(\gamma) - \frac{C3(\gamma)}{n} + \frac{C4(\gamma)}{n}\right\}
\end{aligned}
$$

with $\widetilde{D1}_n$ and $\widetilde{D2}_n$ the approximations to EV and EV^2 given by the right hand sides of (3.34) and (3.35), respectively. The involved moments μ_i should be taken under F_0.

Table 3.6 Simulated expected (observed) false alarm rate using $\widehat{\gamma}$ as estimator and the recommended version of the control limit with correction for the estimation of the mean, the variance and the parameter γ and corresponding first and second approximation assuming the normal power family as model. The unit in the table is 0.001.

F_0	simulation			$App1$	$App2$		
	$n = 100$	$n = 250$	$n = 500$		$n = 100$	$n = 250$	$n = 500$
Φ	1.14	1.05	1.02	1.00	1.06	1.02	1.01
$K_{-0.5}$	0.77	0.93	0.97	1.00	1.06	1.02	1.01
$K_{-0.25}$	1.02	1.01	1.01	1.00	1.08	1.03	1.02
$K_{0.25}$	1.16	1.05	1.02	1.00	1.02	1.01	1.00
$K_{0.5}$	1.18	1.05	1.01	1.00	1.01	1.00	1.00
$K_{0.75}$	1.20	1.05	1.01	1.00	1.03	1.01	1.00
K_1	1.21	1.06	1.01	1.00	1.08	1.03	1.02
T	2.91	3.04	3.03	3.08	2.80	2.99	3.03
RM	2.09	2.14	2.14	2.16	2.02	2.11	2.13
DM	2.13	2.21	2.24	2.28	1.99	2.18	2.22
$TU(-0.1)$	3.12	3.21	3.21	3.25	2.96	3.17	3.19
$TU(0)$	2.24	2.28	2.28	2.33	2.02	2.22	2.26
$TU(0.14)$	1.00	0.90	0.85	0.81	0.98	0.88	0.84
O	0.95	0.78	0.73	0.69	0.86	0.75	0.72

The simulation results, the model error and the first and second approximations are presented in Table 3.6. Table 3.6 shows that the correction for estimation of the parameters works very well, leading to EP (very) close to p for distributions from the normal power family. Comparison with Table 3.4 shows that the corrected control limit based on $\widehat{\gamma}$ performs much better than the corrected standard control limit based on normality. Both the restricted model error and the restricted stochastic error are reduced to an acceptable level by the recommended control limit. *App*2 gives a very good prediction of the simulation, thus making it easy to analyze and predict the behavior of this control chart for all kind of distributions.

The recommended control limit gives a simple explicit control limit with small total error in the normal power family and reasonable total error for distributions outside this family.

3.8 Discussion

We start the discussion by answering the questions posed in Section 3.2. In this section, each question will be treated in a separate subsection.

3.8.1 Normality holds

Firstly, suppose that normality holds, then the questions to discuss are : **How large is the difference between *SE* and *RSE*? Can *RSE* and *SE* be reduced by (simple) corrections? How large are the corrected *SE* and the corrected *RSE*?**

The simulated expected RSE when applying the uncorrected control limit and its (first and second) approximations are found in Table 3.3 on the row containing Φ by subtracting $p = 0.001$ (that is 1 unit) from the corresponding entry. Similarly, again applying the uncorrected control limit the simulated expected SE and its approximations are found in Table 3.5 on the row containing Φ, again subtracting 1 unit. It is seen that without correction indeed SE is larger than RSE. However, by making the appropriate corrections the differences disappear: all the corrected versions are close to 0.

We may conclude that the larger SE in the uncorrected control limit can be accurately repaired by an appropriate correction term.

3.8.2 Observations from the normal power family

Now, suppose that the observations are drawn from the normal power family. In this case, questions like : **How bad can *RME* be? How does this balance with the larger *SE*? Can *RSE* and *SE* be reduced by (simple) corrections? How large are the corrected *RSE* and the corrected *SE*, when the observations are from the normal power family?** are the interesting topics to discuss.

Concerning RME, which has been discussed more extensively in Section 3.2 and 3.5, the conclusion is clear: RME can be quite large. In the situation where no correction takes place we consider the balance between the larger SE and the possibly large RME. Therefore, we compare the results of Table 3.5 with those of Table 3.3 for the distributions $K_{-0.5}, K_{-0.25}, K_{0.25}, K_{0.5}, K_{0.75}$ and K_1. It is seen that RME has a larger effect than SE. Moreover, SE can be corrected, while RME remains.

Recall that RSE refers to the situation where the supposed model is normality. Therefore, the corresponding correction terms are based on this assumption. The correction is not tailored to the normal power family. How well the correction still helps in these kinds of situations is seen by comparing the lines containing $K_{-0.5}, K_{-0.25}, K_{0.25}, K_{0.5}, K_{0.75}$ and K_1 in Table 3.4 and Table 3.3. RSE reduces also for these distributions, but due to the large RME the total error is still (very) large.

The corrected SE is quite small, see the lines containing $K_{-0.5}, K_{-0.25}, K_{0.25}$, $K_{0.5}, K_{0.75}$ and K_1 in Table 3.6.

3.8.3 Outside the normal power family

In the case that the observations are drawn from outside the normal power family, the following questions are quite interesting to discuss: **How large are the uncorrected and corrected *SE* and the uncorrected and corrected *RSE*, when we are outside the normal power family and *ME* is not too big? How is the total error when applying the corrected control limit with as supposed model the normal power family and how does this compare with the total error when applying the corrected control limit with as supposed model normality?**

From the lines containing $T, RM, DM, TU(-0.1), TU(0)$, $TU(0.14), O$ in Tables 3.5 and 3.6 it is seen that the uncorrected SE is substantially improved by application of the correction terms (even although they are derived for the parametric model). Also when supposing normality the correction is very helpful, see the lines containing $T, RM, DM, TU(-0.1)$, $TU(0), O$ in Tables 3.3 and 3.4.

With respect to the total error the lines containing $T, RM, DM, TU(-0.1), TU(0)$, $TU(0.14)$, O in Tables 3.3 – 3.6 show the following. The total error is often very large when using the classical normal control limit, due to a high restricted model error (and a restricted stochastic error, when no correction is applied). As a first step the total error is considerably reduced by application of the normal power family with, secondly, a further great improvement by taking the correction terms into account.

3.8.4 Out-of-control

Finally, we focus on discussing the possibility that we have an out-of-control situation. In this case, we try to answer the following questions: **What is the impact on the out-of-control behavior of the chart? Does the process stop at a reasonable time when it has gone out-of-control?**

Avoiding (apart from a small probability) stopping unexpectedly early during the in-control period is of course desirable, but this should not be achieved by typically stopping much later once the process has gone out-of-control. Thus let X_{n+1} come from a shifted distribution function $F(x - \Delta)$. For simplicity, and without loss of generality, we again let $\mu = 0$ and $\sigma = 1$ and thus work under the standardized F_0. The shift Δ is such that the probability of detecting the shift when the distribution is completely known, that is

$$\widetilde{p} = \overline{F}_0 \left(\overline{F}_0^{-1}(p) - \Delta \right),$$

is no longer extremely small, like p.

The out-of-control behavior is seen from the results in Table 3.7. For the same distributions as used in Tables 3.3 – 3.6 we have performed simulations under $F_0(x - \Delta)$. For each F_0 we have selected two values of Δ such that reasonable values of $\widetilde{p}$ result. Usually, $\Delta = 2$ and $\Delta = 3$ will do, but as γ moves away from 0, the values $\Delta = 1$ or $\Delta = 4$ can become more appropriate.

It is seen from Table 3.7 that repairing the damage during in control does not destroy the out-of-control behavior. Even compared to the situation of exactly knowing the distribution, which rarely occurs in practice, the recommended control chart does not loose that much when considering a shifted distribution. For instance, when normality holds and a shift of 2 standard deviations occurs, the probability of detecting such a shift goes from 0.138 to 0.101 ($n = 100$), 0.120 ($n = 250$) or 0.128 ($n = 500$). With these probabilities correspond average run lengths going from 7.3 to 9.9 ($n = 100$), 8.3 ($n = 250$), or 7.8 ($n = 500$), which indeed is not a tremendous change. For distributions outside the parametric model there may be even a gain w.r.t. the case of completely known distribution, which is due to the model error, see Table 3.6.

From the discussion above as well as the tables presented in Section 3.5 we conclude that the use of a classical normal control limit very often leads to a large total error. This is mainly due to the large restricted model error. Although the correction for estimating the mean and the variance is very useful in reducing the restricted stochastic error (even when normality does not hold), the model error often dominates and causes a large total error.

In comparison, the (corrected) control limits based on the normal power family work very well, also outside the normal power family. The recommended control limit is completely explicit and can be calculated quite straightforwardly. It has a very good approximation (*App*2), which makes it possible to predict in an accurate way the behavior of this control chart under all kind of distributions. The out-of-control behavior is still reasonably good, meaning that the process will stop at a reasonable time when it has gone out-of-control.

Table 3.7 Simulated values of $E_\Delta P$, the expected (observed) probability of producing an out-of-control signal under $F_0(x-\Delta)$, using $\widehat{\gamma}$ as estimator and the recommended version of the control limit with correction for the estimation of the mean, the variance and the parameter γ, with in the column $\widetilde{p}$, the probability of producing an out-of-control signal when the distribution is completely known.

F_0	Δ	$\widetilde{p}$	n		
			100	250	500
Φ	2	0.138	0.102	0.120	0.128
	3	0.464	0.346	0.410	0.436
$K_{-0.5}$	1	0.227	0.144	0.190	0.208
	2	0.500	0.477	0.496	0.499
$K_{-0.25}$	1	0.058	0.040	0.049	0.053
	2	0.355	0.261	0.314	0.335
$K_{0.25}$	2	0.049	0.043	0.046	0.047
	3	0.210	0.168	0.190	0.199
$K_{0.5}$	3	0.082	0.083	0.080	0.080
	4	0.300	0.257	0.292	0.305
$K_{0.75}$	3	0.036	0.046	0.038	0.036
	4	0.119	0.151	0.141	0.129
K_1	3	0.018	0.029	0.021	0.019
	4	0.053	0.097	0.072	0.059
T	2	0.016	0.068	0.072	0.071
	3	0.088	0.254	0.297	0.311
RM	2	0.040	0.081	0.091	0.095
	3	0.217	0.293	0.349	0.372
DM	2	0.045	0.082	0.092	0.097
	3	0.234	0.295	0.353	0.377
$TU(-0.1)$	2	0.012	0.055	0.055	0.054
	3	0.052	0.213	0.237	0.244
$TU(0)$	2	0.036	0.070	0.076	0.078
	3	0.188	0.259	0.306	0.325
$TU(0.14)$	2	0.148	0.099	0.118	0.126
	3	0.481	0.339	0.406	0.431
O	2	0.097	0.062	0.071	0.076
	3	0.340	0.224	0.263	0.280

3.9 Concluding remarks

The failure of the normality assumption imposed on the underlying distribution causes model error and stochastic error. Section 3.2 exposes the problem at hand by discussing the notion of these errors in the restricted and general model. Some questions related to restricted stochastic error (RSE), restricted model error (RME), stochastic error (SE) and model error (ME) are well thought-out in this section. From the discussion and example given in this section it is seen that RME can be quite large. Often RME is several factors larger than the prescribed p. Therefore, if normality fails, there is a need for a larger model, thus reducing the model error.

In general, extending the normal family to a larger model causes some technical difficulties due to the comprehensive conditions and precision required for an appropriate general model. Among the suitable models discussed in Section 3.3, only the normal power family seems to be tractable. Hence this family, which has parameters (μ, σ, γ), is chosen to be the general model and used to construct a new parametric control chart. Section 3.4 presents the characteristics of the new chart. The estimated control limit of this chart is given in (3.8), where γ can be estimated using (3.10) or (3.14). However, our recommended γ estimator is given in (3.15).

As shown in Section 3.5, applying the new parametric model to the control limits manages to reduce most of the model errors substantially. The new model however, increases the stochastic errors due to the estimation of the additional parameter γ. To improve the performance of the parametric chart, a correction term based on the bias criterion is developed in Section 3.6 and applied in the control limit cf. (3.17). Using the first γ estimator, the corrected control limit is given in (3.33), while if we use the second γ estimator, the corrected control limit is given in (3.38). Since these control limits are not so easy to calculate, we recommend a simpler control limit which is given in (3.39).

To evaluate the theoretical results as well as the performance of the recommended chart as compared to the normal chart we perform a simulation study and some numerical calculation in Section 3.7. Firstly, we use normality, followed by a normal power family as a supposed model. Each model is evaluated without and with the corresponding correction term. The results show that the recommended chart works very well both in reducing the model errors and the stochastic errors for distributions from the normal power family as well as for distributions outside the normal power family. Also in this section we calculate numerically two approximations of EP, the expected false alarm rate. The first approximation, $App1$, is obtained by taking into account convergence up to order $o(1)$. The second approximation, $App2$, is obtained by expanding EP up to and including terms of order $O(n^{-1})$. The numerical result shows that although $App1$ gives an impression of EP, it is not very precise. On the other hand, $App2$ gives a very good prediction of the simulation.

Finally, section 3.8 is devoted to discuss the questions posted in Section 3.2. Subsequently we evaluate the case when normality holds, the case when the observations are drawn from the normal power family, the case when the observations are drawn from outside the normal power family, and the case of an out-of-control situation.

From the discussion presented in this section, we conclude that the classical normal control limit is unreliable with often a very large total error, mainly due to the large RME. Although the correction for estimating the mean and the variance is very useful in reducing the RSE (even when normality does not hold), the model error often dominates and causes a large total error.

As a solution to this problem, we propose (corrected) control limits based on the normal power family which turn out to work very well, also outside the normal power family. Our recommended limit as given in (3.39) is completely explicit and can be calculated quite straightforwardly. It has a reliable approximation ($App2$), which makes it possible to predict in an accurate way the behavior of this control chart under all kind of distributions. Moreover, the out-of-control behavior is not dropping off dramatically; the process will stop at a reasonable time when it has gone out-of-control.

Chapter 4

Parametric Approach:
Controlling the exceedance probability

Common control charts assume normality and known parameters. Quite often these assumptions are not valid and large relative errors result in the usual performance characteristics, such as the false alarm rate or the average run length. A fully nonparametric approach can form an attractive alternative to solve such problem, but requires more Phase I observations than are usually available in practice. Sufficiently large parametric families then provide realistic intermediate models.

Continuing what has been done in the previous chapter, the performance of charts based on a normal power family is considered in the present chapter. Exceedance probabilities of the resulting stochastic performance characteristics during in-control process are studied. Corrections are derived to ensure that such probabilities stay within prescribed bounds. Attention is also devoted to the impact of the corrections for an out-of-control process. Simulations are presented both for illustration and to demonstrate that the approximations obtained are sufficiently accurate for use in practice.

4.1 Introduction

In Chapter 3, the problem caused by imposing the normality assumption on the underlying distribution is discussed. In the case that the actual observations are not from the normal family, very large model errors, which reflect the difference between the assumed model and the true model, may occur. The greater the difference between the two models, the larger the model errors that may be produced. To deal with this problem, the previous chapter proposes a parametric control chart having a normal power family as the supposed model. It has been shown in that chapter that the new chart is able to reduce the model errors from various distributions substantially. However, this parametric chart also causes a serious stochastic error as a result of estimation of an additional parameter. Therefore, in order to reduce this stochastic error, a correction term based on a bias criterion has been developed in the same chapter. The corrected parametric chart is shown to perform very well, both in the in-control and in the out-of-control situation.

Briefly reviewing what has been discussed up to now, consider the following standard control chart procedure: the mean of a production process is monitored by means of a normal chart. Each new value is compared with a given upper and lower limit and an out-of-control signal results if either one of these two limits is exceeded. Usually normality of the distribution involved is taken for granted and it only remains to estimate its parameters using Phase I observations. The outcomes are plugged into the expressions for the limits and the estimated chart is expected to behave as if it were based on known values. However, by now it is known that this unfortunately is too optimistic. To explain why this is so, note that the desired value p for the probability of getting a signal while in fact the in-control situation still persists, usually is extremely small, like $p = 0.001$. Now the estimation error based on a common sample size like 50 will in itself be rather small. However, in comparison to a value like 0.001, the error will be unpleasantly large. In other words, the relative errors involved will be much larger than anticipated. This results in an increased rate of very short runs and also a few more very long runs between false signals.

This situation can be repaired by applying suitable corrections to the estimated limits. In the previous chapter, a bias criterion under the following scenario is used to derive a correction term for the control limit. Due to the estimation, p is replaced by a stochastic version P_n, where n is the number of Phase I observations used and typically $Eg(P_n)$ will differ considerably from $g(p)$ for the usual functions of interest (see (2.3)). For each of these three choices, suitable corrections have been derived in Section 2.3. In this way the n required in order to arrive at an acceptably small bias, can be brought down from about 300 to about 40. This is gratifying, but we should realize that the bias criterion is rather mild, as it only corrects the average behavior of the chart over a long series of applications.

If instead we are more interested in what may happen for a single application, we should focus on the distribution of the random variable P_n around p, rather than just look at its average behavior (or that of $g(P_n)$). The fact that this distribution typically is asymptotically normal, clearly implies that $P(P_n > p)$ will tend to $\frac{1}{2}$ as n becomes large. The bias corrections mentioned above merely speed up this convergence and help to avoid that one has to start for smaller n with exceedance probabilities well above this $\frac{1}{2}$. Thus quite naturally, p will be close to the 50%-quantile of the distribution of P_n. Obviously, this can be felt to be much too liberal, and hence that it is important in control chart implementation to correct the estimated limits in such a way that p will be (close to) the upper α-quantile of the distribution of P_n for some sufficiently small α, like 0.1 or 0.2. In other words, it is desired that the distribution of P_n is shifted to the left such that $P(P_n > p) = \alpha$. A slightly relaxed version of this criterion is obtained by allowing the upper α-quantile in question to equal $p(1+\varepsilon)$, rather than p itself, for some small $\varepsilon \geq 0$, like $\varepsilon = 0.1$. In this way, only in a fraction α of the applications, one is faced with a value of P_n which is really too large in the sense that it exceeds not only p, but even $p(1+\varepsilon)$. Using once more the functions g then finally leads to requiring $P(g(P_n) > g(p)(1+\varepsilon)) = \alpha$ for increasing g and $P(g(P_n) < g(p)(1-\varepsilon)) = \alpha$ for decreasing g. Clearly, for $\varepsilon = 0$, g plays no role, as e.g. $P(P_n > p) = P(1/P_n < 1/p)$.

It is good to note that the so obtained exceedance criterion has to be interpreted with some care, as two steps are actually involved. First, one has the probability of getting a signal, which occurs if a limit of the chart is exceeded. As these limits are estimated, this probability P_n will be stochastic, as it depends on the Phase I observations used. The next step involves the form of exceedance on which the criterion focuses: does the P_n obtained in its turn exceed some reasonable bound like $p(1+\varepsilon)$? Also observe that the criterion amply allows to study the behavior of the run length (RL) distribution itself, and not just that of the signal probability P_n. As mentioned above, the (conditional!) expectation $1/P_n$, which is the now stochastic ARL, can be monitored. Or likewise, $P(RL \leq k) = 1 - (1 - P_n)^k$ can be compared to $\{1 - (1-p)^k\}(1+\varepsilon)$.

Adaptations of this second type are obtained in Subsection 2.3.2 for each of the three types of g under consideration. As this second criterion is more strict, it is not surprising that the corresponding corrections are of a larger order of magnitude than those required for the bias case. In fact, the orders involved are $n^{-1/2}$ and n^{-1}, respectively. Consequently, the impact on the out-of-control behavior will also be stronger when using this exceedance probability criterion. Note that this could mean a serious problem: if bringing the in-control behavior under control (i.e. getting $g(P_n)$ sufficiently close to $g(p)$ in a suitable sense) would result in a substantially lowered detection power once the process goes out-of-control, the price for the protection might be judged to be too high. Fortunately, however, the effects during the out-of-control stage will typically be sufficiently moderate. To understand why this is the case, note that during this stage the extremely small p from the in-control situation has increased to some p_1 which may still be small, but no longer extremely so. The change caused by applying a correction to the estimated control limits will be of the same order of magnitude for p and p_1. However, in terms of relative change, the impact on p_1 will be much more mild.

Hence according to the above, the practitioner can choose between a weak and a strong form of protection, at a low or moderate price, respectively, and it may seem that the problem has been satisfactorily solved. However, note that actually we have only repaired the effect of the unwarranted assumption that estimation effects are negligible. The other dubious assumption, according to which the distribution involved is simply normal, still stands. In fact, this second assumption is even more cumbersome. As n increases, eventually the estimation effects will decrease and SE will become negligible. On the other hand, deviations from normality cause an ME that does not go away, no matter how large n is chosen. Again the extremity of the quantiles involved transforms this into a major problem: in the middle of the distribution, the normality assumption may be reasonable, but in the far tail the relative errors caused can be unacceptably large.

A logical next step thus is to acknowledge the possibility of an ME and to search for a compromise which keeps this ME within bounds without letting the SE explode. To see that the latter can easily happen, just go from the simple normal model to the other opposite, the fully nonparametric model. There, using the empirical quantiles, the ME indeed vanishes. But for typical configurations like $p = 0.001$ and $n = 100$, the stochastic error will clearly be overwhelming. Consequently, a family which is larger than the normal, but still parametric, is an attractive type of compromise to look for. Such families were studied in Section 3.4 and there a specific choice, based on the normal power family, was demonstrated to work well (see Section 3.5). For a broad class of distributions, the ME is controlled much better using this choice than under the simple normal model, where unacceptably large errors occur. The price for this improvement is of course a somewhat higher SE: here not only the mean and standard deviation need to be estimated (as in the normal model), but also the best fitting member of the available parametric family. But for such a parametric family as well, corrections can be derived to bring the SE under control. In analogy to the normal case, again two types of criterion suggest themselves: bounding the bias versus bounding the exceedance probability. The focus of Chapter 3 is on the bias reduction and it was shown that with respect to this criterion, accurate control limits can indeed be obtained.

In view of the above it is clear that it is of great interest to study parametric charts when the criterion is based on exceedance probabilities, and this will be the topic of the present chapter. From the four situations, normal chart with bias correction, normal chart with exceedance probability as criterion, parametric control chart with bias correction and parametric chart dealing with exceedance probability, the latter is the most ambitious one. We will now simultaneously try to control the occurrence of unpleasant values of P_n, as well as the occurrence of ME's which are unacceptably large. The question will be whether this still can be achieved for reasonable sample sizes without destroying the power of detection during the out-of-control situation.

The present chapter is organized as follows. In Section 4.2 the characteristics of the parametric chart are briefly reviewed, then correction terms are derived based on the exceedance probability criterion in the following section. Such corrections are indeed larger than those of the bias criterion since the new criterion is more severe. Section 4.4 again specializes to the normal power family and presents a completely specific proposal for that situation. This proposal is subsequently investigated in a simulation study. It turns out to work quite well: without it, the exceedance probabilities are unacceptably large, whereas after correction the values obtained are indeed close to the desired α. Next, Section 4.5 is devoted to studying the impact of the correction on the out-of-control behavior. As expected, it turns out that the effect can be substantial. Guidelines are given to check whether it is acceptable for the values of n, p and α at hand, or adaptations, such as a larger sample size, are called for. Again a simulation study is presented to support and illustrate the recommendations given. The final section summarizes the conclusions.

4.2 Characteristics of the parametric chart

The notion of the parametric control chart has been introduced in Section 3.4. In addition, the advantages and disadvantages of using this chart have been discussed in Section 3.5. The present section provides a brief review on the characteristics of this parametric chart.

Consider i.i.d. r.v.'s $X_1, \ldots, X_n, X_{n+1}$ from some distribution function (d.f.) F. The first n of these r.v.'s come from Phase I and form the basis for the estimation step; the last r.v. belongs to Phase II, the monitoring stage. Clearly, as all $(n+1)$ r.v.'s come from the same F, we have the in-control situation as our starting point. If F is supposed to be $N(\mu, \sigma^2)$, the proper UCL for a certain p simply equals $\mu + u_p\sigma$. The fact that μ and σ are unknown requires these parameters to be replaced by customary estimators leading immediately to the estimator of UCL given in (2.1).

However, if normality is not taken for granted, a more general model should be selected and the question is what choices might be suitable. Along the lines of Section 3.4, we argue as follows. To begin with, the broader model selected should contain the normal family as a submodel. After all, approximate normality always will be the starting point. On the one hand, we want to be protected against the errors caused by this mere approximate validity of the normality assumption, as such errors can be quite serious in view of the tail probabilities we are dealing with. But on the other hand, if normality happens to be perfectly true after all, we want our model to be able to produce a normal distribution as the best fitting candidate. In other words, simply trading in the normal family for some heavier tailed one (like some Student's t with a fixed number of degrees of freedom), is no reasonable option.

A possible alternative might be to let the Phase I data decide between either the normal family or such a heavier tailed alternative family. But this is a rather "jumpy" solution. It seems more attractive to move smoothly from the normal family toward a heavy tailed one. To be explicit, a general model could e.g.

be a mixture (either random or deterministic) of a normal and a t-distributed random variable with a rather small number (say 6) of degrees of freedom. In addition to $\widehat{\mu}$ and σ^*, a third estimator would be required for the mixing parameter. Unfortunately, this proposal leads to unpleasant technicalities when one tries to apply it, stemming from the fact that the normal case lies on the boundary.

Yet another attractive possibility is available, however, for which such problems do not arise. Note that the standardized upper limit $(\widehat{UCL}-\widehat{\mu})/\sigma^*$ in (2.1) simply equals u_p. For heavier tails, this value should increase, while for lighter tails the opposite should happen. This readily suggests to generalize u_p into a power like $u_p^{1+\gamma}$. Obviously $\gamma > 0$ (< 0) corresponds to heavy (light) tails, with the normal case as an interior point for $\gamma = 0$. The so obtained normal power family has intuitive appeal, is flexible, contains the normal, and moreover clearly covers a broad range of distributions. Consequently, it will be the choice we shall adopt in the present chapter as well.

Instead of simply working under the model $X = \mu + \sigma Z$, in which Z has d.f. Φ, with the new model suppose that $X = \mu + \sigma Z_\gamma$, with γ, Z_γ and $c(\gamma)$ as given in Section 3.4. The d.f. of Z_γ is denoted by K_γ and its upper p-quantile $(0 < p < 1/2)$ is given in (3.7). The difference between the normal chart and the normal power chart is that for the latter u_p in (2.1) needs to be replaced by $c(\widehat{\gamma})u_p^{1+\widehat{\gamma}}$, which indeed corresponds to the basic idea suggested above.

Regarding the parameters μ and σ, we used the same estimators as given in Chapter 3, where $\widehat{\mu} = \bar{X}$ and $\breve{\sigma} = S$ are used, respectively to estimate μ and σ. And as concerns the choice for the required estimator of γ, we point out that a kurtosis-type estimator, cf. (3.11), has an obvious appeal when worrying about heavy tails. We investigated this possibility and indeed the resulting $\widehat{\gamma}_1$ performs quite well for a large variety of symmetric distributions.

However, it is not our aim to fit the distribution globally, but only at the (far) right tail. This is done by concentrating on the ordinary tail (where we have observations), assuming that the behavior of the ordinary tail is informative for the far tail. Thus the restriction to symmetry occurring in the normal power family can be taken for granted. Consequently, instead of using all observations, we prefer to concentrate on the upper tail resulting in $\widehat{\gamma}_2$ as given in (3.14). More particularly, we recommend $\widehat{\gamma}$ as given in (3.15). The chosen estimator for γ can then be applied to the estimated control limit of the parametric chart as given in (3.8).

For the performance of the general parametric chart based on the normal power family, in terms of the produced $(R)ME$ and $(R)SE$ see Section 3.5. It turns out that the parametric chart has an advantage over the normal chart, since the new chart manages to reduce RME of the suitable models substantially. Hence, the parametric chart produces sufficiently small ME's for those suitable models compared to the normal chart.

However, we also have to take the opposite side of the picture into account: the comparison of the various $SE's$. In this respect, clearly the parametric chart is at a disadvantage. Not only μ and σ need to be estimated here, but γ as well. The question is whether the effects of the resulting increase of SE are sufficiently small to be outweighed by the above illustrated gain with respect to ME. In Section 3.5 this was shown to be true for the case where bias is the performance criterion. Adequate corrections were derived in Section 3.6, resulting in charts which combine small ME with small expected SE. But in the present chapter we are interested in bounding exceedance probabilities rather than in bounding bias. As argued in the introduction, this type of criterion has a larger impact and, in particular, it will necessitate larger corrections. In the next section we shall analyze how this works out.

4.3 Exceedance probabilities and correction terms

In this section we temporarily take a general parametric family with d.f. K_γ. In order to be able to correct the behavior of the parametric chart, we replace the $\widehat{UCL} = \widehat{\mu} + \breve{\sigma}\overline{K}_{\widehat{\gamma}}^{-1}(p)$ by

$$\widehat{UCL}_c = \widehat{\mu} + \breve{\sigma}\overline{K}_{\widehat{\gamma}}^{-1}(p) + \sigma c_e(\gamma), \tag{4.1}$$

in which $c_e(\gamma)$, or c_e for short, will be an appropriate correction term with the corresponding estimated version $c_e(\widehat{\gamma})$ or $\widehat{c}_e$ for short; different estimators give different functions c_e. Obviously, by letting $c_e = 0$, we will always be able to reproduce the uncorrected charts, as $\widehat{UCL}_0 = \widehat{UCL}$. Before embarking on the precise definition of c_e, it may be illuminating to take a qualitative preview of what to expect. The convergence of $g(P_n)$ toward its limit may be slow, but in the long run any corrections, whether dealing with bias or with exceedance probability, should become negligible. Hence c_e will typically be small in comparison to $\overline{K}_\gamma^{-1}(p)$. Moreover, it will be a deterministic function of the parameters involved: the sample size n, the intended signal probability p, the yet to be specified criterion parameters and, of course, the third model parameter γ. In fact, this latter dependence we already indicated specifically by writing $c_e = c_e(\gamma)$ in (4.1). This was also done to make explicit that, in view of the structure chosen in (4.1), the correction c_e will depend on γ, but not on μ or σ. To allow application of the correction, the third model parameter γ has to be estimated again but, as c_e is small, the error involved in replacing $\sigma c_e = \sigma c_e(\gamma)$ by $\breve{\sigma}\widehat{c}_e = \breve{\sigma} c_e(\widehat{\gamma})$ will be negligible.

Next, we will make explicit what imposing the exceedance criterion on $\widehat{UCL}_c$ from (4.1) means. Let $UCL = \mu + \sigma\overline{K}_\gamma^{-1}(p)$ be the limit in probability of $\widehat{\mu} + \breve{\sigma}\overline{K}_{\widehat{\gamma}}^{-1}(p)$. As P_n equals $\overline{F}(\widehat{UCL}_c)$, the question now is how likely it is that $g(P_n)$, for example for the three choices of g given in (2.3), differs too much from its corresponding limit value $g(\overline{F}(UCL))$. To be more precise, we introduce the relative error

$$W_c = \frac{g(\overline{F}(\widehat{UCL}_c))}{g(\overline{F}(UCL))} - 1, \tag{4.2}$$

with $W = W_0$ corresponding to the uncorrected case. Once more note that we are only dealing with SE here. Note that $ME = g(\overline{F}(UCL)) - g(p)$, which hopefully has been made small by using a larger family, remains a given quantity, no matter how large a sample size we choose. For increasing g (like $g(p) = p$ or $g(p) = 1 - (1-p)^k$), we impose the following exceedance probability criterion: for certain small non-negative ε and small positive α,

$$P(W_c > \varepsilon) \leq \alpha. \tag{4.3}$$

For decreasing g (like $g(p) = 1/p$), instead consider $P(W_c < -\varepsilon)$ in (4.3). However, as in previous chapters, we always assume that g is increasing in the present chapter, unless explicitly stated otherwise.

After specifying the criterion, we can derive c_e as a function of n, p, α, ε and γ. In the present chapter we use notations E_0 and P_0 to denote expectation and probability, respectively, under F_0 (i.e. for the case where we simply have $\mu = 0$ and $\sigma = 1$). Recall that F_0 denotes the d.f. of $(X_{n+1} - \mu)/\sigma$. We assume the natural condition that $(\widehat{\mu} - \mu)/\sigma$, $\breve{\sigma}/\sigma$ and $\widehat{\gamma}$ have the same distribution under F as $\widehat{\mu}$, $\breve{\sigma}$ and $\widehat{\gamma}$ under F_0, respectively. This condition obviously holds for $\widehat{\mu} = \overline{X}$, $\breve{\sigma} = S$ and $\widehat{\gamma}$ or $\widehat{\gamma}_2$. Let

$$b_0 = \overline{F}_0^{-1}(g^{-1}(\{g(\overline{F}_0(\overline{K}_\gamma^{-1}(p)))(1+\varepsilon)\})). \tag{4.4}$$

As ε is supposed to be small, this implies that $b_0 - \overline{K}_\gamma^{-1}(p) \approx -\varepsilon Q$, with

$$Q = g(\overline{F}_0(\overline{K}_\gamma^{-1}(p)))/\{g'(\overline{F}_0(\overline{K}_\gamma^{-1}(p)))f_0(\overline{K}_\gamma^{-1}(p))\}, \tag{4.5}$$

where, as before, f_0 is the corresponding density $f_0 = F_0'$.

In this case we have:

Lemma 4.3.1 *Suppose V from (3.18) satisfies $V/(E_0V^2)^{\frac{1}{2}} \xrightarrow{P_0} N(0,1)$, then $lim_{n\to\infty}\ P(W_c > \varepsilon) = \alpha$ holds for*

$$c_e = (E_0V^2)^{\frac{1}{2}}u_\alpha + b_0 - \overline{K}_\gamma^{-1}(p). \tag{4.6}$$

Proof.
Note that (4.3) translates into $P(\widehat{UCL}_c < b) \leq \alpha$, where b is b_0 from (4.4) with F rather than F_0 and UCL instead of $\overline{K}_\gamma^{-1}(p)$.

Consequently, $b = \mu + \sigma b_0$. From (4.1) it follows that $P(\widehat{UCL}_c < b) = P((\widehat{\mu} - \mu)/\sigma + (\breve{\sigma}/\sigma)\overline{K}_{\widehat{\gamma}}^{-1}(p) + c_e < b_0) = P_0(V + c_e < b_0 - \overline{K}_\gamma^{-1}(p))$, with V as in (3.18). The desired result now is immediate in view of the conditions on V. ■

Typically E_0V and E_0V^2 are of order n^{-1} and hence the "standardization" of V reduces to $V/(E_0V^2)^{\frac{1}{2}}$. Next, we study the behavior of c_e from (4.6) in relation to the underlying parameters. Of main interest in this respect are n and ε.

1) role of n

Implicitly n is present in (4.6) through $(E_0V^2)^{\frac{1}{2}}$, which is typically of order $n^{-\frac{1}{2}}$. Note that this will indeed be the case for the normal power model, with $\widehat{\mu} = \overline{X}$, $\breve{\sigma} = S$ and $\widehat{\gamma}$ or $\widehat{\gamma}_2$. Also observe that this order $n^{-\frac{1}{2}}$ is considerably larger than the order n^{-1} from the bias case (see Section 2.3). Nevertheless, as n increases, this part of c_e will eventually become negligible. The speed of the convergence will decrease as the number of parameters to be estimated increases. For the normal chart from (2.1), (3.18) boils down to $V = \widehat{\mu} + (\breve{\sigma} - 1)u_p$ and thus $E_0V^2 \approx (u_p^2 + 2)/(2n)$. For the normal power family, in addition to μ and σ we need to estimate γ, and E_0V^2 will be larger.

2) role of ε

First of all, for the boundary case where $\varepsilon = 0$, we have from (4.4) that $b_0 = \overline{K}_\gamma^{-1}(p)$ and the second part of (4.6) vanishes, as should be the case. As ε is small, this part can moreover be approximated by $-\varepsilon Q$, with Q given by (4.5). To simplify this a bit further, we can apply a similar argument as in Subsection 2.3.2 as follows. Let $h(p) = g(p)/\{pg'(p)\}$. For $g(p) = p$ we get $h(p) = 1$, for $g(p) = 1/p$ we find $h(p) = -1$, while for $g(p) = 1 - (1 - p)^k$ we obtain $h(p) = (1-p)\left\{(1 - p)^{-k} - 1\right\}/(kp) \approx (e^{kp} - 1)/(kp)$. Usually kp will be small and in the last case $h(p) \approx 1 + (kp)/2$. As a result we obtain as a further approximation

$$c_e = (E_0V^2)^{1/2}u_\alpha - \varepsilon\lambda\overline{F}_0(\overline{K}_\gamma^{-1}(p))/f_0(\overline{K}_\gamma^{-1}(p)), \tag{4.7}$$

where $\lambda = 1$ for $g(p) = p$. For $g(p) = 1/p$, $h(p) = -1$, but we also deal with $1 - \epsilon$ here, rather than with $1 + \epsilon$ because g is decreasing, hence also for this function we get $\lambda = 1$. Finally, for the third function g given in (2.3), $\lambda \approx 1 + (kp)/2$.

The coefficient of ε in (4.7) obviously is negative, which agrees with the fact that increasing ε makes the criterion (4.3) less strict and thus requires a smaller correction. Moreover, (4.7) shows that c_e consists of a small positive term, typically of order $n^{-\frac{1}{2}}$, and a small negative term of order ε. Of course, these two can be tied together by prescribing that actually $\varepsilon = \varepsilon_n = \tilde{c}n^{-\frac{1}{2}}$ for some $\tilde{c}$, but this does not seem to really add anything to the analysis. Clearly, as n becomes huge, for any small positive ε the correction will eventually become negative. This simply reflects the fact that then W_0 tends to be so close to zero that it exceeds ε with a probability less than α. On the other hand, although this situation makes perfect sense, it is also obvious that it will be of little practical interest. Typical applications will exhibit a positive term of order $n^{-\frac{1}{2}}$ which dominates, and is merely moderated a bit by the negative term of order ε. To provide some illustration, we consider the following example.

Example 4.3.1.

If F is indeed contained in the model used, we simply have $F_0 = K_\gamma$ and a factor $p/k_\gamma(\overline{K}_\gamma^{-1}(p))$, with k_γ the density of K_γ, results in the second term from (4.7). If we are moreover in the normal case, this approximately boils down to u_p^{-1} and consequently (4.7) reduces to the approximation (cf. (2.53)),

$$c_e \approx -\frac{\varepsilon\lambda}{u_p} + u_\alpha \left(\frac{u_p^2 + 2}{2n} \right)^{1/2}. \tag{4.8}$$

According to this approximation, correction becomes superfluous once $n \approx \frac{1}{2} u_\alpha^2 u_p^2 (u_p^2 + 2)/(\varepsilon\lambda)^2$. Choosing for example $p = 0.001$, $\varepsilon = 0.1$, $\alpha = 0.25$ and $\lambda = 1$, this produces $n \approx 2500$. Hence for (substantially) smaller n, the correction c_e will be positive and non-negligible. Further numerical illustration was given in Section 2.3. □

3) role of g

The effect of the choice of g, i.e., whether we concentrate on the signal probability itself, or rather on the ARL, or even on the distribution of RL, is represented concisely by the factor λ in (4.7). Note in particular that the fact that $\lambda = 1$ for both $g(p) = p$ and $g(p) = 1/p$ is of course no coincidence. In fact, in the latter case (4.3) gives $P(W_c < -\varepsilon) = P(\overline{F}(UCL)/P_n < 1 - \varepsilon) = P(P_n > \overline{F}(UCL)/(1 - \varepsilon))$, which is nothing but $P(W_c > \tilde{\varepsilon})$ in the case $g(p) = p$, with $\tilde{\varepsilon} = \varepsilon/(1 - \varepsilon) \approx \varepsilon$ for small ε. Hence, any result for the signal probability itself immediately translates into one about the ARL.

As concerns the remaining parameters of c_e in (4.6) or (4.7), we shall be more concise. If α decreases, (4.3) becomes more strict, which is reflected by the fact that the positive term in (4.6) increases linearly in u_α. Moreover, in accordance with our claim in the introduction that very small p cause very large relative errors, $E_0 V^2$ decreases in p. For the normal power family we will see this in the next section, while for the normal family this is already clear from (4.8), where the positive (negative) term increases (decreases) in u_p.

The next step toward application is the estimation of the unknown parts in (4.6) and (4.7). As argued before, the fact that c_e is small to begin with causes these changes to have negligible effects and approximate equality in (4.3) will continue to hold for $\widehat{c}_e$. First, assume that we are in fact within the parametric family, and thus $F_0 = K_\gamma$. Hence, we can simply use $\overline{K}_{\widehat{\gamma}}$ for $\overline{F}_0$, $k_{\widehat{\gamma}}$ for f_0 and also evaluate E_0V^2 under $K_{\widehat{\gamma}}$. Denoting the latter by $\widehat{E_0V^2}$, we arrive from (4.6) at

$$\widehat{c}_e = c_e(\widehat{\gamma}) = (\widehat{E_0V^2})^{1/2}u_\alpha + \overline{K}_{\widehat{\gamma}}^{-1}(g^{-1}(\{g(p)(1+\varepsilon)\})) - \overline{K}_{\widehat{\gamma}}^{-1}(p), \tag{4.9}$$

which reduces through (4.7) to

$$\widehat{c}_e = c_e(\widehat{\gamma}) = (\widehat{E_0V^2})^{1/2}u_\alpha - \frac{\varepsilon\lambda p}{k_{\widehat{\gamma}}(\overline{K}_{\widehat{\gamma}}^{-1}(p))}. \tag{4.10}$$

Hence, with $\widehat{c}_e$ from (4.9) or (4.10), the estimated upper limit from (4.1) as given in (3.17) will now produce approximate equality in (4.3) under the model $K_\gamma((x-\mu)/\sigma)$. This will also be true in the vicinity of this model: as long as F_0 is close to K_γ, estimation of F_0, f_0 and E_0V^2 in $\widehat{c}_e$ through $K_{\widehat{\gamma}}$ will make sense. Actual application in case of the special normal model already requires (see (4.8)) specification of p, ε, α and g, as well as evaluation through $(X_1, \ldots, X_n)$ of $\widehat{\mu}$ and $\breve{\sigma}$. Under a more general model, in addition, K_γ should obviously be defined, as well as an estimator for γ, while also $\widehat{E_0V^2}$ has to be evaluated as a function of this estimator. Especially this last step may require considerable effort. Fortunately, for the normal power model this has already been done in Section 3.6 while deriving the bias correction term, and we can readily use these results here. In the next section we shall investigate the performance of the thus obtained normal power family chart with exceedance probability correction.

4.4 The corrected parametric chart

In the previous section we have uncovered the general structure of the desired correction terms; here we shall demonstrate how (4.9) and (4.10) work out in practice for our prototype example, the normal power family. The first step is:

Lemma 4.4.1 *For the normal power family the results (4.9) and (4.10) specialize to: use*

$$\widehat{UCL}_c = \widehat{\mu} + \breve{\sigma}\{c(\widehat{\gamma})u_p^{1+\widehat{\gamma}} + \widehat{c}_e\}, \tag{4.11}$$

with

$$\widehat{c}_e = (\widehat{E_0V^2})^{1/2}u_\alpha + c(\widehat{\gamma})\{u_{\tilde{p}}^{1+\widehat{\gamma}} - u_p^{1+\widehat{\gamma}}\}, \tag{4.12}$$

where $\tilde{p} = g^{-1}(\{g(p)(1+\varepsilon)\})$*, or with the further simplification*

$$\widehat{c}_e = (\widehat{E_0 V^2})^{1/2} u_\alpha - \varepsilon\lambda p(1+\widehat{\gamma})c(\widehat{\gamma})u_p^{\widehat{\gamma}}/\varphi(u_p). \qquad (4.13)$$

Proof.
In the normal power family $K_\gamma^{-1}(t) = c(\gamma)|\Phi^{-1}(t)|^{1+\gamma} sign(\Phi^{-1}(t))$. Together with (4.9) this readily gives (4.12). As moreover $\{1/k_\gamma(\overline{K}_\gamma^{-1}(1-t))\} = 1/k_\gamma(K_\gamma^{-1}(t)) = \{\overline{K}_\gamma^{-1}(t)\}' = (1+\gamma)c(\gamma)\Phi^{-1}(t)^\gamma/\varphi(\Phi^{-1}(t))$ for $t > \frac{1}{2}$, using (4.10) rather than (4.9) leads to (4.13). ∎

The estimator of γ to be used is again $\widehat{\gamma}$ or $\widehat{\gamma}_2$. Indeed the main obstacle is $\widehat{E_0 V^2}$, which has to be expressed in terms of $\widehat{\gamma}$ or $\widehat{\gamma}_2$ as well. After laborious computations the following result is obtained.

Lemma 4.4.2 *Using* $\widehat{\gamma}$ *as given in (3.15) we arrive at*

$$\{\widehat{E_0 V^2}\}^{1/2} \approx n^{-1/2} A(\widehat{\gamma}, u_p),$$

where

$$A(\widehat{\gamma}, u_p) = -4.00 - 12.54\widehat{\gamma} - 10.02\widehat{\gamma}^2 + 2.91u_p + 6.47\widehat{\gamma}u_p + 4.42\widehat{\gamma}^2 u_p. \qquad (4.14)$$

Proof.
Note that $\widehat{\gamma}$ from (3.15) is nothing but the explicit $\widehat{\gamma} = h_2^{-1}(\widehat{\gamma}_2^*)$, with $\widehat{\gamma}_2^*$ as in (3.12), using again $q = 0.05$ and $r = 0.25$ and $\log(u_{0.05}/u_{0.25})$ replaced by the numerical value 1/1.1218. The computation of $(E_0 V^2)^{\frac{1}{2}}$ is a very technical and rather complicated matter. Fortunately, this task has been performed already in Section 3.6 for $E_0 V$ and $E_0 V^2$, in order to obtain an explicit form of the bias correction. We refer to that section for technical details and merely present here an outline of the steps involved. The basic idea is actually quite straightforward. V from (3.18) is a function $\tilde{H}(\widehat{\mu}, \breve{\sigma}, \widehat{\gamma}_2^*)$, to which a second order expansion around $\tilde{H}(0, 1, \gamma_2^*)$ is applied. After this, the first and second order (mixed) moments of $\widehat{\mu}$, $(\breve{\sigma} - 1)$ and $(\widehat{\gamma}_2^* - \gamma_2^*)$ need to be calculated, all of which are of order n^{-1}. For those involving only $\widehat{\mu}$ and/or $\breve{\sigma}$, this is immediate; as soon as $(\widehat{\gamma}_2^* - \gamma_2^*)$ is involved, the computation is more tedious. Plugging the moment results into the expansion gives as a first version of the desired result an approximation up to $o(n^{-1})$ (see Section 3.6.2). However, as the expansion itself contains complicated coefficients such as the derivative of $\overline{K}_{h_2^{-1}(\gamma_2^*)}^{-1}(p)$ (cf. (3.12) and (3.18)), the result is still quite complicated. Therefore, the asymptotic results are complimented with further numerical approximations for such coefficients leading to more amenable results. Using these two groups of results on $(E_0 V^2)^{\frac{1}{2}}$ produces (4.14). ∎

Note that $\widehat{UCL}_c$ from (4.11) with $\widehat{c}_e$ from (4.12), or the further approximation from (4.13), is now made completely explicit through (4.14) and can be applied in a straightforward manner.

Remark 4.4.1 For $g(p) = p$, we get $\tilde{p} = p(1+\epsilon)$. Hence, using Lemma 4.4.2 and (4.12) we obtain

$$\widehat{UCL}_c = \widehat{\mu} + \breve{\sigma} \left\{ c(\widehat{\gamma}) u_{p(1+\epsilon)}^{1+\widehat{\gamma}} + \frac{A(\widehat{\gamma}, u_p) u_\alpha}{\sqrt{n}} \right\}. \tag{4.15}$$

Since the starting point of a two-sided control chart with prescribed false alarm rate p is a combination of two one-sided control charts each with prescribed false alarm rate $p/2$, we will treat for the two-sided case here the upper and lower limit also separately. This differs slightly from the approach in Chapter 2, see Remark 2.3.5. Hence, for the two sided-case (4.15) translates into

$$\widehat{CL}_c = \widehat{\mu} \pm \breve{\sigma} \left\{ c(\widehat{\gamma}) u_{p(1+\epsilon)/2}^{1+\widehat{\gamma}} + \frac{A(\widehat{\gamma}, u_{p/2}) u_\alpha}{\sqrt{n}} \right\}, \tag{4.16}$$

with the estimator of γ at the upper tail, $\widehat{\gamma}_U$, given by

$$\widehat{\gamma}_U = 1.1218 \left(\frac{X_{([0.95n+1])} - \overline{X}}{X_{([0.75n+1])} - \overline{X}} \right) - 1$$

and the estimator of γ at the lower tail, $\widehat{\gamma}_L$, given by

$$\widehat{\gamma}_L = 1.1218 \left(\frac{\overline{X} - X_{(n-[0.95n])}}{\overline{X} - X_{(n-[0.75n])}} \right) - 1$$

and $A(\widehat{\gamma}, u_{p/2})$ as given in (4.14) with u_p replaced by $u_{p/2}$. □

To see how the proposal in (4.11) actually works out, we shall next perform a simulation study along the lines of Section 3.7. To avoid repetition, however, we shall not be as extensive as in that section. For example, we shall not dwell again on the improvement achieved concerning ME by using a parametric rather than a normal chart. This gain has been amply demonstrated in Section 3.5 and the situation in that respect is exactly the same here. Hence we shall neither consider the normal chart at any point, nor report the ME involved. What matters is whether the present corrections work well for controlling SE under the present criterion. We shall concentrate on g(p) = p; from the derivations given (cf. remark 3 following Lemma 4.3.1) it is evident that completely similar results will hold for the other two choices of g. Moreover, we will always let $p = 0.001$ and use 10,000 repetitions in the simulations.

Sample sizes n involved will range from 250 to 2000. Note that these values are considerably higher than those in Section 3.7, where the values 100, 250 and 500 are considered, with the emphasis on 100. This reflects the fact that in the present situation we are dealing with corrections of order $n^{-1/2}$ rather than of order n^{-1}, thus requiring larger values of n. As concerns the constants α and ε used in setting our criterion cf. (4.3), we shall use $\alpha = 0.2$, and either $\varepsilon = 0$ or $\varepsilon = 0.1$.

The underlying distributions will be the same as in Section 3.7. First of all include the normal d.f. Φ, corresponding to the normal family. Then add K_γ as defined through Z_γ in Section 3.3.3, for the values $\gamma = -0.5, -0.25, 0.25, 0.75$ and 1 ($\gamma = 0$ is already covered by Φ), thus representing the normal power family. Subsequently, it is only fair to add a number of cases outside either family. To begin with, include T, the Student d.f. with 6 degrees of freedom and standardized to unit variance. Next, add the random mixture RM and the deterministic mixture DM as defined in Section 3.3.1 and 3.3.2, respectively. In addition, consider Tukey's λ-family, defined in Section 3.3.4. Include the corresponding d.f.'s for $\lambda = -0.1$, 0 (which corresponds to the standardized logistic d.f.) and 0.14 (which is very close to the standard normal (outside the tails!)). Finally, take the orthonormal family as defined in Section 3.3.5, and apply the corresponding d.f. for $k = 3, \varsigma_1 = \varsigma_2 = -0.1$ and $\varsigma_3 = 0.1$.

Table 4.1a Simulated exceedance probabilities (in %) without ($P(W_0 > \varepsilon)$) correction, using (cf. (4.3)) $\varepsilon = 0$ or 0.1 and $\alpha = 0.2$. The first percentage in each cell corresponds to $\varepsilon = 0$; the second to $\varepsilon = 0.1$.

F_0	$P(W_0 > \varepsilon)$							
	$n = 250$		$n = 500$		$n = 1000$		$n = 2000$	
Φ	52	49	51	46	51	44	50	40
$K_{-0.5}$	50	47	50	46	50	45	51	43
$K_{-0.25}$	51	48	50	46	50	43	50	41
$K_{0.25}$	52	49	51	46	51	43	51	41
$K_{0.5}$	54	50	52	46	51	44	50	39
$K_{0.75}$	53	49	53	47	51	43	50	39
K_1	55	51	53	48	52	44	51	40
T	53	47	51	43	51	39	50	34
RM	53	47	51	43	51	39	50	34
DM	52	47	51	43	50	39	51	37
$TU(-0.1)$	53	47	51	43	51	38	51	34
$TU(0)$	52	47	50	43	51	40	51	37
$TU(0.14)$	51	48	51	47	51	44	51	42
O	52	49	51	46	51	45	51	43

Table 4.1b Simulated exceedance probabilities (in %) with ($P(W_c > \varepsilon)$) correction, using (cf. (4.3)) $\varepsilon = 0$ or 0.1 and $\alpha = 0.2$. The first percentage in each cell corresponds to $\varepsilon = 0$; the second to $\varepsilon = 0.1$.

F_0	$P(W_c > \varepsilon)$							
	$n = 250$		$n = 500$		$n = 1000$		$n = 2000$	
Φ	24	24	23	22	22	22	22	21
$K_{-0.5}$	21	21	19	19	20	20	20	20
$K_{-0.25}$	23	23	22	22	21	21	21	21
$K_{0.25}$	25	25	23	23	23	23	22	22
$K_{0.5}$	25	25	23	23	22	22	21	21
$K_{0.75}$	25	25	23	23	22	22	21	20
K_1	27	27	24	24	22	22	22	22
T	28	26	26	23	26	21	26	19
RM	26	25	24	21	24	21	23	18
DM	25	24	25	22	24	21	23	19
$TU(-0.1)$	29	27	28	24	25	21	26	19
$TU(0)$	27	26	25	23	24	21	24	20
$TU(0.14)$	24	24	22	22	22	22	23	23
O	23	24	22	23	21	23	21	23

The simulation results without correction term are presented in Table 4.1a and those with the correction term in Table 4.1b. As can be seen from these tables, the correction term works quite well. To be more specific, first consider the normal power family when no correction is used. For $\varepsilon = 0$, we are then simply looking at $P(P_n > p)$, which is seen to stabilize around $\frac{1}{2}$. Hence indeed, as remarked in the introduction, p turns out to be close to the 50%-quantile of the distribution of P_n. Note that for distributions outside the normal power family p should be replaced by $\overline{F}(UCL)$. Increasing ε from 0 to 0.1 should help in this respect: as $n \to \infty$, $P(P_n > p(1+\varepsilon)) \to 0$ for $\varepsilon > 0$.

But from Table 4.1a and Table 4.1b we see that apparently, this convergence is quite slow. Even for n as large as 2000, the exceedance probabilities are still larger than 35%. Hence, corrections are certainly in order if values of α well below $\frac{1}{2}$ are desired.

From those tables it is evident that applying such a correction for $\alpha = 0.2$ indeed brings the exceedance probabilities down to values which are close to this desired 20%, both for $\varepsilon = 0$ and $\varepsilon = 0.1$. For $n = 250$, the fluctuations may still be considered to be a bit large, but from $n = 500$ on, the result seems quite satisfactory for practical purposes. Also observe that, although the correction terms are based on the normal power family, they work also rather well outside this family.

4.5 The out-of-control situation

In this section we shall study the impact of the adaptations from the previous sections on the out-of-control behavior of the chart. As we have seen, during the in-control stage, the uncorrected chart tends to stop considerably earlier than anticipated in an unacceptably large fraction of its applications. The corrections serve to bring this fraction down to acceptable proportions ("controlling the in-control behavior") and as such typically delay the moment of stopping somewhat. Clearly, this tendency will be noticeable during out-of-control as well. But there it actually is unwelcome, as it means that detection will take a bit longer as well. Hence, a balance will have to be found between controlling performance during in-control and loss of detection power during out-of-control.

Let X_{n+1} come from a shifted d.f. $F_0(x-\Delta)$, where Δ is such that $\tilde{p} = \overline{F}_0(UCL-\Delta)$ is no longer extremely small, like p. For simplicity, and without loss of generality, we again let $\mu = 0$ and $\sigma = 1$ and thus work under the standardized d.f. F_0. Consequently, the relative errors caused by the replacement of this $\tilde{p}$ by its stochastic counterpart P_n, which in view of (3.17) now equals $\overline{F}_0(\widehat{UCL}_c - \Delta)$, will be much smaller than those during the in-control situation (also cf. Tables 2.13 and 2.15 from Chapter 2). Hence, during out-of-control situation there seems to be no need to use exceedance probabilities again, as a more simple first order expectation approach will already suffice to exhibit the resulting behavior of the chart.

To be specific, let E_Δ denote expectation under $F_0(x-\Delta)$ and introduce

$$\begin{aligned} E(\Delta, \widehat{c}_e) &= E_\Delta g(\overline{F}_0(\widehat{UCL}_c - \Delta)), \\ RC &= \begin{cases} 1 - E(\Delta, \widehat{c}_e)/E(\Delta, 0) & \text{for increasing } g \\ E(\Delta, \widehat{c}_e)/E(\Delta, 0) - 1 & \text{for decreasing } g. \end{cases} \end{aligned} \tag{4.17}$$

Clearly, $E(\Delta, \widehat{c}_e)$ and $E(\Delta, 0)$ stand for $E_\Delta g(P_n)$ with and without correction, respectively. Moreover, RC expresses the relative cost incurred by having to use the correction $\widehat{c}_e$. A simple example explains its meaning: take $g(p) = 1/p$, let $p = 0.001$ and suppose that Δ is such that $\tilde{p} = 0.05$. Then $E(\Delta, 0)$ will be close to 20, and a value for RC of 20% means that using the correction $\widehat{c}_e$ has pushed up the ARL during out-of-control situation by 20% to about 24. Note that of course the notion "cost" is somewhat virtual and should therefore be interpreted with care.

In fact, there are three approaches to be distinguished. In the first case, everything is known, no estimation is needed, and both $g(p)$ during the in-control situation and $g(\tilde{p})$ during the out-of-control situation are achieved effortlessly. The only drawback is that this situation rarely occurs, so one typically has to settle for one of the two remaining ones: to correct or not to correct. In the latter case, $g(\tilde{p})$ indeed is achieved, apart from a usually acceptably small relative error, but, as we saw, the price in terms of exceedance probabilities is unacceptably large with respect to $g(p)$, caused by the very small values of p that are typically used. Hence, the remaining candidate uses the correction, thus repairing the damage during the in-control process (cf. Table 4.1a and Table 4.1b), but at a price under the out-of-control situation as expressed by RC from (4.17).

To obtain an impression of what can be expected during the out-of-control situation, we have repeated the simulations from Section 4.4 for the same choices of p, n, g, α, ε and F_0 as used there. For each F_0 we have selected two values of Δ such that reasonable values of $\tilde{p}$ result, i.e. $\tilde{p}$ considerably larger than $p = 0.001$. Usually, $\Delta = 2$ and $\Delta = 3$ will do, but as γ moves away from 0, the values $\Delta = 1$ or $\Delta = 4$ can become more appropriate. In Table 4.2 the results have been collected: presented are the simulated average P_n under the out-of-control situation when no correction is used, which are close to $\tilde{p}$. In addition, the relative costs RC are given in percentages, using $\varepsilon = 0$ and $\varepsilon = 0.1$ in the correction $\widehat{c}_e$, respectively.

From Table 4.2 it is evident that the values of RC are decreasing nicely as n becomes larger (cf. the much slower decrease of the exceedance probabilities for the case without correction from Table 4.1). Moreover, the percentages occurring can be considerable, and hence the use of such larger values of n can indeed be felt to be necessary. Certainly for $n = 250$, the values are quite large and only from $n = 1000$ on small percentages start to prevail. Note that becoming more liberal in (4.3) by increasing ε from 0 to 0.1 does indeed help, but not so much: the percentages drop, but not dramatically so. Another comment is that things go relatively well for the choices of F_0 outside the normal power family. The most unfavorable cases occur within this family, for large positive values of γ.

All in all, the results in Table 4.2 tend to confirm the expectations expressed in the introduction. The present type of correction is the most ambitious one of the four types considered which are correcting bias or exceedance probabilities, assuming normality or not, and the impact on the out-of-control behavior is no longer negligible. Applying a correction of this type may very well be a good idea (in principle even the best among the available four), but it should not be applied automatically. Some thought should be given to points such as what value of RC is still acceptable, whether the n at hand is sufficiently large (or can be made so) to realize that value, whether increasing α might be an option, etc.

In view of the discussion above, we shall devote the remainder of this section to obtaining a better insight into the behavior of e.g. RC as a function of the rather large number of underlying parameters, such as p, n, α, ε, γ and Δ. To this end we shall consider some further approximations to the quantities from (4.17), with the following result.

Table 4.2 Simulated values of $E(\Delta, 0) = E_\Delta P_n$ (see (4.17)), together with the relative costs due to correction RC (in %), using $\varepsilon = 0$ and $\varepsilon = 0.1$, respectively. Throughout are used: $p = 0.001$, $g(p) = p$ and $\alpha = 0.2$.

F_0	Δ	$n = 250$			$n = 500$			$n = 1000$			$n = 2000$		
Φ	2	0.15	37	34	0.14	27	24	0.14	20	16	0.14	14	10
	3	0.47	23	21	0.46	16	14	0.46	11	9	0.46	8	6
$K_{-0.5}$	1	0.23	24	23	0.23	17	15	0.23	12	10	0.23	8	7
	2	0.50	2	2	0.50	1	1	0.50	0	0	0.50	0	0
$K_{-0.25}$	1	0.065	42	38	0.061	31	28	0.059	23	19	0.059	17	12
	2	0.35	19	17	0.35	13	12	0.35	9	8	0.35	7	5
$K_{0.25}$	2	0.061	47	43	0.055	36	31	0.052	27	22	0.050	20	14
	3	0.24	40	37	0.22	30	26	0.22	22	18	0.21	16	11
$K_{0.50}$	3	0.11	50	45	0.094	39	33	0.088	29	23	0.085	21	15
	4	0.39	47	43	0.36	39	33	0.34	30	24	0.32	22	16
$K_{0.75}$	3	0.054	55	51	0.044	43	38	0.040	33	26	0.037	25	17
	4	0.20	57	52	0.16	45	39	0.14	34	28	0.13	25	18
K_1	3	0.032	57	53	0.024	46	40	0.021	35	28	0.020	26	19
	4	0.11	60	56	0.076	48	42	0.062	37	30	0.057	28	20
T	2	0.091	40	37	0.081	31	27	0.077	23	19	0.074	17	12
	3	0.36	34	31	0.34	26	22	0.34	19	15	0.34	13	9
RM	2	0.12	40	36	0.11	30	26	0.10	22	18	0.10	16	12
	3	0.41	29	26	0.40	21	18	0.40	15	12	0.40	11	7
DM	2	0.12	40	36	0.11	30	26	0.10	22	18	0.10	16	11
	3	0.41	29	26	0.41	21	18	0.41	15	12	0.41	10	7
$TU(-0.1)$	2	0.071	41	37	0.061	31	27	0.057	24	19	0.055	17	12
	3	0.29	39	36	0.27	30	26	0.26	23	18	0.26	17	12
$TU(0)$	2	0.098	41	37	0.088	31	27	0.084	23	18	0.083	17	12
	3	0.37	32	29	0.35	24	21	0.35	18	14	0.35	13	9
$TU(0.14)$	2	0.15	37	34	0.14	28	24	0.14	20	17	0.14	14	11
	3	0.46	23	22	0.46	17	15	0.46	12	10	0.46	8	6
O	2	0.092	44	40	0.086	33	29	0.084	24	19	0.083	18	12
	3	0.32	32	29	0.31	24	20	0.30	17	14	0.30	12	9

Lemma 4.5.1 *RC from (4.17) approximately equals* $\widetilde{RC}$, *where*

$$\widetilde{RC} = \frac{4(1+u_{\tilde{p}})}{5\lambda(1+\gamma)c(\gamma)u_{\tilde{p}}^{\gamma}} A(\gamma, u_p) n^{-1/2} u_\alpha - \varepsilon \frac{(1+u_{\tilde{p}})u_p^{\gamma}}{(1+u_p)u_{\tilde{p}}^{\gamma}}, \tag{4.18}$$

in which $\tilde{p} = \overline{K}_\gamma(\overline{K}_\gamma^{-1}(p) - \Delta)$ *and* $A(\gamma, u_p)$ *as given by (4.14).*

Proof.
Note that $E(\Delta, c_e)$ can be approximated by $g(\overline{F}_0(\overline{K}_\gamma^{-1}(p) + c_e - \Delta))$ and con-

sequently RC to first order equals $\pm\ c_e f_0(\overline{K}_\gamma^{-1}(p)-\Delta)\ (g'/g)(\overline{F}_0(\overline{K}_\gamma^{-1}(p)-\Delta))$ with the $+$ sign for increasing g and the $-$ sign for decreasing g. Using steps for the three choices of g involved similar to those leading to (4.7), it follows that

$$RC \approx c_e f_0(\overline{K}_\gamma^{-1}(p)-\Delta)/\{\lambda \overline{F}_0(\overline{K}_\gamma^{-1}(p)-\Delta)\}, \tag{4.19}$$

where once more, $\lambda = 1$ for either $g(p) = p$ or $g(p) = 1/p$, while $\lambda = 1 + (kp)/2$ for $g(p) = 1-(1-p)^k$. To study the behavior of the expression in (4.19), we may specialize to the family $\{K_\gamma\}$. Note that F_0 for distributions outside the family $\{K_\gamma\}$ (under consideration in this chapter) is well approximated by a suitably chosen member K_γ in this family. Straightforward calculation shows that for $0 < \tilde{p} \le \frac{1}{2}$

$$k_\gamma(\overline{K}_\gamma^{-1}(p)-\Delta)/\overline{K}_\gamma(\overline{K}_\gamma^{-1}(p)-\Delta) = u_{\tilde{p}}^{-\gamma}\kappa(u_{\tilde{p}})/\{(1+\gamma)c(\gamma)\}, \tag{4.20}$$

where again $\tilde{p} = \overline{K}_\gamma(\overline{K}_\gamma^{-1}(p)-\Delta)$ and $\kappa(x)$ as given in Remark 2.4.1. It is well-known that $\kappa(x) \approx x$ for x large. But note that for smaller x, like for example $0 < x \le 3.09 = u_{0.001}$, the function κ can be approximated quite adequately by $4(1+x)/5$ (see Remark 2.4.1). Combination of (4.13) and (4.14), together with (4.19) and (4.20) then leads for $\tilde{p} < \frac{1}{2}$ to the result in (4.18). ■

By way of illustration, consider the following example

Example 4.5.1

Select the values $p = 0.001$, $\lambda = 1$ and $\alpha = 0.2$ which are used throughout Table 4.2 and consider the case $\gamma = 0$, corresponding to $F_0 = \Phi$. Then (4.18) reduces to $\widetilde{RC} = 3.36(4.09-\Delta)n^{-1/2} - \varepsilon(1-\Delta/4.09)$. For $\Delta = 2$ the resulting $7.03n^{-1/2} - 0.51\varepsilon$ produces for $\varepsilon = 0$ the percentages 44, 31, 22 and 16 for $n = 250, 500, 1000$ and 2000, respectively; for $\varepsilon = 0.1$, these percentages should be lowered by 5. For $\Delta = 3$ we have $3.66n^{-1/2} - 0.27\varepsilon$ and the resulting values for $\varepsilon = 0$ are 23, 16, 12 and 8, respectively for $n = 250, 500, 1000$ and 2000. These percentages should be lowered by 2 or 3 for $\varepsilon = 0.1$. Comparison to the simulated values from Table 4.2 for $F_0 = \Phi$ shows a nice agreement, especially for $\varepsilon = 0.1$. □

Incidentally, for the case $\gamma = 0$ considered in the example above, another interesting comparison can be made to the situation where normality is assumed to begin with, and thus estimation of γ is not needed.

Example 4.5.2

Suppose normality has been assumed and a normal chart is used. Then, in the situation of Example 4.5.1, we deal, according to (4.8) with $\{(u_p^2+2)/2\}^{1/2}$, rather than with $A(0, u_p)$. For $p = 0.001$, the former equals 2.40, while the latter equals 4.99. Hence, the fact that the additional parameter γ needs to be estimated calls in this example for a value of n which is $(4.99/2.40)^2 = 4.32$ times as high to reach the same precision. This illustrates that the impact of going from normality to a more general parametric family is indeed far from negligible. □

As intended, the result in (4.18) helps to shed some light on the behavior during out-of- control situation. To begin with, note that the influence of ε is indeed seen to be quite limited, as its coefficient in (4.18) is typically smaller than 1. Hence, from now on we concentrate on the case where $\varepsilon = 0$, i.e. where only the first term in (4.18) is relevant. Clearly, this term decreases in both n and α. The dependence on n and α is rather simple, since they are not mixed up with other parameters.

Furthermore, we can break up the remaining factor into the parts $4(1+u_{\tilde{p}})/\{5(1+\gamma)c(\gamma)u_{\tilde{p}}^{\gamma}\}$ and $A(\gamma, u_p)$. Let γ be fixed, then we observe that it also decreases in p, as $A(\gamma, u_p)$ increases linearly in u_p (cf. (4.14)). The situation with respect to $\tilde{p}$ is slightly more complicated: it decreases in $\tilde{p}$ $(0 < \tilde{p} < \frac{1}{2})$ only as long as $u_{\tilde{p}} > \gamma/(1-\gamma)$.

Finally, the behavior with respect to γ is more variable. On the one hand, $A(\gamma, u_p)$ increases quadratically in γ (for $p = 0.001$, we for example have that it behaves as $(7\gamma + 9)^2/16$), but the effect of $u_{\tilde{p}}^{\gamma}$ will obviously depend on whether $u_{\tilde{p}}$ is larger than or smaller than 1. By way of illustration we mention an upper bound for certain values of γ and $\tilde{p}$ that might by considered reasonable, in the sense that in many applications the given ranges will apply.

Example 4.5.3

It can be checked numerically that for $p = 0.001$ and $\alpha = 0.2$, the coefficient of $n^{-1/2}$ in (4.18) will not exceed 9.43 for γ and $\tilde{p}$ such that $-0.25 \le \gamma \le 0.3$ and $0.1 \le \tilde{p} \le 0.3$. □

A final remark is that the first term in (4.18) can be used in a straightforward manner to derive a lower bound for the n required. Let β be a small positive number, like 0.2 or 0.3, and suppose we want RC to be at most β, then it readily follows that we should let

$$n \ge \left\{ \frac{4(1+u_{\tilde{p}})A(\gamma, u_p)u_\alpha}{5(1+\gamma)c(\gamma)u_{\tilde{p}}^{\gamma}\beta} \right\}^2. \tag{4.21}$$

Example 4.5.4

Continuing Example 4.5.3, it follows that for this case (4.21) boils down to $n \ge \{9.43/\beta\}^2$, which equals 2222 and 988 for $\beta = 0.2$ and $\beta = 0.3$, respectively. Incidentally, if these computations are based, without further approximation steps, on $RC \approx g(\overline{F}_0(\overline{K}_\gamma^{-1}(p) + c_e - \Delta)) \,/\, g(\overline{F}_0(\overline{K}_\gamma^{-1}(p) - \Delta)) - 1$, the corresponding results are 1701 and 678, respectively. □

Note that the values of n obtained in Example 4.5.4 are in line with the impression already created by Table 4.2: the correction c_e works quite well, but considerable sample sizes are needed to avoid effects during out-of-control which might be considered too strong.

4.6 Concluding remarks

By assuming normality for the underlying distribution, standard control charts simply use the control limit given in (2.1). If normality fails, this causes large errors in the sense that during applications of the chart quite often the resulting stochastic signal probability and average run length differ unpleasantly much from the intended values. To bring the corresponding exceedance probabilities within prescribed bounds, it is demonstrated that a more general parametric model, based on the normal power family, offers good results.

The new chart, also called the parametric chart, uses a control limit as given in (3.8) having a γ estimator to select the appropriate type of distribution. In this control limit the required γ estimator can be selected from (3.11) or (3.14). However, the recommended $\widehat{\gamma}$ is given in (3.15). The advantage is that this parametric chart is able to reduce ME substantially. However, the new chart gives a disadvantage by producing a larger SE.

As a remedy for this disadvantage, correction terms based on the exceedance probability criterion (cf. (4.3)) are derived and given in (4.6) and (4.7) with the corresponding estimated versions given in (4.9) and (4.10), respectively; see also (4.12), (4.13) and Lemma 4.4.2. By adding such a correction term to the control limit we get the corrected control limit as given in (4.11). Using this corrected control limit, the exceedance probabilities are indeed adequately controlled (see Table 4.1).

The premium to be paid for this protection during the in-control situation unavoidably is some delay in stopping during the out-of-control situation. An expression for the relative cost RC (cf. (4.17)) caused by applying the correction term is obtained. It turns out that RC indeed becomes small as the sample size increases, (see Table 4.2), but not really fast. This is in agreement with the fact that the correction term offers protection against both non-normality and deviations in single applications of the chart. Hence, if possible, it is definitely advised to use this correction term, but some care is needed. To this end, an approximation $\widetilde{RC}$ for RC is given (see Lemma 4.5.1), which leads to a lower bound on the required sample size in terms of the parameters involved (cf. (4.21)).

Chapter 5

Nonparametric Approach

Due to the extreme quantiles involved, standard control charts are very sensitive to the effects of parameter estimation and non-normality. In the previous chapters more general parametric charts have been devised to deal with the latter complication and corrections have been derived to compensate for the estimation step. The resulting procedures offer a satisfactory solution over a broad range of underlying distributions. However, situations do occur where even such a larger model is inadequate and nothing remains but to consider nonparametric charts. In principle these form ideal solutions, but the problem is that huge sample sizes are required for the estimation step. Otherwise the resulting stochastic error is so large that the chart is very unstable, a disadvantage which seems to outweigh the advantage of avoiding the model error in the parametric case.

In the present chapter we analyze under what conditions nonparametric charts actually become feasible alternatives for their parametric counterparts. In particular, corrected versions are suggested for which a possible change point is reached at sample sizes which are markedly less huge, but still larger than the customary range. These corrections serve to control the behavior during the in-control situation. The price for this protection clearly will be some loss of detection power during the out-of-control situation. A change point comes in view as soon as this loss can be made sufficiently small.

5.1 Introduction

The previous chapters present the parametric approach which offers an interesting solution for problems where the normality assumption in the underlying distribution fails to be fulfilled. The parametric approach, however, is not able to prevent the occurrence of the model errors which come from the distributions other than the normal power family. A nonparametric approach discussed in this chapter, which applies a distribution-free method, provides an alternative solution for such a problem.

In the initial set-up, suppose we control the mean of a production process using a common control chart: an upper and lower limit are prescribed and an out-of-control signal is given as soon as an observation falls outside the interval determined by these two limits. Standard practice is to assume normality of the distribution involved and to estimate its parameters using so-called Phase I observations. The resulting values are plugged in and it is hoped that the estimated chart behaves well. By now, however, we apprehend that often this unfortunately is not the case: both the estimation step and the normality assumption can lead to serious errors.

So far, we have analyzed the resulting picture in a systematic manner. The main source of trouble is the fact that p, the probability of getting an out-of-control signal while the process is actually in-control, typically is chosen to be very small, like $p = 0.001$. Hence, the quantiles involved in the estimation are quite extreme and large relative errors will result, unless the number n of Phase I observations is uncharacteristically large. In Chapter 2 we have analyzed these estimation effects and proposed corrections to get the behavior of the charts under control again. Such remedies work quite well as long as the normality assumption is reasonable. However, quite often there is ample reason to worry about this aspect as well. Again the extremeness of the quantiles involved complicates matters: normality may hold fairly well in the central part of the distribution, but in the tails the relative errors tend to become very large. In Chapters 3 and 4 the second problem has been tackled, using larger parametric models, containing the normal family as a member.

A natural question is why one should stick with such a larger parametric family: the true distribution may still not be in it and the resulting model error may remain unacceptably large. Would it not be better to adopt a fully nonparametric approach? On the one hand, the answer is yes: indeed all problems eventually vanish in a nonparametric chart. A model error is simply not present here, and the stochastic error becomes arbitrarily small as n increases. But on the other hand, for common sample sizes encountered in practice, this is not much help. If we e.g. want to estimate the 0.999-quantile, it is clear that with $n = 100$ we will not get anywhere with a nonparametric estimate. Consequently, it does make sense to look seriously at a larger parametric model, which may successfully bridge the gap between assuming everything (i.e. the distribution simply is normal) or nothing (i.e. the distribution can be anything). Nevertheless, it may happen that after rejection of the normal model, the larger parametric model turns out to be inadequate as well: the actual underlying distribution then is too far away from the model and thus a too large model error results.

In the latter situation we are back once again at the empirical nonparametric approach in which we do not assume any model (except continuity) for the distribution function of the observations and flexibility with respect to n and/or p. How much larger should at least one of these quantities be taken before the estimation effects become more mild and such charts start to behave well? If the increases involved in n or p are indeed tremendous, does it help to look for suitable corrections? In the parametric case such adaptations were seen to be useful in bringing the corresponding charts under control again. But the larger the parametric model, the larger such corrections are for given n and p. Will corrections be feasible as well for the even larger fully nonparametric case? Or will the n and p involved still be unreasonably large before such corrections become sufficiently small in the sense that the out-of-control behavior of the corresponding chart is not affected too much? To answer these and similar questions, we need to study the behavior of nonparametric control charts, and this will precisely be the topic of the present chapter.

We will restrict attention to the obvious choice based on the empirical quantile function, as this will already provide a clear picture of what can be expected in general. Some previous work on closely related charts can be found in Willemain and Runger (1996) and in Ion *et al.* (2000). For a recent overview of nonparametric charts in general, see e.g. Chakraborti *et al.* (2001). Incidentally, as these authors point out, several of the procedures that have been proposed are in fact not truly nonparametric or distribution-free. They are based on a nonparametric estimator, like the Hodges-Lehmann estimator, rather than on $\overline{X}$, but the actual in-control run length distribution involved does depend on the underlying distribution of the observations.

In studying the behavior of the charts, the first thing to note is that, due to the estimation, the usual performance characteristics of the chart become random. Hence p itself is replaced by a stochastic counterpart P_n, and the average run length (ARL) $1/p$ likewise by $1/P_n$. For each characteristic the relative error (e.g. $P_n/p - 1$) can be studied with respect to aspects such as expectation, standard deviation and exceedance probability, each of which in its own way helps to characterize the behavior of the estimated charts. In Section 5.2 we introduce the nonparametric chart, after which the next section is devoted to its expectation and bias. Adaptations are suggested to remove the latter, thus improving the performance of the charts in the long run. Some attention is devoted to the out-of-control behavior as well. However, the resulting procedure still is far from satisfactory, which is mainly due to the large variability involved, as is demonstrated in Section 5.4.

This leads in Section 5.5 to a more rigorous approach toward controlling the performance of the chart, using exceedance probabilities. The idea is as follows: in each given application of the control chart procedure, the Phase I observations produce an estimated chart, and thus a value of P_n, which determines the subsequent behavior of the chart in that application. Instead of merely looking at the average performance of the procedure (i.e. over a long series of different applications), we can also consider the probability with which values of P_n occur that are too unpleasant in a given sense. For example, these are values of P_n which are likely to produce a low run length, even if the underlying process is in control. Due to the variability of the nonparametric chart, such probabilities of unpleasant values will typically be unacceptably large for common n and p.

For this situation as well, corrections can be derived to bring the corresponding exceedance probabilities under control. Again, the impact of such adaptations on the out-of-control behavior is investigated: it is one thing to largely avoid premature stopping during the in-control period, but this should not be achieved by stopping in general much later once the process has gone out of control. Not surprisingly, the effect here is more pronounced than in Section 5.3, as the changes involved are of a larger magnitude. Hence, a considerable price has to be paid in terms of out-of-control performance, in order to guarantee a satisfactory behavior during the in-control phase, unless, again, n or p are sufficiently large. Anyhow, the results obtained allow to strike a proper balance between the following three negative aspects: (i) unacceptable behavior during the in-control phase, (ii) reduced detection power during the out-of-control phase and (iii) using higher n or p than originally intended. Typically, one will figure out how much needs to be sacrificed with respect to (iii) in order to arrive at acceptable results with respect to (i) and (ii). Some examples nicely illustrate what can be expected. For convenience, a short point-by-point description of the algorithm involved is presented in the final section as a conclusion remark for this chapter.

Summarizing, we analyze when and how nonparametric charts can be used in a sensible way in situations between two obvious extremes. The first extreme is that n is astronomical (like $n = 184830$ on page 34 in Willemain and Runger (1996)), which makes any problem disappear, any correction superfluous and any price to be paid negligible. The second extreme is that n is in the customary range of a few hundreds, while p is one (or a few) tenth(s) of a percent, and we are simply out on a limb. The situation in between is much less clear-cut and the clarity which is obtained here helps to make a balanced choice between sticking with a (large) parametric model or switching to a nonparametric approach. The former suffers from the disadvantage of a non-vanishing model error, whereas the latter in view of its higher variability requires larger, and thus more costly, corrections to keep the in-control behavior under control. A next step will be to use the data in deciding whether to use a parametric control chart or a nonparametric one. This latter topic is treated in Chapter 6.

5.2 The nonparametric chart

Using the initial set-up, let X be a r.v. with a continuous d.f. F. For a given, very small p (typically $0.001 \le p \le 0.01$), we need an upper control limit UCL

such that $P(X > UCL) = p$. Hence, the control limit can be written as $UCL = F^{-1}(1-p) = \overline{F}^{-1}(p)$. Usually, F is unknown and some estimation has to be invoked, using a sample $X_1, ..., X_n$ from F (the Phase I observations) as a starting point.

Chapter 2 discusses the normal case, where F is the normal d.f. with expectation μ and variance σ^2 yielding $UCL = \mu + \sigma u_p$, and the corresponding estimated upper control limit given in (2.1). Corrected versions of the control limit are obtained by replacing u_p by $u_p + c$ as given in (2.27). In this chapter, the derivation of the correction terms c was carried out by using the bias criterion and the exceedance probability criterion. Under the first criterion, c is given by (2.33) or by (2.34) if $g(p) = p$ while, under the latter criterion, c is given by (2.47) or by (2.39).

Subsequently, in Chapters 3 and 4 we discuss the parametric case, where the supposed model is based on a larger parametric family. In this case, F has a normal power d.f. (see Subsection 3.3.3) with the upper control limit $\widehat{UCL} = \widehat{\mu} + \overline{K}_{\widehat{\gamma}}^{-1}(p)\sigma^*$ and the corresponding estimated UCL given in (3.8). The parameter γ can be estimated using (3.11) or (3.14). However, our recommended estimator for γ is given in (3.15). Corrections for the control limit are again obtained by adding a suitable correction term c to the standardized quantile $\overline{K}_{\widehat{\gamma}}^{-1}(p)$ leading to the corrected control limit given in (3.17). As in the normal case, we derived the correction terms c in the parametric case using both the bias criterion and the exceedance probability criterion. The corrected control limits, based on the bias criterion, are given in (3.33) and in (3.38), respectively if we use $\widehat{\gamma}_1$ and $\widehat{\gamma}_2$. However, since these control limits are not so easy to calculate, we recommend a simpler control limit which is given in (3.39). The correction terms based on the exceedance probability criterion are derived and given in (4.6) and (4.7), see also (4.9) and (4.10).

In a sense, the nonparametric approach is more transparent than these parametric attempts. Just let $F_n(x) = n^{-1}\#\{X_i \leq x\}$ be the empirical d.f. and F_n^{-1} the corresponding quantile function, i.e. $F_n^{-1}(t) = \inf\{x | F_n(x) \geq t\}$. Then it follows that $F_n^{-1}(t)$ equals $X_{(i)}$ for $(i-1)/n < t \leq i/n$, where $X_{(1)} < ... < X_{(n)}$ are the order statistics corresponding to $X_1, ..., X_n$. Hence, letting $\overline{F}_n^{-1}(t) = F_n^{-1}(1-t)$, we get

$$\widehat{UCL} = \overline{F}_n^{-1}(p) = X_{(n-r)}, \tag{5.1}$$

where $r = [np]$. Note that for ordinary p and n, like $p = 0.001$ and $n = 100$, we will have $r = 0$, and thus $\widehat{UCL} = X_{(n)}$. See Willemain and Runger (1996) and Ion *et al.* (2000) for these or closely related charts.

Remark 5.2.1 For the two-sided case, (5.1) translates into

$$\widehat{CL_L} = \overline{F}_n^{-1}(1-p/2) = X_{(n-[n(1-p/2)])}$$

for the lower control limit and

$$\widehat{CL_U} = \overline{F}_n^{-1}(p/2) = X_{(n-[np/2])}$$

for the upper control limit. □

Just as in the parametric case, in this nonparametric case we need to analyze and possibly correct the behavior of the estimated chart based on $\widehat{UCL}$ from (5.1). Let X_{n+1} be another r.v. from F, then we observe that the role of p in the case of known F will now be played by

$$P_n = P(X_{n+1} > \widehat{UCL}|(X_1, ..., X_n)) = \overline{F}(X_{(n-r)}). \tag{5.2}$$

In what follows we will also write $P_n = P(X_{n+1} > \widehat{UCL})$, without explicitly stating that we work conditionally on $(X_1, ..., X_n)$. Let $U_{(1)} < ... < U_{(n)}$ denote order statistics for a sample of size n from the uniform d.f. on (0,1), then it is immediate from (5.2) that $P_n \cong U_{(r+1)}$, with '$\cong$' denoting 'distributed as'.

To judge the performance of the estimated chart, we propose to consider the relative error

$$W = \frac{g(P_n)}{g(p)} - 1, \tag{5.3}$$

for some suitable function g. Obvious choices are the functions g given in (2.3). The identity function $g(p) = p$ leads to $W_1 = P_n/p - 1$, while $g(p) = 1/p$, producing $W_2 = p/P_n - 1$, is also quite interesting, as it corresponds to the average of the run length RL, given by $ARL = 1/p$. A third choice follows from $g(p) = 1-(1-p)^k = P(RL \leq k)$, where typically $k = [\delta/p]$ for some small δ like 0.1 or 0.2. This possibility we will consider in Section 5.5.

As announced in the introduction, quantities like EW, σ_W and $P(W > \varepsilon)$, can now be studied for some, usually small, ε. For what values of p and n (or r) are these quantities sufficiently close to 0? And, for given n and p, what kind of modification is needed to make each of these quantities behave as desired after all? The following general observation can already be made: the fact that $P_n \cong U_{(r+1)}$ strongly suggests that it will suffice to watch the product np, rather than both n and p separately. For, as long as $r = [np] = 0$, no satisfactory results seem feasible, whereas for really large values of np no problems remain. The point is to make explicit what happens in between. Will e.g. r equal 3 or 4 do? Or do we need an r equal to even 10 and larger? In the next section we shall begin by studying the behavior of EW.

5.3 Expectation and bias

Let $\psi = np - r$ (and thus $0 \leq \psi < 1$), then it follows from $P_n \cong U_{(r+1)}$ that

$$EP_n = \frac{r+1}{n+1} > \frac{r+\psi}{n} = p, \tag{5.4}$$

for $0 \leq \psi < (n-r)/(n+1)$. Hence P_n has a positive bias, unless ψ is very close to 1. It is immediate that $W_1 = P_n/p - 1$ satisfies

$$EW_1 = \frac{\frac{n-r}{n+1} - \psi}{r+\psi} = \frac{\frac{1-\psi}{p} - 1}{n+1}. \tag{5.5}$$

Note the discrete character of this relative error: if we let n increase for some given p, the expression in (5.5) gradually decreases, jumping upwards whenever np becomes integer again (i.e. when $\psi = 0$) to $(n-r)/\{(n+1)r\} = (1-p)/\{(n+1)p\}$. This maximal value behaves like $(1-p)/r \approx 1/r$ (supposing of course that $r > 0$).

Hence, it is straightforward to analyze for which combinations of n and p the chart will start to behave well with respect to (5.5). To begin with, as $n \to \infty$, so does $r = [np]$ for given p, and the bias eventually becomes negligible. Thus, for $p = 0.001$, the relative change is at most 1% as soon as $n \geq 10^5$ (e.g. $n = 184830$), but on the other hand, we require $n \geq 1000$ before r is even positive. The situation in between these two extremes is also still quite transparent. For example, for $n \geq 5000$ we have $r \geq 5$, and then the relative error will be at most 20%, which might for example be reasonable for practical purposes. This is still a very large sample size; an alternative of course is to raise the value of p to e.g. $p = 0.01$. The n required to get the same result then obviously reduces to $n = 500$, which is still not really small. Consequently, even with respect to a rather mild criterion, only concerning the bias involved, reasonable behavior seems to require considerably higher values of n and p than the customary ones.

For the ARL similar conclusions can be drawn. As $E(1/U_{(r+1)}) = n/r$ (and thus $E(1/U_{(1)}) = \infty$, cf. Willemain and Runger (1996), where it is also noted that r has to be at least 1 here), we observe that typically

$$E\left(\frac{1}{P_n}\right) = \frac{n}{r} \geq \frac{n}{r+\psi} = \frac{1}{p}. \tag{5.6}$$

The phenomenon of both P_n and $1/P_n$ being positively biased is already well-known from the parametric case, see Remark 2.2.1 or Quesenberry (1993), p. 245. Clearly, for $W_2 = p/P_n - 1$ we have $EW_2 = \psi/r < r^{-1}$, and precisely the same comments as for W_1 can be given.

Since the behavior for small r indeed turns out to be unsatisfactory, the next question is what types of corrections can be proposed to improve matters. One remedy is to invoke the aforementioned flexibility required with respect to n and/or p in a very simple way: just alter p such that e.g. $p = (r+1)/(n+1)$, for some $r \geq 0$, to arrive at $EP_n = p$. Or, alternatively, set $p = r/n$ for some $r \geq 1$ to obtain $E(1/P_n) = 1/p$. But do realize that this may seem to achieve more than it actually does. To give a simple example, first let $r = 1$ and choose $p = 2/(n+1)$. Then $EP_n = p$, but $E(1/P_n) = n = \{2n/(n+1)\}(1/p)$, i.e. $EW_2 = (n-1)/(n+1) \approx 1$. Hence for $p = 0.001$ and e.g. $n = 1999$, the expected relative error in the ARL is still about 1. If for the present value of $p = 0.001$ we increase n by just 1 to 2000, the picture changes: then r becomes 2, suddenly EW_2 drops to 0, but EW_1 becomes almost 1/2. Hence the chart is very unstable due to its discrete character and, moreover, the errors involved are also very large. Once again, the phenomenon is due to the fact that n is large but r is not. If r is large as well, the problem neatly dissolves: for $p = (r+1)/(n+1)$, we have $EW_1 = 0$ but $EW_2 = (n-r)/\{r(n+1)\}$, while for $p = r/n$ it is precisely the other way around.

For parametric charts the picture is quite different. Obviously, the complications due to discreteness are not present there. In addition, relative errors of a similar magnitude as in the example above do occur as well, but for much smaller sample sizes. First consider the uncorrected normal chart based simply on $\widehat{\mu} + \check{\sigma} u_p$. In this case we e.g. observe from Table 2.1 that for $p = 0.001$ we would obtain $EW_2 = 1$ for $n \approx 65$ and $EW_1 = 1/2$ for $n \approx 70$. Hence for such sample sizes, corrections are called for here as well. For this same p and $n = 500$, the relative errors in the uncorrected chart have already gone down to about 7% for EW_1 and about 8% for EW_2, and the use of corrections has become superfluous. This conforms with the common recommendation to take at least 300 observations for estimating the parameters under normality.

If we go to a larger parametric model and replace u_p by some $\overline{K}_{\widehat{\gamma}}^{-1}(p)$, we can readily obtain an example from Sections 3.4 and 3.5. For the particular choice considered there, which is demonstrated to work well over a broad range of underlying distributions, we observe from Table 3.2, that if normality happens to be true after all, the uncorrected chart leads to an EW_1 of 117%, 43% and 21% for $n = 100$, 250 and 500, respectively. Hence, corrections are certainly needed. From the same table we can see that the uncorrected normal chart would have produced 36%, 13% and 7% (the latter case we already encountered above) for these same sample sizes. Clearly, this is quite a bit better than what is achieved with the more general parametric chart. But the price for this gain is also immediately evident from the table mentioned: if we move towards non-normal distributions, EW_1 varies wildly for the normal chart, for example, $EW_1 > 11$ occurs for the cases considered at $n = 100$ and as this is due to the model error rather than the stochastic error, there is not much improvement for larger n. On the other hand, its parametric counterpart keeps the damage much more limited (e.g., EW_1 is at most 2.5 at $n = 100$). Hence, the premium paid for the latter chart is rather large, but nevertheless well-spent in view of the erratic behavior of the model error for the normal chart outside the normal model.

Note that this discussion of parametric charts helps to better understand what happens in the nonparametric case. In that situation the model error is eliminated by estimating the distribution in a nonparametric way, rather than by just adding a single third parameter to the normal model. When, as we saw, even this latter extension already leads to considerably higher values of the relative errors involved, it becomes less surprising that in the nonparametric case the growth is so tremendous that excessive values of n and p are needed before this approach becomes of potential use. Of course, it should be remarked that we have concentrated on the maximal errors that occur. In the nonparametric case, a minimal error equal to zero can be achieved by just taking a lucky combination of n and p, whereas in the parametric case corrections are always needed until n and p are sufficiently large.

Next, we shall consider a second way of introducing corrections, which will turn out to be of use also in Section 5.5. In the above the bias was removed by adapting p to the n at hand. If we want to stick to a given p, unbiasedness can be achieved in a relatively simple way by randomizing between consecutive order statistics. Let V be independent of $(X_1, \ldots, X_n, X_{n+1}, \ldots)$, with $P(V = 1) = 1 - P(V = 0) = \lambda$ and replace $\widehat{UCL}$ from (5.1) by

$$\begin{aligned} \widehat{UCL}_+ &= (1-V)X_{(n+1-r)} + VX_{(n-r)}, \quad \text{or} \\ \widehat{UCL}_- &= (1-V)X_{(n-1-r)} + VX_{(n-r)}, \end{aligned} \tag{5.7}$$

where we use the convention $X_{(n+1)} = \infty$. Then for $\widehat{UCL}_+$ and $r = 0$, a signal only results with probability λ in case $X_{n+1} > X_{(n)}$.

Remark 5.3.1 Again, for the two-sided case, with $P(V = 1) = 1 - P(V = 0) = \lambda$, (5.7) translates into

$$\begin{aligned} \widehat{CL}_{L+} &= (1-V)X_{(n-[n(1-p/2)]-1)} + VX_{(n-[n(1-p/2)])}, \quad \textit{or} \\ \widehat{CL}_{L-} &= (1-V)X_{(n-[n(1-p/2)]+1)} + VX_{(n-[n(1-p/2)])}, \end{aligned}$$

for the lower control limit and

$$\widehat{CL}_{U+} \;=\; (1-V)X_{(n-[np/2]+1)} + VX_{(n-[np/2])}, \quad or$$

$$\widehat{CL}_{U-} \;=\; (1-V)X_{(n-[np/2]-1)} + VX_{(n-[np/2]))},$$

for the upper control limit. For the lower control limit we use the convention $X_{(0)} = -\infty$, hence for $\widehat{CL}_{L+}$ and $np/2 \leq 1$, a signal only results with probability λ in case $X_{n+1} < X_{(1)}$. Similarly, for the upper control limit we use the convention $X_{(n+1)} = +\infty$, hence for $\widehat{CL}_{U+}$ and $np/2 < 1$, a signal only results with probability λ in case $X_{n+1} > X_{(n)}$. □

Lemma 5.3.1

(i) $EP_n = p$ results for $\psi \leq 1-p$, by taking $\widehat{UCL}_+$ with

$$\lambda = \frac{r+(n+1)\psi}{n} = p + \psi \tag{5.8}$$

and for $\psi > 1-p$ by taking $\widehat{UCL}_-$ with $\lambda = 2-p-\psi$.

(ii) $E(1/P_n) = 1/p$ results for $r \geq 1$, by taking $\widehat{UCL}_-$ with

$$\lambda = \frac{r}{(r+\psi)}(1-\psi). \tag{5.9}$$

Proof. As $P_n = P(X_{n+1} > \widehat{UCL}_+) = (1-V)P(X_{n+1} > X_{(n+1-r)}) + VP(X_{n+1} > X_{(n-r)}) \cong (1-V)U_{(r)} + VU_{(r+1)}$ (we now work conditionally on $(X_1, \ldots, X_n, S)$, cf. the remark following (5.2)), it is immediate that $EP_n = (r+\lambda)/(n+1)$. This equals $p = (r+\psi)/n$ for λ as in (5.8), as long as $\psi \leq 1-p$. For the remaining ψ, the companion result based on $\widehat{UCL}_-$ follows similarly. As concerns $1/P_n$, we observe that for $\widehat{UCL}_-$ we have $1/P_n \cong 1/\{(1-V)U_{(r+2)} + VU_{(r+1)}\} = (1-V)/U_{(r+2)} + V/U_{(r+1)}$, and thus for $r \geq 1$ that $E(1/P_n) = n\{(1-\lambda)/(r+1) + \lambda/r\}$, which equals $1/p = n/(r+\psi)$ for λ as in (5.9), as long as $r \geq 1$. ■

Hence, once more the extremes are clear: for $r = 0$ the corrected chart becomes outright awkward, while for large r it is already intuitively evident that taking a random mixture of two adjacent order statistics, rather than just one of them, will make little difference. In order to further illustrate the situation between these two ends of the scale, we shall now address the interesting question of what happens in the out-of-control situation, with particular attention for the effect of the corrections studied above. Consequently, X_{n+1} now comes from a shifted d.f. $F(x-\Delta)$, where Δ typically is such that $p_1 = \overline{F}(\overline{F}^{-1}(p) - \Delta)$ may still be small, but not extremely so, like p.

Lemma 5.3.2 *Replacement of $\widehat{UCL}$ from (5.1) by $\widehat{UCL}_+$ from (5.7), for some λ, results in a relative change in EP_n approximately equal to $-(1-\lambda)w/\tilde{p}_1$, where*

$$w = \frac{f\left(\overline{F}^{-1}(q) - \Delta\right)}{(n+1)f\left(\overline{F}^{-1}(q)\right)}, \tag{5.10}$$

in which $q = (r+1)/(n+1)$ and $\tilde{p}_1 = \overline{F}(\overline{F}^{-1}(q) - \Delta)$. If we take $\widehat{UCL}_-$ rather than $\widehat{UCL}_+$, only the sign of the relative change is reversed.

Proof. If X_{n+1} has d.f. $F(x-\Delta)$, it follows from (5.2) that P_n corresponding to $\widehat{UCL}$ equals $P_n \cong \overline{F}(\overline{F}^{-1}(U_{(r+1)}) - \Delta)$, and thus EP_n can be approximated by $\tilde{p}_1$. As this $\tilde{p}_1$, just like p_1, is not extremely small, the relative error involved will be reasonably small compared to the relative error during the in-control situation. Now the change in EP_n, caused by replacing $X_{(n-r)}$ by $X_{(n+1-r)}$, will approximately equal $-w$ with w as in (5.10). In (5.7), such a replacement occurs with probability $(1-\lambda)$, and thus the effect will have value $-(1-\lambda)w$, leading to the relative change stated above. ■

The relative change in Lemma 5.3.2 will become reasonably small somewhat sooner than in the in-control situation, where the relative change was seen to behave like r^{-1}. On the other hand, here as well, it is easy to see that, when changing X_{n-r} into X_{n+1-r}, things go wrong as r gets close to or even equal to zero, as in that situation $f(\overline{F}^{-1}(q))$ becomes very small and (5.10) no longer provides a reasonable approximation. As $X_{(n+1-r)} = \infty$ for $r = 0$, we simply have that $\overline{F}(X_{(n+1-r)} - \Delta) = 0$ and thus EP_n is reduced, both in the out-of-control and in the in-control situations, by the factor $\lambda = (n+1)p$.

To be a bit more explicit, as well as to provide some illustration, we shall briefly compare these results to the situation in the parametric case.

Example 5.3.1

Keeping things as simple as possible, just let $F = \Phi$ and use UCL, as given in (2.1). With X_{n+1} coming from $\Phi(x-\Delta)$, we then obtain that $EP_n \approx p_1 = \overline{\Phi}(u_p - \Delta)$, where again the relative error committed is reasonably small because this p_1, just like the one above in the nonparametric case, is supposed not to be (very) small. Hence for this specific choice of F, think of values of Δ between 1 and 3. If we now replace u_p by $u_p + c$ (cf. Section 5.2), the change in EP_n will to first order equal $-c\varphi(u_p - \Delta)$ (cf. e.g. (2.70)). Consequently, the relative change caused by using c will approximately be $-c\kappa(u_p - \Delta)$ (cf. Remark 2.4.1 and (2.72)).

Finally, the correction needed to get $EP_n = p$ during the in-control situation for the normal case equals $c = u_p(u_p^2+2)/(4n)$ (see (2.33), which, in combination with

(2.72), shows that the size of the relative error during the out-of-control situation, due to this correction approximately equals

$$u_p(u_p^2+2)\frac{1+u_p-\Delta}{(5n)}. \tag{5.11}$$

For $p = 0.001$ this boils down to $7.15(4.09 - \Delta)/n$, which clearly becomes reasonably small quite soon. Take for example $\Delta = 1.81$, which leads to a value $\overline{\Phi}(3.09 - 1.81) = 0.10$ for p_1. For this case the relative change is about $16/n$. □

This parametric example is of some interest in its own right, but of course it is primarily meant to provide an explicit comparison with the nonparametric situation, which we shall consider next.

Example 5.3.2

If we again let $F = \Phi$, the expression from (5.10) translates into $w = \varphi(u_q - \Delta)/\{(n+1)\varphi(u_q)\}$. As moreover $\kappa(x) = \varphi(x)/\overline{\Phi}(x)$ implies that $\kappa(u_q) = (n+1)\varphi(u_q)/(r+1)$, it readily follows that the size of the nonparametric relative change simply reduces to approximately $(1-\lambda)w/\tilde{p}_1 = (1-\lambda)\kappa(u_q-\Delta)/\{(r+1)\kappa(u_q)\}$. The final step is to approximate this expression by

$$(1-\lambda)\frac{1+u_q-\Delta}{(r+1)(1+u_q)}. \tag{5.12}$$

The result in (5.12) is very simple and transparent. The general observation above, according to which the relative error "will become reasonably small somewhat sooner than in the in-control situation, where the relative errors were seen to behave like r^{-1}", can now be made explicit for this special case: $(r+1)^{-1}$ is reduced by the factor $\{1+u_q-\Delta\}/(1+u_q)$, which will vary between 1/4 and 1/2 for customary values of n, p and Δ. As concerns the factor $(1-\lambda)$, it is immediate from (5.8) that this will keep returning to (almost) 1 as n increases for given p. To give a numerical example, for the standard $p = 0.001$, let $n = 5000$ and we obtain $r = 5$, $\psi = 0$, $1-\lambda = 1-p \approx 1, u_{6/5001} = 3.04$, which together produces $(1-\Delta/4.04)/6$ as the outcome of (5.12). For $\Delta = 1.81$, which is the value used in the parametric example above, this boils down to 9.2%, which sounds acceptable. But note that this requires n to be as large as 5000; in comparison, the parametric value $16/n$ produces such an outcome already for $n \approx 175$. Indeed, comparing (5.11) and (5.12) shows that both relative errors have a factor $1+u_s-\Delta$ in common (with s either equal to p or to q), but apart from that they differ markedly. The parameter c is small and behaves like n^{-1}, whereas the effect of shifting to the next order statistic introduces a factor $1/\{(n+1)f(\overline{F}^{-1}(q))\}$, which behaves like r^{-1}. □

It is of course quite interesting to study what happens to the ARL as well. But to avoid repetition, we once more point out that, although the analysis above has been given for P_n, it is easily verified that the size of the relative error is (again approximately) the same for $1/P_n$. Just note that $|g'(p)/g(p)| = 1/p$ for both $g(p) = p$ and $g(p) = 1/p$ (cf. (5.3)). Hence, again completely similar conclusions hold for this case.

As indicated earlier, the practical implementation of this nonparametric chart is still questionable for $r = 0$. For example, when $g(p) = p$ then $r = 0$ corresponds to $np < 1$ (, or $np/2 < 1$ for the two-sided limit), which implies that with positive probability we will never get an out-of-control signal. Therefore, we modify the nonparametric control somewhat in case $r = 0$. We replace the definition $X_{(n+1)} = \infty$ by $X_{(n+1)} = X_{(n)} + S$. For the in-control situation $X_{(n)}$ is already rather far in the upper-tail of the distribution of X_{n+1} and adding S gives for most distributions a really large value. However, a substantial shift in the out-of-control case can now be detected with higher probability than $P(V = 1)$ Moreover, it is now no longer decided to keep the process running "forever" when $V = 0$.

With the new setup the modified nonparametric chart is defined as follows:

$$\widehat{UCL} = VX_{(n-r)} + (1-V)X_{(n-r+1)} \text{ for } r \geq 1 \tag{5.13}$$

$$X_{(n)} + (1-V)S \text{ for } r = 0,$$

where, as before, $r = [np]$ and $P(V = 1) = 1 - P(V = 0) = np - r$.

Remark 5.3.2 Using a similar argument as that preceding (5.13), for the two-sided case we replace the definition $X_{(n+1)} = \infty$ by $X_{(n+1)} = X_{(n)} + S$ and $X_{(0)} = -\infty$ by $X_{(0)} = X_{(1)} - S$. For the in-control situation $X_{(1)}$ is already rather far in the lower-tail of the distribution of X_{n+1} and subtracting S gives for most distributions a really small value for the two sided-case. Hence, the modified nonparametric chart for the two-sided limit is defined as follows:

$$\begin{aligned} \widehat{CL}_L &= (1-V)X_{(r)} + VX_{(r+1)} && \text{for } r \geq 1 \\ & \quad X_{(1)} - (1-V)S && \text{for } r = 0, \end{aligned}$$

for the lower control limit and

$$\begin{aligned} \widehat{CL}_U &= VX_{(n-r)} + (1-V)X_{(n-r+1)} && \text{for } r \geq 1 \\ & \quad X_{(n)} + (1-V)S && \text{for } r = 0, \end{aligned}$$

for the upper control limit. In this case, $r = [np/2]$ and $P(V = 1) = 1 - P(V = 0) = np/2 - r$. □

5.4 Variation

In the previous section we have seen that estimation unfortunately introduces bias which is definitely not negligible for the usual small values of r. Adaptations of the chart can easily be devised to remove such bias, but this obviously does not solve the problem to complete satisfaction. Removing bias from P_n means increasing bias in $1/P_n$ and vice versa. Moreover, the effect in the out-of-control situation is also considerable as long as r is small. The source of these troubles is the fact that the estimated nonparametric chart seems to contain too much variation to be very useful for small r. Apparently the balance is lost: the model error from the parametric models has been eliminated, but the resulting stochastic error more than spoils the benefit.

To make this feeling explicit, we shall now consider the standard deviation σ_W of the relative error W from (5.3) for the cases $W_1 = P_n/p - 1$ and $W_2 = p/P_n - 1$. As $\text{var}(U_{(r+1)}) = (r+1)(n-r)/\{(n+2)(n+1)^2\}$ while $\text{var}(1/U_{(r+1)}) = n(n-r)/\{(r-1)r^2\}$, it follows readily that

$$\sigma^2_{W_1} = \frac{n^2(n-r)(r+1)}{(n+1)^2(n+2)(r+\psi)^2}, \qquad \sigma^2_{W_2} = \frac{(n-r)(r+\psi)^2}{n(r-1)r^2}. \tag{5.14}$$

Note that for W_2 we even need $r \geq 2$ now; cf. Willemain and Runger (1996). From (5.14) a similar pattern is observed as in the previous section, where making $EW_1 = 0$ led to $EW_2 = (n-r)/\{(n+1)r\}$ and vice versa. The variances behave like r^{-1} or $(pn)^{-1}$, which is fine for n large and p (not too) small, or n extremely large and p very small, but not otherwise.

Next, we consider parametric charts and show that the situation here is somewhat different. Suppose that, instead of $\widehat{UCL} = X_{(n-r)}$ from (5.1), we use some parametric $\widehat{UCL}$. Then $P_n = \overline{F}(\widehat{UCL})$ in this case leads to

$$\sigma^2_{W_1} \approx \tau^2 \frac{f^2(\overline{F}^{-1}(p))}{p^2}, \tag{5.15}$$

with $\tau^2 = \text{var}\ (\widehat{UCL})$ of order n^{-1}. Now typically $f(\overline{F}^{-1}(p))/p$ may grow somewhat as p becomes small, but not as fast as $p^{-1/2}$. Hence $\sigma^2_{W_1}$ from (5.15) will behave better than $(pn)^{-1}$. As $|g'(p)/g(p)|$ is the same for both $g(p) = p$ and $g(p) = 1/p$ the same holds for $\sigma^2_{W_2}$.

To be more explicit we again give a numerical example.

Example 5.4.1

Once more consider the normal case (cf. (2.1)). From Subsection 2.4.2 we have that $\tau^2 \approx (u_p^2 + 2)/(2n)$ while the fact that $\overline{\Phi}(x) \approx \varphi(x)/x$ for x large shows that

$\varphi(u_p)/p \approx u_p$, thus producing $\sigma^2_{W_1} \approx u_p^2(u_p^2+2)/(2n)$. If we let $p = 0.00135$, we get $u_p = 3$ and thus $\sigma_{W_1} \approx 7n^{-1/2}$, while in the nonparametric case $\sigma_{W_1} \approx (pn)^{-1/2} \approx 27n^{-1/2}$. Indeed these results nicely agree with the simulations from Ion *et al.* (2000) (see Tables 4 and 8), which also exhibit this difference in behavior between the two types of chart. Yet another comparison is obtained by Table 1 from Willemain and Runger (1996), where the ARL case is illustrated and hence σ_{W2} from (5.14) is used. Using $p = 0.00270$, and thus $1/p = 370$, the standard deviation of the ARL is brought down all the way to 16.6, requiring $n = 184830$ and consequently $r = 499$. For $n = 1482$, however, r is merely 4 and the standard deviation equals 214. □

5.5 Exceedance probabilities

To summarize the situation up to now, we know how to remove the bias in P_n (or $1/P_n$), but the variation around the obtained correct expectation p (or $1/p$) remains quite large. Hence, in specific applications of the procedure we must reckon with the occurrence of considerably "wrong" values. If we want to control the frequency with which such accidents occur, exceedance probabilities are the instrument to use.

In fact, for the relative error W from (5.3) we shall now move from EW and σ_W (for $\varepsilon > 0$) to $P(W > \varepsilon)$ for increasing g and $P(W < -\varepsilon)$ for decreasing g. For $W_1 = P_n/p - 1$ this simply produces $P(P_n > p(1+\varepsilon))$, while for $W_2 = p/P_n - 1$ we obtain $P(P_n > p/(1-\varepsilon))$. For ε small, this is virtually the same, as $|g'(p)/g(p)| = 1/p$ in either case. For increasing g we have that $P(W > \epsilon)$ becomes $P(P_n > p(1+\tilde{\varepsilon}))$, with $\tilde{\varepsilon}$ given in (5.16).

Lemma 5.5.1

(i) *For $g(p) = p$ we have $P(W > \epsilon) = P(P_n > p(1+\tilde{\epsilon}))$ with $\tilde{\epsilon} = \epsilon$,*
(ii) *for $g(p) = 1/p$ we have $P(W < -\epsilon) = P(P_n > p(1+\tilde{\epsilon}))$ with $\tilde{\epsilon} = \epsilon/(1-\epsilon)$,*
(iii) *for $g(p) = 1-(1-p)^k$, where $k = [\delta/p]$, we have $P(W > \epsilon) \approx P(P_n > p(1+\tilde{\epsilon}))$ with $\tilde{\epsilon} = (1+\delta/2)\epsilon$, for ϵ and δ small.*

Proof. (i) is trivial, as $W = W_1 = P_n/p - 1$; (ii) is also simple, as for $W_2 = p/P_n - 1$ we obtain $P(W_2 < -\epsilon) = P(P_n > p/(1-\epsilon))$. For (iii) we use that for increasing g

$$\tilde{\varepsilon} = \frac{g^{-1}(g(p)(1+\varepsilon)) - p}{p} \approx \varepsilon \frac{g(p)}{pg'(p)}. \tag{5.16}$$

and thus $\tilde{\varepsilon} = \varepsilon(1-p)\{(1-p)^k - 1\}/(kp) \approx \varepsilon(e^{kp}-1)/(kp) \approx (1+\delta/2)\varepsilon$, for small ϵ and δ (typical values will indeed be 0.1 or 0.2). ■

Again, we start with the standard nonparametric chart based on $\widehat{UCL} = X_{(n-r)}$ (cf. (5.1)). Let $B(n, \tilde{p}, k)$ denote the cumulative binomial probability $P(Y \leq k)$, with $Y \sim bin(n, \tilde{p})$. Moreover, let $Po(\tilde{\lambda}, k)$ stand for the cumulative Poisson probability $P(\widetilde{Y} \leq k)$, where $\widetilde{Y}$ has a Poisson distribution with parameter $\tilde{\lambda}$.

Lemma 5.5.2 *For $\widehat{UCL}$ from (5.1) we have*

$$P(P_n > p(1+\varepsilon)) = B(n, p(1+\varepsilon), r) \approx Po(np(1+\varepsilon), r), \qquad (5.17)$$

Proof. For $\widehat{UCL} = X_{(n-r)}$ we again have that $P_n \cong U_{(n+1)}$. Then it is well-known that $P(U_{(r+1)} > \tilde{p}) = P(Y < r)$. Hence, the left-hand side of (5.17) equals $B(n, p(1+\epsilon), r)$. Since n is typically large, while $r = [np]$ is not, the Poisson approximation in (5.17) will work extremely well. ■

To begin with, we check that the behavior at the two extremes is again as expected. Indeed, as $n \to \infty$ for fixed p, all is well: $P(U_{(r+1)} > p(1+\varepsilon)) \to 0$. According to Hoeffding (1963), we in fact have that $B(n, p(1+\varepsilon), r) \leq \exp\{-2(np(1+\varepsilon) - r)^2/n\} \leq \exp(-2np^2\varepsilon^2)$ and thus the convergence is even exponentially fast. But clearly, in most practical applications $np^2\varepsilon^2$ will be very small, rendering the bound correct but useless. In fact, not only the bound will become large, but the same will hold for the actual probability involved as well. To illustrate the opposite end of the scale, let us now assume that $n < p^{-1}$ and thus $r = 0$. Then the exceedance probability in (5.17) equals $\{1 - p(1+\varepsilon)\}^n \approx \exp\{-np(1+\varepsilon)\}$, which attains some prescribed value α for $n = (\log \alpha^{-1})/(p(1+\varepsilon))$.

Example 5.5.1

Letting $p = 0.001$ again and choosing $\varepsilon = 9$, so that P_n is off by a factor at least 10, we obtain $n = 100 \log \alpha^{-1}$, which e.g. produces $n = 300$ for $\alpha = 0.05$. In the normal case (see p. 33), by comparison we have $n \approx 9.90u_\alpha^2$, leading to $n = 27$ for $\alpha = 0.05$. Hence really huge errors occur in the parametric case as well, but only at relatively small sample sizes. □

The case between the two extremes is still quite easy to analyze. It is less trivial than in the expectation case from Section 5.3, where the maximal bias was seen to behave simply as r^{-1}. But the Poisson probability from (5.17) also readily shows what to expect in a given configuration.

Example 5.5.2

Let, as before, $p = 0.001$ and $n = 5000$, and thus $r = 5$. Then we deal with $Po(5(1+\varepsilon), 5)$, which equals 0.45, 0.19 and 0.067 for ε equal 0.2, 0.6 and 1.0, respectively. Even if we raise n to 10000, the results still do not look very satisfactory: we see that $Po(10(1+\varepsilon), 10)$ equals 0.35, 0.18 and 0.077 for ε equal to 0.2, 0.4 and 0.6, respectively. □

As expected, with respect to the present criterion, the behavior of the uncorrected chart remains unsatisfactory for even larger n than in Section 5.3.

Consequently, there is ample reason to look for suitable corrections. Fortunately, from Section 5.3 it is immediately clear how to modify $\widehat{UCL} = X_{(n-r)}$ such that the associated exceedance probability satisfies some desired upper bound α. Instead of (5.7) we will now use

$$\widehat{UCL}_k = (1-V)X_{(n+k+1-r)} + VX_{(n+k-r)}, \tag{5.18}$$

where k is the some integer ≥ -1.

Lemma 5.5.3 *$P(P_n > p(1+\epsilon)) \approx \alpha$ results by taking $\widehat{UCL}_k$ from (5.18), with k the smallest integer such that $Po(np(1+\varepsilon), r-1-k) \leq \alpha$ and $\lambda = P(V=1)$ such that $(1-\lambda)Po(np(1+\varepsilon), r-1-k) + \lambda Po(np(1+\varepsilon), r-k) = \alpha$.*

Proof. Arguing as in Lemma 5.3.1 we obtain that $P_n \cong (1-V)U_{r-k} + VU_{r+1-k}$. The desired result now readily follows from Lemma 5.5.2. ■

Indeed, the resulting chart which randomizes between $X_{(n+k-r)}$ and $X_{(n+k+1-r)}$ guarantees that relative errors in P_n in excess of ε will only occur in a fraction α of cases. Hence, in effect we copy the solution from Section 5.3, but instead of merely using an adjacent order statistic, we also shift k steps: a superior form of protection requires a stronger correction. Clearly, the price to be paid for such protection will be far from negligible, unless again r is quite large. For example, note that during the in-control situation EP_n will now decrease from $(r+1)/(n+1)$ to $(r+\lambda-k)/(n+1)$. So, rather than removing the positive bias there, we replace it by a typically even larger negative one, in order to be able to satisfy the exceedance criterion.

Remark 5.5.1 For the two-sided case, (5.18) translates into

$$\widehat{CL}_{kL} = (1-V)X_{(r-k)} + VX_{(r-k+1)}$$

for the lower control limit and

$$\widehat{CL}_{kU} = VX_{(n-r+k)} + (1-V)X_{(n-r+k+1)}$$

for the upper control limit. In this case $r = [np/2]$ and k is the smallest integer such that $Po(np(1+\varepsilon)/2, r-1-k) \leq \alpha$ with $\lambda = P(V=1)$ such that $(1-\lambda)Po(np(1+\varepsilon)/2, r-1-k) + \lambda Po(np(1+\varepsilon)/2, r-k) = \alpha$. □

To study the effect of the correction defined through (5.18), let us first look at the extremes again. As $n \to \infty$ for fixed p, a normal approximation e.g. shows that $Po(np(1+\varepsilon), r) \to 0$ and thus k in (5.18) will eventually become 0 (actually, it might even become negative, but by then no risk of dealing with realistic n and

p exists anymore, so we will not bother about this possibility). Turning to the opposite end, a simple example illustrates that for $r = 0$ things are still awkward.

Example 5.5.3

Let $p = 0.001, n = 800$ and $\varepsilon = 0.1$. Then we obtain $Po(0.88, 0) = \exp(-0.88) = 0.42$. Hence, if we choose $\alpha = 0.2$, it means that $\widehat{UCL} = X_{(n)}$, with probability $\lambda = 0.48$ and $\widehat{UCL} = \infty$ otherwise. Hence this adaptation clearly also decreases the expected $1/P_n$ in the out-of-control phase by about a factor 2. In comparison again, for the normal case in Subsection 2.4.2 and the same p, ε and α as used here, the factor by which the ARL is increased equals about 1.3. And that while $n = 100$, rather than the $n = 800$ used here! Clearly the difference is again tremendous. □

For the analysis of what happens between the extremes $r \to \infty$ and $r = 0$, we shall consider the impact of the correction (5.18) on the out-of-control probability.

Lemma 5.5.4 *Replacement of $\widehat{UCL}$ from (5.1) by $\widehat{UCL}_k$ from (5.18), for some k and λ, results in a relative change in EP_n approximately equal to*

$$-\frac{(k+1-\lambda)w}{\tilde{p}_1}, \tag{5.19}$$

with w and $\tilde{p}_1$ as in (5.10).

Proof. Immediate from Lemma 5.3.2. ■

Clearly, due to the presence of k, the expression in (5.19) will become small even slower than its counterpart from Section 5.3. In the latter case only a factor $(1-\lambda)$ occurs, which obviously is at most 1. Some feeling for what actually happens is again obtained by considering a numerical example (for simplicity we shall as much as possible continue with the one from Section 5.3).

Example 5.5.4

Take the usual $p = 0.001$, let $n = 5000$, and thus $r = 5$, and fix ε at 0.2. Then we obtain for $Po(6, z)$ the values 0.45, 0.29, and 0.15 for $z = 5, 4$, and 3, respectively. Hence, if we choose $\alpha = 0.2$, it follows that $k = 1$ and $\lambda = 0.36$: the uncorrected $\widehat{UCL} = X_{(4995)}$ is replaced through (5.18) by $X_{(4996)}$ with probability 0.36 and by $X_{(4997)}$ otherwise. In this way, the realized P_n differs from the intended p by more than 20% in at most 20% of the times this procedure is applied. The price to be paid for this protection will, according to (5.19), roughly be given by $1.64w/\tilde{p}_1$. The behavior of $w/\tilde{p}_1$ was already illustrated in Section 5.3. For example, the 9.2% obtained there will now become 15.1%, which still sounds reasonable. □

Of course, the n involved is again large. In particular, it is much larger than would be needed in the normal case.

Example 5.5.5

We first quote from Section 5.3 (cf. Remark 2.4.1 and (5.11)) that a correction c will produce, with respect to the uncorrected case, a relative error with value approximately $4c(1 + u_p - \Delta)/5$. Next, we obtain from Subsection 2.3.2 (see e.g. (2.53)) that for the present purpose the relevant correction, which thus keeps the exceedance probability below α, approximately satisfies $c = \tau u_\alpha - \varepsilon/u_p$, where (cf. (5.15)) $\tau^2 = var(\widehat{UCL}) \approx (u_p^2 + 2)/(2n)$ for the normal case $\widehat{UCL} = \widehat{\mu} + \breve{\sigma} u_p$. For the values considered above, $p = 0.001$, $\alpha = 0.2$ and $\varepsilon = 0.2$, we find that $c = 2.02n^{-1/2} - 0.065$, and therefore a value $(1.62n^{-1/2} - 0.052)(4.09 - \Delta)$ for the relative error results. Again letting $\Delta = 1.81$, this will wind up to the 15.1% obtained above for $n \approx 190$. This is well below $n = 5000$. □

Example 5.5.6

An intermediate example involving a larger parametric family is easily derived from Example 4.5.1. Since $\widehat{UCL}$ there equals $\widehat{\mu} + \widehat{\sigma}\overline{K}_{\widehat{\gamma}}^{-1}(p)$, a third parameter γ needs to be estimated, which thus leads, for each given n, to an increase of $\tau^2 = var(\widehat{UCL})$ in comparison with the normal case. Following (4.18), it is calculated that $n\tau^2$ actually grows by a factor 4.32 when $p = 0.001$ and $\gamma = 0$. Consequently, the required n also grows by this factor: for $\Delta = 1.81$, the desired 15.1% is now reached for $n \approx 820$, a value which is indeed intermediate between the normal 190 and the nonparametric 5000. □

It is nice to see from the examples above that by using a suitable correction it is indeed possible to have acceptable behavior both while being in-control ("big errors occur rarely") as well as while being out-of-control ("detection power remains close to the uncorrected value"). Clearly, this cannot be achieved for customary n and p, for which r typically equals zero. Substantially larger values of n and/or p are needed than for normal or parametric charts, as the examples have demonstrated. But on the other hand, the pessimism according to which r itself should be really large, is fortunately not realistic. Values like $r = 5$ seem to offer reasonable opportunities, and it matters little whether this is achieved by increasing n, or p, or both.

5.6 Concluding remarks

Although the parametric approach discussed in Chapters 3 and 4 posseses an advantage over the normal approach presented in Chapter 2, that approach is unable to avoid the model errors from distributions (far) outside the normal power family. A nonparametric approach offers a solution for such a problem, but requires huge sample sizes for the estimation step. The present chapter discusses the conditions under which the nonparametric approach becomes a feasible alternative as compared with its parametric counterparts. The following gives the proposed procedure as a summary of the nonparametric approach.

Suppose starting values are given for p and n. The basic nonparametric chart simply uses as an upper limit $\widehat{UCL}$ the order statistic $X_{(n-r)}$, where $r = [np]$. To keep the in-control behavior under control, for example we require that $P(P_n > p(1+\varepsilon)) \leq \alpha$. An alternative is to focus on the ARL and use $P(1/P_n < (1-\varepsilon)/p)$, but this is equivalent: just replace ε in the first criterion by $\varepsilon/(1-\varepsilon)$. Choose values for ε and α. Let $Po(\tilde{\lambda}, k)$ denote the cumulative Poisson probability $P(\widetilde{Y} \leq k)$, where $\widetilde{Y}$ is Poisson distributed with parameter $\tilde{\lambda}$. If $Po(np(1+\varepsilon), r) \leq \alpha$, the requirement is met and the chart can be used without further correction.

If not, somewhat larger values of ε and α could be settled for. But if this is not allowed or does not work either, a correction is applied. Find the smallest integer k for which $Po(np(1+\varepsilon), r-1-k) \leq \alpha$. Note that for $r = 0$ usually k will be 0 as well. Let λ be such that $(1-\lambda)Po(np(1+\varepsilon), r-k-1) + \lambda Po(np(1+\varepsilon), r-k) = \alpha$. Then replace $\widehat{UCL}$ by $X_{(n+k-r)}$ with probability λ and by $X_{(n+k+1-r)}$ with probability $(1-\lambda)$. The requirement is met and the behavior during the in-control situation thus is acceptable.

Next, check whether the price for the protection afforded by this correction, is acceptable during out-of-control as well. Take the normal case as a yardstick and suppose the mean has shifted by Δ standard deviations to the right. Let u_q denote the standard normal upper quantile for any q. Then the effect of the correction will lead to a relative change that approximately equals $(k+1-\lambda)\{1+u_q-\Delta\}/\{(r+1)(1+u_q)\}$, with $q = (r+1)/(n+1)$. For very small r, like $r = 0$, the approximation is not reliable, but then the relative errors involved are too large to be acceptable anyhow. If this expression is suitably small, the corrected chart can indeed be applied.

If not, n and/or p need to be raised. In view of the simplicity of the relative error approximation considered, it is easy to figure out by some trial and error what kind of values will produce a result that is acceptable. Finally, the corrected chart based on the updated n and/or p, is the one to be applied.

Chapter 6

Data Driven Approach:
A combined procedure

In the previous chapters the normal, parametric and non-parametric approaches were proposed to determine a suitable control chart for practical usage. Although each of these approaches has its own advantage, applying these approaches individually causes some disadvantages due to the specific characteristics of the observations encountered in the production process. Therefore, this chapter is devoted to combine the three approaches and let the data, based on some criteria, select either the normal, parametric or nonparametric chart as the most suitable chart for themselves.

The present chapter deals with the question when to switch from the control chart based on normality to a parametric control chart, or even to a nonparametric one. This model selection problem is solved by using the estimated model error as yardstick. We restrict ourselves in this chapter to control charts corrected for the bias. It is shown that the new combined control chart asymptotically behaves as each of the specific control charts in their own domain. Simulations exhibit that the combined control chart performs very well under a great variety of distributions and hence it is recommended as an omnibus control chart, nicely adapted to the distribution at hand. The combined control chart is illustrated by an application on real data. The modified nonparametric control chart, essentially the same as the one introduced in Chapter 5, is an attractive alternative and can be recommended as well.

6.1 Introduction

Classical control charts are based on the assumption that the observations are normally distributed. It has been shown in Chapter 2 that, even if this assumption is fulfilled, we still need many data to get an accurate control limit when estimators of the parameters are simply plugged in. Chapter 2 also provides a comprehensive discussion on how to deal with such a problem.

As a matter of fact, the normality assumption is in practice not always fulfilled and the control chart often is seriously in error when the distributional form of the observations differs from normality. An obvious solution to the problem is to assume a larger parametric model, containing normality as a sub-model, and to produce a control chart in this new setting. This step has been made in Chapters 3 and 4, where the merits of such an approach are comprehensively discussed.

It is clear that, as long as we are (very) close to normality, the control chart based on normality is preferable. The first reason is that in that case the performance of this control chart is better, since it is matched to normality of the observations. The second reason is that the normal control chart is easier and more familiar and hence people prefer to use this chart as long as possible. However, for distributions farther away from normality, but still close to the larger parametric family, the best choice is the parametric control chart. In that case we should no longer stick to the normal control chart, but move toward the parametric chart.

Of course, there also are distributions very far outside the larger parametric family where the parametric control chart is not satisfactory either and a nonparametric approach should be applied. One may ask why not always apply a nonparametric control chart or a parametric control chart in a very large parametric family. The point is that there are two types of error: the model error (due to the distance between the true distribution and the most suitable distribution in the supposed model) and the stochastic error/indexstochastic error (caused by estimating parameters or, in the nonparametric case, an extreme quantile). The larger the parametric model, the smaller the model error (with a vanishing model error in the nonparametric control chart), but the larger the stochastic error. For instance, estimating the 0.999-quantile with 100 observations makes no sense in a nonparametric setting. Chapter 5 provides a detailed analysis on the conditions under which the nonparametric charts become feasible alternatives for their parametric counterparts.

The theme of the present chapter is how to choose between the three control charts: the normal control chart, the parametric control chart as developed for the normal power family and the nonparametric control chart. The idea is to let the data tell what control chart to use. A first idea might be to execute a (standard) goodness of fit test to investigate normality. If normality is not rejected, use the normal control chart. If we do reject it, apply a goodness of fit test for the normal power family. Again, when not rejecting the normal power family, apply the parametric chart and otherwise use the nonparametric chart (if this makes sense).

Although this way of thinking looks attractive, it has a serious drawback. Standard goodness of fit tests are looking at the majority of the data, and as such concentrate on the middle of the distribution, while here we are not interested in this middle part, but in the (extreme) tail. Therefore, standard goodness of fit tests are not appropriate for the situation at hand. For the same reason, less formal methods like "a good look at the data" or "an inspection of a histogram" are completely insufficient to judge the possible normality in the far tail.

In fact, the choice between the three control charts can be seen as a model selection problem. This area in statistics is nowadays in the center of interest and therefore it looks promising to apply such methods not merely for the three charts mentioned here, but for a whole range of (nested) models as well. Unfortunately, under this strategy two problems arise. First of all, it is far from easy to develop control charts in each of these models, see the discussion on several types of models and the corresponding problems in deriving suitable control charts in Section 3.3. Secondly, again the common selection rules are intended for the bulk of the data and not for the (extreme) tail.

The motivation to switch from the normal control chart to the parametric control chart or even to the nonparametric control chart is provided by the model error. As before, let p be the false alarm rate, then the model error is the discrepancy between p and the probability (under the true distribution of X) that X is larger than the "ideal" control limit, being the $(1-p)$-quantile of the "most suitable" distribution in the supposed model. Indeed, it is seen at this point that not the middle of the distribution is coming in, but, due to the (very) small value of p, its far tail. For a more technical description of the model error we refer to Section 3.2.

As this model error is the discriminating quantity in deciding which control chart to use and since moreover the data should tell us what the appropriate model is, it is natural to base the decision between the three control charts on the estimated model error. To avoid too many technical complications and to keep the control chart as simple as possible we use the (standardized) largest observation to choose between the control charts, thus really looking at the tail of the distribution.

In Section 6.2 the three control charts are briefly reviewed. Here also a modified version of the nonparametric control chart is presented. Section 6.3 is devoted to the choice of the model. The three control charts and the decision rule telling us which one to choose among them, in fact together form a new control chart. An application of this combined control chart on real data is performed in Section 6.4. It is shown in detail how to do the calculations. The in-control and out-of-control behavior from a theoretical point of view of the combined chart is worked out in Section 6.5. In order to study the theoretical behavior of the combined control chart we need so called large deviation results for our estimators of the parameters in the normal power family. Therefore, also in this section some large deviation results are presented. In contrast to for instance the normal control chart, the combined control chart is valid for all distributions and its out-of-control behavior is asymptotically as good as when we would know to which of the three classes of distributions the true distribution belongs. In addition to the theoretical results we present some simulation results on the new control chart in Section 6.6. It turns out that the combined control chart behaves very well under a great variety of distributions and therefore it is recommended as an omnibus control chart, nicely adapted to the distribution at hand.

The modified version of the nonparametric control chart shows a good behavior in the in-control situation, as might be expected from a chart close to a real nonparametric chart. This modified chart has also a nice performance in the out-of-control situation in contrast to the standard nonparametric control chart, although under normality some loss has to be accepted w.r.t. the combined control chart. Therefore, this modified nonparametric chart can be considered as an attractive alternative omnibus control chart.

6.2 Three types of control charts

It has been discussed in the previous chapters that applying either the normal, parametric or nonparametric chart individually may cause some disadvantages, although each of them has its own advantages. Therefore, the idea to combine the three charts into one procedure and let the data choose the most suitable charts for themselves seems to be a good proposal.

In this section we briefly review the three control charts to be selected by the data: the normal chart (cf. Chapter 2), the parametric chart (cf. Chapters 3 and 4) and the nonparametric chart (cf. Chapter 5). Here, we use the same notations and estimators as in those chapters. To avoid redundancy, we refer to equations or remarks in those chapters when necessary.

6.2.1 Normal control chart

Let $X_1, \cdots, X_n, X_{n+1}$ be i.i.d. r.v.'s, each with a $N(\mu, \sigma^2)$-distribution. If μ and σ are known and p is the probability of incorrectly concluding that the process is out-of-control, then the UCL of the normal chart equals $\mu + u_p\, \sigma$. However, very often μ and σ are unknown and hence we use the estimators $\widehat{\mu}$ and $\breve{\sigma}$, respectively, resulting in $\widehat{UCL} = \widehat{\mu} + u_p\, \breve{\sigma}$.

The normal control limit devised for $Eg(P_n) \approx g(p)$, with g as given in (2.3), is obtained from the plug-in control limit $\widehat{UCL}$ by adding a term c_N to correct for the bias. That is, the normal control chart is given by

$$X_{n+1} > \widehat{\mu} + (u_p + c_N)\check{\sigma} \ , \quad \text{with} \quad c_N \quad \text{as in Table 6.1.} \tag{6.1}$$

Table 6.1 Correction terms according to the function g.

$g(p)$	c_N
p	$\dfrac{u_p}{4n} + \dfrac{u_p(u_p^2+2)}{4n}$
$\dfrac{1}{p}$	$\dfrac{u_p}{4n} + \dfrac{u_p^2+2}{4n}\left\{u_p - \dfrac{2\varphi(u_p)}{p}\right\}$
$1-(1-p)^k$	$\dfrac{u_p}{4n} + \dfrac{u_p^2+2}{4n}\left\{u_p - \dfrac{(k-1)\varphi(u_p)}{1-p}\right\}$

For the derivation of the normal control chart and its properties we refer to Chapter 2.

6.2.2 Parametric control chart

In order to embed the normal distributions in a larger family with heavier or thinner tails we essentially consider powers of the standard normal quantiles as new quantiles, resulting in the normal power family as given in Subsection 3.3.3. More precisely, the $(1-p)$-quantile in the normal distribution u_p is replaced by $\overline{K}_\gamma^{-1}(p)$, cf. (3.7). The parametric control chart is based on this normal power family with the UCL being equal $\mu + \overline{K}_\gamma^{-1}(p)\,\sigma$.

As in the normal case, the corrected parametric control chart can be derived by adding a correction term c_u to the UCL to correct for the bias. The corrected UCL is given in (3.17). It turns out that adding an extra parameter γ in addition to μ and σ makes it far more complicated to derive control charts. Fortunately, the recommended control chart given in (3.39) is in an explicit form and can be applied straightforwardly. For more detailed discussion on this control chart and its properties we refer to Section 3.6.

6.2.3 Nonparametric control charts

In this Subsection, we present a version of a standard nonparametric control chart slightly different from that of Chapter 5. In addition, we also present the modified version of this chart here. The modified version will be used in the combined chart, in the case that the data reject both the normal and the parametric chart.

Standard nonparametric control chart

We start with a formal definition of the inverse of a distribution function: $F^{-1}(t) = \inf\{x : F(x) \geq t\}$. For $\overline{F} = 1 - F$ we define $\overline{F}^{-1}(p) = F^{-1}(1-p) = \inf\{x : F(x) \geq 1-p\} = \inf\{x : \overline{F}(x) \leq p\}$. Applying this to the empirical distribution function F_n we get

$$\overline{F_n}^{-1}(p) = X_{(n-\widetilde{r})} \text{ for } \frac{\widetilde{r}}{n} \leq p < \frac{\widetilde{r}+1}{n}, \widetilde{r} = 0, \ldots, n-1.$$

Hence we have $\widetilde{r} = [np]$. Estimating the unknown quantile $F^{-1}(p)$ in a non-parametric way by its empirical counterpart gives the uncorrected plug-in control chart

$$X_{n+1} > \overline{F_n}^{-1}(p) = X_{(n-[np])}.$$

This implies that we get unconditional false alarm probabilities EP_n equal to $j/(n+1)$ with $j = 1 + [np] \geq 1$. An obvious way to correct for the estimation error is to replace p by $p(1 \pm \zeta)$ for some suitably chosen ζ. However, due to (a) the fact that we have always at least a false alarm probability $1/(n+1)$ and (b) the discrete character of the possible false alarm probabilities, we will not get satisfactory results unless n is very large or p not too small.

A possible solution is to make the false alarm probabilities continuous by adding a randomization procedure as follows. Let $U_{(1)} \leq \ldots \leq U_{(n)}$ 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 $U_{(n+1)} = 1$. Let g be a monotone function. In particular, we use g given in (2.3). For a monotone increasing g define the integer r with $0 \leq r = r(p) \leq n$ by

$$Eg(U_{(r)}) \leq g(p) < Eg(U_{(r+1)}). \tag{6.2}$$

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

$$X_{n+1} > VX_{(n-r)} + (1-V)X_{(n-r+1)}, \tag{6.3}$$

with

$$P(V = 1) = \frac{g(p) - Eg(U_{(r)})}{Eg(U_{(r+1)}) - Eg(U_{(r)})},$$

where in case $r = 0$ we define $X_{(n+1)} = \infty$. Since $\overline{F}(X_{(n-r)})$ and $\overline{F}(X_{(n-r+1)})$ are distributed as $U_{(r+1)}$ and $U_{(r)}$, respectively, obviously we obtain

$$Eg(P_n) = g(p). \tag{6.4}$$

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

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

$$X_{n+1} > VX_{(n-[p(n+1)])} + (1-V)X_{(n-[p(n+1)]+1)}, \tag{6.5}$$

with

$$P(V=1) = p(n+1) - [p(n+1)].$$

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

$$Eg(U_{(r)}) \ge g(p) > Eg(U_{(r+1)}). \tag{6.6}$$

The control chart is again given by (6.3). 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

$$X_{n+1} > VX_{(n-[np]-1)} + (1-V)X_{(n-[np])},$$

with

$$P(V=1) = \frac{([np]+1)(np-[np])}{np}.$$

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 $E\frac{1}{P_n} = E\frac{1}{U_{(2)}} = n < \frac{1}{p}$.

For further discussion and results on this and similar nonparametric control charts we refer to Chapter 5.

Modified nonparametric control chart

The practical implementation of 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. For example when $g(p) = p$ then $r = 0$ corresponds to $p(n+1) < 1$. Therefore, we modify the nonparametric control chart somewhat in case $r = 0$. We replace the definition $X_{(n+1)} = \infty$ by $X_{(n+1)} = X_{(n)} + S$. For the in-control situation $X_{(n)}$ is already rather far in the tail of the distribution of X_{n+1} and adding S gives for most distributions a really large value. However, a substantial shift in the out-of-control case can now be detected with higher probability than $P(V=1)$, which equals, for instance, 0.251 when $g(p) = p$, $p = 0.001$ and $n = 250$. Moreover, it is now no longer decided to keep the process running "for ever" when $V = 0$.

The modified nonparametric chart is defined by (6.3) if $r \geq 1$ and for $r = 0$ by

$$X_{n+1} > X_{(n)} + (1 - V)S, \tag{6.7}$$

with

$$P(V = 1) = \frac{g(p) - g(0)}{Eg(U_{(1)}) - g(0)}.$$

6.3 Choosing the model

When the observations are close to normality, we want to select the normal control chart. If the departure from normality is too large, we apply the parametric control chart, unless the parametric family also does not fit. In the latter case the (modified) nonparametric control chart comes in. It is argued already in the introduction (but see also below) that the model error is the guide for choosing between the charts. In principle, the model error can be defined both for the in-control and the out-of-control situation. However, because of the validity of the control chart, our main concern lies in the model error for the in-control case. Therefore, when developing rules for choosing between the three control charts we assume that $X_1, \ldots, X_n, X_{n+1}$ are i.i.d. r.v.'s with common distribution function F.

In deciding, for instance, whether the departure from normality is too large, we have to use a measure for the distance between the realized distribution (the true F) and the supposed model (in this case, normality). This "distance" should be chosen in accordance with the problem at hand, that is the difference between the observed false alarm rate P_n and the prescribed false alarm rate p. The total error consists of the model error and the stochastic error. The stochastic error can be reduced by a correction term according to the criterion at hand. The model error should be reduced by an appropriate choice of the model and here the notion of "distance" naturally comes in. Furthermore, the data should tell us whether the model error is (too) large. Therefore, the selection between the three possible control charts is based on a kind of *estimation* of the *model error*. To come to an implementation of the preceding ideas, we use a more technical definition of the model error such as presented in Section 3.2, see (3.5).

6.3.1 Normality

Let us start with considering the question whether to use the normal control chart or not. The supposed distribution function G_θ is that of $\mu + \sigma Z$, with Z having a standard normal distribution. Therefore, the model error equals $\overline{F}\,(\mu + \sigma u_p) - p$ and we want to check, based on our observations $X_1, \ldots, X_n$, the behavior of $(X - \mu)/\sigma$ in the far tail, where p is (very) small. For instance, when $p = 0.001$

and $n < 1000$, the most obvious quantity to look at is the standardized maximum of our observations: $(X_{(n)} - \widehat{\mu})/\breve{\sigma}$. In principle we can use for larger p and/or n other order statistics, but to avoid technical complications and to keep the combined control chart simple we restrict ourselves to $(X_{(n)} - \widehat{\mu})/\breve{\sigma}$.

The next point is the determination of the cut-off points for staying at the normal chart. If $X_1, \ldots, X_n$ are i.i.d. with a standard normal distribution and $\{d_{1N}(n)\}, \{d_{2N}(n)\}$ are sequences of positive numbers satisfying $\lim_{n\to\infty} d_{1N}(n) = \infty, d_{1N}(n) < n$ and $\lim_{n\to\infty} d_{2N}(n) = 0$, then

$$\lim_{n\to\infty} P\left(X_{(n)} < \overline{\Phi}^{-1}\left(\frac{d_{1N}(n)}{n}\right)\right) = 0$$

and

$$\lim_{n\to\infty} P\left(X_{(n)} > \overline{\Phi}^{-1}\left(\frac{d_{2N}(n)}{n}\right)\right) = 0.$$

Therefore, we will prefer the normal control chart when

$$\overline{\Phi}^{-1}\left(\frac{d_{1N}(n)}{n}\right) \le \frac{X_{(n)} - \overline{X}}{S} \le \overline{\Phi}^{-1}\left(\frac{d_{2N}(n)}{n}\right). \tag{6.8}$$

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 3.3 and hence the control chart is invalid. Distributions with thinner tails are conservative in the in-control case with as consequence a loss in the out-of-control situation. Because errors in the in-control situation are more serious than those in the out-of-control case and since a positive model error as large as p or larger can easily occur, whereas the negative model error is at most $-p$, we take the selection rule *unbalanced*. In particular, we will consider $d_{1N}(n) = -0.7 + 0.5\log n, d_{2N}(n) = 5/\sqrt{n}$, leading to

$$\begin{aligned} P\left(X_{(n)} < \overline{\Phi}^{-1}\left(\frac{-0.7 + 0.5\log n}{n}\right)\right) &= \left(1 - \frac{-0.7 + 0.5\log n}{n}\right)^n \\ &\approx \exp\left(0.7 - 0.5\log n\right) \approx \frac{2}{\sqrt{n}} \end{aligned}$$

and

$$P\left(X_{(n)} > \overline{\Phi}^{-1}\left(\frac{5}{n\sqrt{n}}\right)\right) = 1 - \left(1 - \frac{5}{n\sqrt{n}}\right)^n \approx 1 - \exp\left(-\frac{5}{\sqrt{n}}\right) \approx \frac{5}{\sqrt{n}},$$

when $X_1, \ldots, X_n$ are i.i.d. having a standard normal distribution. This choice is based on both the theoretical results of Section 6.5 and our simulation experience, see Section 6.6.

Remark 6.3.1 The first step in the selection procedure is to choose the normal or the parametric chart. One might wonder why not to take $\widehat{\gamma}$ as statistic to choose between these two. After all, the normal distribution is a member of the normal power family with $\gamma = 0$ and we simply have to investigate whether $\gamma = 0$ or not. The estimator $\widehat{\gamma}$ seems to be the obvious measure for that. However, presentation in this form is misleading, since we do not need to escape from the normal chart only if we have a member of the normal power family with a substantially different value of γ, but also for distributions outside the normal power family. So, the selection procedure should not be restricted to the framework of the normal power family.

Indeed, it turns out that for several well known distributions such as the Student distribution pretty small values of $\widehat{\gamma}$ are very common, while nevertheless a rather large model error occurs when assuming normality. For example, the standardized Student distribution with 6 degrees of freedom has 0.118 as limiting value of $\widehat{\gamma}$ (when n tends to infinity), while for $p = 0.001$ its model error equals 3.56 and hence $EP_n \approx 4.56p$, see Table 3.3. $\square$

6.3.2 Normal power family

Next, we address the question how to choose between the parametric control chart and the nonparametric one. Similarly as for the normal chart we prefer the parametric control chart when

$$\frac{X_{(n)} - \overline{X}}{S} \notin \left[\overline{\Phi}^{-1}\left(\frac{d_{1N}(n)}{n}\right), \overline{\Phi}^{-1}\left(\frac{d_{2N}(n)}{n}\right) \right], \tag{6.9}$$
$$\overline{K}_{\widehat{\gamma}}^{-1}\left(\frac{d_{1P}(n)}{n}\right) \le \frac{X_{(n)} - \overline{X}}{S} \le \overline{K}_{\widehat{\gamma}}^{-1}\left(\frac{d_{2P}(n)}{n}\right),$$

where $\{d_{1P}(n)\}$, $\{d_{2P}(n)\}$ are sequences of positive numbers satisfying $\lim_{n\to\infty} d_{1P}(n) = \infty$, $d_{1P}(n) < n$ and $\lim_{n\to\infty} d_{2P}(n) = 0$. In particular, we will consider $d_{1P}(n) = -0.2 + 0.5 \log n$, $d_{2P}(n) = 3/\sqrt{n}$, leading to

$$P\left(X_{(n)} < \overline{K}_{\gamma}^{-1}\left(\frac{-0.2 + 0.5 \log n}{n}\right)\right) = \left(1 - \frac{-0.2 + 0.5 \log n}{n}\right)^n$$
$$\approx \exp\left(0.2 - 0.5 \log n\right) \approx \frac{1.2}{\sqrt{n}}$$

and

$$P\left(X_{(n)} > \overline{K}_{\gamma}^{-1}\left(\frac{3}{n\sqrt{n}}\right)\right) = 1 - \left(1 - \frac{3}{n\sqrt{n}}\right)^n \approx 1 - \exp\left(-\frac{3}{\sqrt{n}}\right) \approx \frac{3}{\sqrt{n}},$$

when $X_1, \ldots, X_n$ are i.i.d. having a normal power family distribution with distribution function K_γ. Note that this specific range has somewhat larger probability of staying at the parametric control chart than the corresponding one for the normal chart, because here the model error is in general lower. The "unbalance-factor" is in both cases approximately the same: $5/2 = 3/1.2$.

6.3.3 Nonparametric

When neither (6.8) nor (6.9) hold, we choose the modified nonparametric control chart. Moreover, when $p(n+1) \geq 1$ we can make some simplifications by ignoring the correction terms in the normal and parametric control chart (since they are rather small in that case) and replacing in the nonparametric chart the stochastic term V by its deterministic counterpart EV.

6.3.4 Combined control chart

In order to avoid too much technicalities we restrict ourselves to $g(p) = p$, which is also considered in the simulation study. Note, however, that with (some) more effort other g's like $g(p) = 1 - (1-p)^k$ or $g(p) = 1/p$ can be analyzed as well.

Consequently, for $g(p) = p$ the combined control chart is defined as follows. Let UL_N, UL_P, UL_{MNP} denote the upper limit of the normal, parametric and modified nonparametric control chart (for $r = 0$), respectively, that is, cf.(2.27), (3.39), and (6.7),

$$UL_N = \widehat{\mu} + (u_p + c_N)\breve{\sigma} \text{ with } c_N \text{ as in Table 1,} \tag{6.10}$$

$$UL_P = \widehat{\mu} + \breve{\sigma}\left\{\overline{K}_{\widehat{\gamma}}^{-1}(p) - C1(\widehat{\gamma})\,C2(\widehat{\gamma}) - \frac{C3(\widehat{\gamma})}{n} + \frac{C4(\widehat{\gamma})}{n}\right\},$$

$$UL_{MNP} = X_{(n)} + (1-V)S \text{ with } P(V=1) = p(n+1)$$

and let

$$IN = \left[\overline{\Phi}^{-1}\left(\frac{d_{1N}(n)}{n}\right), \overline{\Phi}^{-1}\left(\frac{d_{2N}(n)}{n}\right)\right], \tag{6.11}$$

$$IP = \left[\overline{K}_{\widehat{\gamma}}^{-1}\left(\frac{d_{1P}(n)}{n}\right), \overline{K}_{\widehat{\gamma}}^{-1}\left(\frac{d_{2P}(n)}{n}\right)\right],$$

then the combined control chart for $r = 0$ is given by

$$X_{n+1} > UL_N \mathbf{1}\left(\frac{X_{(n)} - \overline{X}}{S} \in IN\right) \tag{6.12}$$

$$+ UL_P \mathbf{1}\left(\frac{X_{(n)} - \overline{X}}{S} \notin IN\right) \mathbf{1}\left(\frac{X_{(n)} - \overline{X}}{S} \in IP\right)$$

$$+ UL_{MNP} \mathbf{1}\left(\frac{X_{(n)} - \overline{X}}{S} \notin IN\right) \mathbf{1}\left(\frac{X_{(n)} - \overline{X}}{S} \notin IP\right),$$

with $\mathbf{1}(A) = 1$ if A holds and 0 otherwise. Writing $\delta = p(n+1) - r$ the combined control chart for $r \geq 1$ is given by

$$X_{n+1} > \left(\overline{X} + u_p S\right) \mathbf{1}\left(\frac{X_{(n)} - \overline{X}}{S} \in IN\right) \tag{6.13}$$

$$+ \left(\overline{X} + \overline{K}_{\widehat{\gamma}}^{-1}(p)\, S\right) \mathbf{1}\left(\frac{X_{(n)} - \overline{X}}{S} \notin IN\right) \mathbf{1}\left(\frac{X_{(n)} - \overline{X}}{S} \in IP\right)$$

$$+ \left\{\delta X_{(n-r)} + (1-\delta) X_{(n-r+1)}\right\} \mathbf{1}\left(\frac{X_{(n)} - \overline{X}}{S} \notin IN\right) \mathbf{1}\left(\frac{X_{(n)} - \overline{X}}{S} \notin IP\right).$$

Returning to both the in- and out-of-control situation, that is $X_1, \ldots, X_n$ have distribution function F and X_{n+1} has distribution function G with $G = F$ for the in-control case, we therefore get (with E referring to the expectation under F)

$$EP_n = E\left\{\overline{G}(UL_N^*)\mathbf{1}\left(\frac{X_{(n)} - \overline{X}}{S} \in IN\right)\right\} \tag{6.14}$$

$$+ E\left\{\overline{G}(UL_P^*)\mathbf{1}\left(\frac{X_{(n)} - \overline{X}}{S} \notin IN\right)\mathbf{1}\left(\frac{X_{(n)} - \overline{X}}{S} \in IP\right)\right\}$$

$$+ E\left\{\overline{G}(UL_{NP}^*)\mathbf{1}\left(\frac{X_{(n)} - \overline{X}}{S} \notin IN\right)\mathbf{1}\left(\frac{X_{(n)} - \overline{X}}{S} \notin IP\right)\right\},$$

where $UL_N^* = UL_N$ if $r = 0$ and $UL_N^* = \overline{X} + u_p S$ if $r \geq 1$, $UL_P^* = UL_P$ if $r = 0$ and $UL_P^* = \overline{X} + \overline{K}_{\widehat{\gamma}}^{-1}(p)\, S$ if $r \geq 1$ and $UL_{NP}^* = UL_{MNP}$ if $r = 0$ and $UL_{NP}^* = \delta X_{(n-r)} + (1-\delta) X_{(n-r+1)}$ if $r \geq 1$. It is easily seen that due to location and scale invariance, without loss of generality, we may take $\mu = EX_1 = 0$ and $\sigma^2 = var(X_1) = 1$.

We want to show the following results:

1. The combined procedure has good in-control behavior under normality ($F = \Phi$), the normal power family ($F = K_\gamma$ for some γ) and outside the normal power family ($F \neq K_\gamma$ for all γ).

2. During out-of-control there is a gain w.r.t. the nonparametric chart if $F = K_\gamma$ for some γ, and hence in particular if $F = \Phi$. There is only a small loss w.r.t. the normal and parametric chart if $F = \Phi$, w.r.t. the parametric chart if $F = K_\gamma$ for some γ, and also only a small loss w.r.t. the nonparametric chart if $F \neq K_\gamma$ for all γ.

From the theoretical point of view this will be shown by demonstrating that the combined control chart in each of the three situations (normality, normal power family, outside the normal power family) asymptotically behaves as the specific corresponding control chart. In the simulations this is established by numerical comparison of the specific and the combined control chart for several distributions.

The main point in the theoretical part is to show that the estimators $\widehat{\mu}$, $\check{\sigma}$ and $\widehat{\gamma}$ can essentially be ignored in the selection part of the procedure. This requires a lot of effort, including results on large deviations for these estimators, which are presented in Section 6.5. Simplifying things a little bit and taking $d_{1N}(n) = -0.7 + 0.5 \log n, d_{2N}(n) = 5/\sqrt{n}, d_{1P}(n) = -0.2 + 0.5 \log n, d_{2P}(n) = 3/\sqrt{n}$, for instance, when normality holds the procedure then looks as follows: choose the normal chart except for a probability $\frac{2}{\sqrt{n}}$ (due to possible thinner tail) $+\frac{5}{\sqrt{n}}$ (due to possible heavier tail); when not taking the normal chart we choose with probability $\frac{0.8}{\sqrt{n}} + \frac{2}{\sqrt{n}}$ the parametric chart and otherwise the nonparametric chart.

In general, the behavior of the combined chart is indeed a mixture of the three charts according to the weights of the selection rule with the following refinement: the contribution of the normal chart is similar to the one in the unselected case, due to almost independence of $X_{(n)}$ and $(\overline{X}, S)$; the contribution of the nonparametric part is lower than in the unselected case, when we enter the nonparametric chart because of large values of $X_{(n)}$, implying that the control limit is also relatively large, and the contribution is larger otherwise; in the parametric part there is some dependence, but not that much. The following simulation results illustrate this phenomenon.

Example 6.3.1

Assume that $X_1, \ldots, X_{1000}$ are i.i.d. r.v.'s with a normal distribution. The probability of choosing the normal chart then is approximately equal to $1 - 7/\sqrt{1000} = 0.78$. Our simulation result based on 100 000 simulations as given in Table 6.3 gives 0.78 as well. The simulated "EP_n" for this part, that is the frequency of the out-of-control signal for these simulations, equals 0.00103, which corresponds with the simulated EP_n using the normal chart for all 100 000 simulations, being also 0.00103.

The probability of choosing the parametric chart, due to large values of $X_{(n)}$ ($EPP1$), is approximately equal to $2/\sqrt{1000} = 0.06$. The simulated value is 0.07. The contribution of this part equals 0.00076, which is somewhat smaller than the simulated EP_n using the parametric chart for all 100 000 simulations, being 0.00113 . The probability of choosing the parametric chart due to small values of $X_{(n)}$ ($EPP2$), is approximately equal to $0.8/\sqrt{1000} = 0.03$, which is also the value in the simulation. The simulated "EP_n" for this part equals 0.00165, which indeed is somewhat larger than 0.00113.

The probability of choosing the nonparametric chart due to large values of $X_{(n)}$ ($EPNP1$), is approximately equal to $3/\sqrt{1000} = 0.09$, which is also the simulated value. The simulated "EP_n" for this part is 0.00007, which is much smaller than 0.00101, being the simulated EP_n using the nonparametric chart for all 100 000 simulations. The probability of choosing the nonparametric chart due to small values of $X_{(n)}$ ($EPNP2$), is approximately equal to $1.2/\sqrt{1000} = 0.04$, while the simulation gives 0.03. Its contribution is 0.00401 and this is (as expected) much higher than 0.00101.

The resulting simulated EP_n for the combined chart thus equals $0.78 \times 0.00103 + 0.07 \times 0.00076 + 0.03 \times 0.00165 + 0.09 \times 0.00007 + 0.03 \times 0.00401 = 0.00103$. Table 6.3 also provides the results using $n = 500$, $n = 1500$ and $n = 2000$.

Table 6.3 In-control behavior of the normal (N), parametric (P) and nonparametric (NP)control charts and the combined control chart (C) for observations drawn from the normal family.

chart	result	$n = 500$	$n = 1000$	$n = 1500$	$n = 2000$
normal	EPN	0.00100	0.00103	0.00102	0.00102
	$\#N$	72,977	78,401	83,661	85,484
parametric	EPP	0.00097	0.00113	0.00109	0.00107
	$EPP1$	0.00046	0.00076	0.00077	0.00080
	$\#P1$	7,448	7,245	4,694	4,087
	$EPP2$	0.00168	0.00165	0.00148	0.00139
	$\#P2$	3,905	2,790	2,221	1,869
nonparametric	$EPNP$	0.00100	0.00101	0.00090	0.00100
	$EPNP1$	0.00010	0.00007	0.00015	0.00057
	$\#NP1$	12,311	8,830	7,148	6,522
	$EPNP2$	0.00386	0.00401	0.00314	0.00265
	$\#NP2$	3,359	2,734	2,276	2,038
combined	EPC	0.00097	0.00103	0.00101	0.00102

□

6.4 Application of the combined chart

We apply the new control chart on a real life example concerning the production of electric shavers by Philips. In an electrochemical process razor heads are formed. The measurements concern the thickness of the razor heads. There are two samples each of 835 measurements. One sample will be used to settle the control chart and then this chart is applied to the second sample. A histogram of the first sample is given in Figure 6.1.

The control limit is obtained in the following steps

1. We calculate $\frac{X_{(n)}-\overline{X}}{S}$. Since $X_{(n)} = 51.66, \overline{X} = 42.366, S = 3.311$, this leads to
$$\frac{X_{(n)} - \overline{X}}{S} = 2.807.$$
2. We calculate the cut-off points for choosing the normal chart. These values are
$$\overline{\Phi}^{-1}\left(\frac{-0.7 + 0.5\log n}{n}\right) = 2.728 \text{ and}$$
$$\overline{\Phi}^{-1}\left(\frac{5}{n\sqrt{n}}\right) = 3.531.$$

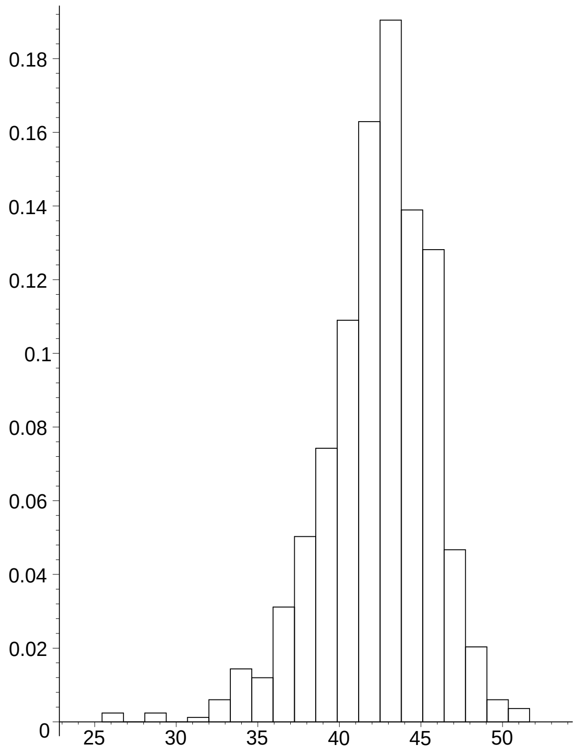

Figure 6.1. Histogram of the thickness of razor heads for the first sample of 835 measurements.

3. Because $2.807 \in [2.728, 3.531]$ we choose the normal chart and therefore we calculate the corresponding control limit. It is given by $\widehat{\mu} + (u_p + c_N)\breve{\sigma}$ with $\widehat{\mu} = \overline{X} = 42.366, u_p = 3.090, c_N = \dfrac{u_p}{4n} + \dfrac{u_p(u_p^2+2)}{4n} = 0.012$ and $\breve{\sigma} = S = 3.311$, see (6.1). As a result the control limit equals 52.635 and an out-of-control signal is given for new observations being larger than 52.635.

We apply the control chart to the second sample. Figure 6.2 shows the result, where the horizontal line gives the control limit 52.635. As is seen there is one observation among the 835 observations which gives an out-of-control signal.

Note that we do not need to calculate the parametric or nonparametric control chart in this case.

If one is interested in controlling the lower values in this example the calculations should be modified in an obvious way. Essentially we consider $-X_1, \ldots, -X_n$, calculate for these r.v.'s the upper control limit and apply these control limit to

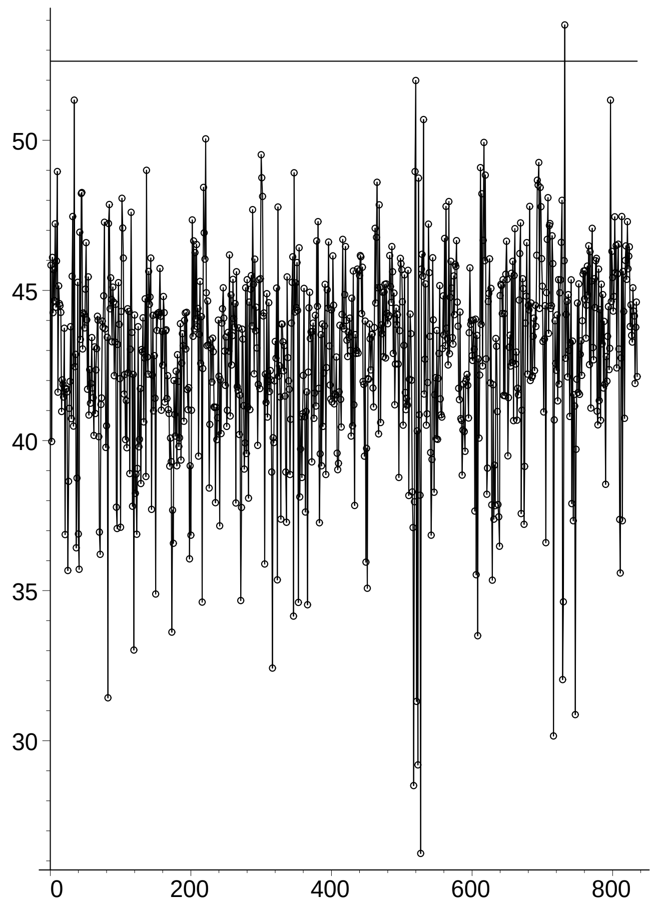

Figure 6.2. Control chart of the second sample of 835 measurements of the thickness of razor heads.

$-X_{n+1}$. An out-of-control signal is given when $-X_{n+1}$ is larger than the obtained upper control limit, or, equivalently, if X_{n+1} is smaller than minus the control limit based on $-X_1, \ldots, -X_n$. Here are the details.

1. We calculate $\frac{\overline{X}-X_{(1)}}{S}$. Since $X_{(1)} = 25.45, \overline{X} = 42.366, S = 3.311$, this leads to

$$\frac{\overline{X} - X_{(1)}}{S} = 5.109.$$

2. The cut-off points for choosing the normal chart are again

$$\overline{\Phi}^{-1}\left(\frac{-0.7 + 0.5\log n}{n}\right) = 2.728 \text{ and } \overline{\Phi}^{-1}\left(\frac{5}{n\sqrt{n}}\right) = 3.531.$$

3. Because $5.109 \notin [2.728, 3.531]$ we do not choose the normal chart. Next, we calculate the cut-off points for the parametric chart. We are dealing with the lower part and hence

$$\widehat{\gamma} = 1.1218\log\left(\frac{\overline{X} - X_{(n-[0.95n])}}{\overline{X} - X_{(n-[0.75n])}}\right) - 1.$$

   This gives $\widehat{\gamma} = 0.352$. The upper cut-off point for the parametric chart equals $\overline{K}_{\widehat{\gamma}}^{-1}\left(\frac{3}{n\sqrt{n}}\right) = 4.957$, implying that in this case the (modified) nonparametric chart is chosen. Hence, an out-of-control signal is given if a new observation is smaller than $X_{(1)} - (1-V)S$, which equals 25.450 if $V = 0$ and 22.139 if $V = 1$, where $P(V = 1) = (n+1)p = 0.836$.

   The minimum of the second sample is 26.25 and therefore, no out-of-control signal is obtained.

6.5 Theoretical behavior of the combined chart

In this section we study both the in-control and the out-of-control behavior of the combined control chart

$$X_{n+1} > UL_N^* \mathbf{1}\left(\frac{X_{(n)} - \overline{X}}{S} \in IN\right)$$

$$+ UL_P^* \mathbf{1}\left(\frac{X_{(n)} - \overline{X}}{S} \notin IN\right) \mathbf{1}\left(\frac{X_{(n)} - \overline{X}}{S} \in IP\right)$$

$$+ UL_{NP}^* \mathbf{1}\left(\frac{X_{(n)} - \overline{X}}{S} \notin IN\right) \mathbf{1}\left(\frac{X_{(n)} - \overline{X}}{S} \notin IP\right),$$

where $UL_N^* = UL_N$ if $r = 0$ and $UL_N^* = \overline{X} + u_p S$ if $r \geq 1$, $UL_P^* = UL_P$ if $r = 0$ and $UL_N^* = \overline{X} + \overline{K}_{\widehat{\gamma}}^{-1}(p)\, S$ if $r \geq 1$ and $UL_{NP}^* = UL_{MNP}$ if $r = 0$ and $UL_N^* = \delta X_{(n-r)} + (1-\delta) X_{(n-r+1)}$ if $r \geq 1$, see also (6.10) and (6.11). We show that the behavior of the combined control chart is asymptotically equivalent to the behavior of the specific control chart for the supposed model. Hence, the combined control chart is valid for all distributions and its out-of-control behavior is asymptotically as good as when we would know to which class of distributions the true distribution belongs.

The large deviation results of Theorem 3.6.1 and the following theorem are used in the proofs of the basic Lemmas 6.5.1, 6.5.4 and 6.5.5 in this section. Theorem 3.6.1 concerns a fixed ϵ, while the following theorem concerns the deviations ϵ_n tending to 0, but at a slow rate. Similarly as in Chapter 3, we use for mathematical convenience the estimator $\widehat{\gamma}_2$ instead of $\widehat{\gamma}$ and also the "exact" form of C_2, see the remarks on it just before Theorem 3.6.1.

Theorem 6.5.1 *Let $X_1, \ldots, X_n$ be i.i.d. r.v.'s with a normal power distribution with parameter γ. Let $\{\varepsilon_n\}$ be a sequence of positive numbers with $\lim_{n\to\infty} \varepsilon_n = 0$ and $\lim_{n\to\infty} n\varepsilon_n^2 = \infty$.*

(i) If $\gamma \leq 1$, then

$$\limsup_{n\to\infty} (n\varepsilon_n^2)^{-1} \log P\left(\left|\overline{X}\right| > \varepsilon_n\right) < 0. \tag{6.15}$$

If $\gamma > 1$, then

$$\limsup_{n\to\infty} \frac{\log P\left(\left|\overline{X}\right| > \varepsilon_n\right)}{\min\{n\varepsilon_n^2, (n\varepsilon_n)^{2/(1+\gamma)}\}} < 0. \tag{6.16}$$

(ii) If $\gamma \leq 0$, then

$$\limsup_{n\to\infty} (n\varepsilon_n^2)^{-1} \log P\left(\left|S^2 - 1\right| > \varepsilon_n\right) < 0. \tag{6.17}$$

If $\gamma > 0$, then

$$\limsup_{n\to\infty} \frac{\log P\left(\left|S^2 - 1\right| > \varepsilon_n\right)}{\min\{n\varepsilon_n^2, (n\varepsilon_n)^{1/(1+\gamma)}\}} < 0. \tag{6.18}$$

(iii) If $\gamma \leq 1$, then

$$\limsup_{n\to\infty} (n\varepsilon_n^2)^{-1} \log P\left(|\widehat{\gamma}_2 - \gamma| > \varepsilon_n\right) < 0. \tag{6.19}$$

If $\gamma > 1$, then

$$\limsup_{n\to\infty} \frac{\log P\left(|\widehat{\gamma}_2 - \gamma| > \varepsilon_n\right)}{\min\{n\varepsilon_n^2, (n\varepsilon_n)^{2/(1+\gamma)}\}} < 0. \tag{6.20}$$

Proof. Denote by $X_{(1)} \leq \ldots \leq X_{(n)}$ the order statistics of $X_1, \ldots, X_n$ and let $U_{(1)} \leq \ldots \leq U_{(n)}$ be the order statistics of the random sample $U_1, \ldots, U_n$ from a uniform distribution on (0,1). Let $\{\varepsilon_n\}$ be a sequence of positive numbers with $\lim_{n\to\infty} \varepsilon_n = 0$ and $\lim_{n\to\infty} n\varepsilon_n^2 = \infty$. Let $0 < s < 1$ be fixed and let $j = j(n)$ satisfy $j/n = s + O(\varepsilon_n)$ as $n \to \infty$. Then, cf.(3.43),

$$P(X_{(j)} > K_\gamma^{-1}(s) + \varepsilon_n) = P\left(\sum_{i=1}^{n} \mathbf{1}_{U_i > K_\gamma(K_\gamma^{-1}(s)+\varepsilon_n)} \geq n - j\right)$$

and by standard large deviation theory, cf. e.g. Feller(1971) p. 553,

$$\limsup_{n\to\infty} (n\varepsilon_n^2)^{-1} \log P(X_{(j)} > K_\gamma^{-1}(s) + \varepsilon_n) < 0.$$

Similarly, we have

$$\limsup_{n\to\infty} (n\varepsilon_n^2)^{-1} \log P(X_{(j)} < K_\gamma^{-1}(s) - \varepsilon_n) < 0$$

and hence

$$\limsup_{n\to\infty} (n\varepsilon_n^2)^{-1} \log P\left(\left|X_{(j)} - K_\gamma^{-1}(s)\right| > \varepsilon_n\right) < 0. \tag{6.21}$$

If $1 + \gamma \leq 2$, again by standard large deviation theory, cf. e.g. Feller(1971) p. 553, it follows that

$$\lim_{n\to\infty} (n\varepsilon_n^2)^{-1} \log P\left(\left|\overline{X}\right| > \varepsilon_n\right) < 0,$$

which completes the proof of (6.15).

For $1+\gamma > 2$, the moment generating function of X_i does not exist. However, by Nagaev (1969), see also Nagaev (1979, formulas (2.31) and (2.32) on page 764), it follows that

$$\limsup_{n\to\infty} \frac{\log P\left(\overline{X} > \varepsilon_n\right)}{\min\{n\varepsilon_n^2, (n\varepsilon_n)^{2/(1+\gamma)}\}} \leq -\min\left\{1/20, \frac{1}{2}\left(\frac{1/2}{c(\gamma)}\right)^{2/(1+\gamma)}\right\},$$

where we have used

$$\lim_{n\to\infty} \frac{\log n}{(n\varepsilon_n)^{2/(1+\gamma)}} = \lim_{n\to\infty} \frac{\log n}{n^{1/(1+\gamma)}(n^{1/2}\varepsilon_n)^{2/(1+\gamma)}} = 0.$$

By symmetry we get

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

which completes the proof of (6.16).

Since

$$\begin{aligned} P\left(\left|S^2-1\right| > \varepsilon_n\right) &\leq P\left(\left|S^2-1\right| > \varepsilon_n, \left|\overline{X}\right| < \varepsilon_n\right) + P\left(\left|\overline{X}\right| > \varepsilon_n\right) \\ &\leq P\left(\left|\sum_{i=1}^{n}\left(X_i^2-1\right)\right| > (n-1)\varepsilon_n - n\varepsilon_n^2 - 1\right) + P\left(\left|\overline{X}\right| > \varepsilon_n\right), \end{aligned}$$

it follows by a similar argument as in the proof of (6.15) and (6.16) that (6.17) and (6.18) hold true.

There exists a constant $c > 0$ such that

$$\begin{aligned} P\left(\left|\widehat{\gamma}_2 - \gamma\right| > \varepsilon_n\right) \leq\ & P\left(\left|X_{([0.95n+1])} - K_\gamma^{-1}(0.95)\right| > c\varepsilon_n\right) \\ &+ P\left(\left|X_{([0.75n+1])} - K_\gamma^{-1}(0.75)\right| > c\varepsilon_n\right) + P\left(\left|\overline{X}\right| > \varepsilon_n\right). \end{aligned}$$

Note that $[n+1-qn]/n - 1 - q = O(1/n) = o(\varepsilon_n)$ for any fixed $0 < q < 1$. The results given in (6.19) and (6.20) now easily follow from (6.21), (6.15) and (6.16). ■

6.5.1 Normality

Suppose that $X_1, \ldots, X_n$ have distribution function $F = \Phi$ and that X_{n+1} is distributed as $X_1 + \Delta$ for some $\Delta \geq 0$. The in-control situation refers to $\Delta = 0$, while for control charts with only an upper limit the out-of-control case corresponds to $\Delta > 0$. Let $\{d_{1N}(n)\}, \{d_{2N}(n)\}$ be sequences of real numbers satisfying $\lim_{n\to\infty} d_{1N}(n) = \infty, d_{1N}(n) < n$ and $\lim_{n\to\infty} d_{2N}(n) = 0$. In view of (6.14) we have (with expectations and probabilities referring to $F = \Phi$)

$$
\begin{aligned}
EP_n &= E\left\{\overline{\Phi}(UL_N^* - \Delta)\mathbf{1}\left(\frac{X_{(n)}-\overline{X}}{S} \in IN\right)\right\} \\
&+ E\left\{\overline{\Phi}(UL_P^* - \Delta)\mathbf{1}\left(\frac{X_{(n)}-\overline{X}}{S} \notin IN\right)\mathbf{1}\left(\frac{X_{(n)}-\overline{X}}{S} \in IP\right)\right\} \\
&+ E\left\{\overline{\Phi}(UL_{NP}^* - \Delta)\mathbf{1}\left(\frac{X_{(n)}-\overline{X}}{S} \notin IN\right)\mathbf{1}\left(\frac{X_{(n)}-\overline{X}}{S} \notin IP\right)\right\} \\
&= E\left\{\overline{\Phi}(UL_N^* - \Delta)\right\} - E\left\{\overline{\Phi}(UL_N^* - \Delta)\mathbf{1}\left(\frac{X_{(n)}-\overline{X}}{S} \notin IN\right)\right\} \\
&+ E\left\{\overline{\Phi}(UL_P^* - \Delta)\mathbf{1}\left(\frac{X_{(n)}-\overline{X}}{S} \notin IN\right)\mathbf{1}\left(\frac{X_{(n)}-\overline{X}}{S} \in IP\right)\right\} \\
&+ E\left\{\overline{\Phi}(UL_{NP}^* - \Delta)\mathbf{1}\left(\frac{X_{(n)}-\overline{X}}{S} \notin IN\right)\mathbf{1}\left(\frac{X_{(n)}-\overline{X}}{S} \notin IP\right)\right\} \\
&= E\left\{\overline{\Phi}(UL_N^* - \Delta)\right\} + E\left\{\mathbf{1}\left(\frac{X_{(n)}-\overline{X}}{S} \notin IN\right) h_N\left(X_{(n)}, \overline{X}, S\right)\right\}
\end{aligned}
$$

with

$$
\begin{aligned}
h_N\left(X_{(n)}, \overline{X}, S\right) &= \left[\overline{\Phi}(UL_P^* - \Delta) - \overline{\Phi}(UL_N^* - \Delta)\right]\mathbf{1}\left(\frac{X_{(n)}-\overline{X}}{S} \in IP\right) \\
&\quad + \left[\overline{\Phi}(UL_{NP}^* - \Delta) - \overline{\Phi}(UL_N^* - \Delta)\right]\mathbf{1}\left(\frac{X_{(n)}-\overline{X}}{S} \notin IP\right)
\end{aligned}
$$

and hence, using $\left|h_N\left(X_{(n)}, \overline{X}, S\right)\right| \leq 1$,

$$
\left|EP_n - E\left\{\overline{\Phi}(UL_N^* - \Delta)\right\}\right| \leq P\left(\frac{X_{(n)}-\overline{X}}{S} \notin IN\right). \tag{6.22}
$$

The following lemma gives asymptotic expressions for the probabilities of the right-hand side of (6.22).

Lemma 6.5.1 *Let $X_1, \ldots, X_n$ be i.i.d. r.v's with a normal distribution. Let $\{d_{1N}(n)\}, \{d_{2N}(n)\}$ be sequences of real numbers satisfying*

$$
\begin{aligned}
&\lim_{n\to\infty} d_{1N}(n) = \infty, \lim_{n\to\infty} d_{1N}(n)\sqrt{\frac{d_{1N}(n)}{n}}\log\left(\frac{d_{1N}(n)}{n}\right) = 0, \\
&\lim_{n\to\infty} d_{2N}(n) = 0, \lim_{n\to\infty} \sqrt{\frac{|\log d_{2N}(n)|}{n}}\log\left(\frac{d_{2N}(n)}{n}\right) = 0.
\end{aligned}
$$

Then

$$P\left(\frac{X_{(n)}-\overline{X}}{S}<\overline{\Phi}^{-1}\left(\frac{d_{1N}(n)}{n}\right)\right)=\left(1-\frac{d_{1N}(n)}{n}\right)^{n}(1+o(1)),$$
$$P\left(\frac{X_{(n)}-\overline{X}}{S}>\overline{\Phi}^{-1}\left(\frac{d_{2N}(n)}{n}\right)\right)=d_{2N}(n)(1+o(1)) \ \textit{as } n\to\infty.$$

Before proving Lemma 6.5.1 we present the following lemma on the behavior in the tail of the standard normal distribution.

Lemma 6.5.2 *Let* $\{x_n\},\{\varepsilon_n\}$ *be sequences of real numbers satisfying* $\lim_{n\to\infty} x_n = \infty$, $\lim_{n\to\infty} x_n^2\varepsilon_n = 0$, *where* ε_n *may be positive as well as negative. Then*

$$\overline{\Phi}\left(x_n\left(1+\varepsilon_n\right)\right)=\overline{\Phi}\left(x_n\right)\left(1+O\left(x_n^2\varepsilon_n\right)\right) \ \textit{as } n\to\infty.$$

If moreover, $\lim_{n\to\infty} n\overline{\Phi}\left(x_n\right)x_n^2\varepsilon_n=0$, *then*

$$\left\{1-\overline{\Phi}\left(x_n\left(1+\varepsilon_n\right)\right)\right\}^n=\left(1-\overline{\Phi}\left(x_n\right)\right)^n(1+o(1)) \ \textit{as } n\to\infty.$$

Proof. By Taylor expansion we get

$$\overline{\Phi}\left(x_n\left(1+\varepsilon_n\right)\right)=\overline{\Phi}\left(x_n\right)-x_n\varepsilon_n\varphi\left(\xi_n\right),$$

with ξ_n between x_n and $x_n(1+\varepsilon_n)$. For $x\to\infty$ we have

$$\overline{\Phi}(x)=\frac{\varphi(x)}{x}(1+o(1))$$

and hence, noting that $\lim_{n\to\infty} x_n^2\varepsilon_n=0$ implies $\lim_{n\to\infty}\frac{\varphi(\xi_n)}{\varphi(x_n)}=1$, we obtain

$$\begin{aligned}\frac{\overline{\Phi}\left(x_n\left(1+\varepsilon_n\right)\right)}{\overline{\Phi}\left(x_n\right)}&=1-x_n^2\varepsilon_n\frac{\varphi\left(\xi_n\right)}{\varphi\left(x_n\right)}(1+o(1))\\&=1+O\left(x_n^2\varepsilon_n\right) \text{ as } n\to\infty.\end{aligned}$$

If, moreover, $\lim_{n\to\infty} n\overline{\Phi}\left(x_n\right)x_n^2\varepsilon_n=0$, then

$$\begin{aligned}n\log\left(1-\overline{\Phi}\left(x_n\left(1+\varepsilon_n\right)\right)\right)&=n\log\left(1-\overline{\Phi}\left(x_n\right)+O\left(\overline{\Phi}\left(x_n\right)x_n^2\varepsilon_n\right)\right)\\&=n\log\left(1-\overline{\Phi}\left(x_n\right)\right)+O\left(n\overline{\Phi}\left(x_n\right)x_n^2\varepsilon_n\right)\\&=n\log\left(1-\overline{\Phi}\left(x_n\right)\right)+o(1),\end{aligned}$$

which completes the proof. ■

Proof of Lemma 6.5.1. For $x \to 0$ we have

$$\overline{\Phi}^{-1}(x) = \sqrt{-2\log x}\,(1 + o(1))$$

and hence

$$\lim_{n\to\infty} d_{1N}(n)\sqrt{\frac{d_{1N}(n)}{n}}\log\left(\frac{d_{1N}(n)}{n}\right) = 0$$

implies

$$\lim_{n\to\infty} d_{1N}(n)\sqrt{\frac{d_{1N}(n)}{n}}\left\{\overline{\Phi}^{-1}\left(\frac{d_{1N}(n)}{n}\right)\right\}^2 = 0.$$

Therefore, there exists a sequence $\{a_n\}$ such that

$$\lim_{n\to\infty} a_n = \infty, \lim_{n\to\infty} a_n d_{1N}(n)\sqrt{\frac{d_{1N}(n)}{n}}\left\{\overline{\Phi}^{-1}\left(\frac{d_{1N}(n)}{n}\right)\right\}^2 = 0. \qquad (6.23)$$

Let

$$\varepsilon_n = a_n\sqrt{\frac{d_{1N}(n)}{n}},$$

then we have $\lim_{n\to\infty} \varepsilon_n = 0$ and $\lim_{n\to\infty} n\varepsilon_n^2 \left(d_{1N}(n)\right)^{-1} = \infty$. By Theorem 6.5.1 we get

$$\lim_{n\to\infty} \frac{P\left(\left|\overline{X}\right| > \varepsilon_n\right)}{\left(1 - \frac{d_{1N}(n)}{n}\right)^n} = 0, \lim_{n\to\infty} \frac{P\left(\left|S^2 - 1\right| > \varepsilon_n\right)}{\left(1 - \frac{d_{1N}(n)}{n}\right)^n} = 0$$

and hence

$$P\left(\frac{X_{(n)} - \overline{X}}{S} < \overline{\Phi}^{-1}\left(\frac{d_{1N}(n)}{n}\right)\right) \qquad (6.24)$$

$$= P\left(\frac{X_{(n)} - \overline{X}}{S} < \overline{\Phi}^{-1}\left(\frac{d_{1N}(n)}{n}\right), \left|S^2 - 1\right| < \varepsilon_n, \left|\overline{X}\right| < \varepsilon_n\right) + o\left(\left(1 - \frac{d_{1N}(n)}{n}\right)^n\right).$$

Writing

$$z_{1n} = \overline{\Phi}^{-1}\left(\frac{d_{1N}(n)}{n}\right)(1 - \varepsilon_n), z_{2n} = \overline{\Phi}^{-1}\left(\frac{d_{1N}(n)}{n}\right)(1 + \varepsilon_n)$$

we obtain, for sufficiently large n,

$$P\left(X_{(n)} < z_{1n}, \left|S^2 - 1\right| < \varepsilon_n, \left|\overline{X}\right| < \varepsilon_n\right) \qquad (6.25)$$

$$\leq P\left(\frac{X_{(n)} - \overline{X}}{S} < \overline{\Phi}^{-1}\left(\frac{d_{1N}(n)}{n}\right), \left|S^2 - 1\right| < \varepsilon_n, \left|\overline{X}\right| < \varepsilon_n\right) \leq P\left(X_{(n)} < z_{2n}\right).$$

Using

$$P\left(X_{(n)} < z_{2n}\right) = P\left(\overline{\Phi}\left(X_{(n)}\right) > \overline{\Phi}\left(z_{2n}\right)\right) = \left(1 - \overline{\Phi}\left(z_{2n}\right)\right)^n$$

and writing $x_n = \overline{\Phi}^{-1}\left(\frac{d_{1N}(n)}{n}\right)$, the conditions of Lemma 6.5.2 are fulfilled, see also (6.23). Application of Lemma 6.5.2 yields

$$P\left(X_{(n)} < z_{2n}\right) = \left(1 - \overline{\Phi}\left(z_{2n}\right)\right)^n = \left(1 - \frac{d_{1N}(n)}{n}\right)^n (1 + o(1)) \text{ as } n \to \infty. \quad (6.26)$$

Because

$$P\left(X_{(n)} < z_{1n}, \left|S^2 - 1\right| < \varepsilon_n, \left|\overline{X}\right| < \varepsilon_n\right) = P\left(X_{(n)} < z_{1n}\right) + o\left(\left(1 - \frac{d_{1N}(n)}{n}\right)^n\right),$$

another application of Lemma 6.5.2 gives

$$P\left(X_{(n)} < z_{1n}, \left|S^2 - 1\right| < \varepsilon_n, \left|\overline{X}\right| < \varepsilon_n\right) = \left(1 - \frac{d_{1N}(n)}{n}\right)^n (1 + o(1)) \text{ as } n \to \infty. \quad (6.27)$$

Combination of (6.24) – (6.27) leads to

$$P\left(\frac{X_{(n)} - \overline{X}}{S} < \overline{\Phi}^{-1}\left(\frac{d_{1N}(n)}{n}\right)\right) = \left(1 - \frac{d_{1N}(n)}{n}\right)^n (1 + o(1)) \text{ as } n \to \infty.$$

Since

$$\lim_{n\to\infty} \sqrt{\frac{|\log d_{2N}(n)|}{n}} \log\left(\frac{d_{2N}(n)}{n}\right) = 0,$$

there exists a sequence $\{b_n\}$ such that

$$\lim_{n\to\infty} b_n = \infty, \lim_{n\to\infty} b_n \sqrt{\frac{|\log d_{2N}(n)|}{n}} \left\{\overline{\Phi}^{-1}\left(\frac{d_{2N}(n)}{n}\right)\right\}^2 = 0. \quad (6.28)$$

Let

$$\eta_n = b_n \sqrt{\frac{|\log d_{2N}(n)|}{n}},$$

then we have $\lim_{n\to\infty} \eta_n = 0$ and $\lim_{n\to\infty} n\eta_n^2 \left(|\log d_{2N}(n)|\right)^{-1} = \infty$. By Theorem 6.5.1 we get

$$\lim_{n\to\infty} \frac{P\left(\left|\overline{X}\right| > \eta_n\right)}{d_{2N}(n)} = 0, \lim_{n\to\infty} \frac{P\left(\left|S^2 - 1\right| > \eta_n\right)}{d_{2N}(n)} = 0$$

and hence

$$P\left(\frac{X_{(n)} - \overline{X}}{S} > \overline{\Phi}^{-1}\left(\frac{d_{2N}(n)}{n}\right)\right) \quad (6.29)$$

$$= P\left(\frac{X_{(n)} - \overline{X}}{S} > \overline{\Phi}^{-1}\left(\frac{d_{2N}(n)}{n}\right), \left|S^2 - 1\right| < \eta_n, \left|\overline{X}\right| < \eta_n\right) + o\left(d_{2N}(n)\right).$$

Writing

$$y_{1n} = \overline{\Phi}^{-1}\left(\frac{d_{2N}(n)}{n}\right)(1+\eta_n), y_{2n} = \overline{\Phi}^{-1}\left(\frac{d_{2N}(n)}{n}\right)(1-\eta_n)$$

we obtain, for sufficiently large n,

$$P\left(X_{(n)} > y_{1n}, \left|S^2 - 1\right| < \varepsilon_n, \left|\overline{X}\right| < \varepsilon_n\right) \tag{6.30}$$
$$\leq P\left(\frac{X_{(n)} - \overline{X}}{S} > \overline{\Phi}^{-1}\left(\frac{d_{2N}(n)}{n}\right), \left|S^2 - 1\right| < \eta_n, \left|\overline{X}\right| < \eta_n\right) \leq P\left(X_{(n)} > y_{2n}\right).$$

Using

$$P\left(X_{(n)} > y_{2n}\right) = P\left(\overline{\Phi}\left(X_{(n)}\right) < \overline{\Phi}(y_{2n})\right) = 1 - \left(1 - \overline{\Phi}(y_{2n})\right)^n$$

and writing $x_n = \overline{\Phi}^{-1}\left(\frac{d_{2N}(n)}{n}\right)$, the conditions of the first part of Lemma 6.5.2 (with ε_n replaced by $-\eta_n$) are fulfilled, see also (6.28). Application of Lemma 6.5.2 yields

$$P\left(X_{(n)} > y_{2n}\right) = 1 - \left(1 - \overline{\Phi}(y_{2n})\right)^n = 1 - \left(1 - \frac{d_{2N}(n)(1+o(1))}{n}\right)^n \tag{6.31}$$
$$= d_{2N}(n)(1+o(1)),$$

as $n \to \infty$. Because

$$P\left(X_{(n)} > y_{1n}, \left|S^2 - 1\right| < \eta_n, \left|\overline{X}\right| < \eta_n\right) = P\left(X_{(n)} > y_{1n}\right) + o\left(d_{2N}(n)\right),$$

another application of Lemma 6.5.2 gives

$$P\left(X_{(n)} > y_{1n}, \left|S^2 - 1\right| < \eta_n, \left|\overline{X}\right| < \eta_n\right) = d_{2N}(n)(1+o(1)) \text{ as } n \to \infty. \tag{6.32}$$

Combination of (6.29) – (6.32) gives

$$P\left(\frac{X_{(n)} - \overline{X}}{S} > \overline{\Phi}^{-1}\left(\frac{d_{2N}(n)}{n}\right)\right) = d_{2N}(n)(1+o(1)) \text{ as } n \to \infty,$$

thus completing the proof of the lemma. ■

In view of (6.22) and Lemma 6.5.1 we get the following result, which expresses that under normality the combined control chart given by (6.12) and (6.13) is asymptotically as good as the normal control chart given in (6.1). The results of Chapter 2 on (corrected) normal control charts ensure that $E\overline{\Phi}(UL_N^*)$ is very close to p.

Theorem 6.5.2 *Let $X_1, \ldots, X_n$ be i.i.d. r.v's with a normal distribution. Let $\{d_{1N}(n)\}, \{d_{2N}(n)\}$ be sequences of real numbers satisfying*

$$\lim_{n\to\infty} d_{1N}(n) = \infty, \lim_{n\to\infty} d_{1N}(n)\sqrt{\frac{d_{1N}(n)}{n}} \log\left(\frac{d_{1N}(n)}{n}\right) = 0,$$
$$\lim_{n\to\infty} d_{2N}(n) = 0, \lim_{n\to\infty} \sqrt{\frac{|\log d_{2N}(n)|}{n}} \log\left(\frac{d_{2N}(n)}{n}\right) = 0.$$

Then

$$\left|EP_n - E\left\{\overline{\Phi}(UL_N^* - \Delta)\right\}\right| \le \left\{\left(1 - \frac{d_{1N}(n)}{n}\right)^n + d_{2N}(n)\right\}(1 + o(1)) \text{ as } n \to \infty.$$

Note that for the choices $d_{1N}(n) = -0.7 + 0.5\log n, d_{2N}(n) = 5/\sqrt{n}$ the conditions of Theorem 6.5.2 are satisfied and we get

$$|EP_n - E\{\overline{\Phi}(UL_N^* - \Delta)\}| \le (\frac{e^{0.7} + 5}{\sqrt{n}})(1 + o(1)),$$

as $n \to \infty$.

6.5.2 Normal power family

The following lemma describes the tail behavior in the normal power family and is in fact a generalization of Lemma 6.5.2.

Lemma 6.5.3 *Let $\{x_n\}, \{\varepsilon_n\}$ be sequences of real numbers satisfying $\lim_{n\to\infty} x_n = \infty$, $\lim_{n\to\infty} x_n^{\frac{2}{1+\gamma}}\varepsilon_n = 0$, where ε_n may be positive as well as negative, then*

$$\overline{K}_\gamma(x_n(1+\varepsilon_n)) = \overline{K}_\gamma(x_n)\left(1 + O\left(x_n^{\frac{2}{1+\gamma}}\varepsilon_n\right)\right) \text{ as } n \to \infty.$$

If, moreover, $\lim_{n\to\infty} n\overline{K}_\gamma(x_n) x_n^{\frac{2}{1+\gamma}}\varepsilon_n = 0$, then

$$\left\{1 - \overline{K}_\gamma(x_n(1+\varepsilon_n))\right\}^n = \left(1 - \overline{K}_\gamma(x_n)\right)^n (1 + o(1)) \text{ as } n \to \infty.$$

Proof. In view of (the proof of) Lemma 6.5.2 we have

$$\overline{K}_\gamma(x_n(1+\varepsilon_n)) = \overline{\Phi}\left(\left(\frac{x_n(1+\varepsilon_n)}{c(\gamma)}\right)^{\frac{1}{1+\gamma}}\right) = \overline{\Phi}\left(\left(\frac{x_n}{c(\gamma)}\right)^{\frac{1}{1+\gamma}}(1+\varepsilon_n)^{\frac{1}{1+\gamma}}\right)$$
$$= \overline{\Phi}\left(\left(\frac{x_n}{c(\gamma)}\right)^{\frac{1}{1+\gamma}}\right)\left(1 + O\left(x_n^{\frac{2}{1+\gamma}}\varepsilon_n\right)\right) = \overline{K}_\gamma(x_n)\left(1 + O\left(x_n^{\frac{2}{1+\gamma}}\varepsilon_n\right)\right),$$

as $n \to \infty$, which gives the first part of the lemma. The proof of the second part is similar to the proof of the second part of Lemma 6.5.2. ■

Note that for $\gamma = 0$, that is the normal distribution, indeed Lemma 6.5.3 gives the result of Lemma 6.5.2.

Suppose that $X_1, \ldots, X_n$ have distribution function $F = K_\gamma$ and that X_{n+1} is distributed as $X_1 + \Delta$ for some $\Delta \geq 0$. If $\gamma = 0$, we are in the previous situation. Therefore, assume that $\gamma \neq 0$. Let $\{d_{1P}(n)\}, \{d_{2P}(n)\}$ be sequences of real numbers satisfying $\lim_{n\to\infty} d_{1P}(n) = \infty, d_{1P}(n) < n$ and $\lim_{n\to\infty} d_{2P}(n) = 0$. In view of (6.14) we have (with expectations and probabilities referring to $F = K_\gamma$)

$$\begin{aligned} EP_n &= E\left\{\overline{K}_\gamma(UL_N^* - \Delta)\mathbf{1}\left(\frac{X_{(n)} - \overline{X}}{S} \in IN\right)\right\} \\ &+ E\left\{\overline{K}_\gamma(UL_P^* - \Delta)\mathbf{1}\left(\frac{X_{(n)} - \overline{X}}{S} \notin IN\right)\mathbf{1}\left(\frac{X_{(n)} - \overline{X}}{S} \in IP\right)\right\} \\ &+ E\left\{\overline{K}_\gamma(UL_{NP}^* - \Delta)\mathbf{1}\left(\frac{X_{(n)} - \overline{X}}{S} \notin IN\right)\mathbf{1}\left(\frac{X_{(n)} - \overline{X}}{S} \notin IP\right)\right\} \\ &= E\left\{\overline{K}_\gamma(UL_P^* - \Delta)\right\} + Eh_P\left(X_{(n)}, \overline{X}, S\right), \end{aligned}$$

with

$$\begin{aligned} h_P\left(X_{(n)}, \overline{X}, S\right) &= \left[\overline{K}_\gamma(UL_N^* - \Delta) - \overline{K}_\gamma(UL_P^* - \Delta)\right]\mathbf{1}\left(\frac{X_{(n)} - \overline{X}}{S} \in IN\right) \\ + \left[\overline{K}_\gamma(UL_{NP}^* - \Delta) - \overline{K}_\gamma(UL_P^* - \Delta)\right]&\mathbf{1}\left(\frac{X_{(n)} - \overline{X}}{S} \notin IP\right)\mathbf{1}\left(\frac{X_{(n)} - \overline{X}}{S} \notin IN\right) \end{aligned}$$

and hence

$$\left|EP_n - E\left\{\overline{K}_\gamma(UL_P^* - \Delta)\right\}\right| \leq P\left(\frac{X_{(n)} - \overline{X}}{S} \in IN\right) + P\left(\frac{X_{(n)} - \overline{X}}{S} \notin IP\right). \tag{6.33}$$

The next two lemmas give the ingredients for the asymptotic expressions of the right-hand side of (6.33).

Lemma 6.5.4 *Let $X_1, \ldots, X_n$ be i.i.d. r.v's with a distribution from the normal power family. Let $\{d_{1P}(n)\}, \{d_{2P}(n)\}$ be sequences of real numbers satisfying for* *some* $\zeta > 0$

$$\lim_{n\to\infty} d_{1P}(n) = \infty,$$

$$\begin{cases} \lim\limits_{n\to\infty} d_{1P}(n)\sqrt{\frac{d_{1P}(n)}{n}}\left|\log\left(\frac{d_{1P}(n)}{n}\right)\right|^{1+\zeta} = 0, \ \textit{if}\ \gamma \le 0 \\ \lim\limits_{n\to\infty} d_{1P}(n)\max\left(\frac{d_{1P}^{1+\gamma}(n)}{n}, \sqrt{\frac{d_{1P}(n)}{n}}\right)\left|\log\left(\frac{d_{1P}(n)}{n}\right)\right|^{1+\zeta} = 0, \ \textit{if}\ \gamma > 0 \end{cases},$$

$$\lim_{n\to\infty} d_{2P}(n) = 0,$$

$$\begin{cases} \lim\limits_{n\to\infty} \sqrt{\frac{|\log d_{2P}(n)|}{n}}\left|\log\left(\frac{d_{2P}(n)}{n}\right)\right|^{1+\zeta} = 0, \ \textit{if}\ \gamma \le 0 \\ \lim\limits_{n\to\infty} \max\left(\frac{|\log d_{2P}(n)|^{1+\gamma}}{n}, \sqrt{\frac{|\log d_{2P}(n)|}{n}}\right)\left|\log\left(\frac{d_{2P}(n)}{n}\right)\right|^{1+\zeta} = 0, \ \textit{if}\ \gamma > 0 \end{cases}.$$

Then

$$P\left(\frac{X_{(n)} - \overline{X}}{S} < \overline{K_{\widehat{\gamma_2}}}^{-1}\left(\frac{d_{1P}(n)}{n}\right)\right) = \left(1 - \frac{d_{1P}(n)}{n}\right)^n (1 + o(1)),$$

$$P\left(\frac{X_{(n)} - \overline{X}}{S} > \overline{K_{\widehat{\gamma_2}}}^{-1}\left(\frac{d_{2P}(n)}{n}\right)\right) = d_{2P}(n)(1 + o(1)) \ \textit{as}\ n \to \infty.$$

Proof. We have

$$\begin{cases} \lim\limits_{n\to\infty} d_{1P}(n)\sqrt{\frac{d_{1P}(n)}{n}}\left|\log\left(\frac{d_{1P}(n)}{n}\right)\right|^{1+\zeta} = 0, \ \text{if}\ \gamma \le 0 \\ \lim\limits_{n\to\infty} d_{1P}(n)\max\left(\frac{d_{1P}^{1+\gamma}(n)}{n}, \sqrt{\frac{d_{1P}(n)}{n}}\right)\left|\log\left(\frac{d_{1P}(n)}{n}\right)\right|^{1+\zeta} = 0, \ \text{if}\ \gamma > 0 \end{cases}$$

and therefore there exists a sequence $\{a_n\}$ such that

$$\lim_{n\to\infty} a_n = \infty, \tag{6.34}$$

$$\begin{cases} \lim\limits_{n\to\infty} a_n d_{1P}(n)\sqrt{\frac{d_{1P}(n)}{n}}\left|\log\left(\frac{d_{1P}(n)}{n}\right)\right|^{1+\zeta} = 0, \ \ \text{if}\ \gamma \le 0 \\ \lim\limits_{n\to\infty} a_n d_{1P}(n)\max\left(\frac{d_{1P}^{1+\gamma}(n)}{n}, \sqrt{\frac{d_{1P}(n)}{n}}\right)\left|\log\left(\frac{d_{1P}(n)}{n}\right)\right|^{1+\zeta} = 0, \ \text{if}\ \gamma > 0. \end{cases}$$

Let

$$\varepsilon_n = \begin{cases} a_n\sqrt{\frac{d_{1P}(n)}{n}}, \ \ \text{if}\ \gamma \le 0 \\ a_n \max\left(\frac{d_{1P}^{1+\gamma}(n)}{n}, \sqrt{\frac{d_{1P}(n)}{n}}\right), \ \text{if}\ \gamma > 0 \end{cases}, \tag{6.35}$$

then we have $\lim_{n\to\infty} \varepsilon_n = 0$ and by Theorem 6.5.1

$$\lim_{n\to\infty} \frac{P\left(\left|\overline{X}\right| > \varepsilon_n\right)}{\left(1-\frac{d_{1P}(n)}{n}\right)^n} = 0, \lim_{n\to\infty} \frac{P\left(\left|S^2-1\right| > \varepsilon_n\right)}{\left(1-\frac{d_{1P}(n)}{n}\right)^n} = 0, \lim_{n\to\infty} \frac{P\left(\left|\widehat{\gamma}_2-\gamma\right| > \varepsilon_n\right)}{\left(1-\frac{d_{1P}(n)}{n}\right)^n} = 0$$

and hence

$$\begin{aligned} &P\left(\frac{X_{(n)}-\overline{X}}{S} < \overline{K_{\widehat{\gamma_2}}}^{-1}\left(\frac{d_{1P}(n)}{n}\right)\right) \quad (6.36)\\ &= P\left(\frac{X_{(n)}-\overline{X}}{S} < \overline{K_{\widehat{\gamma_2}}}^{-1}\left(\frac{d_{1P}(n)}{n}\right), \left|S^2-1\right| < \varepsilon_n, \left|\overline{X}\right| < \varepsilon_n, \left|\widehat{\gamma}_2-\gamma\right| < \varepsilon_n\right)\\ &+ o\left(\left(1-\frac{d_{1P}(n)}{n}\right)^n\right). \end{aligned}$$

Writing

$$\begin{aligned} z_{1n} &= \overline{K}_\gamma^{-1}\left(\frac{d_{1P}(n)}{n}\right)\left(1-\varepsilon_n\overline{\Phi}^{-1}\left(\frac{d_{1P}(n)}{n}\right)^{2\zeta}\right)\\ z_{2n} &= \overline{K}_\gamma^{-1}\left(\frac{d_{1P}(n)}{n}\right)\left(1+\varepsilon_n\overline{\Phi}^{-1}\left(\frac{d_{1P}(n)}{n}\right)^{2\zeta}\right), \end{aligned}$$

we obtain, for sufficiently large n,

$$\begin{aligned} &P\left(X_{(n)} < z_{1n}, \left|S^2-1\right| < \varepsilon_n, \left|\overline{X}\right| < \varepsilon_n, \left|\widehat{\gamma}_2-\gamma\right| < \varepsilon_n\right) \quad (6.37)\\ &\le P\left(\frac{X_{(n)}-\overline{X}}{S} < \overline{K_{\widehat{\gamma_2}}}^{-1}\left(\frac{d_{1P}(n)}{n}\right), \left|S^2-1\right| < \varepsilon_n, \left|\overline{X}\right| < \varepsilon_n, \left|\widehat{\gamma}_2-\gamma\right| < \varepsilon_n\right)\\ &\le P\left(X_{(n)} < z_{2n}\right). \end{aligned}$$

Using

$$P\left(X_{(n)} < z_{2n}\right) = P\left(\overline{K}_\gamma\left(X_{(n)}\right) > \overline{K}_\gamma\left(z_{2n}\right)\right) = \left(1-\overline{K}_\gamma\left(z_{2n}\right)\right)^n$$

and writing $x_n = \overline{K}_\gamma^{-1}\left(\frac{d_{1P}(n)}{n}\right)$, the conditions of Lemma 6.5.3 (with ε_n replaced by $\varepsilon_n\overline{\Phi}^{-1}\left(\frac{d_{1P}(n)}{n}\right)^{2\zeta}$) are fulfilled, see also (6.34) and (6.35). Application of Lemma 6.5.3 yields

$$\begin{aligned} P\left(X_{(n)} < z_{2n}\right) &= \left(1-\overline{K}_\gamma\left(z_{2n}\right)\right)^n \quad (6.38)\\ &= \left(1-\frac{d_{1P}(n)}{n}\right)^n(1+o(1)), \text{ as } n\to\infty. \end{aligned}$$

Because

$$P\left(X_{(n)} < z_{1n}, \left|S^2-1\right| < \varepsilon_n, \left|\overline{X}\right| < \varepsilon_n, |\widehat{\gamma_2}-\gamma| < \varepsilon_n\right) = P\left(X_{(n)} < z_{1n}\right) + o\left(\left(1-\frac{d_{1P}(n)}{n}\right)^n\right),$$

another application of Lemma 6.5.3 gives

$$P\left(X_{(n)} < z_{1n}, \left|S^2-1\right| < \varepsilon_n, \left|\overline{X}\right| < \varepsilon_n\right) = \left(1-\frac{d_{1P}(n)}{n}\right)^n (1+o(1)), \text{ as } n\to\infty. \tag{6.39}$$

Combination of (6.36) – (6.39) leads to

$$P\left(\frac{X_{(n)}-\overline{X}}{S} < \overline{K_{\widehat{\gamma_2}}}^{-1}\left(\frac{d_{1P}(n)}{n}\right)\right) = \left(1-\frac{d_{1P}(n)}{n}\right)^n (1+o(1)), \text{ as } n\to\infty.$$

Since

$$\begin{cases} \lim_{n\to\infty} \sqrt{\frac{|\log d_{2P}(n)|}{n}} \left|\log\left(\frac{d_{2P}(n)}{n}\right)\right|^{1+\zeta} = 0, \text{ if } \gamma\le 0 \\ \lim_{n\to\infty} \max\left(\frac{|\log d_{2P}(n)|^{1+\gamma}}{n}, \sqrt{\frac{|\log d_{2P}(n)|}{n}}\right) \left|\log\left(\frac{d_{2P}(n)}{n}\right)\right|^{1+\zeta} = 0, \text{ if } \gamma > 0, \end{cases}$$

there exists a sequence $\{b_n\}$ such that $\lim_{n\to\infty} b_n = \infty$,

$$\begin{cases} \lim_{n\to\infty} b_n\sqrt{\frac{|\log d_{2P}(n)|}{n}} \left|\log\left(\frac{d_{2P}(n)}{n}\right)\right|^{1+\zeta} = 0, \text{ if } \gamma\le 0 \\ \lim_{n\to\infty} b_n \max\left(\frac{|\log d_{2P}(n)|^{1+\gamma}}{n}, \sqrt{\frac{|\log d_{2P}(n)|}{n}}\right) \left|\log\left(\frac{d_{2P}(n)}{n}\right)\right|^{1+\zeta} = 0, \text{ if } \gamma > 0. \end{cases} \tag{6.40}$$

Let

$$\eta_n = \begin{cases} b_n\sqrt{\frac{|\log d_{2P}(n)|}{n}}, \text{ if } \gamma\le 0 \\ b_n \max\left(\frac{|\log d_{2P}(n)|^{1+\gamma}}{n}, \sqrt{\frac{|\log d_{2P}(n)|}{n}}\right), \text{ if } \gamma > 0 \end{cases}, \tag{6.41}$$

then we have $\lim_{n\to\infty} \eta_n = 0$ and by Theorem 6.5.1 we get

$$\lim_{n\to\infty} \frac{P\left(\left|\overline{X}\right| > \eta_n\right)}{d_{2P}(n)} = 0, \lim_{n\to\infty} \frac{P\left(\left|S^2-1\right| > \eta_n\right)}{d_{2P}(n)} = 0, \lim_{n\to\infty} \frac{P\left(|\widehat{\gamma_2}-\gamma| > \eta_n\right)}{d_{2P}(n)} = 0$$

and hence

$$P\left(\frac{X_{(n)}-\overline{X}}{S} > \overline{K_{\widehat{\gamma_2}}}^{-1}\left(\frac{d_{2P}(n)}{n}\right)\right) \tag{6.42}$$

$$= P\left(\frac{X_{(n)}-\overline{X}}{S} > \overline{K_{\widehat{\gamma_2}}}^{-1}\left(\frac{d_{2P}(n)}{n}\right), \left|S^2-1\right| < \eta_n, \left|\overline{X}\right| < \eta_n, |\widehat{\gamma_2}-\gamma| < \eta_n\right)$$

$$+ o\left(d_{2P}(n)\right).$$

Writing

$$y_{1n} = \overline{K}_\gamma^{-1}\left(\frac{d_{2P}(n)}{n}\right)\left(1-\eta_n\overline{\Phi}^{-1}\left(\frac{d_{2P}(n)}{n}\right)^{2\zeta}\right),$$

$$y_{2n} = \overline{K}_\gamma^{-1}\left(\frac{d_{2P}(n)}{n}\right)\left(1+\eta_n\overline{\Phi}^{-1}\left(\frac{d_{2P}(n)}{n}\right)^{2\zeta}\right)$$

we obtain, for sufficiently large n,

$$P\left(X_{(n)} > y_{1n}, \left|S^2-1\right| < \eta_n, \left|\overline{X}\right| < \eta_n, |\widehat{\gamma_2}-\gamma| < \eta_n\right) \tag{6.43}$$

$$\leq P\left(\frac{X_{(n)}-\overline{X}}{S} > \overline{K_{\widehat{\gamma_2}}}^{-1}\left(\frac{d_{2P}(n)}{n}\right), \left|S^2-1\right| < \eta_n, \left|\overline{X}\right| < \eta_n, |\widehat{\gamma_2}-\gamma| < \eta_n\right)$$

$$\leq P\left(X_{(n)} > y_{2n}\right).$$

Using

$$P\left(X_{(n)} > y_{2n}\right) = P\left(\overline{K}_\gamma\left(X_{(n)}\right) < \overline{K}_\gamma\left(y_{2n}\right)\right) = 1-\left(1-\overline{K}_\gamma\left(y_{2n}\right)\right)^n$$

and writing $x_n = \overline{K}_\gamma^{-1}\left(\frac{d_{2P}(n)}{n}\right)$, the conditions of the first part of Lemma 6.5.3 (with ε_n replaced by $-\eta_n\overline{\Phi}^{-1}\left(\frac{d_{2P}(n)}{n}\right)^{2\zeta}$) are fulfilled, see also (6.40) and (6.41). Application of Lemma 6.5.3 yields

$$P\left(X_{(n)} > y_{2n}\right) = 1-\left(1-\overline{K}_\gamma\left(y_{2n}\right)\right)^n = 1-\left(1-\frac{d_{2P}(n)(1+o(1))}{n}\right)^n \tag{6.44}$$

$$= d_{2P}(n)(1+o(1)) \text{ as } n\to\infty.$$

Because

$$P\left(X_{(n)} > y_{1n}, \left|S^2-1\right| < \eta_n, \left|\overline{X}\right| < \eta_n, |\widehat{\gamma_2}-\gamma| < \eta_n\right)$$
$$= P\left(X_{(n)} > y_{1n}\right) + o\left(d_{2P}(n)\right),$$

another application of Lemma 6.5.3 gives

$$P\left(X_{(n)} > y_{1n}, \left|S^2-1\right| < \eta_n, \left|\overline{X}\right| < \eta_n, |\widehat{\gamma_2}-\gamma| < \eta_n\right) = d_{2P}(n)(1+o(1)), \tag{6.45}$$

as $n \to \infty$. Combination of (6.42) – (6.45) gives

$$P\left(\frac{X_{(n)} - \overline{X}}{S} > \overline{K_{\widehat{\gamma_2}}}^{-1}\left(\frac{d_{2P}(n)}{n}\right)\right) = d_{2P}(n)(1 + o(1)), \text{ as } n \to \infty,$$

thus completing the proof of the lemma. ■

Lemma 6.5.5 *Let $X_1, \dots, X_n$ be i.i.d. r.v's with distribution function K_γ from the normal power family. Let $\gamma > 0$ and let $\{d_{2N}(n)\}$ be a sequence of real numbers satisfying for some $0 < \zeta < \gamma$*

$$\lim_{n\to\infty} d_{2N}(n) = 0, \lim_{n\to\infty} \frac{|\log d_{2N}(n)|}{(\log n)^{1+\gamma-\zeta}} = 0,$$

then, for some $c > 0$,

$$P\left(\frac{X_{(n)} - \overline{X}}{S} \le \overline{\Phi}^{-1}\left(\frac{d_{2N}(n)}{n}\right)\right) = O\left(\exp\left(-cn^{1/(1+\gamma)}\right)\right), \text{ as } n \to \infty.$$

Let $\gamma < 0$ and let $\{d_{1N}(n)\}$ be a sequence of real numbers satisfying

$$\lim_{n\to\infty} d_{1N}(n) = \infty, \lim_{n\to\infty} d_{1N}(n)\sqrt{\frac{d_{1N}(n)}{n}}\log\left(\frac{d_{1N}(n)}{n}\right) = 0,$$

then, for some $c^* > 0$,

$$P\left(\frac{X_{(n)} - \overline{X}}{S} \ge \overline{\Phi}^{-1}\left(\frac{d_{1N}(n)}{n}\right)\right) = O\left(\exp\left\{-c^*(\log n)^{\frac{1}{1+\gamma}}\right\}\right), \text{ as } n \to \infty.$$

Proof. Consider first $\gamma > 0$. Let $\varepsilon > 0$. It follows from Theorem 3.6.1 that for some $c > 0$

$$\begin{aligned} &P\left(\frac{X_{(n)} - \overline{X}}{S} \le \overline{\Phi}^{-1}\left(\frac{d_{2N}(n)}{n}\right)\right) \qquad (6.46)\\ &\le P\left(X_{(n)} \le \sqrt{1+\varepsilon}\,\overline{\Phi}^{-1}\left(\frac{d_{2N}(n)}{n}\right) + \varepsilon\right) + O\left(\exp\left(-cn^{1/(1+\gamma)}\right)\right)\\ &= \left\{1 - \overline{\Phi}\left(\left(\frac{\sqrt{1+\varepsilon}\,\overline{\Phi}^{-1}\left(\frac{d_{2N}(n)}{n}\right) + \varepsilon}{c(\gamma)}\right)^{\frac{1}{1+\gamma}}\right)\right\}^n + O\left(\exp\left(-cn^{1/(1+\gamma)}\right)\right). \end{aligned}$$

Since

$$\lim_{n\to\infty} d_{2N}(n) = 0, \lim_{n\to\infty} \frac{|\log d_{2N}(n)|}{(\log n)^{1+\gamma-\zeta}} = 0$$

and for $x \to 0$ we have

$$\overline{\Phi}^{-1}(x) = \sqrt{-2\log x}\,(1 + o(1)),$$

we obtain, for sufficiently large n,

$$\frac{\sqrt{1+\varepsilon}\overline{\Phi}^{-1}\left(\frac{d_{2N}(n)}{n}\right) + \varepsilon}{c(\gamma)} \leq \sqrt{(\log n)^{1+\gamma-\zeta}}$$

and hence, taking $0 < \eta < 1 - \frac{1}{1+\gamma}$, we get for sufficiently large n

$$\left\{1 - \overline{\Phi}\left(\left(\frac{\sqrt{1+\varepsilon}\overline{\Phi}^{-1}\left(\frac{d_{2N}(n)}{n}\right) + \varepsilon}{c(\gamma)}\right)^{\frac{1}{1+\gamma}}\right)\right\}^n \leq \left\{1 - \overline{\Phi}\left((\log n)^{\frac{1+\gamma-\zeta}{2(1+\gamma)}}\right)\right\}^n \tag{6.47}$$

$$\leq \left\{1 - \exp\left\{-(\log n)^{1-\frac{\zeta}{2(1+\gamma)}}\right\}\right\}^n \leq \left\{1 - n^{-\eta}\right\}^n \leq \exp\left(-n^{1-\eta}\right)$$
$$= O\left(\exp\left(-cn^{1/(1+\gamma)}\right)\right).$$

Combination of (6.46) and (6.47) yields

$$P\left(\frac{X_{(n)} - \overline{X}}{S} \leq \overline{\Phi}^{-1}\left(\frac{d_{2N}(n)}{n}\right)\right) = O\left(\exp(-cn^{1/(1+\gamma)}\right).$$

Next, let $\gamma < 0$. It follows from Theorem 3.6.1 that, for some $c > 0$,

$$P\left(\frac{X_{(n)} - \overline{X}}{S} \geq \overline{\Phi}^{-1}\left(\frac{d_{1N}(n)}{n}\right)\right) \tag{6.48}$$
$$\leq P\left(X_{(n)} \geq \sqrt{1-\varepsilon}\overline{\Phi}^{-1}\left(\frac{d_{1N}(n)}{n}\right) - \varepsilon\right) + O\left(\exp\left(-cn^{1/(1+\gamma)}\right)\right)$$
$$= 1 - \left\{1 - \overline{\Phi}\left(\left(\frac{\sqrt{1-\varepsilon}\overline{\Phi}^{-1}\left(\frac{d_{1N}(n)}{n}\right) - \varepsilon}{c(\gamma)}\right)^{\frac{1}{1+\gamma}}\right)\right\}^n + O\left(\exp\left(-cn^{1/(1+\gamma)}\right)\right).$$

Since

$$\lim_{n\to\infty} d_{1N}(n) = \infty, \lim_{n\to\infty} d_{1N}(n)\sqrt{\frac{d_{1N}(n)}{n}}\log\left(\frac{d_{1N}(n)}{n}\right) = 0,$$

we obtain, for sufficiently large n,

$$\frac{\sqrt{1-\varepsilon}\overline{\Phi}^{-1}\left(\frac{d_{1N}(n)}{n}\right) - \varepsilon}{c(\gamma)} \geq \sqrt{\tilde{c}\log n},$$

for some $\widetilde{c} > 0$. Hence, taking $0 < \eta < 1 - \frac{1}{1+\gamma}$, we get, for sufficiently large n,

$$1 - \left\{ 1 - \overline{\Phi}\left(\left(\frac{\sqrt{1-\varepsilon}\,\overline{\Phi}^{-1}\left(\frac{d_{1N}(n)}{n}\right) - \varepsilon}{c(\gamma)} \right)^{\frac{1}{1+\gamma}} \right) \right\}^n \tag{6.49}$$

$$\leq 1 - \left\{ 1 - \overline{\Phi}\left(\left(\sqrt{\widetilde{c}\log n} \right)^{\frac{1}{1+\gamma}} \right) \right\}^n$$

$$\leq 1 - \left\{ 1 - \exp\left\{ -\frac{1}{2}\left(\widetilde{c}\log n\right)^{\frac{1}{1+\gamma}} \right\} \right\}^n \leq \exp\left\{ -c^* \left(\log n\right)^{\frac{1}{1+\gamma}} \right\}$$

for some $c^* > 0$. Combination of (6.48) and (6.49) yields

$$P\left(\frac{X_{(n)} - \overline{X}}{S} \geq \overline{\Phi}^{-1}\left(\frac{d_{1N}(n)}{n} \right) \right) = O\left(\exp\left\{ -c^* \left(\log n\right)^{\frac{1}{1+\gamma}} \right\} \right). \quad \blacksquare$$

Note that

$$\lim_{n\to\infty} d_{2N}(n) = 0, \lim_{n\to\infty} \frac{|\log d_{2N}(n)|}{(\log n)^{1+\gamma-\zeta}} = 0$$

implies

$$\lim_{n\to\infty} d_{2N}(n) = 0, \lim_{n\to\infty} \sqrt{\frac{|\log d_{2N}(n)|}{n}} \log\left(\frac{d_{2N}(n)}{n} \right) = 0.$$

In view of (6.33), Lemma 6.5.4 and Lemma 6.5.5 we get the following result, which expresses that within the normal power family the combined control chart given by (6.12) and (6.13) is asymptotically as good as the parametric control chart for the normal power family given in (3.39). Note that

$$\lim_{n\to\infty} \max\left(\frac{|\log d_{2P}(n)|^{1+\gamma}}{n}, \sqrt{\frac{|\log d_{2P}(n)|}{n}} \right) \left| \log\left(\frac{d_{2P}(n)}{n} \right) \right|^{1+\zeta_2} = 0$$

implies

$$\lim_{n\to\infty} \frac{|\log d_{2P}(n)|^{1+\gamma}}{n} = 0$$

and hence

$$\exp\left(-cn^{1/(1+\gamma)} \right) = o\left(d_{2P}(n) \right), \text{ as } n \to \infty.$$

The results of Chapter 3 on (corrected) parametric control charts ensure that $E\overline{K}_\gamma(UL_P^*)$ is very close to p.

Theorem 6.5.3 *Let $X_1, \ldots, X_n$ be i.i.d. r.v's with a distribution from the normal power family. Let $\{d_{1N}(n)\}$, $\{d_{2N}(n)\}$ be sequences of real numbers satisfying, for some $0 < \zeta_1 < \gamma$,*

$$\lim_{n\to\infty} d_{1N}(n) = \infty, \lim_{n\to\infty} d_{1N}(n)\sqrt{\frac{d_{1N}(n)}{n}} \log\left(\frac{d_{1N}(n)}{n}\right) = 0, \ \text{if } \gamma > 0,$$
$$\lim_{n\to\infty} d_{2N}(n) = 0, \lim_{n\to\infty} \frac{|\log d_{2N}(n)|}{(\log n)^{1+\gamma-\zeta_1}} = 0, \ \text{if } \gamma > 0,$$

and let $\{d_{1P}(n)\}, \{d_{2P}(n)\}$ be sequences of real numbers satisfying, for some $\zeta_2 > 0$,

$$\lim_{n\to\infty} d_{1P}(n) = \infty,$$
$$\begin{cases} \lim\limits_{n\to\infty} d_{1P}(n)\sqrt{\frac{d_{1P}(n)}{n}} \left|\log\left(\frac{d_{1P}(n)}{n}\right)\right|^{1+\zeta_2} = 0, \ \text{if } \gamma \le 0 \\ \lim\limits_{n\to\infty} d_{1P}(n) \max\left(\frac{d_{1P}^{1+\gamma}(n)}{n}, \sqrt{\frac{d_{1P}(n)}{n}}\right) \left|\log\left(\frac{d_{1P}(n)}{n}\right)\right|^{1+\zeta_2} = 0, \ \text{if } \gamma > 0 \end{cases},$$
$$\lim_{n\to\infty} d_{2P}(n) = 0,$$
$$\begin{cases} \lim\limits_{n\to\infty} \sqrt{\frac{|\log d_{2P}(n)|}{n}} \left|\log\left(\frac{d_{2P}(n)}{n}\right)\right|^{1+\zeta_2} = 0, \ \text{if } \gamma \le 0 \\ \lim\limits_{n\to\infty} \max\left(\frac{|\log d_{2P}(n)|^{1+\gamma}}{n}, \sqrt{\frac{|\log d_{2P}(n)|}{n}}\right) \left|\log\left(\frac{d_{2P}(n)}{n}\right)\right|^{1+\zeta_2} = 0, \ \text{if } \gamma > 0. \end{cases}$$

Then

$$\begin{aligned}\left|EP_n - E\left\{\overline{K}_\gamma(UL_P^* - \Delta)\right\}\right| \le & \left\{\left(1 - \frac{d_{1P}(n)}{n}\right)^n + d_{2P}(n)\right\}(1 + o(1)) \\ & + \left[O\left(\exp\left\{-c^* (\log n)^{\frac{1}{1+\gamma}}\right\}\right) \ \text{if } \gamma < 0\right] \ \text{as } n \to \infty.\end{aligned}$$

Note that for the choices $d_{1N}(n) = -0.7 + 0.5\log n, d_{2N}(n) = 5/\sqrt{n}, d_{1P}(n) = -0.2 + 0.5\log n, d_{2P}(n) = 3/\sqrt{n}$ the conditions of Theorem 6.5.3 are satisfied and we get

$$|EP_n - E\{\overline{K}_\gamma(UL_P^* - \Delta)\}| \le (\frac{e^{0.2} + 3}{\sqrt{n}})(1 + o(1)),$$

as $n \to \infty$.

6.5.3 Outside the normal power family

Suppose that $X_1, \ldots, X_n$ have distribution function $F \ne K_\gamma$ for all γ and that X_{n+1} is distributed as $X_1 + \Delta$ for some $\Delta \ge 0$. As before, we have $EX_1 = 0$ and $var(X_1) = 1$. In view of (6.14) we have (with expectations and probabilities referring to F)

$$\begin{aligned}
EP_n &= E\left\{\overline{F}(UL_N^* - \Delta)\mathbf{1}\left(\frac{X_{(n)} - \overline{X}}{S} \in IN\right)\right\} \\
&+ E\left\{\overline{F}(UL_P^* - \Delta)\mathbf{1}\left(\frac{X_{(n)} - \overline{X}}{S} \notin IN\right)\mathbf{1}\left(\frac{X_{(n)} - \overline{X}}{S} \in IP\right)\right\} \\
&+ E\left\{\overline{F}(UL_{NP}^* - \Delta)\mathbf{1}\left(\frac{X_{(n)} - \overline{X}}{S} \notin IN\right)\mathbf{1}\left(\frac{X_{(n)} - \overline{X}}{S} \notin IP\right)\right\} \\
&= E\left\{\overline{F}(UL_{NP}^* - \Delta)\right\} + Eh_{NP}\left(X_{(n)}, \overline{X}, S\right),
\end{aligned}$$

with

$$\begin{aligned}
h_{NP}\left(X_{(n)}, \overline{X}, S\right) &= \left[\overline{F}(UL_N^* - \Delta) - \overline{F}(UL_{NP}^* - \Delta)\right]\mathbf{1}\left(\frac{X_{(n)} - \overline{X}}{S} \in IN\right) \\
&+ \left[\overline{F}(UL_P^* - \Delta) - \overline{F}(UL_{NP}^* - \Delta)\right]\mathbf{1}\left(\frac{X_{(n)} - \overline{X}}{S} \notin IN\right)\mathbf{1}\left(\frac{X_{(n)} - \overline{X}}{S} \in IP\right)
\end{aligned}$$

and hence

$$\left|EP_n - E\left\{\overline{F}(UL_{NP}^* - \Delta)\right\}\right| \le P\left(\frac{X_{(n)} - \overline{X}}{S} \in IN\right) + P\left(\frac{X_{(n)} - \overline{X}}{S} \in IP\right). \tag{6.50}$$

The following result expresses that outside the normal power family the combined control chart given by (6.12) and (6.13) is asymptotically as good as the (modified) nonparametric control chart given by (6.3) if $r \geq 1$ and for $r = 0$ by (6.7) (with $g(p) = p$).

Theorem 6.5.4 *Let $X_1, \ldots, X_n$ be i.i.d. r.v's with distribution function F. Let $\{d_{1N}(n)\}, \{d_{2N}(n)\}, \{d_{1P}(n)\}, \{d_{2P}(n)\}$ be sequences of real numbers satisfying*

$$\begin{aligned}
&\lim_{n\to\infty} d_{1N}(n) = \infty, \lim_{n\to\infty} \frac{\log d_{1N}(n)}{\log n} = 0, \\
&\lim_{n\to\infty} d_{2N}(n) = 0, \lim_{n\to\infty} \frac{|\log d_{2N}(n)|}{\log n} = 0, \\
&\lim_{n\to\infty} d_{1P}(n) = \infty, \lim_{n\to\infty} \frac{\log d_{1P}(n)}{\log n} = 0, \\
&\lim_{n\to\infty} d_{2P}(n) = 0, \lim_{n\to\infty} \frac{|\log d_{2P}(n)|}{\log n} = 0
\end{aligned} \tag{6.51}$$

Let γ be defined as the limit of the estimator $\widehat{\gamma_2}$ under F, that is by

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

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

$$\left|EP_n - E\left\{\overline{F}(UL_{NP}^* - \Delta)\right\}\right| \leq \min\left\{m_1, m_2\right\} + \min\left\{m_3, m_4\right\},$$

where

$$m_1 = F\left(\left(\sqrt{1+\varepsilon_1} + \zeta_1\right)\sqrt{2\log n}\right)^n + P\left(\left|\overline{X}\right| > \eta_1\right) + P\left(\left|S^2 - 1\right| > \varepsilon_1\right),$$

$$m_2 = 1 - F\left(\left(\sqrt{1-\varepsilon_2} - \zeta_2\right)\sqrt{2\log n}\right)^n + P\left(\left|\overline{X}\right| > \eta_2\right) + P\left(\left|S^2 - 1\right| > \varepsilon_2\right),$$

$$m_3 = F\left(\left(\sqrt{\log n}\right)^{1+\gamma+2\zeta_3}\right)^n + P\left(\left|\overline{X}\right| > \eta_3\right) + P\left(\left|S^2 - 1\right| > \varepsilon_3\right) + P\left(\left|\widehat{\gamma}_2 - \gamma\right| > \zeta_3\right),$$

$$m_4 = 1 - F\left(\left(\sqrt{\log n}\right)^{1+\gamma-2\zeta_4}\right)^n + P\left(\left|\overline{X}\right| > \eta_4\right) + P\left(\left|S^2 - 1\right| > \varepsilon_4\right) + P\left(\left|\widehat{\gamma}_2 - \gamma\right| > \zeta_4\right).$$

Proof. We have for each $\varepsilon, \eta > 0$

$$P\left(\frac{X_{(n)} - \overline{X}}{S} \leq \overline{\Phi}^{-1}\left(\frac{d_{2N}(n)}{n}\right)\right)$$
$$\leq P\left(X_{(n)} \leq \sqrt{1+\varepsilon}\,\overline{\Phi}^{-1}\left(\frac{d_{2N}(n)}{n}\right) + \eta\right) + P\left(\left|\overline{X}\right| > \eta\right) + P\left(\left|S^2 - 1\right| > \varepsilon\right).$$

Since

$$\lim_{n\to\infty} d_{2N}(n) = 0, \lim_{n\to\infty} \frac{\left|\log d_{2N}(n)\right|}{\log n} = 0,$$

we get

$$\overline{\Phi}^{-1}\left(\frac{d_{2N}(n)}{n}\right) = \sqrt{-2\log\left(\frac{d_{2N}(n)}{n}\right)}(1 + o(1)) = \sqrt{2\log n}(1 + o(1))$$

and hence, for each $\zeta > 0$, we get, for sufficiently large n,

$$\sqrt{1+\varepsilon}\,\overline{\Phi}^{-1}\left(\frac{d_{2N}(n)}{n}\right) + \eta \leq \left(\sqrt{1+\varepsilon} + \zeta\right)\sqrt{2\log n}.$$

Therefore, we obtain, for each $\varepsilon, \eta, \zeta > 0$ and sufficiently large n,

$$P\left(\frac{X_{(n)} - \overline{X}}{S} \leq \overline{\Phi}^{-1}\left(\frac{d_{2N}(n)}{n}\right)\right)$$
$$\leq F\left(\left(\sqrt{1+\varepsilon} + \zeta\right)\sqrt{2\log n}\right)^n + P\left(\left|\overline{X}\right| > \eta\right) + P\left(\left|S^2 - 1\right| > \varepsilon\right).$$

Similarly, we get, for each $\varepsilon, \eta, \zeta > 0$ and sufficiently large n,

$$P\left(\frac{X_{(n)} - \overline{X}}{S} \geq \overline{\Phi}^{-1}\left(\frac{d_{1N}(n)}{n}\right)\right)$$
$$\leq 1 - F\left(\left(\sqrt{1-\varepsilon} - \zeta\right)\sqrt{2\log n}\right)^n + P\left(\left|\overline{X}\right| > \eta\right) + P\left(\left|S^2 - 1\right| > \varepsilon\right).$$

Using

$$\overline{K}_\gamma^{-1}(t) = c(\gamma)\left\{\overline{\Phi}^{-1}(t)\right\}^{1+\gamma}, \text{ for } 0 < t < \frac{1}{2},$$

we get, for each $0 < \zeta < 1 + \gamma$ (with ξ between γ and $\widetilde{\gamma}$), as $t \to 0$

$$\sup_{|\widetilde{\gamma}-\gamma|\leq\zeta} \overline{K_{\widetilde{\gamma}}}^{-1}(t)$$
$$= \sup_{|\widetilde{\gamma}-\gamma|\leq\zeta} \left\{\overline{K}_\gamma^{-1}(t) + (\widetilde{\gamma} - \gamma)\left[c'(\xi)\left\{\overline{\Phi}^{-1}(t)\right\}^{1+\xi} + c(\xi)\left\{\overline{\Phi}^{-1}(t)\right\}^{1+\xi}\log\left(\overline{\Phi}^{-1}(t)\right)\right]\right\}$$
$$\leq \left\{\overline{K}_\gamma^{-1}(t)\right\}^{1+\frac{1.5\zeta}{1+\gamma}}.$$

We obtain, for sufficiently large n,

$$P\left(\frac{X_{(n)} - \overline{X}}{S} \leq \overline{K_{\widehat{\gamma}_2}}^{-1}\left(\frac{d_{2P}(n)}{n}\right)\right)$$
$$\leq F\left(\left(\sqrt{\log n}\right)^{1+\gamma+2\zeta}\right)^n + P\left(\left|\overline{X}\right| > \eta\right) + P\left(\left|S^2 - 1\right| > \varepsilon\right) + P\left(\left|\widehat{\gamma}_2 - \gamma\right| > \zeta\right).$$

Similarly, we get, for sufficiently large n,

$$P\left(\frac{X_{(n)} - \overline{X}}{S} \geq \overline{K_{\widehat{\gamma}_2}}^{-1}\left(\frac{d_{1P}(n)}{n}\right)\right) \leq 1 - F\left(\left(\sqrt{\log n}\right)^{1+\gamma-2\zeta}\right)^n$$
$$+ P\left(\left|\overline{X}\right| > \eta\right) + P\left(\left|S^2 - 1\right| > \varepsilon\right) + P\left(\left|\widehat{\gamma}_2 - \gamma\right| > \zeta\right).$$

■

Theorem 6.5.4 makes only sense if F differs from the normal power family in the sense that, for some $\varepsilon > 0$,

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

(heavier tail than the normal distribution) or

$$\lim_{n\to\infty}\left[F\left((1-\varepsilon)\sqrt{2\log n}\right)\right]^n = 1$$

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

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

(heavier tail than K_γ) or

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

(thinner tail than K_γ).

Note that for the choices $d_{1N}(n) = -0.7 + 0.5\log n, d_{2N}(n) = 5/\sqrt{n}, d_{1P}(n) = -0.2 + 0.5\log n, d_{2P}(n) = 3/\sqrt{n}$ the conditions (6.51) are satisfied.

6.6 Simulation

In this section we show by means of simulation the performance for a finite sample size of the new control chart, given in (6.10) – (6.13) with $d_{1N}(n) = -0.7 + 0.5\log n$, $d_{2N}(n) = 5/\sqrt{n}$, $d_{1P}(n) = -0.2+0.5\log n$ and $d_{2P}(n) = 3/\sqrt{n}$. We take $g(p) = p$ and hence compare EP_n to p in the in-control situation, while also in the out-of-control case EP_n is our criterion. We choose $p = 0.001$ and for n we take $250, 500, 1000, 1500, 2000$. The number of repetitions in the simulation study equals 100 000.

The theoretical results from Section 6.5 show that the behavior of the new control chart is asymptotically the same as that of the specific control chart adjusted for the distribution at hand. In a simulation study the behavior of the new control chart is investigated to see how well this property continues to hold in the finite sample case. Therefore, we compare the combined (C) control chart with the normal (N) control chart, the parametric (P) chart (both charts with correction for $r = 0$ and without correction for $r \geq 1$) and the nonparametric (NP) control chart, in its modified version for $r = 0$ and for $r \geq 1$ with the stochastic term V replaced by its deterministic counterpart EV, all of this just as in composing the combined control chart, see (6.10) – (6.13).

All procedures are location and scale invariant and hence we take for all distributions involved in the simulation the expectation equal to 0 and the variance equal to 1. The distributions can be classified as follows.

1. **The standard normal distribution** ($F_1 = \Phi$). For this distribution the control chart based on normality will obviously be the favorite one. It is expected that the in-control behavior of all control charts under consideration is sufficiently good. We want to see how much we loose in the out-of-control

case when applying the parametric, the nonparametric and, in particular, the combined control chart.

2. **Distributions from the normal power family** ($F_2 = K_{-0.5}$, $F_3 = K_{0.5}$, $F_4 = K_1$). It is seen from Table 3.3 that the normal control chart behaves very badly for these distributions, in the sense that for the in-control situation EP_n differs much from p. In contrast, the parametric, the nonparametric and the combined control chart are expected to behave well when the observations are in-control. With respect to the out-of-control behavior, the interest is in the loss of the nonparametric and especially the combined control chart in comparison with the parametric control chart.

3. **Distributions outside the normal power family**. We take as F_5 the Student distribution with 6 degrees of freedom and standardized to unit variance. Its model error, see (3.5), when the supposed model is the normal power family, equals 2.08. Furthermore, we consider as F_6 the random mixture RM with model error 1.16. As F_7 we take the Normal Inverse Gaussian $(2, 1.5, 0, 1)$ distribution, cf. Barndorff-Nielsen (1996). This distribution shows that the member of the normal power family chosen by our estimator does not fit the distribution globally, but it does at the right tail and that is exactly what we want; see Figure 6.3. Nevertheless, its model error w.r.t. the normal power family still equals $1.93p$. Note that the model error w.r.t. the normal distribution is $14.65p$! So, the normal power family gives a great improvement for this distribution.

   The Normal Inverse Gaussian $(0.5, 0, 0, 1)$ distribution, denoted by F_8, has model error $2.31p$ w.r.t. the normal power family. Finally, as an example of a (very) negative model error we take for F_9 the (standardized) Beta $(3, 3.75)$ distribution. Its model error w.r.t. the normal power family equals $-0.996p$.

   It is already known from Table 3.3 that the normal control chart may behave very badly for several of the distributions involved here. Table 3.5 shows that the parametric control chart gives reasonable total error for these distributions, thus improving the normal control chart tremendously, but still the model errors are not vanishing and considerable deviances remain. It is expected that the nonparametric and the combined control chart behave better when the observations are in-control. Especially for the combined control chart it is of interest to compare the out-of-control results with those of the nonparametric chart for the distributions in this class.

It is seen in Table 6.3 that indeed the combined control chart has good in-control behavior under all distributions. Recall that the ideal value equals 1.00. The normal control chart cannot be used unless the distribution is very close to normality. The parametric control chart is often a great improvement w.r.t. the normal chart and gives reasonable results, while the nonparametric control chart behaves very well under all distributions.

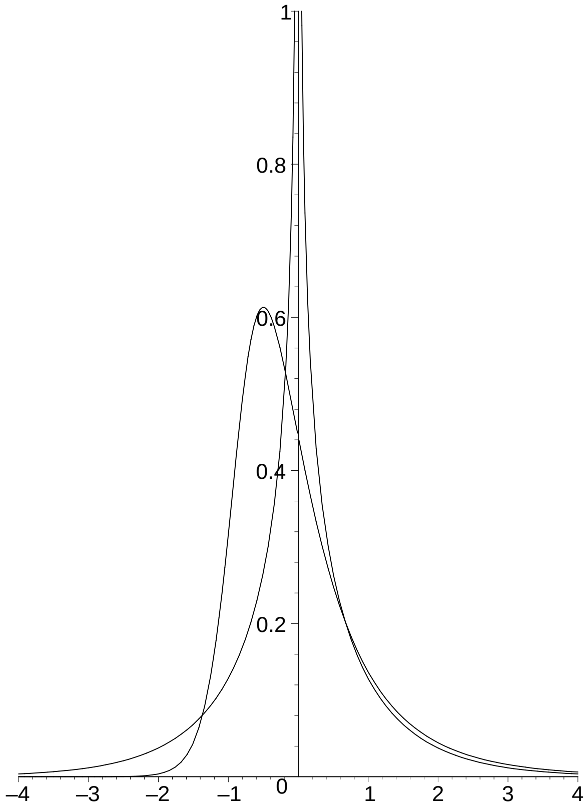

Figure 6.3. Density of the Normal Inverse Gaussian $(2, 1.5, 0, 1)$ and the corresponding density of the normal power family with $\gamma = 0.77$.

In the next two tables the out-of-control situation is presented. We consider shift alternatives, that is $X_1, \ldots, X_n$ have distribution function $F(x)$ and X_{n+1} has distribution function $F(x - \Delta)$. For each of the nine distributions F we have selected two values of Δ such that reasonable values of the probability of an out-of-control signal result.

When examining Table 6.4 one should take into account the results in Table 6.3, because a high value in Table 6.3 may be caused by an unduly high and thus incorrect value in Table 6.3. In case of a simulated expected observed false alarm rate in the in-control situation, at least twice as large as p, we do not give the (simulated) probability of an out-of-control signal. Such situations are denoted by * in Table 6.4. The following table presents the simulation results for the in-control situation.

Table 6.3 In-control behavior of the normal (N), parametric (P) and nonparametric (NP) control charts and the combined control chart (C). Presented is the simulated expected observed false alarm rate, i.e. the simulated EP_n multiplied by 1000.

n	chart	F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8	F_9
250	N	1.00	0.00	6.63	10.67	4.60	2.80	16.09	7.91	0.01
	P	1.05	0.94	1.05	1.06	3.02	2.14	2.98	3.34	0.16
	NP	1.12	1.00	1.67	2.20	1.81	1.58	2.32	2.10	1.00
	C	0.97	0.75	1.51	1.21	2.19	1.81	1.92	2.28	0.31
500	N	1.00	0.00	6.61	10.52	4.59	2.79	15.88	7.84	0.01
	P	1.02	0.97	1.02	1.02	3.04	2.15	2.95	3.31	0.08
	NP	1.03	1.00	1.21	1.39	1.32	1.24	1.44	1.38	1.00
	C	0.97	0.86	1.25	1.01	1.79	1.60	1.71	1.72	0.46
1000	N	1.03	0.00	6.69	10.53	4.64	2.83	15.92	7.86	0.00
	P	1.13	1.21	1.12	1.11	3.23	2.29	3.17	3.48	0.06
	NP	1.01	1.00	0.99	1.00	1.03	1.00	1.01	1.00	1.00
	C	1.03	1.14	1.17	1.08	1.48	1.40	1.89	1.45	0.70
1500	N	1.02	0.00	6.66	10.47	4.62	2.82	15.82	7.83	0.00
	P	1.09	1.14	1.08	1.08	3.19	2.25	3.08	3.43	0.04
	NP	0.90	0.91	0.89	0.89	0.88	0.87	0.88	0.89	0.92
	C	1.01	1.08	1.12	1.04	1.28	1.29	1.88	1.33	0.72
2000	N	1.02	0.00	6.64	10.45	4.60	2.81	15.78	7.81	0.00
	P	1.07	1.10	1.06	1.06	3.16	2.23	3.04	3.40	0.03
	NP	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	C	1.02	1.09	1.10	1.05	1.33	1.35	1.99	1.39	0.80

Table 6.4 Out-of-control behavior of N, P, NP and C control charts. Presented are the simulated EP_n with $X_1, \ldots, X_n$ having distribution function $F(x)$ and X_{n+1} distribution function $F(x-\Delta)$.

		F1		F2		F3		F4		F5
n	chart	$\Delta=2$	$\Delta=3$	$\Delta=1$	$\Delta=2$	$\Delta=3$	$\Delta=4$	$\Delta=3$	$\Delta=4$	$\Delta=2$
250	N	.133	.450	.000	.157	∗	∗	∗	∗	∗
	P	.120	.410	.191	.496	.080	.293	.021	.071	∗
	NP	.090	.311	.071	.330	.115	.344	∗	∗	.039
	C	.115	.396	.145	.437	.107	.299	.031	.082	∗
500	N	.135	.457	.000	.164	∗	∗	∗	∗	∗
	P	.128	.436	.208	.499	.080	.305	.020	.060	∗
	NP	.096	.327	.120	.370	.088	.294	.030	.110	.027
	C	.122	.417	.179	.462	.093	.294	.021	.063	.040
1000	N	.139	.464	.000	.172	∗	∗	∗	∗	∗
	P	.140	.464	.227	.501	.088	.338	.021	.062	∗
	NP	.121	.412	.208	.499	.077	.287	.020	.065	.019
	C	.132	.443	.218	.500	.090	.321	.021	.065	.032
1500	N	.138	.464	.000	.172	∗	∗	∗	∗	∗
	P	.140	.464	.227	.501	.086	.328	.020	.059	∗
	NP	.120	.418	.211	.500	.071	.272	.017	.051	.016
	C	.134	.451	.222	.500	.088	.318	.020	.059	.027
2000	N	.138	.464	.000	.172	∗	∗	∗	∗	∗
	P	.139	.464	.227	.501	.085	.322	.020	.057	∗
	NP	.129	.437	.218	.500	.079	.303	.019	.058	.018
	C	.137	.459	.225	.501	.087	.320	.020	.058	.027

		F5	F6		F7		F8		F9	
n	chart	$\Delta=3$	$\Delta=2$	$\Delta=3$	$\Delta=4$	$\Delta=5$	$\Delta=3$	$\Delta=4$	$\Delta=1$	$\Delta=2$
250	N	∗	∗	∗	∗	∗	∗	∗	.016	.148
	P	∗	∗	∗	∗	∗	∗	∗	.019	.149
	NP	.157	.058	.219	∗	∗	∗	∗	.023	.140
	C	∗	.076	.285	.089	.237	∗	∗	.021	.161
500	N	∗	∗	∗	∗	∗	∗	∗	.017	.151
	P	∗	∗	∗	∗	∗	∗	∗	.020	.161
	NP	.119	.048	.190	.063	.175	.050	.176	.033	.174
	C	.170	.068	.262	.076	.218	.067	.232	.025	.174
1000	N	∗	∗	∗	∗	∗	∗	∗	.017	.155
	P	∗	∗	∗	∗	∗	∗	∗	.023	.175
	NP	.093	.040	.180	.040	.115	.030	.121	.052	.257
	C	.139	.060	.236	.085	.252	.053	.204	.039	.220
1500	N	∗	∗	∗	∗	∗	∗	∗	.017	.154
	P	∗	∗	∗	∗	∗	∗	∗	.023	.174
	NP	.079	.034	.166	.033	.094	.098	.327	.052	.259
	C	.121	.054	.222	.084	.251	.183	.394	.043	.234
2000	N	∗	∗	∗	∗	∗	∗	∗	.017	.154
	P	∗	∗	∗	∗	∗	∗	∗	.023	.174
	NP	.092	.040	.195	.037	.107	.027	.112	.053	.263
	C	.127	.057	.243	.087	.263	.046	.188	.047	.244

Table 6.5 Out-of-control behavior of N, P, NP and C control charts. Presented are $(100\,(EP_n))/\widetilde{p}_n$ with $X_1, \ldots, X_n$ having distribution function $F(x)$ and X_{n+1} distribution function $F(x-\Delta)$.

		F1		F2		F3		F4		F5
n	chart	$\Delta = 2$	$\Delta = 3$	$\Delta = 1$	$\Delta = 2$	$\Delta = 3$	$\Delta = 4$	$\Delta = 3$	$\Delta = 4$	$\Delta = 2$
250	N	96	97	*	*	*	*	*	*	*
	P	85	87	85	99	94	94	107	126	*
	NP	62	65	31	66	92	65	*	*	108
	C	85	86	67	87	93	68	137	125	*
500	N	98	99	*	*	*	*	*	*	*
	P	92	93	92	100	97	100	103	111	*
	NP	69	70	53	74	92	83	113	141	113
	C	90	91	81	92	95	81	114	119	111
1000	N	99	99	*	*	*	*	*	*	*
	P	96	97	96	100	98	103	102	104	*
	NP	87	89	91	100	94	96	106	124	112
	C	94	95	93	100	97	93	104	112	113
1500	N	99	100	*	*	*	*	*	*	*
	P	97	98	97	100	99	103	101	103	*
	NP	91	93	94	100	96	100	104	112	113
	C	97	97	96	100	98	96	102	106	116
2000	N	100	100	*	*	*	*	*	*	*
	P	98	98	98	100	99	102	101	102	*
	NP	93	94	96	100	97	101	103	110	109
	C	98	98	97	100	99	98	101	104	112

		F5	F6		F7		F8		F9	
n	chart	$\Delta = 3$	$\Delta = 2$	$\Delta = 3$	$\Delta = 4$	$\Delta = 5$	$\Delta = 3$	$\Delta = 4$	$\Delta = 1$	$\Delta = 2$
250	N	*	*	*	*	*	*	*	66	81
	P	*	*	*	*	*	*	*	53	69
	NP	82	82	68	*	*	*	*	42	52
	C	*	92	80	115	104	*	*	51	69
500	N	*	*	*	*	*	*	*	74	87
	P	*	*	*	*	*	*	*	64	78
	NP	92	90	72	117	112	135	114	59	65
	C	90	94	81	114	111	135	105	56	71
1000	N	*	*	*	*	*	*	*	83	92
	P	*	*	*	*	*	*	*	77	87
	NP	102	100	83	113	117	124	132	93	96
	C	94	97	80	113	113	134	122	78	86
1500	N	*	*	*	*	*	*	*	87	93
	P	*	*	*	*	*	*	*	81	90
	NP	108	103	88	108	113	116	127	96	97
	C	98	97	81	112	113	131	126	86	91
2000	N	*	*	*	*	*	*	*	86	93
	P	*	*	*	*	*	*	*	84	91
	NP	104	100	89	107	110	112	121	96	98
	C	98	97	85	109	110	124	121	90	94

The comparison of the out-of-control behavior of the several charts is obscured by the differences in the in-control behavior. Indeed, when the in-control rate of chart 1 is $1.3p$ and that of chart 2 equals p, we may expect also a higher out-of-control rate for chart 1. Therefore, we have made a more "fair" comparison in Table 6.5. Let E_0P_n be the in-control rate of Table 6.3. The probability of detecting the shift Δ with this in-control rate, when the distribution is completely known, equals

$$\widetilde{p}_n = \overline{F}\left(\overline{F}^{-1}\left(E_0P_n\right) - \Delta\right).$$

This $\widetilde{p}_n$ serves as a bench mark of what can be obtained in the out-of-control case with in-control rate E_0P_n. The way the out-of-control rate is related to the in-control rate is given by presenting EP_n of Table 6.4 as a percentage of $\widetilde{p}_n$. Firstly, this indicates how well the chart performs in the out-of-control case on its own and secondly, it makes comparison between several charts more fair. However, still we have to realize that in the first place the in-control behavior should be controlled. Therefore, again a * denotes when a simulated expected observed false alarm rate under in-control is at least twice as large as p. Also when the simulated in-control rate E_0P_n equals (exactly) 0 a * is given.

The limiting value $\widetilde{p}_n$ indicates the out-of-control probability that can be obtained for the distribution at hand if the distribution is completely known. It is seen in Table 6.5 that as a rule all charts and hence in particular the combined control chart and the (modified) nonparametric control chart compare very well w.r.t. this bench mark. Moreover, Tables 6.3 and 6.4 show that

1. if $F_1 = \Phi$, the combined control chart has a substantial gain w.r.t. the nonparametric control chart and only a small loss w.r.t. the normal and parametric control chart;

2. if $F = K_\gamma$ for some γ, i.e. $F = F_1,\ F_2$ or F_3, the combined control chart, has for not too large n and γ, a considerable gain w.r.t. the nonparametric control chart and in general only a small loss w.r.t. the parametric control chart. The normal control chart cannot be applied unless γ is very close to 0; for positive γ's this is due to its bad in-control behavior and for negative γ's it has a very low alarm rate for the out-of-control situation;

3. outside the normal power family, the combined control chart exhibits only a small loss w.r.t. the nonparametric control chart, while the normal control chart cannot be applied in this case due to its bad in-control behavior or its low alarm rate (unless the distribution is very close to normality); the same holds for the parametric control chart, albeit to a much smaller extent.

6.7 Concluding remark

Although the normal, parametric and non-parametric approaches proposed in the previous chapters offer some advantages, it has been shown that applying these approaches individually also bears some disadvantages. Therefore, in this chapter the three approaches are combined into one procedure in which the data, based on some criteria, will select one of the three charts that is most suitable to their characteristics.

In Section 6.2 we briefly review the three charts to be combined in the new procedure. The following section discusses this procedure by determining the criteria used to select one of the charts. In addition, Section 6.3 presents the combined control chart for $g(p) = p$ (cf. (6.10) – (6.13)). Section 6.4 demonstrates an application of the combined chart on real life data sets. Theoretical behavior of the combined control chart discussed in Section 6.5 shows that the new chart asymptotically behaves as each of the specific control charts in their own domain. Using various distributions, the behavior of the combined chart both in the in-control and out-of-control situation are simulated and given in Section 6.6.

From the theory and the simulation results presented in this chapter as well as from additional simulations that we have performed, we conclude that the new combined control chart can be recommended as an omnibus control chart with a good performance under a great variety of distributions, both for the in-control and for the out-of-control situation. The modified nonparametric chart can be recommended as well, although under normality a price has to be paid for its simpler form.

Chapter 7

Implementation:
A user guide

In the previous chapters we discussed a number of methods to maintain a company's production process in good condition. As a continuation, the present chapter intends to provide a user with a guideline for implementing those concepts in a real life situation. It presents a brief review of the concepts in the form of ready-made formulas as well as flowcharts that describe actual steps to be performed by the user in the field.

7.1 Introduction

It is important to implement the methods developed in the previous chapters to deal with real life problems. Very often we experience that, although a concept has been developed very well, without a detailed guideline, the user may still have difficulty to implement the concept in the field. The reason for this is that those who work in the field may not be familiar with terms or symbols used to explain the concept. Another reason is that an experienced practitioner is used to rely on his intuition more than on a sophisticated method that sounds complicated to solve his problem. For the same reason practitioners in the field quite often ignore the assumptions required by the method; regardless the validity of the assumptions, as long as they are satisfied with the result of a common method, they consider that method acceptable.

Since on the one hand correct application of the method is an important issue not to be neglected by practitioners, and on the other hand the practicality and simplicity of the method is also an important factor to be considered by those who develop the method, the present chapter is devoted to provide a methodology which consists of a simple guideline to implement the concept in the field. The methodology can ignore the complicated theories on which it is based, except for the assumptions involved. Moreover, the methodology has to be described clearly in terms of the inputs required, the steps to be executed and the output to be produced. It is also preferable for the developed method to use mainly common terms already familiar to the practitioners who carry out the job in the field.

To achieve the objectives mentioned above, this chapter firstly presents a brief review on the concepts, followed by the implementation of these concepts in detail. Although in the previous chapters we discussed the one-sided limits, here we present the formulas of the two-sided limits, since they are more important in practice.

In Section 7.2, three types of control charts as components of the combined chart are presented in explicit formulas. For all three types of charts, two forms of corrected control charts are designed, associated with two performance criteria. The corrections are needed because quantiles of the distributions should be estimated and the estimation process causes stochastic errors. The first form concerns bias removal for the control chart as a whole. The bias criterion deals with the long run behavior of charts. Even with a small bias a single chart may be in error, since there still may be large variability due to the estimation step. The second form is devoted to correct the influence of the estimators in a more stringent way, aiming at sensible behavior not merely in the long run, but for a single chart. We call this the exceedance probability criterion. A more detailed description of these criteria is given in Section 1.2.

Following the review, Section 7.3 provides a guideline to construct and apply the combined chart. In this section an algorithm which consists of preliminary and main steps to be performed by the user, is given. The preliminary step highlights the type of data the user needs to collect before applying the concept. Once these data are ready, the user can continue with the main steps. Here, the actual steps to be performed in the field are described systematically using flowcharts. In the form of flowchart, the methodology can be easily transformed into software. At the end of the chapter a summary of the material presented throughout the chapter is given.

7.2 Three types of control chart : a brief review

Recall from the previous chapters that X is a random variable to be monitored in a production process. Let $X_1, \ldots, X_n$ be a sample of the Phase I observations. It is assumed that these observations are in-control and we use them for estimating unknown quantities in the basic control limits and in addition for making necessary corrections after plugging in the estimators. Let X_{n+1} be a new observation from Phase II. The upper and lower limit depend on the observations $X_1, \ldots, X_n$.

Let P_n be the observed alarm rate, given $X_1, \ldots, X_n$. In the in-control situation, where X_{n+1} has the same distribution as $X_1, \ldots, X_n$, we want P_n to be close to the prescribed false alarm rate p. If not stated otherwise we consider in this section the in-control situation. As P_n is random, there is no unique way to express the discrepancy. Two popular approaches are as follows. The first is the mildest one: look at the bias $EP_n - p$; if it is sufficiently small, the chart is considered to be acceptable. To the same category belongs the more general comparison of $Eg(P_n)$ with for instance the function $g(p) = 1/p$ corresponding to the ARL. When considering the bias we will not take the ARL into account, e.g. because, contrary to simple intuition, $1/P_n$ has a positive bias (due to the occurrence of extremely long runs, which although having small probability nevertheless strongly determine the expectation). As a consequence, the correction for estimating brings the control limits closer to each other, while one has the feeling that one should "pay " something for estimating the parameters and not "gain" something. Several authors, for example Roes (1995, page 34), therefore remark that $E\,(1/P_n)$ does not adequately summarize the run length properties of the chart, cf. also Quesenberry (1993, page 242) and Jones *et al.* (2004, page 100).

The bias criterion is satisfactory if we are interested in the behavior of a long series of applications of charts. However, if the bias of P_n is indeed small, but its variability is still large, very long runs are mixed up with unwelcome very short ones. Therefore, for an application to a single chart the second criterion is more appropriate. It concerns the probability that P_n exceeds p by more than a given percentage (for example 10%). If this exceedance probability is sufficiently small (e.g. 10%), the chart is O.K. in this more strict way. So, with only a small probability P_n is in error with more than 10%, say. Since the starting point of a two-sided control chart with prescribed false alarm rate p is a combination of two one-sided control charts each with prescribed false alarm rate $p/2$, we will treat the upper and lower limit here also separately. Therefore, we present corrected lower control limits $\widehat{LCL}$ such that P_{nL} exceeds p (for a one-sided control chart) or exceeds $p/2$ (for a two-sided control chart) by more than a given percentage only with some given small probability, where

$$P_{nL} = \Pr\left(X_{n+1} < \widehat{LCL}\,|X_1, \ldots, X_n\right)$$

and similarly for P_{nU}, with

$$P_{nU} = \Pr\left(X_{n+1} > \widehat{UCL}\,|X_1, \ldots, X_n\right).$$

To be more precise, we require that the intended false alarm rate p (for the one-sided control chart) or $p/2$ (for the two-sided control chart) may be exceeded by the outcome P_{nL} and P_{nU} by more than a fraction ε in at most $100\alpha\%$ of the applications. The next two formulas are for the two-sided control chart. The first one is for a control chart with a lower control limit, while the second one is for that with an upper control limit. For the one-sided control chart, $p/2$ in the formulas is simply replaced by p and the word "and" is replaced by "or".

$$\Pr\left(P_{nL} > \frac{p}{2}(1+\varepsilon)\right) \leq \alpha \text{ and } \Pr\left(P_{nU} > \frac{p}{2}(1+\varepsilon)\right) \leq \alpha. \qquad (7.1)$$

We consider also the more general situation of $g(P_{nL})$ exceeding $g(p)$ and similarly for $g(P_{nU})$. The problems with $g(p) = 1/p$ in the bias case do not occur here, since we deal with probabilities and hence extremely long runs are not dominating. In the exceedance case, we require that unpleasant values of $1/P_{nL}$ and $1/P_{nU}$ should be sufficiently rare, leading to

$$\Pr\left(\frac{1}{P_{nL}} < \frac{1}{p/2}(1-\varepsilon)\right) \leq \alpha \text{ and } \Pr\left(\frac{1}{P_{nU}} < \frac{1}{p/2}(1-\varepsilon)\right) \leq \alpha. \qquad (7.2)$$

It is easily seen that (7.2) is the same as (7.1) with ε replaced by $\varepsilon/(1-\varepsilon)$.

The exceedance criterion given in (7.2) can also be interpreted as follows. The conditional ARL's due to the lower limit, $1/P_{nL}$, and due to the upper limit, $1/P_{nU}$, depend on the Phase I distributions $X_1, \ldots, X_n$. Consider these ARL's as random variables. The $100\alpha^{\text{th}}$ percentile of those random variables is at least $(1-\varepsilon)/(p/2)$. For instance with $p/2 = 0.001, \varepsilon = 0.1, \alpha = 0.1$ this gives that the 10^{th} percentile of the ARL's $1/P_{nL}$ and $1/P_{nU}$ is at least 900. It may be better to settle this kind of percentiles of the ARL than to base oneself on the bias (the expectation) of the ARL, see also Jones *et al.* (2004, page 102).

In the following subsections, subsequently the three types of control charts, that compose the combined chart with the appropriate corrections, are reviewed briefly. The corresponding control limits are presented in explicit formulas.

7.2.1 The normal chart

The normal approach which is producing the normal control chart was discussed in Chapter 2. Here we simply present the results with a slightly different criterion than that of in Chapter 2. There, the exceedance probability criterion for the two side case is $Pr(P_{nL} + P_{nU} > p(1+\epsilon)) < \alpha$, see Remark 2.3.5, while here we use $Pr(P_{nL} > \frac{p}{2}(1+\epsilon)) \leq \alpha$ and $Pr(P_{nU} > \frac{p}{2}(1+\epsilon)) \leq \alpha$, respectively for the lower and upper control limit. Table 7.1 presents the two-sided control limits for the (corrected) normal control chart. The lower control limit of the normal chart is denoted by $\widehat{LCL}_N$, while the upper limit is denoted by $\widehat{UCL}_N$. In case of the lower limit $\widehat{LCL}_N$ the $-$ sign should be read and for $\widehat{UCL}_N$ the $+$ sign is used.

Application to the example given in Section 6.4 gives for $p/2 = 0.001, \varepsilon = 0.1, \alpha = 0.1$ the following two-sided control limits. The sample mean $\overline{X}$ equals 42.366 and

Table 7.1. Two-sided (corrected) normal control limits.

Aim	$\widehat{LCL}_N$ and $\widehat{UCL}_N$
$EP_n = p$	$\overline{X} \pm u_{p/2} S \left\{ 1 + \frac{u_{p/2}^2 + 3}{4n} \right\}$
$\Pr\left(P_{nL} > \frac{p}{2}(1+\varepsilon)\right) \leq \alpha$, $\Pr\left(P_{nU} > \frac{p}{2}(1+\varepsilon)\right) \leq \alpha$	$\overline{X} \pm u_{p/2} S \left\{ 1 + \frac{u_\alpha \left(\frac{1}{2} + u_{p/2}^{-2}\right)^{1/2}}{\sqrt{n}} - \frac{\varepsilon}{u_{p/2}^2} \right\}$
$\Pr\left(\frac{1}{P_{nL}} < \frac{1}{p/2}(1-\varepsilon)\right) \leq \alpha$, $\Pr\left(\frac{1}{P_{nU}} < \frac{1}{p/2}(1-\varepsilon)\right) \leq \alpha$	$\overline{X} \pm u_{p/2} S \left\{ 1 + \frac{u_\alpha \left(\frac{1}{2} + u_{p/2}^{-2}\right)^{1/2}}{\sqrt{n}} - \frac{\varepsilon}{u_{p/2}^2 (1-\varepsilon)} \right\}$

the sample standard deviation $S = 3.311$, while $u_{p/2} = 3.090, u_\alpha = 1.282$ and $n = 835$. We get

$$\begin{aligned} &\text{bias FAR} && \widehat{LCL}_N = 32.096, \widehat{UCL}_N = 52.635 \\ &\text{exceedance } FAR && \widehat{LCL}_N = 31.889, \widehat{UCL}_N = 52.842 \\ &\text{exceedance } ARL && \widehat{LCL}_N = 31.901, \widehat{UCL}_N = 52.830. \end{aligned} \tag{7.3}$$

7.2.2 The parametric chart

The parametric approach with the corresponding parametric control chart was discussed in Chapters 3 and 4. Here, we only present the results in explicit formulas using the following notations, where $ent(x)$ is the integer part of x, for instance $ent(0.95 \times 835 + 1) = 794$.

$$\begin{aligned} c(\gamma) &= \pi^{1/4} 2^{-(1+\gamma)/2} \Gamma(\gamma + 3/2)^{-1/2}, \\ C1(\gamma, u_{p/2}) &= -1.23 - 0.63\gamma + 0.73\gamma^2 + 0.74 u_p - 0.08\gamma u_{p/2} - 0.14\gamma^2 u_{p/2}, \\ C2(\gamma) &= \left(\frac{u_{a_n}}{u_{b_n}}\right)^{1+\gamma} - 2.4387^{1+\gamma} \\ \text{with } a_n &= 1 - \frac{ent\,(0.95n+1)}{n+1}, b_n = 1 - \frac{ent\,(0.75n+1)}{n+1}, \\ C3(\gamma, u_{p/2}) &= -76.37 - 120.12\gamma - 81.93\gamma^2 + 35.53 u_{p/2} + 53.71\gamma u_{p/2} \\ &\quad + 37.18\gamma^2 u_{p/2}, \\ A(\gamma, u_{p/2}) &= -4.00 - 12.54\gamma - 10.02\gamma^2 + 2.91 u_{p/2} + 6.47\gamma u_{p/2} + 4.42\gamma^2 u_{p/2}. \end{aligned}$$

The notations are for the two-sided limit. For the one-sided limit, $p/2$ in the

Table 7.2. Two-sided (corrected) parametric control limits based on the normal power family.

Aim	$\widehat{LCL}_P$ and $\widehat{UCL}_P$
$EP_n = p$	$\overline{X} \pm S\left\{c(\widehat{\gamma})u_{p/2}^{1+\widehat{\gamma}} - C1(\widehat{\gamma}, u_{p/2})C2(\widehat{\gamma}) + \frac{C3(\widehat{\gamma}, u_{p/2})}{n}\right\}$
$\Pr\left(P_{nL} > \frac{p}{2}(1+\varepsilon)\right) \leq \alpha$, $\Pr\left(P_{nU} > \frac{p}{2}(1+\varepsilon)\right) \leq \alpha$	$\overline{X} \pm S\left\{c(\widehat{\gamma})u_{p(1+\varepsilon)/2}^{1+\widehat{\gamma}} + \frac{A(\widehat{\gamma}, u_{p/2})u_\alpha}{\sqrt{n}}\right\}$
$\Pr\left(\frac{1}{P_{nL}} < \frac{1}{p/2}(1-\varepsilon)\right) \leq \alpha$, $\Pr\left(\frac{1}{P_{nU}} < \frac{1}{p/2}(1-\varepsilon)\right) \leq \alpha$	$\overline{X} \pm S\left\{c(\widehat{\gamma})u_{p/\{2(1-\varepsilon)\}}^{1+\widehat{\gamma}} + \frac{A(\widehat{\gamma}, u_{p/2})u_\alpha}{\sqrt{n}}\right\}$

notations are simply replaced by p. Moreover, the estimator of γ at the upper tail, $\widehat{\gamma}_U$, is given by

$$\widehat{\gamma}_U = 1.1218 \log\left(\frac{X_{(ent(0.95n+1))} - \overline{X}}{X_{(ent(0.75n+1))} - \overline{X}}\right) - 1.$$

Similarly, the estimator of γ at the lower tail, $\widehat{\gamma}_L$, is given by

$$\widehat{\gamma}_L = 1.1218 \log\left(\frac{\overline{X} - X_{(n-ent(0.95n))}}{\overline{X} - X_{(n-ent(0.75n))}}\right) - 1.$$

The two-sided control limits for the (corrected) parametric control chart are given in Table 7.2. The lower limit is denoted by $\widehat{LCL}_P$, while the upper limit is denoted by $\widehat{UCL}_P$. In case of the lower limit $\widehat{LCL}_P$ the $-$ sign should be read and $\widehat{\gamma} = \widehat{\gamma}_L$ should be inserted, while for $\widehat{UCL}_P$ the $+$ sign is used and $\widehat{\gamma} = \widehat{\gamma}_U$.

Application to the example mentioned before gives for $p/2 = 0.001, \varepsilon = 0.1, \alpha = 0.1$ the following two-sided control limits. We have $\overline{X} = 42.366$, $S = 3.311$, $n = 835$, $X_{(n-ent(0.95n))} = 36.54$, $X_{(n-ent(0.75n))} = 40.62$, $\widehat{\gamma}_L = 0.352$, $X_{(ent(0.95n+1))} = 47.03$, $X_{(ent(0.75n+1))} = 44.54$, $\widehat{\gamma}_U = -0.144$ while $u_{p/2} = 3.090$, $u_{p(1+\varepsilon)/2} = 3.062$, $u_{p/\{2(1-\varepsilon)\}} = 3.059$ and $u_\alpha = 1.282$. We get

$$\begin{array}{rl} \text{bias } FAR & \widehat{LCL}_P = 29.100, \widehat{UCL}_P = 51.606 \\ \text{exceedance } FAR & \widehat{LCL}_P = 28.306, \widehat{UCL}_P = 52.001 \\ \text{exceedance } ARL & \widehat{LCL}_P = 28.324, \widehat{UCL}_P = 51.994. \end{array} \qquad (7.4)$$

By comparing the parametric control limits in (7.4) with the normal control limits in (7.3) we can see that the lower control limits of the parametric chart are smaller than those of the normal chart. The reason for that is that we have a heavy lower tail and indeed $\widehat{\gamma}_L > 0$ (see Figure 6.1). As a consequence $\widehat{LCL}_P$ is much lower than $\widehat{LCL}_N$. In the upper tail the situation is reversed. We have a thin upper tail (see Figure 6.1) giving $\widehat{\gamma}_U < 0$ and hence the upper control limits are somewhat smaller in the parametric model than those in the normal control charts.

7.2.3 The nonparametric chart

The nonparametric approach with the corresponding standard nonparametric chart was presented in Section 5.2. A modified version of the standard nonparametric chart was discussed in Section 5.3. In Subsection 6.3.3, a different version of the modified non-parametric chart was presented and used as one of the components of the combined chart. Here, we use another version using the following notations. The given notations are for the two-sided control chart. For the one-sided control chart, $p/2$ in the control limits is simply replaced by p.

$$r = ent[(n+1)p/2], \delta = (n+1)p/2 - r,$$

$k(\varepsilon)$: smallest integer such that $Po\,(n(1+\varepsilon)p/2, r-1-k) \le \alpha$,
where $Po\,(\mu, x) = \Pr(Y \le x)$ with $Y \sim Poisson(\mu)$,

$$\lambda(\varepsilon) = \frac{\alpha - Po\,(n(1+\varepsilon)p/2, r-1-k(\varepsilon))}{Po\,(n(1+\varepsilon)p/2, r-k(\varepsilon)) - Po\,(n(1+\varepsilon)p/2, r-1-k(\varepsilon))},$$

$$X_{(0)} = X_{(1)} - S, X_{(n+1)} = X_{(n)} + S.$$

The two-sided control limits for the nonparametric control chart are given in Table 7.3.

The lower control limit is denoted by $\widehat{LCL}_{NP}$, while the upper control limit is denoted by $\widehat{UCL}_{NP}$.

Application to the example mentioned before gives for $p/2 = 0.001, \varepsilon = 0.1, \alpha = 0.1$, the following one-sided control limits. We have $r = 0, \delta = 0.836, r_1 = 1, \delta_1 = 0.835, k(0.1) = 0, k(0.1/(1-0.1)) = 0, \lambda(0.1) = 0.251, \lambda(0.1/(1-0.1)) = 0.253, X_{(1)} = 25.45, X_{(2)} = 26.54, X_{(384)} = 51.04, X_{(835)} = 51.66$ and $S = 3.311$. We get

Table 7.3. Two-sided nonparametric control limits

Aim	$\widehat{LCL}_{NP}$ and $\widehat{UCL}_{NP}$
$EP_n = p$	$\widehat{LCL}_{NP}: \; r = 0: \begin{cases} X_{(1)} - S & \text{with probability } 1-\delta \\ X_{(1)} & \text{with probability } \delta \end{cases}$ $r \geq 1: \; (1-\delta)X_{(r)} + \delta X_{(r+1)}$ $\widehat{UCL}_{NP}: \; r = 0: \begin{cases} X_{(n)} & \text{with probability } \delta \\ X_{(n)} + S & \text{with probability } 1-\delta \end{cases}$ $r \geq 1: \; \delta X_{(n-r)} + (1-\delta)X_{(n-r+1)}$
$\Pr\left(P_{nL} > \frac{p}{2}(1+\varepsilon)\right) \leq \alpha,$ $\Pr\left(P_{nU} > \frac{p}{2}(1+\varepsilon)\right) \leq \alpha$	$\widehat{LCL}_{NP} \begin{cases} X_{(r-k(\varepsilon))} & \text{with probability } 1-\lambda(\varepsilon) \\ X_{(r+1-k(\varepsilon))} & \text{with probability } \lambda(\varepsilon) \end{cases}$ $\widehat{UCL}_{NP} \begin{cases} X_{(n+k(\varepsilon)-r)} & \text{with probability } \lambda(\varepsilon) \\ X_{(n+k(\varepsilon)-r+1)} & \text{with probability } 1-\lambda(\varepsilon) \end{cases}$
$\Pr\left(\frac{1}{P_{nL}} < \frac{1}{p/2}(1-\varepsilon)\right) \leq \alpha,$ $\Pr\left(\frac{1}{P_{nU}} < \frac{1}{p/2}(1-\varepsilon)\right) \leq \alpha$	$\widehat{LCL}_{NP} \begin{cases} X_{(r-k(\varepsilon/(1-\varepsilon)))} & \text{with prob. } 1-\lambda(\varepsilon/(1-\varepsilon)) \\ X_{(r+1-k(\varepsilon/(1-\varepsilon)))} & \text{with prob. } \lambda(\varepsilon/(1-\varepsilon)) \end{cases}$ $\widehat{UCL}_{NP} \begin{cases} X_{(n+k(\varepsilon/(1-\varepsilon))-r)} & \text{with prob. } 1-\lambda(\varepsilon/(1-\varepsilon)) \\ X_{(n+k(\varepsilon/(1-\varepsilon))-r+1)} & \text{with prob. } \lambda(\varepsilon/(1-\varepsilon)) \end{cases}$

$$
\begin{aligned}
\text{bias } FAR \quad \widehat{LCL}_{NP} &= \begin{cases} 22.139 & \text{with prob. } 0.164 \\ 25.45 & \text{with prob. } 0.836 \end{cases} \\
\widehat{UCL}_{NP} &= \begin{cases} 51.66 & \text{with prob. } 0.836 \\ 54.971 & \text{with prob. } 0.164 \end{cases} \\
\text{exceedance } FAR \quad \widehat{LCL}_{NP} &= \begin{cases} 22.139 & \text{with prob. } 0.749 \\ 25.45 & \text{with prob. } 0.251 \end{cases} \\
\widehat{UCL}_{NP} &= \begin{cases} 51.66 & \text{with prob. } 0.251 \\ 54.971 & \text{with prob. } 0.749 \end{cases} \\
\text{exceedance } ARL \quad \widehat{LCL}_{NP} &= \begin{cases} 22.139 & \text{with prob. } 0.747 \\ 25.45 & \text{with prob. } 0.253 \end{cases} \\
\widehat{UCL}_{NP} &= \begin{cases} 51.66 & \text{with prob. } 0.253 \\ 54.971 & \text{with prob. } 0.747 \end{cases} .
\end{aligned} \tag{7.5}
$$

Comparing the numerical results with the previous control limits for the normal and normal power families, it is clearly seen that in particular the acceptance regions here are much wider than those for the normal and normal power families. Moreover, it is seen that in all situations the exceedance criterion is a more restrictive criterion than the bias criterion, leading to smaller lower and larger upper limits. The difference between FAR and ARL in the exceedance case is fortunately very small, implying that protection against an unwanted large FAR gives at the same time protection against an unpleasantly long ARL.

7.3 Implementing the combined approach

From the discussion presented throughout Chapters 2-5, we concluded that although every approach offers some advantages if it is used individually, it also bears some disadvantages, especially when the true distribution of the observations is (very) far from the supposed model being used. Since in real situations the true distribution of the observations is most likely unknown, we propose to combine those approaches and **let the data select the most suitable one**.

The setting is as follows. When the observations are close to normality, we want to select the normal control chart. If the departure from normality is too large, we apply the parametric control chart, unless the parametric family also does not fit. In the latter case the nonparametric control chart comes in.

From this setting we know that the critical point for selecting the most appropriate chart for the data depends on the determination of the "distance" between the true model and the supposed model. The distance represents the model error caused by replacing the true model of the observations with the supposed model. The question that remains is to decide the cut-off point for choosing a specific model. For this, a set of rules has been developed in Section 6.3 and is included in the following algorithm.

7.3.1 Algorithm

In this subsection, an algorithm to implement the combined approach in the field is presented in two parts. The first part is mainly concerned with the preliminary step which explains why some preparation is needed to monitor a production process. The second part, which is considered as the main step, will be given in the form of flowcharts which are ready to be translated into software. Once the practitioners decide to apply the combined approach, this algorithm provides them with a detailed guideline how to perform the steps. Although in the present chapter the algorithm is developed for the combined approach, essentially, the algorithm is also applicable for the individual approach contained in the combined approach: normal, parametric and nonparametric approach.

Preliminary step

In order to implement the concepts for monitoring the production process, the user needs to provide a set of Phase I observations, that is outcomes of independent r.v.'s $X_1, \cdots, X_n$ with common d.f. F. Once the "in-control" data are available, they can be used to construct the control limit. At this point the user may wonder, how many in-control data are needed to calculate such a control limit. Essentially, the answer of this question can be inferred from the discussion throughout Chapters 2 - 5. Without knowing the distribution of the observations, the user needs to have very many data to get an accurate control limit. As a general rule, the more data available to set up the control limit, the more accurate the control limit will be. Of course, the exact number depends on the criteria (bias or exceedance probability), types of function g as well as other parameters, such as p, ϵ and α. For example, if $g(p) = p$ with $p = 0.001$ and the bias criterion was applied, then around 40 observations would suffice for applying the normal chart, if normality holds. For the same case, at least 300 observations are required for applying the parametric chart. For applying the nonparametric approach, however, we need to have very many data. For example, for $p = 0.001$ and $n < 1000$, with probability as high as 0.499, the out-of-control signal will never occur, see (5.7) and (5.8). Hence, for applying the nonparametric chart, we have to either use much larger n for the same p, or to increase the value of p to, say, 0.01.

Main step

Having provided a set of in-control data from Phase I observations, the practitioners can continue with the main step described here. The main step is presented in the form of flowcharts that show actions to be performed systematically. The well-known basic structure of the algorithm which consists of two stages is given as a starting point in Figure 7.1. As seen in this figure, basically the structure can be divided into an estimation stage and a monitoring stage.

In the estimation stage, firstly n in-control data (from the preliminary step) are recorded and utilized as inputs of the flowchart. The next step is using these data to calculate the combined control limits. These limits determine a permissible area for new observations. The calculation of the upper control limit of the combined control chart in the estimation stage is given as a flowchart in Figure 7.2. This flowchart shows that first of all, a quantity which represents the distance between the true model and the actual model is calculated. As discussed in Section 6.3, for the quantity at the lower part we use $me_L = (\overline{X} - X_{(1)})/S$, while at the upper part we use $me_U = (X_{(n)} - \overline{X})/S$. Next, the so called normal area, which is the area in the interval $[u_{(-0.7+0.5\log n)/n},\ u_{5/\left(n\sqrt{n}\right)}]$, is calculated. If me_U is in this area then the normal chart can be applied. Otherwise, the so called parametric area, which is the area in the interval $[c\,(\widehat{\gamma})\,u^{1+\widehat{\gamma}}_{(-0.2+0.5\log n)/n},\ c\,(\widehat{\gamma})\,u^{1+\widehat{\gamma}}_{3/\left(n\sqrt{n}\right)}]$, is calculated. If me_U is inside this parametric area then the parametric chart is to be applied. Otherwise, since me_U is neither in the normal nor in the parametric area, then the nonparametric chart has to be applied. The normal, parametric and nonparametric charts to be selected are taken from Tables 7.1 - 7.3 according to the required aim: bias with FAR or exceedance probability criterion with FAR or ARL. For calculating the lower control limit of the combined chart we can apply the same flowchart with me_U is replaced by me_L.

Chapter 6 discusses the properties of the combined control chart extensively, both

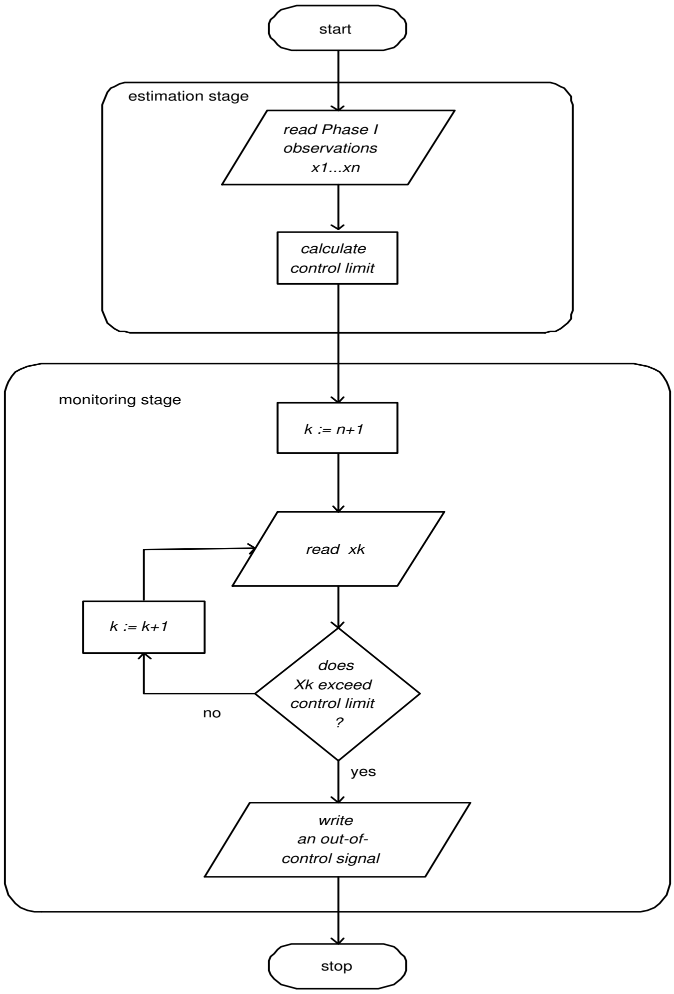

Figure 7.1. The basic structure of the algorithm.

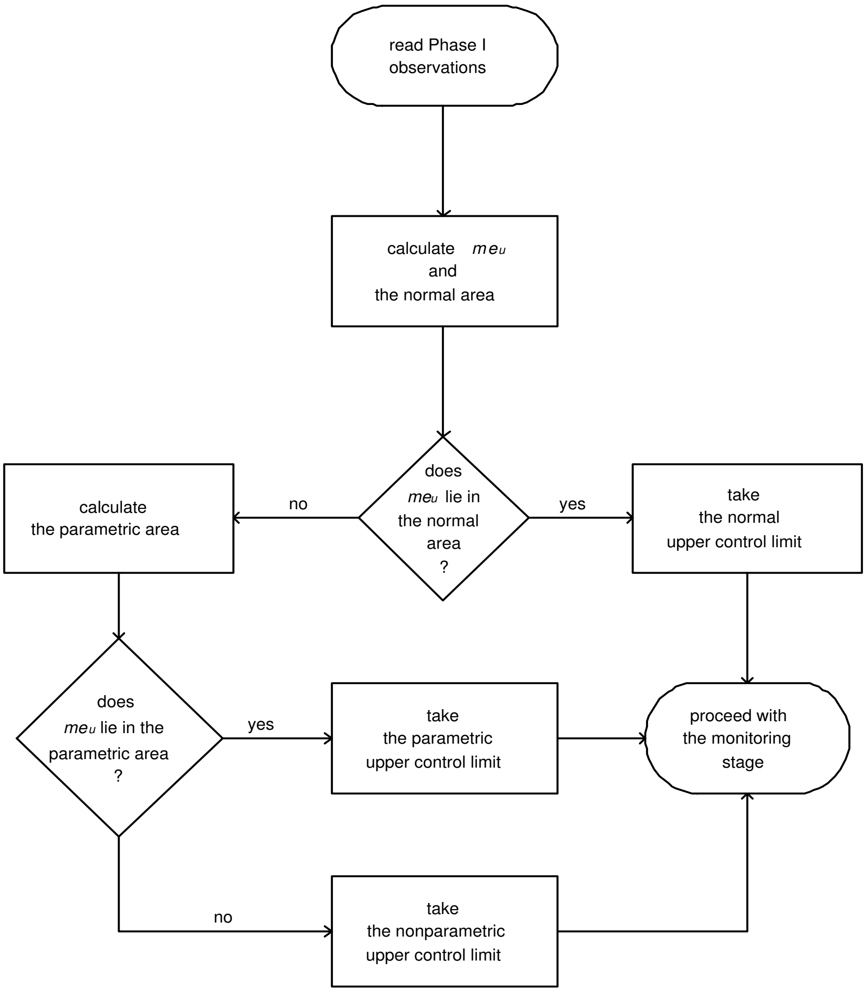

Figure 7.2. The calculation of the combined upper control limit in the estimation stage.

from a theoretical point of view and by simulations to see the finite sample behavior. Here are the conclusions.

1. Theoretically it is shown that the combined control chart in each of the three situations (normality, normal power family, outside the normal power family) asymptotically behaves as the specific corresponding control chart.
2. If normality holds, the combined control chart has a substantial gain w.r.t. the nonparametric control chart and only a small loss w.r.t. the normal and parametric control chart.
3. If the true distribution belongs to the normal power family, the combined control chart has, for not too large n and γ, a pretty gain w.r.t. the nonparametric control chart and in general only a small loss w.r.t. the parametric control chart; the normal control chart cannot be applied in that case unless γ is very close to 0; for positive γ's this is due to its bad in-control behavior and for negative γ's it has a very low alarm rate under an out-of-control situation.
4. Outside the normal power family, the combined control chart exhibits only a small loss w.r.t. the nonparametric control chart, while the normal control chart cannot be applied in this case due to its bad in-control behavior or its low alarm rate (unless the distribution is very close to normality); the same holds for the parametric control chart, albeit to a much smaller extent.

The control limit produced in the estimation stage then is used in the monitoring stage, see Figure 7.1. At this point the sketch of the calculated control chart can be drawn for the next stage. In the monitoring stage, continuously each new observation as a product of the running process, is compared with the calculated control limit. In Figure 7.1, the new observation is denoted by X_k; firstly we have X_{n+1}, then X_{n+2}, etc. Only if the observation exceeds the limit an out-of-control signal will be produced. Otherwise, the process is supposed to be in-control and hence it can be kept running. Note that the control limit is not changed (although in the mean time possibly a lot of observations are marked as "in-control" or, may be better expressed as observations not marked as "out-of-control"). The occurrence of the out-of-control signal requires the user to stop the process and find out the cause of this problem. Of course, it may also be a false alarm. The process can be restarted again provided that the problem has been solved.

7.4 Closing remarks

In this final section, the contents of the present chapter are summarized. In the first section, the background as well as the objective of this chapter are introduced. The following section provides the user with a brief review of the three types of control charts. To avoid technical difficulties, only some ready-made formulas are given in this section. Next, criteria to select the most suitable control charts for the observations are given. The selection process is depicted in flowcharts which can be implemented in software.

In-control data from the Phase I observations are used as inputs of the software. Based on these data, the software constructs the most suitable control limit for the observations. The control limits determine the permissible area for a new observation. During the monitoring process, the software continuously records and evaluates the new observation. Only if there exists an observation which falls outside the permissible control limit, an out-of-control signal will be produced. The occurrence of this signal presumably indicates an out-of-control situation, and the production process needs to be stopped to allow some fixation before it is restarted again.

Summary

Maintaining a production process in good condition is an important issue for practitioners. For that reason it is necessary to monitor such a process so that as soon as it starts to act differently from the required conditions a signal will be produced. With the occurrence of this signal, the production process will be stopped for an investigation or an adjustment before it is restarted again. This monitoring task is commonly based on the Shewhart $\overline{X}$-control chart which assumes that the observations are normally distributed. Although the Shewhart chart is widely used in practice because of its simplicity, applying this control chart to monitor the mean of a process may lead to two types of problems. The first concerns the typically unknown parameters involved in the distribution, while the second concerns the validity of the assumption of normality itself. The objective of the research is to study and find solutions for these problems. More specifically, our goal is to determine the most suitable control chart to be used in practice. For this, subsequently, so called normal, parametric, nonparametric and combined approaches are considered, leading to corresponding control charts.

In Chapter 2, we investigate the effect of parameter estimation in applying the Shewhart chart. In the case that the assumption is true that the observations are normally distributed but with unknown parameters, one has to estimate the parameters based on observations from the past. Since the estimation usually involves p-quantiles for very small p, it may result in large relative errors, unless (very) large sample sizes are applied. The results show that although the observations are normally distributed, we still need too large sample sizes to construct accurate control limits. Since such large sample sizes are often not available in practice, we need to find a way to correct the control limits in order to have accurate limits for commonly used sample sizes. It turns out that by adding suitable correction terms to the control limits, we are able to reduce substantially the sample sizes needed to construct accurate control limits. We apply both bias and exceedance probability criteria for setting the control limits. The bias criterion aims at reducing the systematic error between the actual result and the true value of the performance characteristic, thus dealing with the long term behavior of the chart in a series of separate applications, while the exceedance probability criterion is applied for controlling the probability of having a relative error larger than

a given value occurring in a single application.

In practice, the distribution of the observations can be different from the normal family. Hence, the normality assumption made for the observations often fails, which causes a model error. For such a case, we can use either a parametric or a nonparametric approach. In the parametric approach we have in mind that the distribution of the observations is in a neighborhood of a larger parametric model containing the normal family as a submodel. Several possible models have been considered for this purpose, such as random or deterministic mixtures. However, for a number of reasons we choose a so called normal power family as a suitable extension of the normal family. Based on the assumption that the true distribution of the observations is from the normal power family, we can construct the parametric chart. It turns out that this parametric chart is able to reduce the model errors of many distributions considerably. On the other hand, as more parameters are needed to be estimated in the normal power family, larger stochastic errors may be expected. Therefore, the next step to be taken is to control these stochastic errors by developing correction terms for the control limits. With the addition of the correction terms to the control limits, the corrected parametric chart performs much better since it reduces the stochastic errors substantially. As in the normal approach, here we apply the bias and exceedance probability criteria for setting the parametric control limits. The bias case is discussed in Chapter 3, while the exceedance probability case is presented in Chapter 4.

The parametric approach based on the normal power family offers a satisfactory solution over a broad range of underlying distributions. This approach, however, is not able to prevent the occurrence of unacceptably large model errors in cases where one has distributions far outside the normal power family. In such a situation nothing remains but to consider a nonparametric approach. The problem is left that huge sample sizes are required for the estimation step. Otherwise, the resulting stochastic error is so large that the control limits are very unstable, a disadvantage which seems to outweigh the advantage of avoiding the model error from the parametric case. In Chapter 5 we analyze under what conditions the nonparametric approach actually becomes feasible as compared to the parametric approach. In particular, corrected versions are suggested for which a possible change point is reached at sample sizes which are markedly less huge, but still larger than the customary range. These corrections serve to control the behavior during the in-control situation. The price for this protection clearly will be some loss of detection power during the out-of-control situation. A change point comes in view as soon as this loss can be made sufficiently small.

Although each of the three approaches has its own advantages, applying these approaches individually causes some disadvantages due to the specific characteristics of the observations encountered in the production process. Therefore, in Chapter 6 the three approaches are combined into one procedure in which the data, based on some criteria, will select the most suitable approach by themselves. As a result, a combined control chart can be recommended to be used in practice, since it performs very well under a great variety of distributions. Theoretical behavior of the combined control chart shows that the combined chart asymptotically behaves as each of the specific control charts in their own domain. From the theory and the simulation results, as well as from additional simulations that we have performed, we conclude that the new combined control chart can be recommended as an omnibus control chart with a good performance under a great variety of distributions, both for the in-control and for the out-of-control situation. Finally, Chapter 7 provides a detailed guideline to implement the combined chart in real life situations.

Samenvatting

Het is in de praktijk belangrijk een productieproces in de juiste toestand te houden. Daarom is het noodzakelijk het proces te volgen, zodat een signaal wordt gegeven wanneer het proces niet meer aan de gestelde voorwaarden voldoet. Op dit signaal zal het productieproces dan gestopt worden om het te onderzoeken en eventueel bij te stellen voordat het weer gestart wordt. Dit volgen is gewoonlijk gebaseerd op de Shewhart $\overline{X}$-controlekaart, waarvoor wordt aangenomen dat de waarnemingen normaal verdeeld zijn. Hoewel de Shewhart controlekaart in de praktijk veel gebruikt wordt vanwege zijn eenvoud, kan het toepassen van deze controlekaart om de verwachte waarde van een proces variable te controleren tot twee typen problemen leiden. Het eerste type betreft de parameters van de verdeling, die in het algemeen niet bekend zijn, terwijl het tweede type de geldigheid van de aanname van normaliteit zelf betreft. Het doel van het onderzoek is deze twee typen problemen te bestuderen en op te lossen. Meer specifiek willen we de meest geschikte controlekaart bepalen voor ieder gegeven praktijkgeval. Hiertoe beschouwen we achtereenvolgens de zogenaamde normale, parametrische, niet-parametrische en gecombineerde aanpak, die tot overeenkomstige controlekaarten leiden.

In Hoofdstuk 2 onderzoeken we het effect van parameterschattingen op het toepassen van de Shewhartkaart. In het geval dat men terecht aanneemt dat de waarnemingen normaal verdeeld zijn, maar met onbekende parameterwaarden, moet men de parameters schatten gebaseerd op waarnemingen uit het verleden. Omdat de schatters meestal p-kwantielen voor zeer kleine waarden van p betreffen, kunnen grote relatieve fouten optreden, tenzij (zeer) grote steekproeven worden gehanteerd. De resultaten laten zien dat, hoewel de waarnemingen normaal verdeeld zijn, er toch nog te grote steekproeven nodig zijn om nauwkeurige controlelimieten te construeren. Omdat in de praktijk grote steekproeven vaak niet beschikbaar zijn, moet daarom een manier gevonden worden om de controlelimieten te corrigeren, zodat nauwkeurige limieten resulteren voor steekproeven van een gebruikelijke omvang. Er blijkt dat door het toevoegen van geschikte correctietermen aan de controlelimieten dit doel bereikt kan worden. Om de controlelimieten vast te leggen gebruiken wij zowel het “bias” als het “exceedance probability” criterium. Het bias criterium beoogt de systematische afwijking tussen resulterende waarde en werkelijke waarde van de prestatiekarakteristiek te reduceren. Op deze wijze wordt het lange termijn gedrag van de kaart behandeld in een reeks van applicaties. Het exceedance probability criterium wordt toegepast om de kans op een relatieve fout groter dan een gegeven waarde onder controle te houden voor iedere afzonderlijke applicatie.

In de praktijk kan het zijn dat de verdeling van de waarnemingen niet tot de familie van de normale verdelingen behoort. De aanname van normaliteit van de waarnemingen is vaak onterecht, waarmee een modelfout wordt gemaakt. In zo'n geval kunnen we hetzij een parametrische, hetzij een niet-parametrische aanpak kiezen. In de parametrische aanpak stellen we ons voor dat de verdeling van de waarnemingen in de buurt ligt van een meer omvangrijk parametrisch model, waarvan de familie van normale verdelingen een sub-model is. Verschillende mogelijke modellen zijn hiertoe beschouwd, zoals random of deterministische mengvormen. Om een aantal redenen kiezen we echter een zogenaamde "normal power" familie als geschikte uitbreiding van de familie van normale verdelingen. We construeren dan de parametrische kaart onder de aanname dat de ware verdeling van de waarnemingen tot de normal power familie behoort. Het blijkt dat deze parametrische kaart de modelfouten van veel verdelingen aanzienlijk kan reduceren. Aan de andere kant is te verwachten dat er grotere stochastische fouten zullen optreden, omdat er meer parameters moeten worden geschat in de normal power familie. Een volgende stap is daarom deze stochastische fouten in de hand te houden door het ontwikkelen van geschikte correctietermen voor de controlelimieten. Met het toevoegen van correctietermen aan de controlelimieten verbetert de gecorrigeerde parametrische kaart aanzienlijk omdat de stochastische fouten substantieel gereduceerd worden. Net als voor de normale aanpak, passen we hier de criteria van bias en exceedance probability toe om de parametrische controlelimieten vast te leggen. Het geval waarin bias wordt beschouwd wordt besproken in Hoofdstuk 3, het geval van exceedance probability in Hoofdstuk 4.

De parametrische aanpak is gebaseerd op de normal power familie en biedt een tot tevredenheid strekkende oplossing voor een breed bereik van onderliggende verdelingen. Echter, deze aanpak kan niet verhinderen dat er onacceptabel grote modelfouten optreden bij verdelingen die duidelijk niet tot de normal power familie behoren. In een dergelijke situatie zit er niets anders op dan het volgen van de niet-parametrische aanpak. Er resteert dan het probleem dat voor de schattingsstap zeer grote steekproeven nodig zijn. Anders is de resulterende stochastische fout zo groot dat de controlelimieten niet stabiel zijn, een nadeel dat groter lijkt dan het voordeel dat de modelfout bij de parametrische aanpak vermeden wordt. In Hoofdstuk 5 wordt geanalyseerd onder welke condities de niet-parametrische aanpak hanteerbaar wordt in vergelijking met de parametrische aanpak. In het bijzonder worden gecorrigeerde versies voorgesteld waarvoor het omslagpunt bereikt wordt bij een duidelijk kleinere grootte van de steekproef dan voor de niet gecorrigeerde versie, maar die toch nog groter is dan gebruikelijk. Deze correcties dienen om het "in-control" gedrag in de hand te houden. De prijs voor deze bescherming is uiteraard verlies van detectievermogen in de "out-of-control" situatie. Een omslagpunt komt in zicht zodra dit verlies voldoende klein kan worden gehouden.

Hoewel ieder van de drie manieren van aanpak zijn eigen voordelen heeft, veroorzaakt het volgen van elk van deze aanpakken apart enige nadelen als gevolg van het specifieke karakter van de waarnemingen die we tijdens het productieproces tegen komen. Daarom worden in Hoofdstuk 6 de drie manieren van aanpak gecombineerd tot een procedure waarin de data, afhandelijk van de gestelde criteria, de meest geschikte aanpak zelf bepalen. Het resultaat, dat we de gecombineerde controlekaart noemen, kan aanbevolen worden in de praktijk, omdat deze zeer goed presteert voor een grote diversiteit van verdelingen. Het theoretische gedrag van de gecombineerde controlekaart laat zien dat dit asymptotisch overeenkomt met dat van elk van de specifieke controlekaarten voor verdelingen waarop zij van toepassing zijn. Zowel theoretische beschouwingen als simulatieresultaten, beschouwd in deze dissertatie alsook nog daarbuiten, leiden tot de conclusie dat de nieuwe gecombineerde controlekaart aanbevolen kan worden als een omnibus controlekaart met een goede prestatie voor een grote diversiteit van verdelingen, zowel voor de in-control als voor de out-of-control situatie. Tenslotte geeft Hoofdstuk 7 aanwijzingen voor de implementatie van de gecombineerde kaart in praktijksituaties.

Bibliography

[1] Albers, W. and Kallenberg, W.C.M. (2003). *New corrections for old control charts.* Memorandum 1694, Department of Applied Mathematics, University of Twente.

[2] Albers, W. and Kallenberg, W.C.M. (2004a). Estimation in Shewhart control charts. *Metrika* **59**, 207 - 234.

[3] Albers, W. and Kallenberg, W.C.M. (2004b). Are estimated control charts in control? *Statistics* **38**, 67 - 79.

[4] Albers, W. and Kallenberg, W.C.M. (2004c). Empirical nonparametric control charts: estimation effects and corrections. *J. Appl. Stat.* **31**, 345-360.

[5] Albers, W., Kallenberg, W.C.M. and Nurdiati, S. (2002). *Exceedance probabilities for parametric control charts.* Memorandum 1650, Faculty of Mathematical Sciences, University of Twente.

[6] Albers, W., Kallenberg, W.C.M. and Nurdiati, S.(2004a). Parametric control charts. *J. Statist. Plann. Inference* **124**, 159-184.

[7] Albers, W., Kallenberg, W.C.M. and Nurdiati, S.(2004b). Data driven choice of control charts. To appear in *J. Statist. Plann. Inference.*

[8] Andrews, D.F., (1973). A general method for the approximation of tail areas. *Ann. Statist.* **1**, 367-372.

[9] Bahadur, R.R. (1971). *Some Limit Theorems in Statistics.* SIAM, Philadelphia

[10] Barndorff-Nielsen, O.E. (1996). Normal inverse Gaussian distributions and stochastic volatility modelling. *Scand. J. Statist.* **24**, 1–13.

[11] Bickel, P.J. and Doksum, K.A. (2001). *Mathematical Statistics I*, 2nd ed., Pren-

tice Hall.

[12] Burr, I.W. (1979). *Elementary Statistical Quality Control.* Marcel Dekker, New York and Basel.

[13] Chakraborti, S. (2000). Run length, average run length and false alarm rate of Shewhart $\overline{X}$ chart: exact derivations by conditioning. *Commun. Statist. Simul. Comput.* **29**, 61-81.

[14] Chakraborti, S., van der Laan, P. and Bakir, S.T. (2001). Nonparametric statistical process control: an overview and some results. *J. Quality Technol.* **33**, 304-315.

[15] Chan, L.K., Hapuarachchi, K.P. and Macpherson, B.D. (1988). Robustness of $\overline{X}$ and R charts. *IEEE Trans. Reliability* **37**, 117-123.

[16] 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.

[17] de Haan, L., Sinha, A.K. (1999). Estimating the probability of a rare event. *Ann. Statist.* **27**, 732-759.

[18] Dekkers, A.L.M., de Haan L., (1989). On the estimation of the extreme-value index and large quantile estimation. *Ann. Statist.* **17**, 1795-1832.

[19] 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.

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

[21] Feller, W. (1971). *An Introduction to Probability Theory and Its Applications*, vol. II, 2nd ed., Wiley, New York.

[22] 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.

[23] Hall, P., Weissman, I., (1997). On the estimation of extreme tail probabilities. *Ann. Statist.* **25**, 1311-1326.

[24] Hillier, F.S., (1969). $\overline{X}$- and R-chart control limits based on a small number of subgroups. *J. Quality Technol.* **1**, 17-26.

[25] Huškova, M., (1988). Adaptive parameter estimation for generalized Tukey's λ-family. *Statistics* **19**, 15-26.

[26] Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Ass.* **58**, 13-29.

[27] 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.

[28] Jones, L.A., Champ, C.W. and Rigdon, S.E. (2004). The run length distribution of the CUSUM with estimated parameters. *J. Quality Technol.* **36**, 95-108.

[29] Lehmann, E.L. (1986). *Testing Statistical Hypotheses.* 2nd ed. Wiley, New York.

[30] Nagaev, A.V. (1969). Integral limit theorems for large deviations when Cramer's condition is not fulfilled I and II. *Theor. Probability Appl.* **14**, 51-64 and 193-208.

[31] Nagaev, A. V. (1979). Large deviations of sums of independent random variables. *Ann. Probab.* **7**, 745-789.

[32] Nedumaran, G., Pignatiello, J.J.Jr., 2001. On estimating $\overline{X}$ control chart limits. *J. Quality Technol.* **33**, 206-212.

[33] Pappanastos, E.A. and Adams, B.M. (1996). Alternative designs of the Hodges-Lehmann control chart. *J. Quality Technol.* **28**, 213-223.

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

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

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

[37] Roes, C.B. (1995). *Shewhart-type Charts in Statistical Process Control.* PhD-thesis, University of Amsterdam.

[38] 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.

[39] Willemain, T.R. and Runger, G.C. (1996). Designing control charts using an empirical reference distribution. *J. Quality Technol.* **28**, 31-38.

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

[41] Yang, C.H. and Hillier, F.S. (1970). Mean and variance control chart limits based on a small number of subgroups. *J. Quality Technol.* **2**, 9-16.

Index

Acknowledgement

All praises belong to God Almighty, Most Gracious and Most Merciful. Only with His will I got a lot of help, support and encouragement from individuals as well as from organizations so that I could finish my study in the Department of Applied Mathematics, Faculty of Electrical Engineering, Mathematics and Computer Science at the University of Twente on time. For that, I would like to sincerely express my gratitude to those who helped me in accomplishing this thesis. May God bless all of them.

First of all, I will always be very grateful to my promoter prof. dr. Wim Albers and my co-promoter dr. Wilbert Kallenberg, who are also my supervisors, for the opportunity to work in this project as well as for their very many ideas, great support and incessant help during the execution of this project.

Next, I would like to cordially thank the Netherlands Organization for Scientific Research (NWO) for the financial support of this project. I would also like to thank dr.ir. J. Praagman from CQM Eindhoven for providing the electric shaver data used in this thesis, and to thank ir. Ramun Kho from Philips Semiconductor, Nijmegen for sharing practical information related to this project.

A special word of thanks goes to prof. dr. Kees Hoede for giving moral support throughout the difficult times and for doing the "samenvatting" of this thesis. I also received help from the side of members of the Department (or formerly the Faculty) of Applied Mathematics. In particular I should mention secretarial support from Thyra, Anja, Mirande, Carla and Ellen and other practical and technical support from Johan, Michel, Henk and Henri.

Many others helped me in some way or another, but it is impossible to name all these people. However, I would like to deeply thank all my friends in IMEA, PPIE and Anugrah's catering for making my life in Enschede more enjoyable and memorable. Also many thanks go to all my friends in IAMS-N and KZIS for giving me encouragement and support in finishing this program.

Finally, I would like to dedicate this thesis to my family who constantly pray for me. Their love, patience and understanding have lightened up my spirit to finish this thesis.

Curriculum Vitae

Sri Nurdiati was born in Malang, Indonesia, on November 26, 1960. After completing her secondary school at the Sekolah Menengah Atas Negeri 3 (SMAN 3) Malang in 1980, she continued her study at the Bogor Agricultural University, Indonesia. She has always been interested in education since she loves teaching. Because of that, she directly accepted an offer to be a teaching assistant when she got her Bachelor's degree in Statistics in 1984. Since then, she has been a staff member of the Department of Mathematics at that university.

While doing the assistantship in 1985-1987, she attended the Master Program in Applied Statistics at the same university. In 1988 - 1991, she joined the MSc. Program in Computer Science at the University of Western Ontario, Canada. After finishing her Master Program, she returned to Indonesia, where shortly after that she got a part time position as a Management Information System consultant in a private company in Jakarta. She worked there for 6 years (1992 - 1998), while at the same time she was a lecturer at the Bogor Agricultural University.

From September 2000, she has been a PhD student in the group of Statistics and Probability, in the Department of Applied Mathematics at the University of Twente. The results of her research which are supervised by her promoter prof. dr. Wim Albers and her co-promoter dr. Wilbert Kallenberg are presented in this thesis. After her graduation she plans to return to her position in Indonesia as a lecturer at the Department of Mathematics, Bogor Agricultural University.