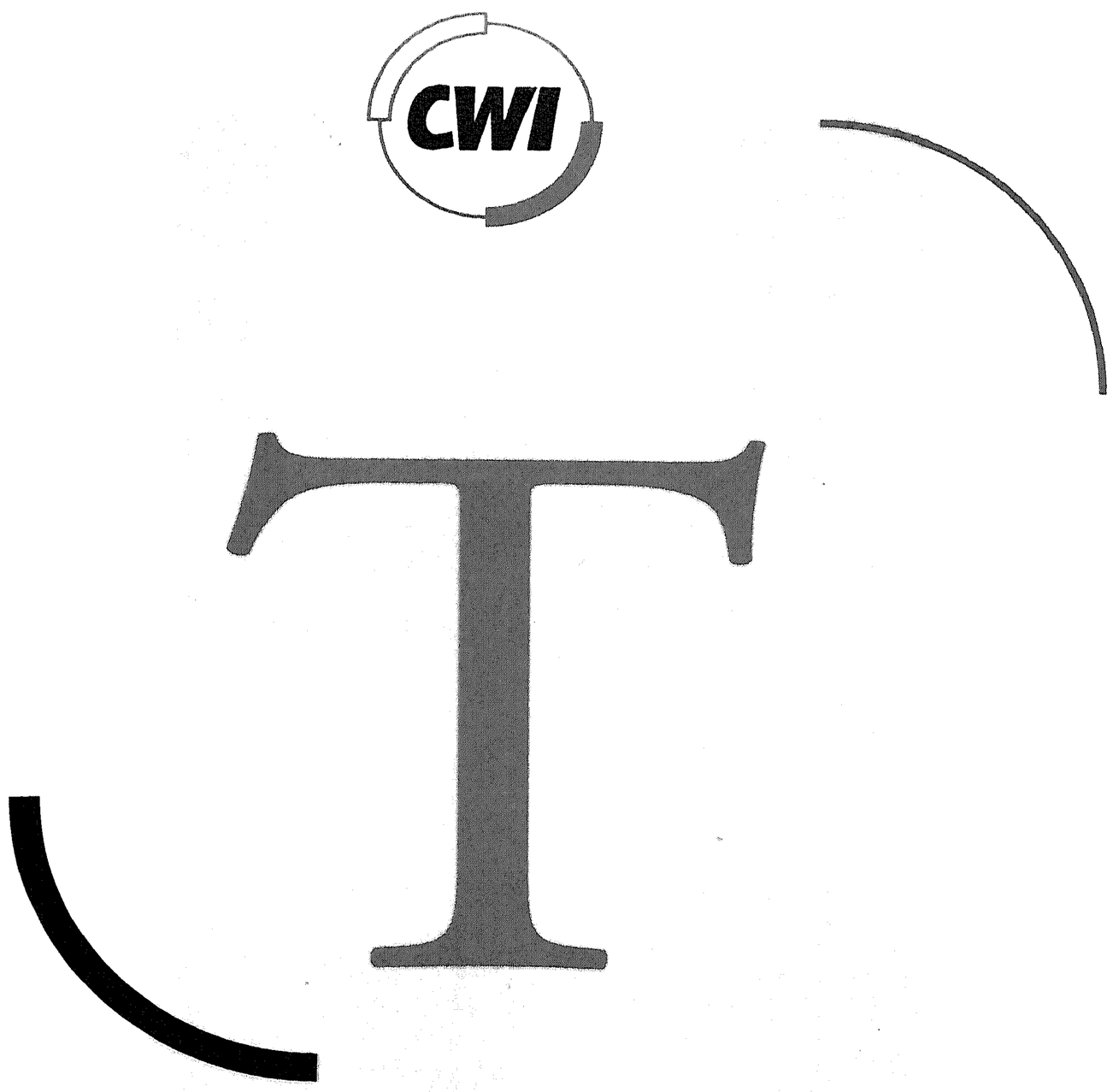


Statistical test limits in quality control

G.D. Otten

116

Centrum voor Wiskunde en Informatica
CWI TRACT

CWI Tracts

CWI
P.O. Box 94079, 1090 GB Amsterdam, The Netherlands
Telephone + 31 - 20 592 9333
Telefax + 31 - 20 592 4199
URL `http://www.cwi.nl`

CWI is the nationally funded Dutch institute for research in Mathematics and Computer Science.

1991 Mathematics Subject Classification: 62N10, 62E20, 62G35
ISBN 90 6196 469 5
NUGI-code: 811

Contents

Acknowledgement

There are a lot of people who contributed to this tract. During my workplacement at Philips Semiconductors Nijmegen I got acqainted with the subject of this tract, and in a very pleasant and effective cooperation with Wim Albers en Wilbert Kallenberg I have been working on this subject then at the Faculty of Applied Mathematics of the University of Twente. I am particularly grateful for the accuracy with which Wim and Wilbert commented on my ideas and contributions. In addition, it was nice to work in the pleasant atmosphere at the faculty.
I have enjoyed the motivating interest of Ronald Does.

I visited Philips Semiconductors Nijmegen several times during the period of research, and have very much appreciated the discussions with Ger van Boxem. The data for the example in chapter 3 have been provided by Ger. I would like to thank Maarten Hasselaar (CQM, Philips Semiconductors Nijmegen) for the data for the example in chapter 6.

Introduction

For certain types of products requirements as to their quality are formulated by fixing specification limits for a number of characteristics. If one of the characteristics does not satisfy the specification, the producer should not deliver that part to the customer. In complicated production processes it often occurs that a (small) fraction of the manufactured parts does not satisfy the specifications, despite efforts to reduce the variability within the production process. To ensure that this fraction is not shipped to the customer, the products are inspected on the characteristics specified. A typical example is the manufacturing of Integrated Circuits (IC's), for which specification limits are agreed on for a large number of characteristics and each part is tested before delivery.

In general, the measurements will not give the true value of the characteristic. Due to measurement errors the measurement could wrongly suggest that the characteristic is conforming. To prevent accepting too many nonconforming products in this way, a somewhat more stringent test limit is set by the producer. The more stringent this test limit, the smaller the number of accepted nonconforming products, but, at the same time, the larger the number of products that are needlessly rejected. Usually, it is considered unnecessary to make sure that not a single nonconforming product passes the test. In particular, when products are manufactured in very large quantities, as is the case with IC's, it is acceptable if a small fraction has been passed wrongly. It is common practice to agree with the customer on a bound on the probability of wrongly accepting a product, denoted by γ, together with a specification limit.

So, given a specification limit and a value of γ, a test limit has to be determined. Besides by the value of γ, the distance between test limit and specification limit is determined by two more factors. Obviously, the magnitude of the measurement error is of importance. The smaller the measurement error, the closer the test limit can be set to the specification limit. Furthermore, the fraction of products for which the characteristic is around the specification limit plays a part. If there are simply no products of which the characteristic is around the specification limit, then there is no risk of wrongly accepting a product either.

Although the problem has been given some attention in the literature, in practice test limits are determined mainly in an informal way. Often the test

limit is taken simply 'somewhat' stricter, in which 'somewhat' is not further specified. The so called '3σ-limit' is often applied as well. Then the distance between test and specification limit is set at three times the estimated standard deviation of the measurement error. Naturally, if test limits are determined in one of these ways, the three factors mentioned (the bound γ, the magnitude of the measurement error and the fraction of products around the specification limit) are hardly taken into account. It turns out that test limits determined in these ways are often too strict and, consequently, an unnecessary loss of yield is suffered.

In this book procedures are derived to determine accurate test limits. This means that procedures are derived for which the obtained test limit gives the largest yield feasible, while the value of γ (the bound on the probability of a product being both nonconforming and accepted) is observed. In particular, if products are manufactured in very large numbers, a slightly less strict test limit already leads to important improvements in the yield.

To determine test limits, the inspected characteristic and the measurement error are considered random variables. In the literature appeared so far attention has been paid only to the situation in which both the characteristic and the measurement error are normally distributed. The first practical complication, which has not yet been studied, is that even if the assumption of normality is justified, the corresponding parameters (mean, variance) will be unknown. Estimation itself of the parameters is not a problem, however, the accuracy of the obtained test limit depends on the accuracy of the estimators. In chapter 3 of this book ample attention is paid to the complication when for the determination of the test limits estimators of the parameters have to be used. Test limits which allow for the fact that they are based on estimators are derived.

In many situations in practice the assumption of normality for the measurement error and/or the characteristic is not justified. In chapter 4 test limits are derived for the situation in which the measurement error can still be assumed to be normally distributed, but in which the characteristic is not necessarily normally distributed. Chapter 5 is dedicated to a specific part of the determination of test limits. To obtain an accurate test limit, it is essential to estimate the density of the characteristic at the specification limit correctly. In chapter 5 several ways to estimate the density in one point are discussed and compared. In chapter 6 test limits are derived for the general situation: normality for the measurement error is no longer assumed either.

In chapter 2 the results are summarized for application in practice. The test limits for the different situations are presented in the form of a manual.

Chapter 1

Test limits and consumer loss

In many production processes large numbers of parts have to be inspected on various characteristics, for each of which a specification limit has been set. The specification limit is the largest or smallest value of a characteristic for which a product may be shipped to the customer. In semiconductor industry, for example, integrated circuits (IC's) have specified values for many measurements. Usually, these values are agreed on with the customer or published in presentations. This is,

In the eighties a trend away from such inspection procedures has been started. Many industries have made great effort to improve the control of manufacturing processes and to establish working conditions for continuous improvements. In this respect, statistical process control and design of experiments receive much attention. The focus in these fields is on better control of variation of the product characteristics. In areas such as semiconductor industry, however, inspection of products is still inevitable. In the manufacturing of IC's, relations between process characteristics and product characteristics are not always very well known. Therefore process control by itself will not give sufficient safeguards, and IC's are still 100% tested on the characteristics specified.

Typically, the measurement process used during inspection will not be infallible, although one tries to reduce its variability. It is therefore common practice to apply a somewhat more stringent test limit, instead of the specification limit itself, to decide whether to accept a product. By accepting only those products which satisfy this test limit, the producer hopes that the consumer loss, which is the probability of obtaining products that are both bad and accepted, will stay below a prescribed bound γ. Such bounds are often required to be very small (10 - 100 ppm, parts per million), especially when many characteristics are measured separately and parts are supposed to pass all of these tests, as is the case with IC's, for example.

The choice of the test limit is an important practical problem. If it is chosen too close to the specification limit, the bound γ on the consumer loss may be

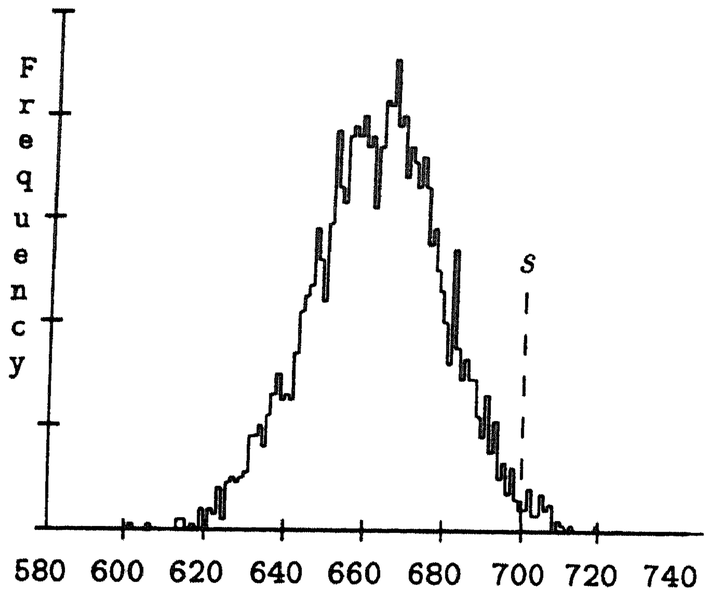
violated. On the other hand, if the limit is chosen too conservative, one will be faced with an unnecessary large producer loss, which is the probability of obtaining products that are both conforming and rejected. Clearly this will lower the yield – the probability of obtaining products that are accepted and thus delivered. In production processes of the type mentioned before, a loss of yield of 0.1% can already be quite costly. Nevertheless, it is still quite common to set test limits in the informal way described previously – just 'somewhat' more stringent than the specification limit.

1.1 The problem of finding correct test limits

By way of example, consider the following situation. The specification limit for a certain characteristic is agreed on at 700.0 and products should only be approved if the characteristic is smaller than this specification limit. A 100% inspection is carried out by the producer. Together with the specification limit a bound γ on the consumer loss is fixed.

Suppose that the following histogram represents the measurements from production of the inspected characteristic.

Figure 1.1.1 Histogram for measurements from production



Because the measurements are contaminated by a (small) measurement error, the production data cannot be considered as the true values of the characteristic. Besides measurement errors, imprecision, for example due to round off errors, may contaminate production data. Consequently, testing on the basis of the specification limit will lead to products which are nonconforming but accepted, and thus a test limit is placed on the left side of s to make sure that this will happen not too often. The problem is where to put this limit.

By setting the test limit far left of the specification limit, perhaps not one nonconforming product will be accepted. There is no need, however, to be more strict than required by the bound on the consumer loss. It is clear that to maximize the yield, the test limit should be as large as possible. The location of the test limit also depends on the magnitude of the measurement error. The smaller the measurement error, the closer the test limit can be set to the specification limit. Finally, the fraction of products around the specification limit is of importance, because these are products that can be judged incorrectly. The smaller this fraction, the less stringent the test limit may be.

There are several, more or less, informal ways to determine the location of the test limit. For a clear discussion it is useful to formulate the problem more precisely. We define the random variable X as the true value of the inspected characteristic, but because of a small random measurement error U, the production data are actually observations from

$$\tilde{X} = X + U. \tag{1.1.1}$$

We assume that the specification limit, denoted by s, is given. The limit is such that products should be rejected if the characteristic has a value larger than s. It is often convenient to use the following quantity,

$$\pi = P(X > s), \tag{1.1.2}$$

which is the probability of a product being nonconforming. To fix ideas, in many applications π will be in the range from 0.001 to 0.15.

The test limit we are looking for should be as large as possible to maximize the yield, defined as

$$\mathit{Yield} = P(\tilde{X} < t), \tag{1.1.3}$$

which is the probability of accepting a product. On the other hand, the consumer loss, defined as

$$CL = P(X > s,\ \tilde{X} < t),$$

should not exceed the prescribed bound γ. Hence the optimal choice for t is obtained by letting $t = t_e$, where t_e is such that the consumer loss equals γ exactly, that is

$$P(X > s,\ \tilde{X} < t_e) = \gamma. \tag{1.1.4}$$

From this definition it is immediately clear that it is rather difficult to actually find the test limit we are looking for. This is probably also the main reason why informal methods to set the test limit are still popular.

Perhaps the most informal way to set a test limit, is to take simply

$$t = s - b, \tag{1.1.5}$$

for some small constant b. In the example above we could choose to take $t = 695$. This approach, of course, fully ignores the behavior of the measurement error and it takes not into account the bound on the consumer loss.

A small improvement is to use the standard deviation of the measurement error, which we denote by σ_U, and take

$$t = s - k \cdot \sigma_U, \tag{1.1.6}$$

for some given k. Common choice is $k = 3$ leading to the so called '3σ-limit'.

It is relatively easy to find a conservative approximation to the value of t_e in which all of the three relevant factors — the bound γ, the measurement error and the location of the specification limit in relation to the production data — are involved.

From (1.1.4) and (1.1.2) it follows that

$$P(\tilde{X} < t_e \mid X > s) = \gamma/\pi. \tag{1.1.7}$$

In view of (1.1.1) this leads to

$$\gamma/\pi = P(U < t_e - X \mid X > s) \leq P(U < t_e - s \mid X > s) = P(U < t_e - s),$$

where we tacitly assumed that X and U are independent. Following standard practice, let us make the assumption that U is normally distributed with mean 0 and variance σ_U^2, which we denote by $U \sim N(0, \sigma_U^2)$. Let Φ denote the standard normal distribution function and Φ^{-1} its inverse, then it follows that $\gamma/\pi \leq \Phi\left(\frac{t_e - s}{\sigma_U}\right)$, hence that $\Phi^{-1}(\gamma/\pi) \leq \frac{t_e - s}{\sigma_U}$ and, consequently, that by

$$t_c = s - \Phi^{-1}(1 - \gamma/\pi) \cdot \sigma_U \tag{1.1.8}$$

a conservative approximation to t_e is obtained. The advantage of this approximation is its simplicity, but in the derivation the simplifying assumption is made that each nonconforming part has $X = s$, rather than $X > s$.

When besides U also the characteristic is normally distributed, $X \sim N(\mu_X, \sigma_X^2)$, the measurement $\tilde{X}$ is $N(\mu_X, \sigma_X^2 + \sigma_U^2)$ distributed (assuming X and U are independent) and $(X, \tilde{X})$ has a bivariate normal distribution, $(X, \tilde{X}) \sim N(\mu_X, \mu_X, \sigma_X^2, \sigma_X^2 + \sigma_U^2, \rho)$, where the correlation coefficient ρ equals $(1 + \sigma_U^2/\sigma_X^2)^{-1/2}$. In this situation the consumer loss is thus simply a function of the specification limit s, the test limit t and the three parameters μ_X, σ_X and σ_U. Consequently, the test limit t_e for which the consumer loss equals γ exactly, can be determined numerically.

This approach has been proposed by Mullenix (1990). A first complication is that even if X and U are normally distributed, in practical situations only s and γ are given while μ_X, σ_X^2 and σ_U^2 often have to be estimated. When a very large number of observations is available for estimation, one may expect that the test limit obtained by plugging in the estimates instead of the true values will indeed be close to t_e. In practical situations especially the number of

observations to estimate σ_U is often not very large, however. It is not guaranteed that for moderate sample sizes the obtained test limit is close to t_e and, more importantly, that the resulting consumer loss is close to γ.

1.2 Improvements in yield

We have discussed several methods to set a test limit. An interesting question is what gains in yield are achieved if a more advanced test limit is applied, rather than a crude approximation. Because, if it were negligible, why should we bother?

With the test limits from (1.1.5) and (1.1.6) the consumer loss is not at all in control and hence not suitable, but with t_c from (1.1.8) the bound γ is certainly not violated, at least, if U is normally distributed, X and U are independent and the parameters π, cf. (1.1.2), and σ_U are known. For the sake of illustration let us assume $X \sim N(0,1)$ and $U \sim N(0,\sigma^2)$, and compare the yield obtained when using t_e rather than t_c. In table 1.2.1 we present the gains achieved in yield for various choices of π, γ and σ.

Table 1.2.1 The gains in yield by using t_e, rather than t_c
The gain in yield is given in %, for various combinations of π, γ and σ.

	(π,γ)			
σ	(0.10,10ppm)	(0.05,20ppm)	(0.01,40ppm)	(0.0025,100ppm)
0.01	0.29	0.18	0.06	0.02
0.05	1.11	0.70	0.20	0.07
0.10	2.14	1.38	0.40	0.12
0.20	4.32	3.04	0.98	0.27

It is easily shown that for $X \sim N(\mu_X, \sigma_X^2)$ and $U \sim N(0, \sigma_U^2)$ with $\sigma_U = \sigma \cdot \sigma_X$ the same figures follow. From table 1.2.1 we conclude that using the simple approximation t_c, rather than t_e, can lead to a considerable loss of yield. Therefore, with respect to further investigations, for example the consequences of estimating parameters, we need approximations to t_e which are much more accurate than t_c.

1.3 Literature on test limits

The problem of controlling the misclassification probabilities – incorrectly accepting a part and incorrectly rejecting a part – while carrying out a 100% inspection has been discussed over a long period of time and from various points of view. All of the contributions have in common the assumption that both the characteristic and measurement error are normally distributed. In the problem

of setting test limits no attention has been paid to the consequences of using estimators of parameters rather than the true values.

One of the first papers in which application of a test limit relative to a specification limit is discussed, comes from Grubbs and Coon (1954). They consider, among other things, determination of test limits if the consumer and producer loss should be equal or if the sum of the two should be minimal. In both cases the test limit appears to be often less stringent than the specification limit. Nowadays customers seem to be more demanding, however.

Mee, Owen and Shyu (1986) give confidence bounds for the misclassification probabilities for the situation in which not a separate test limit is set, but the specification limit itself is applied.

Easterling, Johnson, Brement and Nachtsheim (1991) discuss various measures of consumer risk, for example the consumer loss, and evaluate the measures with respect to computational convenience of the corresponding test limit, information requirements and economic considerations.

Many further references are found in these papers.

1.4 Estimation of parameters and nonnormality

To get an impression of the consequences involved by estimation of parameters, assume again that both the characteristic and measurement error are normally distributed and independent, $X \sim N(\mu_X, \sigma_X^2)$ and $U \sim N(0, \sigma_U^2)$, but μ_X, σ_X and σ_U are unknown. If we have to rely on estimators of the parameters, trivially, the obtained estimates cannot be considered as the true values of the parameters. Compare this complication with the following standard textbook problem. We have a random variable $Z \sim N(\mu, \sigma^2)$, where both μ and σ are unknown. Based on a sample $Z_1, \ldots, Z_n$ a confidence interval for μ has to be determined. We compute sample mean $\bar{Z}$ and sample variance S^2. Then, although we still have that $(\bar{Z} - \mu)/(\sigma/\sqrt{n}) \sim N(0,1)$, we should use instead that $(\bar{Z} - \mu)/(S/\sqrt{n})$ has Student's distribution to find the confidence interval, as σ is unknown. Simply plugging in S for σ and continuing as if it were the true value, is not correct.

While the difference between the standard normal distribution and Student's distribution with n degrees of freedom is negligible for practical purposes if $n > 30$, it turns out that to determine test limits much more observations are needed before the estimates of μ_X, σ_X and σ_U may be identified with the true values. In fact, in chapter 3 it is shown that if t_e is computed by plugging in estimates which are based on moderate sample sizes (cf. the discussion at the end of section 1.1), average consumer losses up to 2γ can occur.

If X or U is not normally distributed, the relation between the consumer loss and the parameters μ_X, σ_X and σ_U does not hold. Certain quantities remain to be estimated, however, and similar problems occur. It is very well known that

for several reasons (outliers, skewness, heavy tails, bimodality etc.) normality fails quite often.

1.5 Test limits which allow for estimation

When the test limit is based on estimators of parameters, the test limit itself and the resulting consumer loss actually are random variables. Since the true values of the parameters are unknown, we should no longer try to find a test limit which leads to a consumer loss exactly equal to γ, as for instance, we should not try to find an estimator $\hat{\mu}_X$ exactly equal to μ_X.

We will derive two types of test limits for the case where estimators are used. The first test limit is to be applied if the consumer loss should be equal to γ in expectation. If estimates are obtained regularly, for example for each new batch of products, the long run average of the consumer loss then will tend to γ, notwithstanding the (considerable) variation between batches. If one and the same consumer receives all batches this will be quite satisfactory. In general, a consumer loss with average γ (and a lot of spread) is not advisable. For, we can hardly expect that a consumer who complains about receiving 215 ppm defectives rather than the 100 ppm which was agreed, will be soothed much if he is told that his competitor received only 30 ppm, thus making the average more correct! In this kind of situation a somewhat conservative test limit is appropriate, such that the bound γ is violated only with a small probability. We will derive such a test limit as well.

1.6 Outline of this book

In this book the investigations are presented on the basis of various assumptions on the distributions of the characteristic and measurement error. In chapter 2 the results are summarized in the form of a manual.

The following four situations have been considered.

- Both the measurement error and the characteristic are normally distributed (chapter 3).
- The measurement error is normally distributed but the type of distribution of the characteristic is unknown (chapter 4).
- The characteristic is normally distributed but the type of distribution of the measurement error is unknown (chapter 6).
- Both the measurement error and characteristic have an unknown type of distribution (chapter 6).

It is not surprising that in the course of finding correct test limits, several more or less isolated problems were encountered. A problem which was studied extensively emerged from the following. If the distribution of X is not normal, especially the reciprocal of the density of X at the specification limit is of

interest. Since estimation of this quantity is not a problem which has received much attention in the literature, a lot of effort has been put in finding the way in which this quantity can be estimated optimally. Several nonparametric estimators have been investigated (cf. chapter 4). Moreover, the most suitable nonparametric estimator has been compared to several parametric estimators. The results, which are heavily based on simulations, are presented in chapter 5.

Chapter 2

Implementation in practice

Based on the theory in the next chapters we will provide here the procedures to set test limits as they can be applied in practice. Although from a theoretical point of view it is obvious, we want to emphasize that the data used to compute the test limit must be obtained from a stable process. The computed test limit has lost its accuracy if a modification occurs in the production process.

Corresponding to most practical situations we consider a one-sided specification interval. The specification limit, denoted by s, is given and the test limit, denoted by t, is somewhat more stringent than s. The following random variables are used,

- X refers to the true value of the inspected characteristic,
- U stands for the measurement error,
- $\tilde{X} = X + U$ is the observed value of the inspected characteristic.

By γ we denote the prescribed bound on the consumer loss.

To apply the procedures, the following two conditions on X and U should be satisfied.

- X and U are mutually independent,
- the standard deviation of U is much smaller than the standard deviation of X. (As a rule, less than one third.)

In four sections we will consider the following situations.

- Both the measurement error and the characteristic are normally distributed (section 2.1)
- The measurement error is normally distributed but the type of distribution of the characteristic is unknown (section 2.2)
- The characteristic is normally distributed but the type of distribution of the measurement error is unknown (section 2.3)
- Both the measurement error and characteristic have an unknown type of distribution (section 2.4)

In each of these sections expressions for two test limits are given, one test limit in order to obtain a consumer loss which is in expectation equal to the prescribed bound γ (referred to as 'unbiased estimation') and a test limit in order to obtain a consumer loss which exceeds γ with a probability α only (referred to as 'confidence interval approach'). The test limits are denoted by t_u and t_i, respectively.

There is a slight difference in the formulas between the situation in which products are nonconforming if $X > s$ and the situation in which products are nonconforming if $X < s$. Therefore, two expressions are given for each type of the test limit.

The accuracy of the test limits as well as the question whether the chosen estimators are optimal is not a point of discussion here. For that we like to refer the reader to the relevant chapters in this book where it is discussed extensively.

For the purpose of review the following sections are divided into seven parts:

1. A short description of the **situation**.
2. Expressions for the **estimators** of the parameters.
3. Expressions for the **test limits** t_u and t_i.
4. Some remarks on the actual **computation** of the test limits.
5. A **numerical example**.
6. A guide for the **number of observations** for estimation.
7. A **reference** to the chapter in which the mathematically interesting aspects and formal justification of the procedures are treated.

All calculations to compute the test limits can be performed on a pocket calculator, we assume however that either a table or a numerical procedure is available to evaluate the standard normal distribution.

About the notation, we remark that a " ˆ " on top of a parameter or a function refers to an estimator of the parameter or a function of estimators. The standard normal distribution and its density are denoted by Φ and ϕ, respectively.

2.1 Both the characteristic and measurement error are normally distributed

Situation The measurement error is $N(0, \sigma_U^2)$-distributed.
The characteristic is $N(\mu_X, \sigma_X^2)$-distributed.
The parameters σ_U, μ_X and σ_X are not known.

Estimators The estimator of σ_U is based on n repeated measurements of the characteristic with two replications, $\tilde{X}_{i1}$ and $\tilde{X}_{i2}$ $(i = 1, \ldots, n)$, and is given by

$$\hat{\sigma}_U^2 = \frac{1}{2n} \sum_{i=1}^{n} (\tilde{X}_{i1} - \tilde{X}_{i2})^2.$$

With respect to estimation of μ_X and σ_X^2 we consider two possibilities. There are either $m - n$ additional observations $\tilde{X}_{i1}$ $(i = n+1, \ldots, m)$ from production available or not. If the extra observations are not available, the estimators of μ_X and σ_X^2 are,

$$\hat{\mu}_X = \frac{1}{n} \sum_{i=1}^{n} \tilde{X}_{i\bullet},$$

$$\hat{\sigma}_X^2 = \frac{1}{n-1} \sum_{i=1}^{n} (\tilde{X}_{i\bullet} - \hat{\mu}_X)^2 - \frac{1}{2} \hat{\sigma}_U^2,$$

where $\tilde{X}_{i\bullet} = (\tilde{X}_{i1} + \tilde{X}_{i2})/2$.
In the other situation,

$$\hat{\mu}_X = \frac{1}{m} \sum_{i=1}^{m} \tilde{X}_{i1},$$

$$\hat{\sigma}_X^2 = \frac{1}{m-1} \sum_{i=1}^{m} (\tilde{X}_{i1} - \hat{\mu}_X)^2 - \hat{\sigma}_U^2.$$

Test limits In case of unbiased estimation and when items should be approved if $X < s$, the test limit is given by

$$\hat{t}_u = s - (\hat{a}_1 - \hat{c} + \hat{c}_u) \cdot \hat{\sigma}_U,$$

$$\text{with} \quad a_1 \text{ such that } g_1(a_1) = \frac{\gamma}{\sigma_U f(s)},$$

$$\text{where} \quad g_1(a) = \phi(a) - a(1 - \Phi(a)),$$

$$f(s) = \frac{1}{\sigma_X}\phi(\bar{s})$$

$$\text{with } \bar{s} = \frac{s - \mu_X}{\sigma_X}$$

$$c = \frac{\bar{s}}{2}\frac{\sigma_U}{\sigma_X}\left(a_1^2 + 1 - a_1 k(a_1)\right),$$

$$\text{where} \quad k(a) = \frac{\phi(a)}{1 - \Phi(a)},$$

$$c_u = \frac{k(a_1)\left\{2a_1 k(a_1) + 1 - a_1^2\right\}}{4n} + \frac{(\bar{s}^4 + 4\bar{s}^2 + 1)(k(a_1) - a_1)}{4m}.$$

The test limit is computed with the estimated values $\hat{a}_1$, $\hat{c}$ and $\hat{c}_u$. These are obtained from a_1, c and c_u, respectively, by replacing the parameters by their estimators everywhere.

If the specification limit s is such that items should be approved if the value of the characteristic is larger than s, the test limit becomes

$$\hat{t}_u = s + (\hat{a}_1 + \hat{c} + \hat{c}_u) \cdot \hat{\sigma}_U.$$

In case the consumer loss may exceed the prescribed bound γ with probability α only, the correction term c_u above is replaced by c_i which is defined as,

$$c_i = u_\alpha \sqrt{\frac{k(a_1)^2}{2n} + \frac{(k(a_1) - a_1)^2(\bar{s}^4 + 1)}{2m}},$$

where $u_\alpha = \Phi^{-1}(1 - \alpha)$, the upper α-quantile of the standard normal distribution. Hence, if items should be approved if $X < s$, the test limit becomes

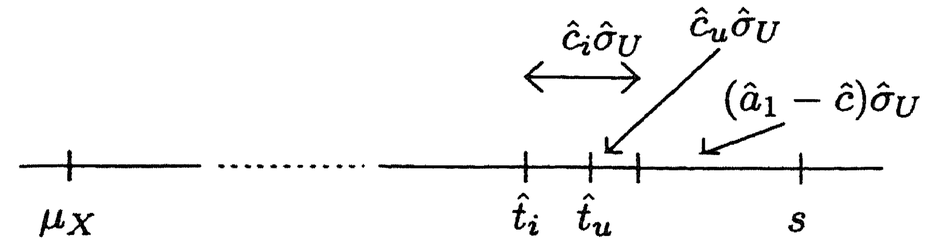
$$\hat{t}_i = s - (\hat{a}_1 - \hat{c} + \hat{c}_i) \cdot \hat{\sigma}_U,$$

otherwise

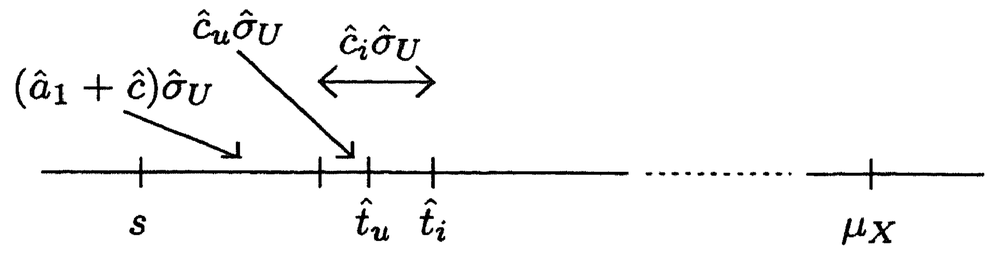
$$\hat{t}_i = s + (\hat{a}_1 + \hat{c} + \hat{c}_i) \cdot \hat{\sigma}_U.$$

In a figure:

Test limits when product is nonconforming if $X > s$:



Test limits when product is nonconforming if $X < s$:



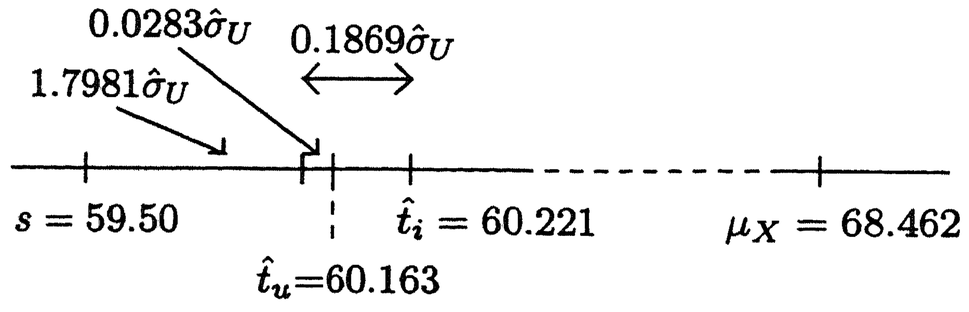
Computation The expression for a_1 is not explicit, but as $g_1(a)$ is strictly decreasing in a, a_1 is uniquely defined. Its value is determined either by using a simple numerical root-finding procedure or with the help of table 2.1.1 at the end of this chapter (page 30). In this table numerical values of $g_1^{-1}(b)$ are given. The step-size of b is taken such that linear interpolation will lead to errors in a_1 of 0.001, at most, which is amply accurate for practical purposes.

The first part of the correction terms c_u and c_i, with n in the denominator, is due the estimation of σ_U, the second part is due to the estimation of μ_X and σ_X. In the situation where no data from production is available, we should read $m = n$. In the special case where μ_X and σ_X are known or m is very large (over a few thousand) the second part of the correction terms may be omitted. However, omitting c_u for moderate sample sizes (for example $n = 40$) leads to consumer losses which in expectation can be equal up to 2γ.

Numerical example Suppose the value of a characteristic should not be below $s = 59.50$ and the consumer loss should not be larger than $\gamma = 40$ ppm (parts per million). In order to estimate σ_U, $n = 120$ products have been measured twice, leading to $\hat{\sigma}_U = 0.3631$. The estimates of μ_X and σ_X are obtained on the basis of a production sample of 2781 observations, resulting in $\hat{\mu}_X = 68.462$ and $\hat{\sigma}_X = 4.017$. With the estimates we first compute

$\hat{a}_1$.

We find $\gamma/(\hat{\sigma}_U \hat{f}(s)) = 1.3362 \cdot 10^{-2}$ and $\hat{a}_1 = g_1^{-1}(1.3362 \cdot 10^{-2}) \approx 1.8264$, from table 2.1.1. Using $k(1.8264) = 2.2204$, we find $\hat{c} = -0.0283$, $\hat{c}_u = 0.0267 + 0.0016 = 0.0283$ and $\hat{c}_i = 0.1869$. We obtain $\hat{t}_u = 60.163$ and $\hat{t}_i = 60.221$.
In a figure:



Number of observations

If unbiased estimation is sufficient, but on the other hand consumer losses larger than $\gamma(1 + \delta_0)$ should only occur with probability α (for example $\delta_0 = 0.20$ and $\alpha = 0.10$), the number of observations required to estimate σ_U is approximately

$$n \approx \frac{1}{2}\left(\frac{u_\alpha}{\delta_0}\frac{k(a_1)}{(k(a_1) - a_1)}\right)^2,$$

with $u_\alpha = \Phi^{-1}(1 - \alpha)$, the upper α-quantile of the standard normal distribution. The factor $k(a_1)/\{k(a_1) - a_1\}$ is approximately 3 for $a_1 = 1$ and 6.4 for $a_1 = 2$. For $\alpha = 0.05$ or 0.10 ($u_\alpha = 1.65$, 1.28, respectively) and $\delta_0 = 0.1$ or 0.2, the factor $u_\alpha^2/(2\delta_0^2)$ varies from 20 to 140. Together this leads to quite large values of n.

The average yield in the confidence interval approach is smaller than in case of unbiased estimation (because $c_i > c_u$). The number of observations required to limit the reduction of yield to a value β_0 (for example $\beta_0 = 0.1\%$) strongly depends on the true (but unknown) value of $\sigma_U f(s)$. In fact, in certain situations there is practically no loss of yield at all, even for very small sample sizes, whereas in other situations it is possible to achieve a certain value of β_0 only for extremely large sample sizes. With $\beta_0 = \sigma_U f(s)/K$ for some constant K, the required sample size is approximately

$$n \approx \frac{1}{2}\{u_\alpha k(a_1) K\}^2,$$

The factor $k(a_1)$ is between a_1 and $a_1 + 1$. It is clear that for this guide to be of use, some idea of the value $\sigma_U f(s)$ is needed.

Section 3.4.3 contains more information.

Reference Chapter 3.

2.2 The measurement error only is normally distributed

Situation The measurement error is $N(0, \sigma_U^2)$-distributed, but σ_U is unknown.
The type of distribution of the characteristic is unknown.

Estimators The estimator of σ_U here is the same as in the previous situation. It is based on n repeated measurements of the characteristic with two replications, $\tilde{X}_{i1}$ and $\tilde{X}_{i2}$ $(i = 1, \ldots, n)$, and it is given by

$$\hat{\sigma}_U^2 = \frac{1}{2n} \sum_{i=1}^{n} (\tilde{X}_{i1} - \tilde{X}_{i2})^2.$$

The test limits are further based on estimators of the density of $\tilde{X}$ and its derivative at the specification limit. We write $g(s)$ and $g'(s)$ to denote this density and its derivative, respectively.

To estimate $g(s)$ Rosenblatt's estimator is applied. The estimator is based on an additional sample $\tilde{X}_1, \ldots, \tilde{X}_m$ from production and it is defined in the following way,

$$\hat{g}(s) = \frac{1}{2mh} \sum_{i=1}^{m} Z_i,$$

where $$Z_i = \begin{cases} 1 & \text{if } \tilde{X}_i \in [s-h, s+h] \\ 0 & \text{otherwise} \end{cases},$$

$$h = \hat{\sigma}_{\tilde{X}} \left\{ m \cdot \phi \left(\frac{s - \hat{\mu}_X}{\hat{\sigma}_{\tilde{X}}} \right) \right\}^{-1/2},$$

in which $\hat{\mu}_X$ and $\hat{\sigma}_{\tilde{X}}$ are the sample mean and sample standard deviation of $\tilde{X}$ (we have $\mu_X = \mu_{\tilde{X}}$). So, estimation of $g(s)$ boils down to simply counting the number of observations around s. By h, called the bandwidth, the area around s is specified.

The estimator of $g'(s)$ is of a similar form as the one of $g(s)$,

$$\hat{g}'(s) \;=\; \frac{1}{m\bar{h}^2}\sum_{i=1}^{m}\bar{Z}_i,$$

$$\text{where}\quad \bar{Z}_i = \begin{cases} -1 & \text{if } \tilde{X}_i \in [s-\bar{h}, s] \\ 1 & \text{if } \tilde{X}_i \in (s, s+\bar{h}] \\ 0 & \text{otherwise} \end{cases},$$

$$\bar{h} \;=\; \hat{\sigma}_{\tilde{X}}\left\{m\cdot\phi\left(\frac{s-\hat{\mu}_X}{\hat{\sigma}_{\tilde{X}}}\right)\right\}^{-1/4}.$$

Test limits In case of unbiased estimation and when items should be approved if $X < s$, the test limit is given by

$$\hat{t}_u \;=\; s - (\hat{a}_1 - \hat{c} + \hat{c}_u)\cdot\hat{\sigma}_U,$$

$$\text{with}\quad a_1 \text{ such that } g_1(a_1) = \frac{\gamma}{\sigma_U g(s)},$$

$$\text{where}\quad g_1(a) \;=\; \phi(a) - a(1-\Phi(a)),$$

$$c \;=\; -\frac{\sigma_U}{2}\frac{g'(s)}{g(s)}\left(a_1^2 + 1 - a_1 k(a_1)\right),$$

$$\text{where}\quad k(a) \;=\; \frac{\phi(a)}{1-\Phi(a)},$$

$$c_u \;=\; \frac{k(a_1)\left\{2a_1 k(a_1) + 1 - a_1^2\right\}}{4n} + (k(a_1) - a_1)\left\{\frac{1}{2mhg(s)} - \frac{1}{m}\right\}.$$

If the specification limit s is such that items should be approved if $X > s$, the test limit is

$$\hat{t}_u \;=\; s + (\hat{a}_1 + \hat{c} + \hat{c}_u)\cdot\hat{\sigma}_U.$$

In case the consumer loss may exceed γ with probability α only, the correction term c_u above is replaced by c_i which is defined here as,

$$c_i \;=\; u_\alpha\sqrt{\frac{k(a_1)^2}{2n} + \frac{(k(a_1)-a_1)^2}{2mhg(s)}},$$

where $u_\alpha = \Phi^{-1}(1-\alpha)$, the upper α-quantile of the standard normal distribution. Hence, in the situation where items

should be approved if $X < s$, the test limit is

$$\hat{t}_i \;=\; s - (\hat{a}_1 - \hat{c} + \hat{c}_i) \cdot \hat{\sigma}_U,$$

otherwise,

$$\hat{t}_i \;=\; s + (\hat{a}_1 + \hat{c} + \hat{c}_i) \cdot \hat{\sigma}_U.$$

Computation Given a value of $\gamma/\{\hat{\sigma}_U \hat{g}(s)\}$ computation of $\hat{a}_1$ is the same as in section 2.1. The parts in the correction terms c_u and c_i with n in the denominator are due to the estimation of σ_U. They are identical to those in section 2.1. The parts with m are due to the estimation of g.

With respect to the estimation of $g(s)$ it may occur that there are no observations in the interval $[s-h, s+h]$. This happens especially in situations where s is in the tail of the distribution and only a small number of measurements is available. The best thing to do in this case is to try to get more observations. If this is not possible, one could apply the test limit from the previous section. However, this test limit is unreliable if the characteristic is not normally distributed. An alternative is to take simply the conservative test limit t_c from (1.1.8). That is, $\hat{t}_c = s - \Phi^{-1}(1-\gamma/\hat{\pi}) \cdot \hat{\sigma}_U$ if products should be rejected if $X > s$ and $\hat{t}_c = s + \Phi^{-1}(1-\gamma/\hat{\pi}) \cdot \hat{\sigma}_U$ otherwise. The parameter π (either $P(X > s)$ or $P(X < s)$) is estimated by the fraction of observations which does not satisfy the specification.

Numerical example Just as in the example from the previous section, suppose the consumer loss should not be larger than $\gamma = 40$ ppm (parts per million), the specification limit is $s = 59.50$ and the value of the characteristic should not be below s. Based on $n = 120$ repeated measurements, $\hat{\sigma}_U = 0.3631$ is found. From production $m = 2781$ measurements of the characteristic are available.

To estimate the density of $\tilde{X}$ and its derivative, we first need the estimates of μ_X and $\sigma_{\tilde{X}}$ to determine the bandwidths h and $\bar{h}$. The sample mean and sample standard deviation yield $\hat{\mu}_X = 68.462$ and $\hat{\sigma}_{\tilde{X}} = 4.0334$, say. We find $h = 0.4161$ and $\bar{h} = 1.2954$. Now, suppose in the interval $[s-h, s+h]$ there are 19 observations of $\tilde{X}$, and in the intervals $[s-\bar{h}, s]$ and $[s, s+\bar{h}]$ there are 21 and 43 observations, respectively. We then obtain $\hat{g}(s) = 8.211 \cdot 10^{-3}$ and $\hat{g}'(s) = 4.714 \cdot 10^{-3}$.

Having all the estimates we find $\hat{a}_1 = g_1^{-1}(1.3417 \cdot 10^{-2}) \approx$

1.8248 (table 2.1.1) and $\hat{c} = -0.0293$, $\hat{c}_u = 0.0473$ and $\hat{c}_i = 0.2171$. The test limits become $\hat{t}_u = 60.169$ and $\hat{t}_i = 60.231$.

Number of observations The guidelines in the previous section for the number n of repeated measurements apply to the present situation as well (reading g instead of f where necessary). When using one of these directions here, the number m of additional production data should be much larger than n, as a rule $m \approx n^2$.

Reference Chapter 4.

2.3 The characteristic only is normally distributed

Situation The characteristic is $N(\mu_X, \sigma_X^2)$-distributed, but μ_X and σ_X are unknown.
The type of distribution of the measurement error is unknown. The mean of the measurement error is denoted by μ. If there is no systematic measurement error μ is simply 0, otherwise the value of μ is assumed unknown.

Estimators Regarding the measurement error, observations $U_1, \ldots, U_n$ of the measurement error itself are needed. This means that carrying out some repeated measurements does not give the necessary information. Observations on the measurement error are usually obtained from a comparison between standard measurement results and precise laboratory measurements of the same objects. The measurement error on the laboratory measurements is negligible and hence the difference between the two measurements yields observations from the measurement error in the factory.

In the computation of the test limits we will encounter sample functions $r_k(d)$ and $l_k(d)$ which are defined by

$$r_k(d) = \frac{1}{n}\sum_{i=1}^{n}(U_i - d)^k \cdot I_{(U_i-d)>0}, \quad k = 0, 1, 2 \text{ and}$$

$$l_k(d) = \frac{1}{n}\sum_{i=1}^{n}(-U_i - d)^k \cdot I_{(U_i+d)<0}, \quad k = 0, 1, 2,$$

where $I_{x>0} = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases}$.

If there is a systematic measurement error ($\mu \neq 0$), the pa-

rameter μ is estimated by the sample mean

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} U_i \ .$$

The estimators of μ_X and σ_X^2 are based on an additional sample from production, $\tilde{X}_1, \ldots, \tilde{X}_m$. The estimator of μ_X is defined as

$$\hat{\mu}_X \ = \ \frac{1}{m}\sum_{i=1}^{m} \tilde{X}_i \ .$$

The estimator of σ_X^2 here is simply the sample variance. It is defined as

$$\hat{\sigma}_X^2 \ = \ \frac{1}{m-1}\sum_{i=1}^{m} (\tilde{X}_i - \hat{\mu}_X)^2 .$$

Test limits First we consider the situation where products should be rejected if $X > s$.
In case of unbiased estimation the test limit is

$$\hat{t}_u \ = \ s - (\hat{d}_l - \hat{c}_l + \hat{c}_{u,l}),$$

$$\text{with} \quad d_l \text{ such that } l_1(d_l) = \frac{\gamma}{f(s+\mu)},$$

$$\text{where } f(s+\mu) \ = \ \frac{1}{\sigma_X}\phi(\bar{s}),$$

$$\text{with } \bar{s} \ = \ \frac{s+\mu-\mu_X}{\sigma_X},$$

$$c_l \ = \ \frac{\bar{s}}{2\sigma_X}\frac{l_2(d_l)}{l_0(d_l)},$$

$$c_{u,l} \ = \ \frac{l_1(d_l)}{l_0(d_l)}\left(\frac{1-l_0(d_l)}{n\ l_0(d_l)} + \frac{\bar{s}^4+4\bar{s}^2+1}{4m}\right).$$

In the case where the consumer loss may exceed γ with probability α only, the correction term $c_{u,l}$ above is replaced by $c_{i,l}$, and the test limit becomes

$$\hat{t}_i \ = \ s - (\hat{d}_l - \hat{c}_l + \hat{c}_{i,l}),$$

with

$$c_{i,l} = \frac{l_1(d_l)}{l_0(d_l)}\ u_\alpha \sqrt{\frac{l_2(d_l)/l_1(d_l)^2 - 1}{n} + \frac{\bar{s}^4+1}{2m}},$$

where $u_\alpha = \Phi^{-1}(1-\alpha)$, the upper α-quantile of the standard normal distribution.

If the specification limit is such that products should be rejected if $X < s$, then in case of unbiased estimation the test limit is given by

$$\hat{t}_u = s + (\hat{d}_r - \hat{c}_r + \hat{c}_{u,r}),$$

with d_r such that $r_1(d_r) = \dfrac{\gamma}{f(s+\mu)}$,

$$\text{where } f(s+\mu) = \frac{1}{\sigma_X}\phi(\bar{s}),$$

$$\text{with } \bar{s} = \frac{s+\mu-\mu_X}{\sigma_X},$$

$$c_r = -\frac{\bar{s}}{2\sigma_X}\frac{r_2(d_r)}{r_0(d_r)},$$

$$c_{u,r} = \frac{r_1(d_r)}{r_0(d_r)}\left(\frac{1-r_0(d_r)}{n\, r_0(d_r)} + \frac{\bar{s}^4+4\bar{s}^2+1}{4m}\right),$$

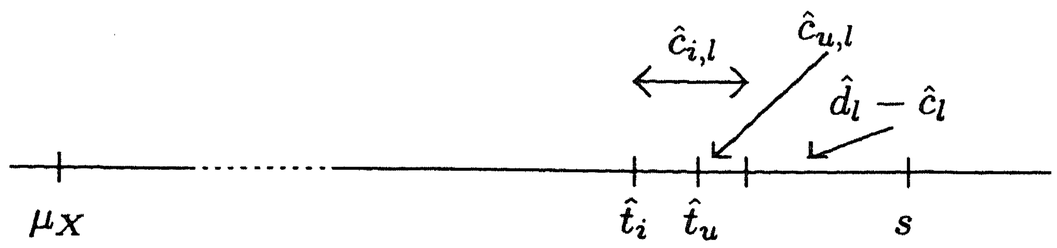
while in case the test limit should be such that the consumer loss may exceed γ with probability α only, the correction term $c_{u,r}$ above is replaced by $c_{i,r}$, and the test limit becomes

$$\hat{t}_i = s + (\hat{d}_r - \hat{c}_r + \hat{c}_{i,r}),$$

with $c_{i,r} = \dfrac{r_1(d_r)}{r_0(d_r)}\, u_\alpha \sqrt{\dfrac{r_2(d_r)/r_1(d_r)^2 - 1}{n} + \dfrac{\bar{s}^4+1}{2m}}.$

In a figure:

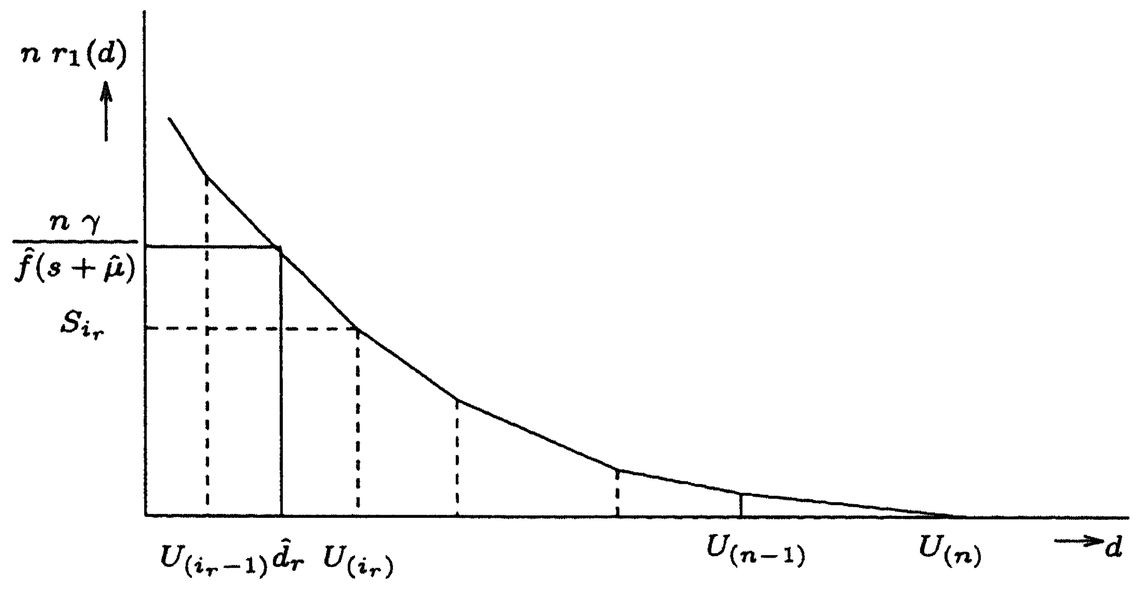
Test limits when product is nonconforming if $X > s$:



Test limits when product is nonconforming if $X < s$:

Computation The expressions for d_l and d_r are not explicit. The easiest way to determine them is to consider the ordered observations of the measurement error. Let $U_{(1)}$ denote the smallest observation, $U_{(2)}$ the second smallest, etc., until $U_{(n)}$.

The following figure shows the function $n\, r_1(d)$.



From this figure it is clear that to compute $\hat{d}_r$ we have to find the value i_r, which is defined as

$$i_r \;=\; \max\left\{ i \;\middle|\; S_i \geq n\frac{\gamma}{\hat{f}(s+\hat{\mu})} \right\} + 1,$$

$$\text{with} \quad S_i = n\, r_1(U_{(i)}) = \sum_{j=i+1}^{n} (n-j+1)(U_{(j)} - U_{(j-1)})$$

$$\hat{f}(s+\hat{\mu}) = \frac{1}{\hat{\sigma}_X}\phi\left(\frac{s+\hat{\mu}-\hat{\mu}_X}{\hat{\sigma}_X}\right)$$

(by definition $S_n = 0$; if there is no systematic measurement error, $\mu = 0$ and we do not estimate μ).
Then $\hat{d}_r$ is given by

$$\hat{d}_r = U_{(i_r)} - \frac{n\gamma/\hat{f}(s+\hat{\mu}) - S_{i_r}}{n - i_r + 1}.$$

The corresponding algorithm is straightforward:

$i := n;\ \ S_i := 0;$
found:= $S_i > n\ \dfrac{\gamma}{\hat{f}(s+\hat{\mu})}$;
WHILE (NOT found) DO
 BEGIN
  $S_{i-1} := S_i + (n-i+1)\cdot(U_{(i)} - U_{(i-1)})$;
  found:= $S_{i-1} > n\ \dfrac{\gamma}{\hat{f}(s+\hat{\mu})}$;
  $i := i - 1$;
 END;
$i_r := i + 1$;
$\hat{d}_r := U_{(i_r)} - \dfrac{n\gamma/\hat{f}(s+\hat{\mu}) - S_{i_r}}{n - i_r + 1}$;

Computation of $\hat{d}_l$ is analoguous. We have to find the value i_l which is defined as,

$$i_l \;=\; \min\left\{ i \,\middle|\, S_i \geq n\frac{\gamma}{\hat{f}(s+\hat{\mu})} \right\} - 1,$$

with $S_i = n\, l_1(U_i) = \sum_{j=1}^{i-1} j\ (U_{(j+1)} - U_{(j)})$ (in this case $S_1 = 0$).
Now $\hat{d}_l$ is given by

$$\hat{d}_l = -U_{(i_l)} - \frac{n\gamma/\hat{f}(s+\hat{\mu}) - S_{i_l}}{i_l}.$$

The corresponding algorithm is as follows.

$i := 1;\quad S_i := 0;$
found:= $S_i > n\ \dfrac{\gamma}{\hat{f}(s+\hat{\mu})}$;
WHILE (NOT found) DO

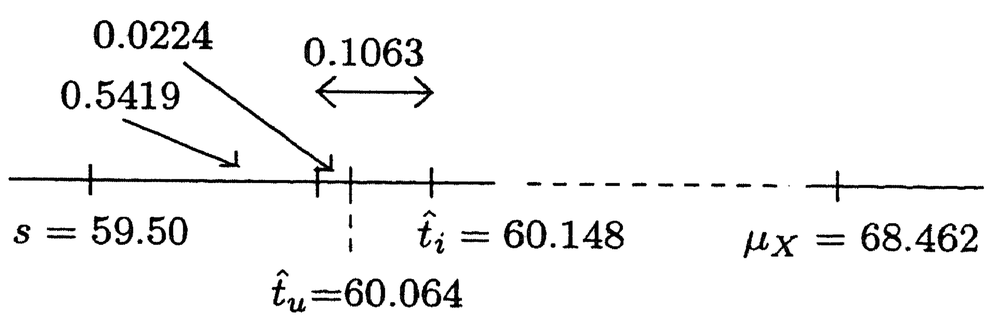
BEGIN
$S_{i+1} := S_i + i \cdot (U_{(i+1)} - U_{(i)})$;
found:= $S_{i+1} > n \dfrac{\gamma}{\hat{f}(s+\hat{\mu})}$;
$i := i + 1$;
END;
$i_l := i - 1$;
$\hat{d}_l := -U_{(i_l)} - \dfrac{n\gamma/\hat{f}(s+\hat{\mu}) - S_{i_l}}{i_l}$;

Once $\hat{d}_r$ or $\hat{d}_l$ is computed, computation of the test limit is straightforward.

Numerical example

Suppose the consumer loss should not exceed $\gamma = 40$ ppm (parts per million), the value of the inspected characteristic should not be below $s = 59.50$ and the estimates $\hat{\mu}_X = 68.462$ and $\hat{\sigma}_X = 4.0334$ are obtained from a production sample of $m = 2781$ observations. With these estimates it is found $\hat{f}(s) = 0.008379$. (It is known there is no systematic measurement error, $\mu = 0$).

Suppose of $n = 120$ products both laboratory measurements and standard measurements are carried out. The differences between the laboratory and standard measurements are the observations of the measurement error and suppose the six largest observations are $U_{(115)} = 0.5015$, $U_{(116)} = 0.5658$, $U_{(117)} = 0.5737$, $U_{(118)} = 0.5965$, $U_{(119)} = 0.6388$ and $U_{(120)} = 0.9551$. We have $n\gamma/\hat{f}(s) = 0.5729$ and to compute $\hat{d}_r$, first we determine i_r. The values of S_i are $S_{119} = 0.3163$, $S_{118} = 0.4009$, $S_{117} = 0.4693$, $S_{116} = 0.5009$ and $S_{115} = 0.8224$, hence, $i_r = 116$ and we find $\hat{d}_r = 0.5514$. Further computation is straightforward, $\hat{c}_r = 0.0095$, $\hat{c}_{u,r} = 0.0224$ and $\hat{c}_{i,r} = 0.1063$ and the test limits are $\hat{t}_u = 60.064$ and $\hat{t}_i = 60.148$.
In a figure:



Number of observations

About the computation of $\hat{d}_r$ and $\hat{d}_l$, it is important that $i_r \leq n-2$, respectively, $i_l \geq 2$. If this is not the case, one should try to obtain more observations on the measurement error because the computed test limit is unreliable in this kind of situation. Application of the test limit from section 2.1 if the measurement error is not normally distributed, is not to be recommended.

If unbiased estimation is sufficient, but on the other hand consumer losses larger than $\gamma(1+\delta_0)$ should only occur with probability α (for example $\delta_0 = 0.20$ and $\alpha = 0.10$), the number of observations required on the measurement error is approximately

$$n \approx \left(\frac{u_\alpha}{\delta_0}\right)^2 \frac{l_2(d_l) - l_1^2(d_l)}{l_1^2(d_l)}$$

when products are nonconforming if $X > s$, or

$$n \approx \left(\frac{u_\alpha}{\delta_0}\right)^2 \frac{r_2(d_r) - r_1^2(d_r)}{r_1^2(d_r)}$$

when products are nonconforming if $X < s$. The number m of additional production data should be at least of the same order of magnitude.

The average yield in the confidence interval approach is smaller than in case of unbiased estimation (because the correction terms $c_{i,r}$, $c_{i,l}$ are larger than $c_{u,r}$, $c_{u,l}$, respectively). The number of observations required on the measurement error to limit the reduction of yield to a value β_0 (for example $\beta_0 = 0.1\%$) is approximately

$$n \approx \left(u_\alpha \frac{f(s+\mu)}{\beta_0}\right)^2 \frac{l_2(d_l) - l_1^2(d_l)}{l_0^2(d_l)}$$

when products are nonconforming if $X > s$, or

$$n \approx \left(u_\alpha \frac{f(s+\mu)}{\beta_0}\right)^2 \frac{r_2(d_r) - r_1^2(d_r)}{r_0^2(d_r)},$$

when products are nonconforming if $X < s$. The number m of additional production data should be again at least of the same order of magnitude.

It is clear that for these guides to be of use, some idea of the value of $f(s+\mu)$ and of d_l, d_r is needed. The guidelines are based on the theory in section 3.4.3.

Reference Chapter 6.

2.4 Both the characteristic and measurement error have unknown types of distribution

Situation The type of distribution of the characteristic is unknown. The type of distribution of the measurement error is unknown. The mean of the measurement error is denoted by μ. If there is no systematic measurement error μ is simply 0, otherwise the value of μ is assumed unknown.

Estimators As in the situation of section 2.3 observations $U_1, \ldots, U_n$ of the measurement error itself are needed and we will again encounter the sample functions $r_k(d)$ and $l_k(d)$. If there is a systematic measurement error ($\mu \neq 0$) we need to estimate μ.

With respect to the unknown distribution of the characteristic we merely need to estimate the density of $\tilde{X}$ and its derivative at the specification limit s if $\mu = 0$ or at $s + \mu$ if $\mu \neq 0$. The density and its derivative are denoted by g and g', respectively. The estimators $\hat{g}$ and $\hat{g}'$ are based on an additional sample $\tilde{X}_1, \ldots, \tilde{X}_m$ from production and are as defined in section 2.2 (reading $s + \mu$ instead of s when necessary).

Test limits First we consider the situation where products should be rejected if the value of the characteristic is larger than the specification limit.

In case of unbiased estimation the test limit in this situation is

$$\hat{t}_u = s - (\hat{d}_l - \hat{c}_l + \hat{c}_{u,l}),$$

with d_l such that $l_1(d_l) = \dfrac{\gamma}{g(s+\mu)}$,

$$c_l = -\frac{1}{2}\frac{g'(s+\mu)}{g(s+\mu)}\frac{l_2(d_l)}{l_0(d_l)},$$

$$c_{u,l} = \frac{l_1(d_l)}{l_0(d_l)}\left(\frac{1-l_0(d_l)}{n\ l_0(d_l)} + \frac{1}{2mhg(s+\mu)} - \frac{1}{m}\right),$$

In case the consumer loss may exceed γ with probability α only, the correction term $c_{u,l}$ above is replaced by $c_{i,l}$, and the test limit becomes

$$\hat{t}_i = s - (\hat{d}_l - \hat{c}_l + \hat{c}_{i,l}),$$

with

$$c_{i,l} = \frac{l_1(d_l)}{l_0(d_l)}\, u_\alpha \sqrt{\frac{l_2(d_l)/l_1(d_l)^2 - 1}{n} + \frac{1}{2mhg(s+\mu)} - \frac{1}{m}}$$

where $u_\alpha = \Phi^{-1}(1-\alpha)$, the upper α-quantile of the standard normal distribution.

If the specification limit s is such that products are non-conforming if the value of the characteristic is smaller than s, then in case of unbiased estimation the test limit is given by

$$\hat{t}_u \;=\; s + (\hat{d}_r - \hat{c}_r + \hat{c}_{u,r}),$$

$$\text{with} \quad d_r \;=\; \text{such that } r_1(d_r) = \frac{\gamma}{g(s+\mu)},$$

$$c_r \;=\; \frac{1}{2}\frac{g'(s+\mu)}{g(s+\mu)}\frac{r_2(d_r)}{r_0(d_r)},$$

$$c_{u,r} \;=\; \frac{r_1(d_r)}{r_0(d_r)}\left(\frac{1-r_0(d_r)}{n\, r_0(d_r)} + \frac{1}{2mhg(s+\mu)} - \frac{1}{m}\right),$$

while in case the estimated test limit should be such that consumer loss may exceed γ with probability α only, the correction term $c_{u,r}$ above is replaced by $c_{i,r}$, and the test limit becomes

$$\hat{t}_i \;=\; s + (\hat{d}_r - \hat{c}_r + \hat{c}_{i,r}),$$

with

$$c_{i,r} = \frac{r_1(d_r)}{r_0(d_r)}\, u_\alpha \sqrt{\frac{r_2(d_r)/r_1(d_r)^2 - 1}{n} + \frac{1}{2mhg(s+\mu)} - \frac{1}{m}}$$

where $u_\alpha = \Phi^{-1}(1-\alpha)$, the upper α-quantile of the standard normal distribution.

Computation Given a value of $\gamma/\hat{g}(s+\hat{\mu})$ (or $\gamma/\hat{g}(s)$ if $\mu = 0$), computation of $\hat{d}_r$ and $\hat{d}_l$ here is carried out in the way which is described in the previous section. If there are no observations in the interval $[s+\hat{\mu}-h,\ s+\hat{\mu}+h]$, hence $\hat{g}(s+\hat{\mu}) = 0$, the best thing to do is to try to get more observations. If this is not possible, the test limits from the previous section may be applied, however, these limits are unreliable if X is not normally

distributed.

An alternative is to apply a conservative approach similar to the one in section 2.2. Instead of $\hat{d}_l$ take $\hat{d}_{l,c} = \min\{d \mid l_0(d) \leq \gamma/\hat{\pi}\}$ with $\hat{\pi}$ the fraction of products for which the measurement is larger than $s + \hat{\mu}$, and instead of $\hat{d}_r$ take $\hat{d}_{r,c} = \min\{d \mid r_0(d) \leq \gamma/\hat{\pi}\}$ with $\hat{\pi}$ the fraction of products for which the measurement is smaller than $s + \hat{\mu}$. The correction terms are omitted.

Numerical example

Suppose the consumer loss should not exceed $\gamma = 40$ ppm (parts per million) and the value of the inspected characteristic should not be below $s = 59.50$. On $n = 120$ products both laboratory measurements and standard measurements are carried out and from production $m = 2781$ measurements of the characteristic are available. It is known there is no systematic measurement error ($\mu = 0$).

To estimate the density of $\tilde{X}$ and its derivative, we first need the estimates $\hat{\mu}_X$ and $\hat{\sigma}_{\tilde{X}}$ to determine the bandwidths h and $\bar{h}$. From the production data the estimates $\hat{\mu}_X = 68.462$ and $\hat{\sigma}_{\tilde{X}} = 4.0334$ are obtained, leading to $h = 0.4161$ and $\bar{h} = 1.2954$. In the interval $[s - h, s + h]$ there are 19 observations of $\tilde{X}$, and in the intervals $[s - \bar{h}, s]$ and $[s, s + \bar{h}]$ there are 21 and 43 observations, respectively. We obtain $\hat{g}(s) = 8.211 \cdot 10^{-3}$ and $\hat{g}'(s) = 4.714 \cdot 10^{-3}$.

The differences between the laboratory measurements and the standard measurements are the observations of the measurement error. Suppose the six largest observations are $U_{(115)} = 0.5015$, $U_{(116)} = 0.5658$, $U_{(117)} = 0.5737$, $U_{(118)} = 0.5965$, $U_{(119)} = 0.6388$ and $U_{(120)} = 0.9551$. Further computation is as in the example in the previous section. We have $n\gamma/\hat{g}(s) = 0.5846$ and to compute $\hat{d}_r$ we determine i_r first. We find $i_r = 116$, $\hat{d}_r = 0.5491$, $\hat{c}_r = 0.0101$, $\hat{c}_{u,r} = 0.0285$ and $\hat{c}_{i,r} = 0.1120$. The test limits are $\hat{t}_u = 60.067$ and $\hat{t}_i = 60.151$.

Number of observations

About the computation of $\hat{d}_r$ and $\hat{d}_l$, it is important that $i_r \leq n - 2$, respectively, $i_l \geq 2$. If this is not the case, one should try to obtain more observations on the measurement error because the computed test limit is unreliable in this kind of situation. Application of the test limit from section 2.2 if the measurement error is not normally distributed, is not to be recommended. When the mean μ of the measurement error is estimated, it is important that the number m of additional production data is much larger than n.

The guideline in the previous section to prevent obtaining too large values of the consumer loss in the case of unbiased estimation, and the guideline with respect to the loss of yield in the confidence interval approach both apply to the present situation (reading g instead of f where necessary). However, when using one of these directions here, the number m of additional production data should be much larger than n, as a rule $m \approx n^2$.

Reference Chapter 6

Table 2.1.1 Values of $g_1^{-1}(b)$

The values of b are such that linear interpolation lead to errors of 0.001, at most.

b	$g_1^{-1}(b)$	b	$g_1^{-1}(b)$	b	$g_1^{-1}(b)$	b	$g_1^{-1}(b)$
3.5E-01	0.1020	3.9E-02	1.3719	3.0E-03	2.3674	1.4E-04	3.2748
3.1E-01	0.1926	3.4E-02	1.4341	2.6E-03	2.4150	1.2E-04	3.3155
2.8E-01	0.2659	3.0E-02	1.4897	2.3E-03	2.4553	1.1E-04	3.3382
2.5E-01	0.3449	2.6E-02	1.5521	2.0E-03	2.5007	9.2E-05	3.3846
2.2E-01	0.4308	2.3E-02	1.6044	1.7E-03	2.5527	7.7E-05	3.4303
2.0E-01	0.4929	2.0E-02	1.6631	1.5E-03	2.5923	6.5E-05	3.4734
1.8E-01	0.5597	1.8E-02	1.7065	1.3E-03	2.6371	5.5E-05	3.5154
1.6E-01	0.6323	1.6E-02	1.7544	1.1E-03	2.6887	4.6E-05	3.5599
1.5E-01	0.6711	1.4E-02	1.8077	9.3E-04	2.7398	3.9E-05	3.6007
1.4E-01	0.7120	1.3E-02	1.8370	7.9E-04	2.7889	3.3E-05	3.6416
1.3E-01	0.7552	1.2E-02	1.8682	6.7E-04	2.8378	2.8E-05	3.6814
1.2E-01	0.8010	1.1E-02	1.9019	5.7E-04	2.8852	2.4E-05	3.7185
1.1E-01	0.8499	9.5E-03	1.9578	4.8E-04	2.9349	2.0E-05	3.7619
9.6E-02	0.9245	8.2E-03	2.0130	4.1E-04	2.9800	1.7E-05	3.8003
8.4E-02	0.9957	7.1E-03	2.0661	3.5E-04	3.0248	1.5E-05	3.8297
7.4E-02	1.0615	6.1E-03	2.1212	3.0E-04	3.0679	1.3E-05	3.8630
6.5E-02	1.1273	5.3E-03	2.1714	2.6E-04	3.1076	1.1E-05	3.9017
5.7E-02	1.1923	4.6E-03	2.2212	2.2E-04	3.1534	9.1E-06	3.9451
5.0E-02	1.2556	4.0E-03	2.2697	1.9E-04	3.1932	7.6E-06	3.9861
4.4E-02	1.3160	3.5E-03	2.3154	1.6E-04	3.2393	6.3E-06	4.0283

Chapter 3

The normal case

In this chapter the situation is treated where both the inspected characteristic and the measurement error are normally distributed.

3.1 Introduction and notations

Let X denote the true value of the characteristic we want to measure. Because of a measurement error, denoted by U, we measure in fact $\tilde{X} = X + U$. We write $X \sim N(\mu_X, \sigma_X^2)$ that is, X is normally distributed with mean μ_X and variance σ_X^2. Likewise, we write $U \sim N(0, \sigma_U^2)$. We assume that X and U are independent and that σ_U is small compared to σ_X, which is typically the case in situations of practical interest.

The specification limit, denoted by s, is given. We will consider the situation where products are nonconforming if $X > s$ and we are typically interested in situations where the probability of a product being nonconforming, which is denoted by π,

$$\pi = P(X > s) = 1 - \Phi\left(\frac{s - \mu_X}{\sigma_X}\right),$$

is smaller than 0.5 (implying $s > \mu_X$). At the end of section 3.2 we discuss the difference with the situation where products are nonconforming if $X < s$ (and hence $s < \mu_X$).

Given a test limit t, the consumer loss is defined by

$$CL = P(X > s,\ X + U < t).$$

The consumer loss should not exceed the prescribed bound γ. By t_e the test limit is denoted for which $CL = \gamma$. We use Φ and ϕ to denote the standard normal distribution function and density, respectively.

In section 1.1 we argued that in this situation the consumer loss is a function of the specification limit s, the test limit t and the three parameters μ_X,

σ_X and σ_U. (Note that $(X, \tilde{X})$ has a bivariate normal distribution, $(X, \tilde{X}) \sim N(\mu_X, \mu_X, \sigma_X^2, \sigma_X^2 + \sigma_U^2, \rho)$ with $\rho = (1 + \sigma_U^2/\sigma_X^2)^{-1/2}$.) If these latter parameters are known, the limit t_e can be determined numerically. Usually these parameters are unknown, however.

First, to appraise the considerable effect of plugging in estimators rather than the true values, and then to subsequently correct for this effect, we need a more explicit relation between the test limit, γ, the specification limit and the three parameters μ_X, σ_X and σ_U. In the next section this is established. Accurate approximations to t_e as functions of μ_X, σ_X and σ_U are derived.

3.2 The case of known parameters

Fortunately, simple but nevertheless accurate approximations to the bivariate normal distribution can be obtained from Cox and Wermuth (1991) and Albers and Kallenberg (1994). The latter paper specializes to large ρ, the correlation coefficient involved. Since σ_U is assumed to be much smaller than σ_X, ρ is typically large and we shall use some of the results from Albers and Kallenberg (1994).

Let

$$a = \frac{s-t}{\sigma_U}, \quad a_e = \frac{s-t_e}{\sigma_U}, \tag{3.2.1}$$

then $t = s - a \cdot \sigma_U$. Note that a plays a role similar to that of k in (1.1.6) and $\Phi^{-1}(1-\gamma/\pi)$ in (1.1.8). In applications we will typically be interested in nonnegative a. Hence in what follows we shall tacitly assume $a > 0$, just like we assumed $s > \mu_X$. Many of the results we are going to formulate remain true for, or can be adapted to, negative a but it does not seem interesting to bother about this.

Let

$$\sigma = \frac{\sigma_U}{\sigma_X}, \quad Y = -\frac{U}{\sigma_U}, \quad \bar{X} = \frac{X-\mu_X}{\sigma_X}, \quad \bar{s} = \frac{s-\mu_X}{\sigma_X}, \tag{3.2.2}$$

then $\sigma \ll 1$, $\bar{X}$ and Y are independent and standard normal, while

$$\begin{aligned} CL &= P(X > s,\ X - \sigma_U Y < t) \\ &= P(Y > a,\ \bar{s} < \bar{X} < \bar{s} + \sigma(Y-a)). \end{aligned} \tag{3.2.3}$$

For the consumer loss we can write

$$CL = \int_a^\infty \{\Phi(\bar{s} + \sigma(y-a)) - \Phi(\bar{s})\}\, \phi(y)\mathrm{d}y, \tag{3.2.4}$$

and expand this integral in powers of σ. In doing so, we encounter the functions

$$g_k(a) = \int_a^\infty (y-a)^k \phi(y)\mathrm{d}y, \tag{3.2.5}$$

for $k \geq 0$. It is easily verified that, for $k \geq 1$,

$$\begin{aligned} g_{k+1}(a) &= k \cdot g_{k-1}(a) - a \cdot g_k(a), \\ g_k'(a) &= -k \cdot g_{k-1}(a). \end{aligned} \tag{3.2.6}$$

Moreover,

$$\begin{aligned} g_1(a) &= \phi(a) - a(1 - \Phi(a)), \\ g_2(a) &= (a^2 + 1)(1 - \Phi(a)) - a\phi(a). \end{aligned} \tag{3.2.7}$$

The consumer loss can thus be approximated by

$$CL = \sigma\phi(\bar{s})g_1(a) + \tfrac{1}{2}\sigma^2\phi'(\bar{s})g_2(a) + \dots . \tag{3.2.8}$$

As the g_k $(k \geq 0)$ decrease, we can define

$$a_1 \;=\; g_1^{-1}\left(\frac{\gamma}{\sigma\phi(\bar{s})}\right). \tag{3.2.9}$$

With a_1 we obtain a first order approximation $t_1 = s - a_1\sigma_U$ to t_e. The following lemma gives information on the performance of this limit.

Lemma 3.2.1 *For a_1 from (3.2.9) we have that*

$$0 \leq \frac{\gamma - P(X > s,\ \tilde{X} < s - a_1\sigma_U)}{\gamma} \leq \tfrac{3}{4}\sigma \max\{\bar{s}, 1\}. \tag{3.2.10}$$

Proof. Observe that

$$\Phi(\bar{s} + \sigma(y - a)) - \Phi(\bar{s}) - \sigma\phi(\bar{s})(y - a) = \tfrac{1}{2}\sigma^2\phi'(\xi_y) \cdot (y - a)^2, \tag{3.2.11}$$

with ξ_y between $\bar{s}$ and $\bar{s} + \sigma(y - a)$. As for $z \geq \bar{s} \geq 0$ we have that $0 \leq -\phi'(z) \leq -\phi'(\max\{\bar{s}, 1\})$, it follows from (3.2.4), (3.2.11) and (3.2.5) that

$$\begin{aligned} &\tfrac{1}{2}\sigma^2\phi'\left(\max\{\bar{s}, 1\}\right) \cdot g_2(a) \\ &\quad \leq\ P(X > s,\ \tilde{X} < s - a\sigma_U) - \sigma\phi(\bar{s}) \cdot g_1(a)\ \leq\ 0, \end{aligned} \tag{3.2.12}$$

which in its turn, together with (3.2.9), leads to

$$\tfrac{1}{2}\sigma^2\phi'\left(\max\{\bar{s}, 1\}\right) \cdot g_2(a_1) \leq P(X > s,\ \tilde{X} < s - a_1\sigma_U) - \gamma \leq 0. \tag{3.2.13}$$

Now, according to (3.2.6), $g_3(a) = 2g_1(a) - ag_2(a)$. As all g_k are nonnegative, it follows that $0 \leq g_3 \leq 2g_1$. Moreover the fact that $z^2 \leq \frac{1}{2}(z + z^3)$ for $z \geq 0$, implies in view of (3.2.5) that $g_2 \leq \frac{1}{2}(g_1 + g_3) \leq \frac{3}{2}g_1$. Consequently, the left-hand side of (3.2.13) is bounded from below by $3\sigma\gamma\phi'\left(\max\{\bar{s}, 1\}\right)/(4\phi(\bar{s}))$, which is at least $-\frac{3}{4}\sigma\gamma \max\{\bar{s}, 1\}$. □

It follows from lemma 3.2.1 that t_1 is conservative and that a_1 is an upper bound for a_ϵ. But it is also evident from the lemma that the relative error in the consumer loss caused by using a_1, is under control. In fact, for values of σ in the range of a few percents, the upper bound in (3.2.10) is probably already sufficiently small for practical purposes. However, for larger values of σ or $\bar{s}$, improvement is certainly called for.

The way to achieve this improvement already suggests itself in (3.2.8). By analogy with a_1 we introduce $\tilde{a}$ such that

$$g_1(\tilde{a}) - \tfrac{1}{2}\sigma\bar{s}g_2(\tilde{a}) = \frac{\gamma}{\sigma\phi(\bar{s})}. \tag{3.2.14}$$

Note that $\tilde{a}$ is not necessarily defined uniquely. While the functions g_k are monotone, the function as a whole on the left-hand side of (3.2.14) not necessarily is. The following lemma gives a simple necessary and sufficient condition.

Introduce the function

$$k(x) = \frac{\phi(x)}{1-\Phi(x)}, \tag{3.2.15}$$

then we formulate

Lemma 3.2.2 *As $k(x)-x$ decreases and $\lim_{x\to-\infty} k(x)-x = \infty$, $\lim_{x\to\infty} k(x)-x = 0$, there is a unique x_0 such that*

$$k(x_0) - x_0 = \frac{1}{\sigma\bar{s}}, \tag{3.2.16}$$

for given σ and $\bar{s}$. The function $g_1(x) - \frac{1}{2}\sigma\bar{s}g_2(x)$ decreases for $x > x_0$.

Proof. Combining (3.2.7) and (3.2.15), we see that $k(x) - x = -g_1(x)/g_1'(x)$. Hence the derivative of $k(x)-x$ equals $h(x)/g_1'(x)^2$, where $h(x) = g_1(x)g_1''(x) - g_1'(x)^2$. Now $h' = g_1 g_1^{(3)} - g_1'g_1''$, and thus $h'(x) = x\phi(x)g_1(x) - \phi(x)g_1'(x) = \phi(x)g_2(x) > 0$. As moreover g_1, g_1', g_1'' and hence also $h(x)$ tend to 0 as $x \to \infty$, it follows that $h(x) < 0$. Consequently, $k(x) - x$ decreases.

The derivative of $g_1(x) - \frac{1}{2}\sigma\bar{s}g_2(x)$ equals $g_1'(x) + \sigma\bar{s}g_1(x) = -g_1'(x)\{-1 + \sigma\bar{s}(k(x)-x)\}$, which is negative iff $k(x)-x < 1/(\sigma\bar{s})$. This will hold for $x > x_0$, with x_0 as in (3.2.16), because $k(x) - x$ is decreasing. □

The lemma shows that for typical values of γ, σ and $\bar{s}$ no monotonicity problems can occur. In fact, to obtain $x_0 = 0$ as solution of (3.2.16) it is required to have $\sigma\bar{s}$ as large as 1.25, which bound will hardly ever be violated. Hence we shall implicitly assume in the remainder of this chapter that the parameters are chosen such that $\tilde{a}$ is well above $x_0 = x_0(\sigma, \bar{s})$ from (3.2.16). Analogously to lemma 3.2.1 we have

Lemma 3.2.3 *For $\bar{s} \geq 1$ we have for $\tilde{a}$ from (3.2.14) that*

$$0 \leq \frac{P(X > s,\ \tilde{X} < s - \tilde{a}\sigma_U) - \gamma}{\gamma} \leq \tfrac{1}{3}\sigma^2 \frac{\max(\bar{s}^2 - 1, 2)}{1 - \frac{3}{4}\sigma\bar{s}}. \tag{3.2.17}$$

Proof. For $z \geq \bar{s} \geq 1$ we have that $0 \leq \phi''(z) \leq \phi''(\max\{\bar{s}, \sqrt{3}\})$, which leads to (cf. (3.2.13))

$$0 \leq P(X > s,\ \tilde{X} < s - \tilde{a}\sigma_U) - \gamma \leq \tfrac{1}{6}\sigma^3\phi''\left(\max\{\bar{s}, \sqrt{3}\}\right) g_3(\tilde{a}). \tag{3.2.18}$$

Now $g_2 \leq \frac{3}{2}g_1$, and therefore $g_1(a)(1 - \frac{3}{4}\sigma\bar{s}) \leq g_1(a) - \frac{1}{2}\sigma\bar{s}g_2(a)$. As moreover $g_3 \leq 2g_1$, it follows that $g_3(a) \leq \dfrac{2\left(g_1(a) - \frac{1}{2}\sigma\bar{s}g_2(a)\right)}{(1 - \frac{3}{4}\sigma\bar{s})}$ and thus $g_3(\tilde{a}) \leq \dfrac{2\gamma}{\sigma\phi(\bar{s})(1 - \frac{3}{4}\sigma\bar{s})}$. In combination with (3.2.18) this leads to (3.2.17), which proves the lemma. □

It follows from the lemma that $\tilde{a}$ is a lower bound for a_e. Hence we now have

$$\tilde{a} \leq a_e \leq a_1. \tag{3.2.19}$$

Moreover, $\tilde{a}$ is a second order approximation, and as such more precise than a_1 which can be seen from comparing the bounds in (3.2.10) and (3.2.17). In principle it is possible to continue in this fashion, by e.g. giving a third order upper bound. There are several reasons not to pursue this. In the first place the results become more complicated. Moreover, the lower bound on $\bar{s}$ will increase. Finally, the first and second order bound seem adequate for practical purposes: for small σ and $\bar{s}$ the first order result may suffice, whereas for the remaining cases the second order result will be sufficiently precise. It seems more rewarding from a practical point of view to look for a more convenient version of the second order approximation.

We introduce therefore

$$a_2 = a_1 - \frac{\sigma\bar{s}}{2}\left(a_1^2 + 1 - a_1 k(a_1)\right), \tag{3.2.20}$$

with k as in (3.2.15). Then we have

Lemma 3.2.4 *The approximation a_2 from (3.2.20) and $\tilde{a}$ agree to second order, while moreover*

$$\tilde{a} \leq a_2 \leq a_1. \tag{3.2.21}$$

Proof. From (3.2.9) and (3.2.14) it follows that

$$g_1(\tilde{a}) - \frac{\sigma\bar{s}}{2} g_2(\tilde{a}) = g_1(a_1).$$

Letting $\delta = a_1 - \tilde{a}$, we note that $\delta \geq 0$ and moreover $\delta = -\frac{\sigma\bar{s}}{2} g_2(\tilde{a})/g_1'(\tilde{a}+\theta\delta)$, where $0 \leq \theta \leq 1$. As $\frac{1}{-g_1'(\tilde{a}+\theta\delta)} = \frac{1}{1-\Phi(\tilde{a}+\theta\delta)} \geq \frac{1}{-g_1'(\tilde{a})}$, this leads to

$$\delta \geq -\frac{\sigma\bar{s}}{2}\frac{g_2(\tilde{a})}{g_1'(\tilde{a})} = \frac{\sigma\bar{s}}{2}\left(\tilde{a}^2 + 1 - \tilde{a}k(\tilde{a})\right), \tag{3.2.22}$$

where we use (3.2.6) and (3.2.7) for the second step in (3.2.22). Comparison of (3.2.20) and (3.2.22) shows that in order to prove (3.2.21) it suffices to make clear that $-g_2(x)/g_1'(x)$ is decreasing in x and $\lim_{x\to\infty} \frac{-g_2(x)}{g_1'(x)} = 0$. To show this, observe that $(-g_2/g_1')' = \left(-g_2'g_1' + g_2g_1''\right)/(g_1')^2$, which is negative iff $g_2' - g_2g_1''/g_1' < 0$. This in turn is negative iff $2g_1(x) - g_2(x)k(x) > 0$ (use (3.2.6), (3.2.7) and (3.2.15)), which is equivalent to $\{2 + xk(x)\}g_1(x) > -k(x)g_1'(x) = \phi(x)$. This translates into $(1 + xk(x))\,\phi(x) > (2 + xk(x))\,x\,(1 - \Phi(x))$, which is equivalent to

$$k(x) > x + \frac{x}{1 + xk(x)}. \tag{3.2.23}$$

For $x > 0$, we obtain from Kotz and Johnson (1985), p. 505, the inequality $k(x) > l(x) = (3x + \sqrt{x^2+8})/4$. Hence it suffices to prove (3.2.23) with k replaced by l on both sides. Straightforward evaluation shows the latter inequality to be equivalent to $x^6 + 12x^4 + 36x^2 + 32 > x^6 + 12x^4 + 36x^2$, which is indeed true (and quite sharp for larger x!).

The remaining assertion of the lemma, concerning the agreement to second order between $\tilde{a}$ and a_2, is evident from the proof given. □

In comparing a_2 to $\tilde{a}$, we note to begin with that $\tilde{a}$ is really a lower bound for a_e, whereas, a_2 merely shares with a_e the property of falling between $\tilde{a}$ and a_1. On the other hand, a lower bound is less attractive in itself than an upper bound as it leads to anti-conservative test limits. The attractiveness of $\tilde{a}$ lies mostly in its being a second order approximation. Arguing heuristically, we may expect a_2 to be an improvement in that sense, for, $\tilde{a}$ also serves as a lower bound and as such its error always has the same sign. Such a side condition does not hold for a_2. Of course this argument can be misleading. A thorough check would require comparison of the leading terms in the errors with respect to the consumer loss caused by using a_2 and $\tilde{a}$, respectively. Since this is not difficult but on the other hand tedious, we shall not bother to carry it out here and postpone our judgement to the numerical results in section 3.3.

Having accurate approximations to a_e we will now consider the yield and the producer loss. The yield is the simple univariate probability

$$\textit{Yield} = P(\tilde{X} < t) = \Phi\left(\frac{s - a\sigma_U - \mu_X}{\sqrt{\sigma_X^2 + \sigma_U^2}}\right) = \Phi\left(\frac{\bar{s} - a\sigma}{\sqrt{1+\sigma^2}}\right). \tag{3.2.24}$$

For given a_e the yield can thus be obtained exactly. Typically, we will not be using the exact a_e but an approximation. To judge the error committed in this way, note that for $0 \leq b_1 \leq b_2 \leq \bar{s}/\sigma$

$$\phi\left(\frac{\bar{s} - b_1\sigma}{\sqrt{1+\sigma^2}}\right) \leq \frac{\Phi\left(\frac{\bar{s} - b_1\sigma}{\sqrt{1+\sigma^2}}\right) - \Phi\left(\frac{\bar{s} - b_2\sigma}{\sqrt{1+\sigma^2}}\right)}{\frac{(b_2 - b_1)\sigma}{\sqrt{1+\sigma^2}}} \leq \phi\left(\frac{\bar{s} - b_2\sigma}{\sqrt{1+\sigma^2}}\right). \tag{3.2.25}$$

This shows in the first place that for the very accurate approximation a_2 of a_e, the error in *Yield* will be negligible. The numerical examples shown in table 1.2.1 (section 1.2), thus extend to a_2.

Closely related to the consumer loss and the yield is the producer loss, which is defined by

$$PL = P(X < s,\ \tilde{X} > t). \tag{3.2.26}$$

The producer loss is the probability of needlessly rejecting a product. It is not difficult to verify the following relation.

$$PL = (1 - \pi) - P(X < s,\ \tilde{X} < t) = CL + (1 - \pi) - \textit{Yield}. \tag{3.2.27}$$

Combination of (3.2.27), (3.2.8) and (3.2.24) leads to a particularly simple approximation for the producer loss,

$$PL \approx \sigma\phi(\bar{s})\left(a + \frac{\sigma\bar{s}}{2}(a^2 + 1)\right). \tag{3.2.28}$$

This approximation has to be handled with some care as it involves expansion of $(\bar{s} - a\sigma)/\sqrt{1+\sigma^2}$ around $\bar{s}$. The expansion gets inaccurate for not too large $\bar{s}$ in combination with larger a and σ. Moreover, it contains simplifications $g_1(a) + a \approx a$ and $g_2(a) + (a^2 + 1) \approx a^2 + 1$, which are not adequate for very small a.

At the end of this section we briefly pay attention to the case in which products should be rejected if $X < s$.

Given a test limit t, the consumer loss is then given by

$$CL = P(X < s,\ X + U > t).$$

The test limit here is typically larger than the specification limit, therefore we now define (cf. (3.2.1))

$$a = \frac{t - s}{\sigma_U},\quad a_e = \frac{t_e - s}{\sigma_U}, \tag{3.2.29}$$

and consequently $t = s + a \cdot \sigma_U$. We are again interested in nonnegative a.

We define

$$\sigma = \frac{\sigma_U}{\sigma_X}, \quad Y = \frac{U}{\sigma_U}, \quad \bar{X} = -\frac{X - \mu_X}{\sigma_X}, \quad \bar{s} = -\frac{s - \mu_X}{\sigma_X}, \tag{3.2.30}$$

where Y, $\bar{X}$ and $\bar{s}$ here differ from (3.2.2) by a minus-sign. Again we have that Y and $\bar{X}$ are independent and standard normal, and since we assumed $s < \mu_X$, we also have in the present case $\bar{s} > 0$.

For the consumer loss we write

$$\begin{aligned} CL &= P(X < s,\ X + \sigma_U Y > s + a\sigma_U) \\ &= P(\bar{X} > \bar{s},\ \bar{X} - \sigma Y < \bar{s} - a\sigma) \\ &= P(Y > a,\ \bar{s} < \bar{X} < \bar{s} + \sigma(Y - a)), \end{aligned}$$

which is exactly the same expression as (3.2.3). Hence, all the results which have been derived in this section apply to the present case, however with alternative definitions of a and $\bar{s}$.

3.3 Numerical results

In this section we numerically investigate the quality of the various approximations proposed in section 3.2. By the triple σ, $\bar{s}$ (hence π) and γ the situation is completely specified. The range of (σ, π, γ)-values considered is

$$\begin{aligned} \sigma &= 0.01, 0.05, 0.10, 0.20, 0.30 \\ \pi &= 0.0025, 0.01, 0.05, 0.10, 0.15 \\ \gamma &= 1, 5, 10, 20, 100\text{ppm} \end{aligned} \tag{3.3.1}$$

The true a_e is obtained through numerical evaluation. As it runs from roughly -1 to 4, it is felt that the region of values of practical interest is amply covered by (3.3.1).

We evaluate the upper bound a_1 from (3.2.9), the lower bound $\tilde{a}$ from (3.2.14) and the approximation a_2 from (3.2.20), together with the corresponding consumer losses. In table 3.3.1 on the next page we present these results for all σ and a choice of (π, γ)-values from (3.3.1), This choice is representative, the picture for all 5^3 possibilities is completely similar.

The foremost conclusion form table 3.3.1 is that in particular a_2 performs strikingly well. In fact, $(CL(a_2) - \gamma)/\gamma$ is completely negligible for practical purposes and a_2 can be assumed identical to a_e! Closer inspection of the results provides some additional information along the following lines. In the first place, for the cases considered all four a's behave as they should in the sense that each increases in σ and π and decreases in γ. Moreover, we indeed have that $\tilde{a} \leq a_e$, $a_2 \leq a_1$ (cf (3.2.19) and (3.2.21)), while the first order approximation a_1 is less

Table 3.3.1 Accuracy of a_1, $\tilde{a}$ and a_2.
The approximations a_1, $\tilde{a}$ and a_2 are compared to the true value a_e, for various (σ, π, γ). For each approximation, the realized consumer loss (in ppm) is given as well.

$(\pi, \gamma) = (0.15, 1 \text{ ppm})$

σ	a_1	$CL(a_1)$	$\tilde{a}$	$CL(\tilde{a})$	a_2	$CL(a_2)$	a_e
0.01	2.9672	1.00	2.9664	1.00	2.9664	1.00	2.9664
0.05	3.4026	0.99	3.3994	1.00	3.3994	1.00	3.3993
0.10	3.5773	0.98	3.5712	1.00	3.5713	1.00	3.5712
0.20	3.7454	0.95	3.7338	1.00	3.7341	1.00	3.7339
0.30	3.8409	0.93	3.8241	1.00	3.8247	1.00	3.8244

$(\pi, \gamma) = (0.05, 20 \text{ ppm})$

σ	a_1	$CL(a_1)$	$\tilde{a}$	$CL(\tilde{a})$	a_2	$CL(a_2)$	a_e
0.01	1.6758	19.88	1.6733	20.00	1.6733	20.00	1.6733
0.05	2.2803	19.49	2.2712	20.01	2.2714	20.00	2.2714
0.10	2.5106	19.05	2.4943	20.03	2.4948	20.00	2.4948
0.20	2.7272	18.24	2.6971	20.12	2.6989	20.00	2.6989
0.30	2.8483	17.50	2.8046	20.26	2.8084	20.01	2.8086

$(\pi, \gamma) = (0.01, 40 \text{ ppm})$

σ	a_1	$CL(a_1)$	$\tilde{a}$	$CL(\tilde{a})$	a_2	$CL(a_2)$	a_e
0.01	0.6708	39.54	0.6638	40.00	0.6638	40.00	0.6638
0.05	1.4895	38.24	1.4688	40.06	1.4695	40.00	1.4695
0.10	1.7801	36.85	1.7445	40.23	1.7465	40.02	1.7468
0.20	2.0458	34.44	1.9815	40.84	1.9883	40.11	1.9894
0.30	2.1918	32.36	2.0979	41.88	2.1123	40.27	2.1148

$(\pi, \gamma) = (0.0025, 100 \text{ ppm})$

σ	a_1	$CL(a_1)$	$\tilde{a}$	$CL(\tilde{a})$	a_2	$CL(a_2)$	a_e
0.01	-1.2365	97.34	-1.2771	100.05	-1.2756	99.95	-1.2763
0.05	0.3240	92.57	0.2654	100.51	0.2695	99.93	0.2690
0.10	0.7603	87.66	0.6713	101.64	0.6814	99.96	0.6812
0.20	1.1317	79.67	0.9801	105.91	1.0098	100.30	1.0114
0.30	1.3273	73.22	1.1042	113.72	1.1665	100.95	1.1714

accurate than the second order approximation $\tilde{a}$, which (almost always) is less accurate than a_2 (cf. the remarks following lemma 3.2.4). It is also evident that the errors grow as σ and $\bar{s}$ increase, which is in agreement with the bounds from (3.2.10) and (3.2.17). As concerns the quality of these bounds, for the cases investigated the upper bound (3.2.10) for the first order approximation performs reasonably well in the sense that it is of the right order of magnitude. On the average it differs from the actual error by a factor 2.5. The upper bound (3.2.17) for the second order approximation is relatively less precise. This is not surprising, as the errors involved can be extremely small. For large σ and $\bar{s}$, where non-negligible errors occur, the bound again is of the right order of

magnitude and differs by a factor of about 5 on the average from the actual error.

Having established in the above that when using a_2 the CL is under almost perfect control, we conclude that also with a_2 the gains in *Yield* as described in section 1.2 are attained.

3.4 Estimation of parameters

We have obtained a very accurate approximation to t_e in which the role of s, γ and μ_X, σ_X and σ_U has become more explicit. In practical situations we shall often have to rely on corresponding estimators $\hat{\mu}_X$, $\hat{\sigma}_X$ and $\hat{\sigma}_U$. We will write $\hat{t} = t(\hat{\mu}_X, \hat{\sigma}_X, \hat{\sigma}_U)$ and $\hat{a} = a(\hat{\mu}_X, \hat{\sigma}_X, \hat{\sigma}_U)$, and by $CL(\hat{t})$ we denote the probability of a new item to be nonconforming and accepted, that is $P(X > s,\ \tilde{X} < \hat{t})$, given $\hat{t}$.

3.4.1 Second order unbiased test limits

We will look for $\hat{t}_u = t_u(\hat{\mu}_X, \hat{\sigma}_X, \hat{\sigma}_U)$ such that

$$\mathrm{E}CL(\hat{t}_u) = \gamma, \tag{3.4.1}$$

to sufficient precision for practical purposes.

Arguing as in (3.2.2)-(3.2.4), we obtain that (3.4.1) can be translated into

$$\mathrm{E}\int_0^\infty \{\Phi(\bar{s}+\sigma y) - \Phi(\bar{s})\}\phi(y + \hat{\sigma}_U \hat{a}_u/\sigma_U)\mathrm{d}y = \gamma, \tag{3.4.2}$$

where in analogy to (3.2.1), we have $\hat{a}_u = a_u(\hat{\mu}_X, \hat{\sigma}_X, \hat{\sigma}_U)$ defined by

$$\hat{a}_u = \frac{s - \hat{t}_u}{\hat{\sigma}_U}. \tag{3.4.3}$$

To determine the function $a_u = a_u(\mu_X, \sigma_X, \sigma_U)$ we can use an expansion similar to (3.2.8), leading from (3.4.2) to

$$\mathrm{E}\left\{g_1\left(\frac{\hat{\sigma}_U}{\sigma_U}\hat{a}_u\right) - \frac{\sigma\bar{s}}{2}g_2\left(\frac{\hat{\sigma}_U}{\sigma_U}\hat{a}_u\right) + \ldots\right\} = \frac{\gamma}{\sigma\phi(\bar{s})}, \tag{3.4.4}$$

again to sufficient precision. Comparison of (3.4.4) with (3.2.14) suggests that the use of consistent estimators for the three parameters involved will lead to an a_u which is closely related to $\tilde{a}$, and hence to a_2 from (3.2.20). The difference will consist of an appropriate term

$$c_u = a_u - a_2 \tag{3.4.5}$$

to correct for the effect of having estimated instead of being able to use the known values.

For arbitrary $(\hat{\mu}_X, \hat{\sigma}_X, \hat{\sigma}_U)$ the explicit expression for this difference will be unduly complicated. Therefore we first present two choices for these estimators which are likely to be used in practice.

Typically, estimation of σ_U will entail a series of repeated measurements. Suppose we consider n parts and apply p replications,

$$\tilde{X}_{ij} = X_i + U_{ij}, \quad i = 1, \ldots, n, \; j = 1, \ldots, p. \tag{3.4.6}$$

Denoting averaging over an index by a '$\bullet$', we have (cf. e.g. Scheffé (1959, p. 228)) the three independent statistics

$$\begin{aligned} &\tilde{X}_{\bullet\bullet} \\ &S_w^2 = \frac{1}{n(p-1)} \sum_{i=1}^{n} \sum_{j=1}^{p} (\tilde{X}_{ij} - \tilde{X}_{i\bullet})^2 \\ &S_b^2 = \frac{1}{n-1} \sum_{i=1}^{n} (\tilde{X}_{i\bullet} - \tilde{X}_{\bullet\bullet})^2. \end{aligned} \tag{3.4.7}$$

As moreover

$$\begin{aligned} \tilde{X}_{\bullet\bullet} &\sim N\left(\mu_X, \frac{\sigma_X^2}{n} + \frac{\sigma_U^2}{np}\right) \\ \frac{n(p-1)S_w^2}{\sigma_U^2} &\sim \chi^2_{n(p-1)} \\ \frac{(n-1)S_b^2}{\sigma_X^2 + \sigma_U^2/p} &\sim \chi^2_{n-1}, \end{aligned}$$

this leads to the UMVU (uniformly minimum variance unbiased) estimators (cf. Lehmann (1983, p. 198))

$$\begin{aligned} \hat{\mu}_X &= \tilde{X}_{\bullet\bullet} \\ \hat{\sigma}_X^2 &= S_b^2 - S_w^2/p \\ \hat{\sigma}_U^2 &= S_w^2. \end{aligned} \tag{3.4.8}$$

It is straightforward to obtain (co)variances. We e.g. have

$$\begin{aligned} \mathrm{VAR}(\hat{\sigma}_X^2) &= \frac{2(\sigma_X^2 + \sigma_U^2/p)}{n-1} + \frac{2\sigma_U^4}{np^2(p-1)} \\ \mathrm{COV}(\hat{\sigma}_X^2, \hat{\sigma}_U^2) &= -\frac{2\sigma_U^4}{np(p-1)}. \end{aligned}$$

Since we have that $\sigma = \sigma_U/\sigma_X \ll 1$, we may simplify to (equalities holding up to $O(\sigma^2)$)

$$\begin{aligned} &n\mathrm{VAR}\left(\frac{\hat{\mu}_X}{\sigma_X}\right) = \frac{n-1}{2}\mathrm{VAR}\left(\frac{\hat{\sigma}_X^2}{\sigma_X^2}\right) = \frac{n(p-1)}{2}\mathrm{VAR}\left(\frac{\hat{\sigma}_U^2}{\sigma_U^2}\right) = 1 + O(\sigma^2) \\ &\mathrm{COV}\left(\frac{\hat{\sigma}_X^2}{\sigma_X^2}, \frac{\hat{\sigma}_U^2}{\sigma_U^2}\right) = O(\sigma^2). \end{aligned} \tag{3.4.9}$$

Given a budget of np observations to base the estimators on, the question remains how to choose p. Inspection of (3.4.9) suggests strongly to let $p = 2$. Indeed, for $\theta = \sigma_X^2/\sigma_U^2 \geq 1$, which is clearly the case here, Scheffé (1959, p. 238) shows that this is the optimal choice. Hence we shall always let p equal 2 in what follows. Summarizing our first choice, we observe that the estimators defined through (3.4.8) are unbiased, and to the order considered, independent with variances n^{-1}, $2n^{-1}$ and $2n^{-1}$ for $\hat{\mu}_X/\sigma_X$, $\hat{\sigma}_X^2/\sigma_X^2$ and $\hat{\sigma}_U^2/\sigma_U^2$, respectively.

Also of practical interest is the following design. Quite often the test data described above form part of a typically much larger set of production data. Hence we have a sample $\tilde{X}_1, \ldots, \tilde{X}_m$ from $\tilde{X}$, with $m > n$ (and usually $m \gg n$). A subset of size n is measured twice, giving rise to $\tilde{X}_{i,1} = \tilde{X}_{i1}$ and a replication $\tilde{X}_{i,2} = \tilde{X}_{i2}$ for $i = 1, \ldots, n$. The $\tilde{X}_{ij}$ are again of the form (3.4.6), with $p = 2$. For σ_U^2 the situation is the same as before, just as in (3.4.8) it is estimated by (using that $p = 2$)

$$\hat{\sigma}_U^2 = \frac{1}{2n}\sum_{i=1}^{n}(\tilde{X}_{i2} - \tilde{X}_{i1})^2. \tag{3.4.10}$$

But the estimators for μ_X and σ_X^2 can be improved by using the non-replicated $m - n$ observations from production. In fact, the mean $\tilde{X}_\bullet^{(m-n)}$ of these latter data is normally distributed with mean μ_X and variance $(\sigma_X^2 + \sigma_U^2)/(m-n)$ and $\tilde{X}_\bullet^{(m-n)}$ can be combined optimally with $X_{\bullet\bullet}$ from (3.4.7). It is immediately clear, however, that to $O(\sigma^2)$ it is also optimal to use simply $\hat{\mu}_X = \tilde{X}_\bullet$, for, the gain in variance reduction by using the replications X_{i2} is utterly negligible, especially when besides $\sigma_U \ll \sigma_X$ we also have $m \gg n$. For $\hat{\sigma}_X^2$ completely analogous arguments hold. Hence in this situation we suggest

$$\begin{aligned} \hat{\mu}_X &= \tilde{X}_\bullet, \\ \hat{\sigma}_X^2 &= \frac{1}{m-1}\sum_{i=1}^{m}(\tilde{X}_i - \tilde{X}_\bullet)^2 - \hat{\sigma}_U^2, \end{aligned} \tag{3.4.11}$$

with $\hat{\sigma}_U^2$ as in (3.4.10). Again the estimators are unbiased, and to the order considered independent, but now with variances m^{-1}, $2m^{-1}$ and $2m^{-1}$ for $\hat{\mu}_X/\sigma_X$, $\hat{\sigma}_X^2/\sigma_X^2$ and $\hat{\sigma}_U^2/\sigma_U^2$, respectively.

In the following theorem it is proved that with the estimators as in (3.4.8) or as in (3.4.10) in combination with (3.4.11) second order unbiasedness is obtained when taking (cf. (3.4.3))

$$\begin{aligned} c_u &= a_u - a_2 = \\ &\frac{k(a_1)\{2a_1k(a_1) + 1 - a_1^2\}}{4n} + \frac{(\bar{s}^4 + 4\bar{s}^2 + 1)(k(a_1) - a_1)}{4m}, \end{aligned} \tag{3.4.12}$$

with a_2 from (3.2.20), a_1 from (3.2.9) and k as in (3.2.15). In the case where no production data are available we should read $m = n$.

Theorem 3.4.1 *Let $\bar{s}$ be bounded and suppose $n \to \infty$ and $\sigma \to 0$, $\gamma \to 0$ such that a_1 is bounded.*
For $\hat{t}_u = s - \hat{a}_u \hat{\sigma}_U$, with $\hat{a}_u = a_u(\hat{\mu}_X, \hat{\sigma}_X, \hat{\sigma}_U)$ from (3.4.12) and $\hat{\mu}_X$, $\hat{\sigma}_X$, $\hat{\sigma}_U$ as in (3.4.8), or as in (3.4.10) and (3.4.11), we have, for some $C > 0$

$$\mathrm{E}CL(\hat{t}_u) = \gamma\{1 + O(n^{-2} + \sigma^2)\} + O\left(\mathrm{e}^{-Cn}\right). \tag{3.4.13}$$

Proof. Once the theorem has been proved for the choice of (3.4.8), the case where $m > n$ follows automatically. hence we will assume $m = n$. Moreover, let

$$V = \frac{\hat{\mu}_X - \mu_X}{\sigma_X}, \quad W = \frac{\hat{\sigma}_X^2}{\sigma_X^2} - 1, \quad Z = \frac{\hat{\sigma}_U^2}{\sigma_U^2} - 1. \tag{3.4.14}$$

For the set

$$A = \{\max(|V|,\ |W|,\ |Z|) < \epsilon\}, \tag{3.4.15}$$

we have for every $\epsilon > 0$ that $P(A^c) = O\left(\mathrm{e}^{-Cn}\right)$ for some $C > 0$. Since probabilities are bounded, $\mathrm{E}\left(CL(\hat{t}_u) - \gamma | I_{A^c}\right) = O\left(\mathrm{e}^{-Cn}\right)$, where I_{A^c} denotes the indicator function of the set A^c. Hence the inclusion of the last term in (3.4.13) allows us to restrict attention to A in what follows, i.e. to assume that each estimator is close to the parameter it corresponds to, which in its turn makes it possible to apply expansions.

Let $d = c_u + (a_2 - a_1) = a_u - a_1$, $\hat{d} = d(\hat{\mu}_X, \hat{\sigma}_X, \hat{\sigma}_U)$, $\hat{a}_1 = a_1(\hat{\mu}_X, \hat{\sigma}_X, \hat{\sigma}_U)$ etc. From (3.4.12) and (3.2.20) it follows that $d = O(n^{-1} + \sigma)$. Together with (3.4.2)-(3.4.4) this gives that the expected consumer loss $\mathrm{E}CL(\hat{t}_u)$ equals

$$\mathrm{E}\left\{\sigma\phi(\bar{s})\left[g_1\left(\frac{\hat{\sigma}_U}{\sigma_U}\hat{a}_1\right) + \frac{\hat{\sigma}_U}{\sigma_U}\hat{d}\ g_1'\left(\frac{\hat{\sigma}_U}{\sigma_U}\hat{a}_1\right)\right] - \frac{\sigma^2\bar{s}}{2}\phi(\bar{s})\ g_2\left(\frac{\hat{\sigma}_U}{\sigma_U}\hat{a}_1\right)\right\} \cdot I_A$$
$$+ O\left(\gamma(n^{-2} + \sigma^2) + \mathrm{e}^{-Cn}\right). \tag{3.4.16}$$

As $\hat{a}_1 = g_1^{-1}\left(\gamma / \left\{\frac{\hat{\sigma}_U}{\hat{\sigma}_X}\phi\left(\frac{s - \hat{\mu}_X}{\hat{\sigma}_X}\right)\right\}\right)$, we observe that the expression above requires expansion of $m(X, Y) = g_1(X g_1^{-1}(bY/X))$ around $b = m(1, 1)$, where

$$b = \frac{\gamma}{\sigma_U \phi(\bar{s})/\sigma_X} = g_1(a_1),\ X = \frac{\hat{\sigma}_U}{\sigma_U},\ Y = \frac{\hat{\sigma}_X}{\sigma_X} \frac{\phi(\bar{s})}{\phi\left(\frac{s - \hat{\mu}_X}{\hat{\sigma}_X}\right)}. \tag{3.4.17}$$

To evaluate the expectation of the resulting expansion, we need $\mathrm{E}(X-1)^k(Y-1)^l$, for integers k and l with $0 \le k,\ l \le 4$. In view of (3.4.14) we have $X - 1 = (1+Z)^{1/2} - 1 = \frac{1}{2}Z - \frac{1}{8}Z^2 + \ldots$. From (3.4.8) it is known that $\mathrm{E}Z = 0$, $\mathrm{E}Z^2 = 2n^{-1}$. Moreover, it is also easily verified that $\mathrm{E}Z^3$ and $\mathrm{E}Z^4$ are both $O(n^{-2})$. Hence $\mathrm{E}(X-1) = -(4n)^{-1}$ and $\mathrm{E}(X-1)^2 = (2n)^{-1}$, apart from the contribution to the remainder in (3.4.13), while $\mathrm{E}(X-1)^3$ and $\mathrm{E}(X-1)^4$ only contribute to the remainder. For $(Y-1)$ we observe that (3.4.14) and (3.4.17) give

$$Y = \frac{\sqrt{1+W}\,\phi(\bar{s})}{\phi((\bar{s} - V)/\sqrt{1+W})}, \tag{3.4.18}$$

which can be expanded in terms of V and W. Proceeding as above we obtain that to the desired order $\mathrm{E}(Y-1) = \dfrac{(\bar{s}^4 + 4\bar{s}^2 + 1)(k(a_1) - a_1)}{4n}$. The remaining (mixed) moments can either be relegated to the remainder, or will not occur in the expansion (this is the case with $\mathrm{E}(Y-1)^2$). In fact, the expansion of $m(X,Y) - m(1,1)$ up to and including second order terms equals

$$\begin{aligned} &-\phi(a_1)\cdot(X-1) + g_1(a_1)\cdot(Y-1) + \tfrac{1}{2}a_1\phi(a_1)(2k(a_1) - a_1)\cdot(X-1)^2 \\ &\quad - a_1 k(a_1) g_1(a_1)\cdot(X-1)(Y-1). \end{aligned} \tag{3.4.19}$$

Since terms like $|\mathrm{E}(X-1)\cdot 1_{A^c}| \le \mathrm{E}(X-1)^4 + \mathrm{E}1_{A^c} = O(n^{-2} + e^{-Cn})$ we may use the evaluated moments. Using (3.4.19) we now obtain from (3.4.16) that $ECL(\hat{t}_u) - \gamma$ equals

$$\begin{aligned} &\sigma\phi(\bar{s})g_1'(a_1)\left\{-\frac{k(a_1)\{2a_1k(a_1) + 1 - a_1^2\}}{4n} - \frac{(\bar{s}^4 + 4\bar{s}^2 + 1)(k(a_1) - a_1)}{4n} + d\right\} \\ &-\frac{\sigma^2\bar{s}}{2}\phi(\bar{s})g_2(a_1) + O\left(\gamma(n^{-2} + \sigma^2) + \mathrm{e}^{-Cn}\right). \end{aligned} \tag{3.4.20}$$

But from the definition of d it is immediate that the choice for c_u in (3.4.12) will cancel all the terms in (3.4.20), apart from the remainder. □

The correction term c_u in (3.4.12) consists of two parts: the part with n in the denominator corrects for estimation of σ_U, while the other part, with m in the denominator, corrects for estimation of μ_X and σ_X. Since $k(x) \ge x$, both terms in (3.4.12) are nonnegative and thus $a_u > a_2$. This agrees with our intuition as $a_u - a_2$ stands for the complication of having to estimate the parameters, rather than knowing them. Moreover, note that the behavior of the correction terms in $a_u - a_2$ differs from that of $a_2 - a_1$ (cf. (3.2.20)). The coefficient $a_1^2 + 1 - a_1k(a_1)$ of $\sigma\bar{s}/2$ is quite small (e.g. it decreases from 1 for

$a_1 = 0$ to 0.26 for $a_1 = 2$). But in (3.4.12) the coefficient of $(4n)^{-1}$ increases considerably as a_1 grows (e.g. from 0.8 for $a_1 = 0$ to 15.4 for $a_1 = 2$), while the factor $(\bar{s}^4 + 4\bar{s}^2 + 1)$ in the coefficient of $(4m)^{-1}$ also clearly increases steeply in $\bar{s}$.

To obtain an impression whether replacement of a_2 by a_u indeed compensates for the effect of using estimators for μ_X, σ_X and σ_U, a simulation study has been carried out. The following sample sizes have been taken, $n = 40, 80$ and $m = 40, 80, 2500$. For the case $m = 2500$ we expect that we may identify the estimators of μ_X and σ_X with their true values and thus that we may omit the second part of correction. To study the effect of the separate corrections, the consumer loss has been computed as well if a_2 is corrected for the estimation of σ_U only, hence omitting the second part of c_u. In table 3.4.1 below the study is summarized.

Table 3.4.1. Accuracy of $\hat{a}_u$

Three test limits are computed when using estimators of μ_X, σ_X and σ_U. Based on a simulation with 10000 replications, the average consumer losses are shown.

- CL_e denotes the average consumer loss if simply a_2 is computed.
- CL_n refers to the consumer loss if a_2 is corrected for estimating σ_U only.
- CL_{nm} refers to the consumer loss if there is corrected for the estimation of μ_X and σ_X as well.

	$n = 40$			$n = 80$		
	$(\sigma, \pi, \gamma) = (0.01, 0.15, 20$ ppm$)$					
m	CL_e	CL_n	CL_{nm}	CL_e	CL_n	CL_{nm}
40	26.4	21.2	20.3		-	
80	25.7	20.6	20.2	23.2	20.6	20.2
2500	25.5	20.4	20.4	22.7	20.2	20.2
	$(\sigma, \pi, \gamma) = (0.10, 0.15, 20$ ppm$)$					
40	36.9	22.1	21.2		-	
80	34.9	20.7	20.3	27.3	20.4	20.0
2500	35.2	20.8	20.8	26.9	20.1	20.0
	$(\sigma, \pi, \gamma) = (0.20, 0.15, 20$ ppm$)$					
40	40.5	21.6	20.7		-	
80	39.7	21.1	20.7	29.6	20.7	20.3
2500	39.5	20.9	20.9	29.5	20.6	20.6

Table 3.4.1. (Continued) Accuracy of $\hat{a}_u$

Three test limits are computed when using estimators of μ_X, σ_X and σ_U. Based on a simulation with 10000 replications, the average consumer losses are shown.

- CL_e denotes the average consumer loss if simply a_2 is computed.
- CL_n refers to the consumer loss if a_2 is corrected for estimating σ_U only.
- CL_{nm} refers to the consumer loss if there is corrected for the estimation of μ_X and σ_X as well.

	$n = 40$			$n = 80$		
	$(\sigma, \pi, \gamma) = (0.01, 0.10, 40$ ppm$)$					
m	CL_e	CL_n	CL_{nm}	CL_e	CL_n	CL_{nm}
40	49.0	43.3	40.2		-	
80	47.8	42.0	40.7	44.5	41.6	40.2
2500	45.7	40.0	39.9	43.0	40.1	40.1
	$(\sigma, \pi, \gamma) = (0.10, 0.10, 40$ ppm$)$					
40	65.1	45.3	42.2		-	
80	61.2	42.1	40.9	50.9	41.5	40.2
2500	59.7	40.8	40.8	49.1	39.9	39.8
	$(\sigma, \pi, \gamma) = (0.20, 0.10, 40$ ppm$)$					
40	69.8	43.9	41.0		-	
80	67.6	42.1	40.9	53.8	41.4	40.1
2500	65.9	40.8	40.8	52.3	40.1	40.1
	$(\sigma, \pi, \gamma) = (0.01, 0.01, 40$ ppm$)$					
40	156.7	156.1	82.8		-	
80	121.2	120.6	97.8	122.5	122.1	98.9
2500	101.3	100.6	100.1	100.7	100.3	99.8
	$(\sigma, \pi, \gamma) = (0.10, 0.01, 100$ ppm$)$					
40	173.2	163.6	94.6		-	
80	132.2	122.3	100.9	126.8	121.8	100.3
2500	110.3	100.4	99.9	105.0	100.0	99.5
	$(\sigma, \pi, \gamma) = (0.20, 0.01, 100$ ppm$)$					
40	187.3	171.4	100.0		-	
80	140.5	124.7	103.5	130.5	122.5	101.2
2500	114.8	99.4	99.0	107.3	99.5	99.0

The results clearly indicate that for the sample sizes considered, the result $\mathrm{E}CL(\hat{t}_u) = \gamma$ is realized with sufficient precision for practical purposes. More important, the results also show that this is absolutely not the case if we neglect the correction to a_2. Without correction average consumer losses up to 2γ are obtained and we conclude that the assertions which we made in section 1.4 are justified. Plugging in estimates and continuing as if it were the true parameter values can lead to averages which are much larger than γ indeed.

We see that in many situations correcting for estimation of σ_U only already gives a great improvement. However, when s is in the tail of the distribution, hence when $\bar{s}$ is large (cf. the case $\pi = 0.01$), correction for the estimation of

μ_X and σ_X is necessary as well. Note that this is perfectly in correspondence with the theory, as in these situations $(\bar{s}^4 + 4\bar{s}^2 + 1)$ is large. Finally, note that the table indicates that for the situation of $m = 2500$ correction for estimation of σ_U only is sufficient.

3.4.2 Test limits for which γ is violated with small probability

It is gratifying to note in the previous section that our second order asymptotics work well in the sense that $CL(\hat{t}_u)$ on the average is close to the prescribed γ. However, simulations also show that $CL(\hat{t}_u)$ varies widely around γ (e.g. leading to upper and lower 5% quantiles like 215 and 30 ppm, respectively, when $\gamma=$ 100 ppm, $n = 40$ and even $m = 2500$). It is not difficult to identify the cause of this variation. A glance at the proof of theorem 3.4.1 reveals that in the expansion of $CL(\hat{t}_u)$ factors like $(\hat{\sigma}_U/\sigma_U - 1)$ occur (cf. (3.4.17) and (3.4.19)). While $\mathrm{E}(\hat{\sigma}_U/\sigma_U - 1)$ is of order n^{-1}, we have that $(\hat{\sigma}_U/\sigma_U - 1)$ itself is of order $n^{-1/2}$ in probability. As we noted in the previous section that terms of order n^{-1} are not at all negligible in correcting the expectation for intermediate n, it stands to reason that terms of order $n^{-1/2}$ will be quite important. In fact, let

$$l(x) = \frac{k(x)}{k(x) - x}, \tag{3.4.21}$$

with k as in (3.2.15). We can show that

Lemma 3.4.1 *Under the conditions of theorem 3.4.1 we have,*

$$\mathrm{E}\left(\frac{CL(\hat{t}) - \gamma}{\gamma}\right)^2 = \frac{l^2(a_1)}{2n} + \frac{\bar{s}^4 + 1}{2m} + O(n^{-2} + \sigma^4 + \gamma^{-2}\mathrm{e}^{-Cn}), \tag{3.4.22}$$

both for $\hat{t} = s - \hat{a}_u\hat{\sigma}_U$ *and for* $\hat{t} = s - \hat{a}_2\hat{\sigma}_U$.

Proof. Using (3.4.1)-(3.4.4) in combination with (3.4.12), (3.4.16) and (3.4.19) we obtain that (on the set A, cf. (3.4.15))

$$\begin{aligned} & CL(\hat{t}_u) - \gamma \\ &= \sigma\phi(\bar{s})g_1'(a_1)\left\{(X-1)\cdot k(a_1) - (Y-1)\cdot(k(a_1) - a_1) + c_u\right\} \\ &\quad + O\left(\gamma\{(X-1)^2 + (Y-1)^2 + n^{-2} + \sigma^2\}\right). \end{aligned} \tag{3.4.23}$$

Now $\sigma\phi(\bar{s}) = \gamma/b = \gamma/g_1(a_1)$, and therefore $\sigma\phi(\bar{s})g_1'(a_1) = -\gamma/(k(a_1) - a_1)$, which in its turn leads to, using (3.4.21),

$$\begin{aligned} \frac{CL(\hat{t}_u) - \gamma}{\gamma} &= -(X-1)l(a_1) + (Y-1) - c_u l(a_1)/k(a_1) \\ &\quad + O\left(n^{-2} + \sigma^2 + (X-1)^2 + (Y-1)^2\right). \end{aligned} \tag{3.4.24}$$

Through (3.4.18) it can be shown that $\mathrm{E}(Y-1)^2 = (\bar{s}^4+1)/(4m)$ to the desired order. We already know that $\mathrm{E}(X-1)^2 = (2n)^{-1}$ and that all other moments involved are $O(n^{-2})$. Hence the result in (3.4.22) follows for $t = s - a_u\sigma_U$.

As c_u only plays a role in the remainder, it is immediate that for $t = s - a_2\sigma_U$ the result is true as well. □

A first conclusion from this lemma is that the correction step from a_2 to a_u in (3.4.12), which in the previous section was seen to be vital in controlling the unbiasedness, is unimportant in the context of the mean squared (relative) error. More interesting it is to note that (3.4.22) and (3.4.24) are quite illuminating, as these formulas provide a simple explanation for the wide variability in the values of $CL(\hat{t}_u)$ mentioned at the beginning of this section. In going from $\hat{\sigma}_U$ to $CL(\hat{t}_u)$ the relative error is inflated by a factor $l(a_1)$ which rapidly increases in a_1 (e.g. $a_1 = 1$ gives 3, while for $a_1 = 2$ it equals 6.4). Next, the way from $\hat{\mu}_X$ and $\hat{\sigma}_X$ to $CL(\hat{t}_u)$ goes through $(Y-1)$. Now $(Y-1)$ lacks an inflating factor, but from (3.4.18) we see that here the effect is hidden inside, leading to the factor $\bar{s}^4+1$ in (3.4.22). Together, these effects explain why a heuristic argument, according to which the fact that for moderately large n and m the estimates $\hat{\mu}_X$, $\hat{\sigma}_X$ and $\hat{\sigma}_U$ will be reasonably close to μ_X, σ_X and σ_U, respectively, should imply that $CL(\hat{t}_u)$ is reasonably close to γ as well, turns out to be quite misleading.

Fortunately, using lemma 3.4.1 it is now rather straightforward to find a correction c_i to a_2, leading to $a_i = a_2 + c_i$ and $t_i = s - a_i\sigma_U$ such that

$$P(CL(\hat{t}_i) > \gamma) = \alpha \tag{3.4.25}$$

with sufficient precision, for some small, given, probability α. If c_i is small it is seen from (3.4.23) that under the conditions of theorem 3.4.1 $CL(\hat{t}_i)$ is asymptotically normal,

$$\frac{CL(\hat{t}_i)-\gamma}{\gamma} \text{ is } AN(\mu_{CL},\ \sigma^2_{CL}), \tag{3.4.26}$$

with (neglecting terms of order n^{-1} and m^{-1} in μ_{CL})

$$\begin{aligned} \mu_{CL} &= -c_i\frac{l(a_1)}{k(a_1)} \\ \sigma^2_{CL} &= \frac{l^2(a_1)}{2n} + \frac{(\bar{s}^4+1)}{2m}. \end{aligned} \tag{3.4.27}$$

By taking the correction term c_i such that $\mu_{CL} = -\sigma_{CL}\Phi^{-1}(1-\alpha)$, (3.4.25) is realized with sufficient precision. We obtain

$$c_i = a_i - a_2 = u_\alpha\sqrt{\frac{k^2(a_1)}{2n} + \frac{(k(a_1)-a_1)^2(\bar{s}^4+1)}{2m}}, \tag{3.4.28}$$

where $u_\alpha = \Phi^{-1}(1-\alpha)$ is the upper α-quantile of the standard normal distribution and a_2, a_1 and k are given in (3.2.20), (3.2.9) and (3.2.15), respectively.

With this choice of c_i we may indeed apply (3.4.23) and we have proved the following theorem.

Theorem 3.4.2 *Under the conditions of theorem 3.4.1 we have for* $\hat{a}_i = a_i(\hat{\mu}_X, \hat{\sigma}_X, \hat{\sigma}_U)$ *from (3.4.28),* $\hat{\mu}_X$, $\hat{\sigma}_X$, $\hat{\sigma}_U$ *as in (3.4.8), or as in (3.4.10) and (3.4.11), and* $\hat{t}_i = s - \hat{a}_i\hat{\sigma}_U$, *that*

$$P(CL(\hat{t}_i) > \gamma) = \alpha + o(1). \tag{3.4.29}$$

Note that the correction term c_i is of order $n^{-1/2}$ (if $m \geq n$) and hence much larger than the correction term c_u. We introduced a negative bias and we have to first order that

$$\mathrm{E}CL(\hat{t}_i) = \gamma(1-\delta), \tag{3.4.30}$$

with $\delta = u_\alpha\sqrt{l^2(a_1)/(2n) + (\bar{s}^4+1)/(2m)}$. Thus to ensure that γ is violated with probability at most α, the average consumer loss has to be lowered to $\gamma(1-\delta)$. In passing we note that for $\alpha = \frac{1}{2}$ we have $\delta = 0$ and $CL(\hat{t}_i)$ agrees to $CL(\hat{t}_u)$ to first order.

Numerical results which give some insight into the accuracy of the results will be presented in the next section.

3.4.3 Choosing the number of observations

Application of the stronger criterion $P(CL(\hat{t}) > \gamma)$ instead of $\mathrm{E}CL(\hat{t}) = \gamma$ will lead to a reduction of the yield. A sensible criterion now requires that the consumer loss violates the bound γ with probability α only, while the average loss of yield is limited to β_0, for example $\beta_0 = 0.1\%$.

Consider the test limit $t_i = s - a_i\sigma_U$ (cf. (3.4.25) and (3.4.28)). The corresponding yield is (cf. (3.2.24))

$$Yield = P(\tilde{X} < t_i) = \Phi\left(\frac{s - a_i\sigma_U - \mu_X}{\sqrt{\sigma_X^2 + \sigma_U^2}}\right) = \Phi\left(\frac{\bar{s} - a_i\sigma}{\sqrt{1+\sigma^2}}\right). \tag{3.4.31}$$

Since both $a_i - a_u$ and $a_i - a_2$ are of order $n^{-1/2}$ (cf. (3.4.12) and (3.4.28)) we expand (3.4.31) around $(s-a_2)/\sqrt{1+\sigma^2}$, leading to

$$\Phi\left(\frac{\bar{s} - a_2\sigma}{\sqrt{1+\sigma^2}}\right) - \Phi\left(\frac{\bar{s} - a_i\sigma}{\sqrt{1+\sigma^2}}\right) \approx \frac{c_i\sigma}{\sqrt{1+\sigma^2}}\phi\left(\frac{\bar{s} - a_2\sigma}{\sqrt{1+\sigma^2}}\right) = \beta. \tag{3.4.32}$$

From (3.2.25) it is seen that this approximation actually is a lower bound (if $\bar{s} > (a_2 + c_i)\sigma$), but also that the approximation in itself is quite accurate. To limit the average reduction in yield to β_0, it follows for c_i that

$$c_i < \beta_0 \left\{ \frac{\sigma}{\sqrt{1+\sigma^2}} \phi\left(\frac{\bar{s} - a_2\sigma}{\sqrt{1+\sigma^2}} \right) \right\}^{-1}.$$

To get an impression of the sample sizes involved, we specialize to the case where $m \gg n$, then the second term on the right-hand side of (3.4.28) can be neglected and it follows that

$$n > n_i = \frac{1}{2} \left\{ \frac{u_\alpha k(a_1)\sigma}{\beta_0\sqrt{1+\sigma^2}} \phi\left(\frac{\bar{s} - a_2\sigma}{\sqrt{1+\sigma^2}} \right) \right\}^2. \tag{3.4.33}$$

Note that for $\alpha = \frac{1}{2}$, $n_i = u_\alpha = 0$ there is no loss of yield to first order. In practice u_α varies between 1.28 (α=0.10) and 1.65 (α=0.05) and $k(a_1)$ varies between 1 and 3 (using that $a \leq k(a) \leq a+1$ for $a > 0$). Together, this leads to a factor between $1.28^2 = 1.6$ and $(1.65 \cdot 3)^2 = 24$ for n_i. Usually the value of σ is between 0.01 and 0.20 and $\phi(\bar{s})$ ranges from 0.03 for $\bar{s} = 2.33$ ($\pi = 0.01$) to 0.28 for $\bar{s} = 0.84$ ($\pi = 0.20$). We see that with choices of β_0 between 0.1% and 1%, the value of n_i varies widely. For example, the combination $\sigma = 0.20$, $\bar{s} = 1$ and $\beta_0 = 0.1\%$ gives $\sigma\phi(\bar{s})/\beta_0 \approx 48$, hence leads to a factor of about 2300 for n_i. Taking $\beta_0 = 1\%$ instead, the factor reduces to about $4.8^2 = 23$. Obviously, it is not feasible to choose a value of β_0 irrespective of the true (but unknown) value of $\sigma\phi(\bar{s})$. Therefore it may be helpful to write $\beta_0 = \sigma\phi(\bar{s})/K$ for some constant K. Then we have $n_i > \frac{1}{2}\{u_\alpha k(a_1)K\}^2$.

Remark 3.4.1 It may seem more realistic to consider the estimated version $\hat{t}_i = s - \hat{a}_i\hat{\sigma}_U$ in relation to $s - \hat{a}_2\hat{\sigma}_U$. In a similar way as the proof of theorem 3.4.1 it is proved that

$$\begin{aligned} &\mathrm{E}\left\{ \Phi\left(\frac{s - \hat{a}_2\hat{\sigma}_U - \mu_X}{\sqrt{\sigma_X^2 + \sigma_U^2}} \right) - \Phi\left(\frac{s - \hat{a}_i\hat{\sigma}_U - \mu_X}{\sqrt{\sigma_X^2 + \sigma_U^2}} \right) \right\} \\ &= \frac{c_i\sigma}{\sqrt{1+\sigma^2}} \phi\left(\frac{\bar{s} - a_2\sigma}{\sqrt{1+\sigma^2}} \right) + O\left(\sigma^2 n^{-1} + n^{-2} + \mathrm{e}^{-Cn} \right), \end{aligned}$$

for some $C > 0$ as $\sigma \to 0$, $n \to \infty$ and assuming $m > n$. □

In case of unbiased estimation there is a loss of yield as well, caused by using $a_2 + c_u$ instead of a_2. The losses are much smaller however, since c_u is of order n^{-1}. In case of unbiased estimation another point of view may apply. Sometimes it is sufficient if the consumer loss on the average equals γ, but large deviations from γ should not occur very often. We will consider the number of observations required for estimation to ensure the consumer loss exceeds $\gamma(1+\delta_0)$ with probability α only. In the previous subsections we have seen that

moderate sample sizes m and n allow second order unbiasedness, but much larger sample sizes are needed if not only the mean, but also the variation has to be controlled. As a consequence, if the sample size is such that the variation is under control, the bias correction c_u actually has lost its importance.

Because $(CL(\hat{t}_u) - \gamma)/\gamma$ is asymptotically normal, cf. (3.4.26), it follows that $\sigma_{CL} u_\alpha$ (where $u_\alpha = \Phi^{-1}(1-\alpha)$) should be smaller than δ_0, hence that n and m should satisfy

$$\frac{l(a_1)^2}{n} + \frac{\bar{s}^4 + 1}{m} \leq \frac{2\delta_0^2}{u_\alpha^2}. \tag{3.4.34}$$

To get an impression of the sample sizes involved here, let us specialize again to the case where $m \gg n$. Then the second term on the left-hand side of (3.4.34) can be neglected and it follows that

$$n \geq n_u = \frac{1}{2}\left(\frac{u_\alpha l(a_1)}{\delta_0}\right)^2. \tag{3.4.35}$$

For $\alpha = 0.05$ or 0.10 and $\delta_0 = 0.1$ or 0.2, the factor $u_\alpha^2/(2\delta_0^2)$ varies from 20 to 140. Together with values of $l(a_1)$ of about 6, it is clear that n_u is very large. This is perfectly in correspondence with the observation at the beginning of section 3.4.2.

Table 3.4.2 on the next page gives an impression of the accuracy of the first order asymptotics from this and the previous section. The approximation β (cf. (3.4.32)), is compared to the true loss of yield, and the relative error $(\gamma - CL(t_i))/\gamma$ is compared to the approximation δ from (3.4.30).

We conclude that for the situations considered, the approximation β is very accurate. The approximation δ is reasonably precise.

3.5 An application in semiconductor industry

As an example of the theory of sections 3.2 and 3.4 we consider a stereo decoder TDA1543/N2 manufactured at Philips' consumer IC plant at Nijmegen. There are 85 characteristics measured, from which we choose one that has to fall above a specification $s = 1.935$. Current practice is to set the test limit t at $3\hat{\sigma}_U$ from the specification limit ($\hat{\sigma}_U^2$ is the estimated measurement variance). Available are a bulk of production data and some test data to estimate the measurement error. The production sample size is so large (m=3099) that we may identify the resulting estimates for $\mu_{\tilde{X}} = \mu_X$ and $\sigma_{\tilde{X}}$ with these parameters (cf. the discussion at the end of section 3.4.1).

Table 3.4.2 Accuracy of the first order asymptotics

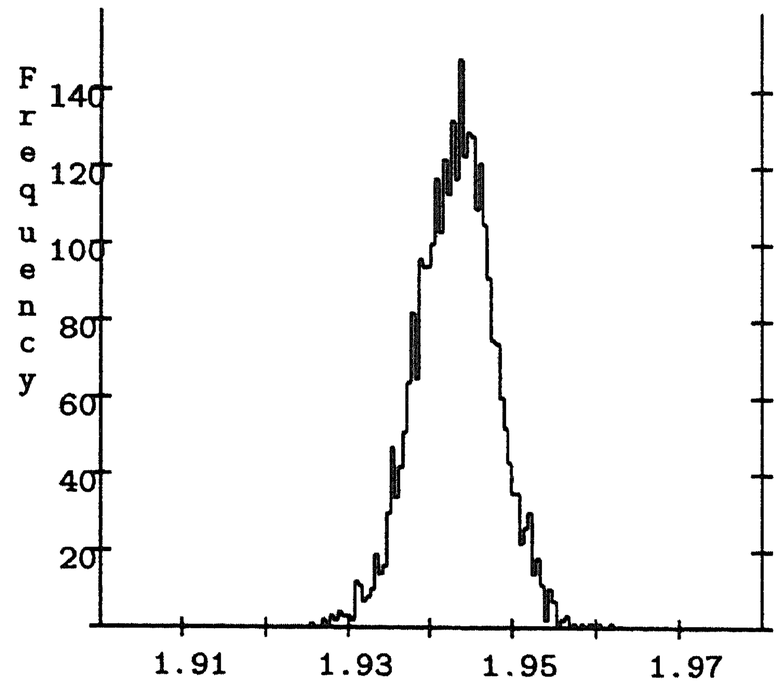
The approximation β from (3.4.32) is compared to the true loss of yield. The approximation δ from (3.4.30) is compared to the true relative error in the consumer loss.
For several choices of (σ, π, γ) a_2 from (3.2.20) is computed, in addition a_i from (3.4.28) is computed for $\alpha = 0.1$ and $n = 400, 1600$ and $m \to \infty$. In view of remark 3.4.1 the true value of σ is used to compute a_i and a_2.
$yield(a_2)$ denotes the yield obtained with $t = s - a_2\sigma$. *loss* denotes $yield(a_2) - yield(a_i)$.

		$n = 400$				$n = 1600$			
σ	$yield(a_2)$	β	*loss*	$\frac{\gamma - CL_i}{\gamma}$	δ	β	*loss*	$\frac{\gamma - CL_i}{\gamma}$	δ
				($\pi = 0.15, \gamma$= 20 ppm)					
0.01	0.845	0.0%	0.0%	0.253	0.287	0.0%	0.0%	0.135	0.144
0.10	0.776	0.4%	0.4%	0.374	0.462	0.2%	0.2%	0.207	0.231
0.20	0.669	1.0%	1.1%	0.407	0.516	0.5%	0.5%	0.229	0.258
				($\pi = 0.10, \gamma$= 40 ppm)					
0.01	0.897	0.0%	0.0%	0.198	0.218	0.0%	0.0%	0.104	0.109
0.10	0.849	0.3%	0.3%	0.324	0.386	0.1%	0.1%	0.176	0.193
0.20	0.769	0.8%	0.8%	0.358	0.440	0.4%	0.4%	0.198	0.220
				($\pi = 0.01, \gamma$=100 ppm)					
0.01	0.990	0.0%	0.0%	0.047	0.048	0.0%	0.0%	0.024	0.024
0.10	0.985	0.0%	0.0%	0.165	0.183	0.0%	0.0%	0.086	0.092
0.20	0.975	0.1%	0.1%	0.200	0.231	0.1%	0.1%	0.104	0.116

We obtain $\mu_X = 1.942531$ and $\sigma_{\tilde{X}} = 0.004857$. To estimate σ_U, $n = 40$ products have been measured twice, leading to an estimate $\hat{\sigma}_U = 0.0001043$. Because $\sigma_X^2 = \sigma_{\tilde{X}}^2 - \sigma_U^2$, it follows that σ_X is estimated by 0.004856. Since $\sigma_{\tilde{X}}$ is assumed to be known and σ_U is so much smaller than $\sigma_{\tilde{X}}$, we see that we can reasonably assume $\sigma_X = 0.004856$ (Of course, we drag along more digits here than strictly justified, but this enables us to show where $\sigma_{\tilde{X}}$ and σ_X start to differ).

Before applying the results we check whether the normality assumptions for X and U are sufficiently reasonable in our example. The histogram for the production data, figure 3.5.1, suggests that using our model for X is quite reasonable. Moreover, the kurtosis and skewness of the production data turn out to be 0.20 and -0.09, respectively. For U only 40 observations are available and we perform the Shapiro-Wilk test. This leads to a p-value of 0.4. Hence we see that once more the assumption of normality is reasonable.

Figure 3.5.1 Histogram for production data



Now, to apply the results we standardize to $\hat{\sigma} = \hat{\sigma}_U/\sigma_X = 0.021$ and $\bar{s} = -(s - \mu_X)/\sigma_X) = 1.551$ (cf. (3.2.30)), leading to $\pi = 0.0605$. For $\gamma = 100$ ppm we find $\hat{a}_2 = 1.367$ (cf. (3.2.20)), whereas taking the effect of estimating σ into account and thus using the correction term c_u (cf. (3.4.12)), leads to $\hat{a}_u = 1.415$. For the conservative approximation from (1.1.8) we have $\Phi^{-1}(1-\gamma/\pi) = 2.938$. For $\gamma = 20$ ppm we obtain $\hat{a}_2 = 2.029$, $\hat{a}_u = 2.128$ and $\Phi^{-1}(1-\gamma/\pi) = 3.405$. Assuming by way of illustration that the true σ equals the observed $\hat{\sigma}$, then the test limit currently used, $\hat{t} = s + 3\hat{\sigma}_U$, leads to $CL(\hat{t}) = 1.0$ ppm and *Yield*= 0.9314. We obtain in case of $\gamma = 100$ ppm *Yield*=0.9359 using both a_2 and a_u, and *Yield*=0.9316 and *CL*=1.2 ppm using $\Phi^{-1}(1-\gamma/\pi)$. For $\gamma = 20$ ppm the corresponding results for the yield are 0.9341, 0.9338 and 0.9302 using a_2, a_u and $\Phi^{-1}(1-\gamma/\pi)$, respectively. Hence in this example a substantial increase of yield of about 0.5% can be obtained.

Finally, using the more conservative approach (3.4.25) leads to $\hat{a}_i = 1.629$ for $\gamma = 100$ ppm and $\hat{a}_i = 2.373$ for $\gamma = 20$ ppm, both for $\alpha = 0.10$. Note that this stays well below the choice of the current practice, according to which $a = 3$, regardless of σ, γ and s.

3.6 Sampling schemes

For the selection of a criterion for a test limit, in section 3.4 we considered one single new item to be tested on the basis of our estimated test limit (cf. the introduction of section 3.4). In practice we deal of course with a whole series N, with typically $N \gg n$, of new items to be judged through a single test limit which is based on n (and m) earlier observations. Intuitively it is clear that n should increase in N: the longer one is going to use a certain test limit, the better it should be.

To make this intuitive feeling more precise we argue as follows. As always, we use a test limit $\hat{t}$ for which $\mathrm{E}CL(\hat{t}) = \gamma$ (at least) to first order. In this section we write $\widehat{CL}_n = CL(\hat{t})$. Let G_N be the average realized consumer loss over the N new items, then to first order $\mathrm{E}G_N = \gamma$ and

$$\begin{aligned} \mathrm{VAR}G_N &= (1 - N^{-1})\mathrm{VAR}(\widehat{CL}_n) + N^{-1}\gamma(1-\gamma) \\ &= \mathrm{VAR}(\widehat{CL}_n) + N^{-1}\gamma(1-\gamma), \end{aligned} \tag{3.6.1}$$

once more to first order, using that $\mathrm{VAR}(\widehat{CL}_n) = O(\gamma^2)$. This suggests that $\mathrm{VAR}(\widehat{CL}_n)$ should be of the order of magnitude $N^{-1}\gamma(1-\gamma)$. Taking both terms in (3.6.1) equal and using lemma 3.4.1, we obtain

$$\frac{l^2(a_1)}{n} + \frac{\bar{s}^4 + 1}{m} = \frac{2(1-\gamma)}{\gamma N} \approx \frac{2}{\gamma N}. \tag{3.6.2}$$

Specializing as before to the case where $m \gg n$ for the sake of illustration, (3.6.2) reduces to $n \approx n_0 = \frac{1}{2}l^2(a_1)\gamma N$. Having again values of $l(a_1)$ of about 6, letting e.g. $\gamma \in (10^{-5}, 10^{-4})$ and $N \sim 10^6$ (which is implied by speaking about parts per million), we see again large sample sizes n arising. Combination with (3.4.35) leads to

$$n \geq \frac{1}{2}l^2(a_1)\max\{u_\alpha^2/\delta_0^2, \gamma N\}. \tag{3.6.3}$$

We conclude this section with the following consideration, inspired by the fact that large sizes n (and m) are often required. Without loss of generality assume that n, m and N are multiples of a given positive integer k. Suppose that the whole process described before is now divided into k subgroups each leading to a $G_{N/k}^{(j)}$, $j = 1, \ldots, k$. From (3.6.1) it follows that $\mathrm{VAR}G_{N/k}^{(j)} = \mathrm{VAR}\widehat{CL}_{n/k} + kN^{-1}\gamma(1-\gamma)$. In view of (3.4.22) it is clear that $\mathrm{VAR}\widehat{CL}_{n/k} = k\mathrm{VAR}\widehat{CL}_n$ to first order. Hence $\tilde{G}_N = k^{-1}\sum_{j-1}^{k} G_{N/k}^{(j)}$ has

$$\mathrm{VAR}\tilde{G}_N = \mathrm{VAR}\widehat{CL}_n + N^{-1}\gamma(1-\gamma). \tag{3.6.4}$$

Comparison of 3.6.4) with (3.6.1) shows that $\mathrm{VAR}\tilde{G}_N \approx \mathrm{VAR}G_N$. Hence there is no loss in working with a number of smaller samples, each leading to its own estimates as compared to using one single, very large sample, Moreover, from the perspective of robustness, it is even quite attractive to work with a number of separate steps, as this will provide better protection against deviations from the assumption that the production process is stationary.

Chapter 4

Normality assumption on the measurement error only

The assumption of normality for both the characteristic and the measurement error often is not met in practice. In this chapter we will assume the measurement error only to be normally distributed. We have two reasons not to involve nonnormality of the measurement error here as well. First of all, in many situations where the characteristic is not normally distributed, the measurement error may. Secondly, in chapter 3 we have seen that the effects of estimating σ_U and of estimating μ_X and σ_X can be treated separately. In fact, we will see that the results derived in this chapter can be adapted directly to the more general case in which the distribution of the measurement error is unknown as well.

Denoting the procedures derived under the assumption of a normally distributed characteristic and measurement error as normal test limits, the first question is whether the normal test limits are still (approximately) correct. The extensive literature on robustness in statistics shows that typically classical procedures, which are optimal under normality, are quite sensitive to even small deviations from the normality assumption, especially in cases of heavy-tailedness or outliers.

Both theoretically and by simulations it is investigated here when test limits are sensitive for deviations of the normality assumption for the characteristic. As it turns out that the actual consumer loss may be much larger than the prescribed bound γ, there is need for robust test limits, indeed. We will derive two robust test limits. One limit for which the expected consumer loss is equal to γ and a limit for which γ is violated only with a small, given, probability, both to sufficient precision.

4.1 The normal test limit under nonnormal distributions

With respect to the inspected characteristic X we simply assume that it has density f_X and that $f_X(s) > 0$, with s denoting the specification limit. We measure $\tilde{X} = X + U$ where the measurement error U is $N(0, \sigma^2)$-distributed and independent of X. We assume that the standard deviation σ of the measurement error is much smaller than the standard deviation of X. (In this chapter we omit the subscript $_U$ of σ_U.)

By

$$\pi = P(X > s) \tag{4.1.1}$$

we denote the probability that a product is nonconforming. By t_e we again denote the test limit for which the consumer loss (CL) equals γ exactly (cf. (1.1.4)).

In this section we study the behavior of the normal test limit when the shape of the distribution of X is not normal. We consider standardized densities in the sense that

$$\mathrm{E}X = 0,\ \mathrm{VAR}X = 1. \tag{4.1.2}$$

Let t_e^N denote the solution of (1.1.4) if $f_X = \phi$ (the standard normal density), corresponding to the situation where the characteristic is normally distributed. We investigate

$$CL(t_e^N) = P(X > s,\ \tilde{X} < t_e^N), \tag{4.1.3}$$

for various f_X, thus studying how the normal test limit behaves under nonnormal distributions. In this way we avoid a mix up of errors due to the estimation process and the nonnormality. Moreover, while estimation errors can be eliminated by taking a large amount of observations, the nonnormality error remains.

We assume that γ is of order $\sigma\phi(s)$ as $\sigma \to 0$. The distribution function of X is denoted by F_X and the standard normal distribution function and its density by Φ and ϕ, respectively. The following theorem shows the behavior of the CL as $\sigma \to 0$.

Theorem 4.1.1 *If F_X is differentiable at s with $F_X'(s) = f_X(s)$, then*

$$CL(t_e^N) = \gamma \frac{f_X(s)}{\phi(s)}(1 + o(1)), \tag{4.1.4}$$

as $\sigma \to 0$.

Proof. Let $a_e^N = (s - t_e^N)/\sigma$. From the proof of lemma 3.2.1 it follows that (cf. (3.2.12))

$$\tfrac{1}{2}\sigma^2\phi'(\max\{s, 1\})g_2(a_e^N) \le \gamma - \sigma\phi(s)g_1(a_e^N) \le 0.$$

As $g_2 \leq \frac{3}{2}g_1$ it follows that

$$a_e^N = g_1^{-1}\left(\frac{\gamma}{\sigma\phi(s)}\right) + O(\sigma), \quad \text{as } \sigma \to 0, \tag{4.1.5}$$

and a_e^N is bounded. Analogous to (3.2.4) we now write

$$CL(t_e^N) = \int_{a_e^N}^{\infty} \{F_X(s + \sigma(y - a_e^N)) - F_X(s)\}\phi(y)\mathrm{d}y. \tag{4.1.6}$$

Since

$$0 \leq F_X(s + \sigma(y - a_e^N)) - F_X(s) \leq 1$$

and

$$1 - \Phi(|\log\ \sigma|) = o\left(\mathrm{e}^{-\frac{1}{2}|\log\ \sigma|^2}\right) = o(\sigma^k)$$

for any k, we restrict attention in (4.1.6) to $y < |\log\ \sigma|$. Because a_e^N is bounded, $\sigma|\log\ \sigma| \to 0$ as $\sigma \to 0$ and F_X is differentiable in s, it follows that

$$\begin{aligned} CL(t_e^N) &= \int_{a_e^N}^{|\log\ \sigma|} f_X(s)\sigma(y - a_e^N)\phi(y)\mathrm{d}y(1 + o(1)) + o(\sigma^k) \\ &= \sigma f_X(s) g_1(a_e^N)(1 + o(1)) \\ &= \gamma \frac{f_X(s)}{\phi(s)}(1 + o(1)), \text{ as } \sigma \to 0. \end{aligned}$$

This completes the proof of the theorem. □

It is clear from this theorem that the behavior of $f_X(s)/\phi(s)$ determines the performance of the normal test limit. Note that this differs from many classical results on questions of robustness, where bad behavior under nonnormality of methods based on normality assumptions is caused by outliers.

Of course, the ratio $f(s)/\phi(s)$ for a density f can be anything if there are no restrictions on this density. Therefore we investigate

$$\sup_{f \in \mathcal{F}} \frac{f(s)}{\phi(s)} = \frac{1}{\phi(s)} \sup_{f \in \mathcal{F}} f(s)$$

for some interesting classes $\mathcal{F}$ of densities f. In view of (4.1.2) all densities are standardized in the sense that

$$\int_{-\infty}^{\infty} x f(x)\mathrm{d}x = 0 \text{ and } \int_{-\infty}^{\infty} x^2 f(x)\mathrm{d}x = 1.$$

The first class $\mathcal{F}_1$ under consideration is the class of symmetric, unimodal densities.

Lemma 4.1.1 *For all* $s > 0$

$$\sup_{f \in \mathcal{F}_1} f(s) = \frac{1}{2} \min \left\{ \frac{1}{s}, \frac{3}{s^3} \right\}. \tag{4.1.7}$$

Proof. If $f \in \mathcal{F}_1$ and $s > 0$, then

$$f(x) \geq f(s) \text{ for all } |x| \leq s$$

and hence

$$1 \geq \int_{-s}^{s} f(x)\mathrm{d}x \geq 2sf(s). \tag{4.1.8}$$

Moreover,

$$1 = \int_{-\infty}^{\infty} x^2 f(x)\mathrm{d}x \geq \int_{-s}^{s} x^2 f(x)\mathrm{d}x \geq \frac{2}{3}s^3 f(s),$$

which in combination with (4.1.8) leads to

$$\sup_{f \in \mathcal{F}_1} f(s) \leq \frac{1}{2} \min \left\{ \frac{1}{s}, \frac{3}{s^3} \right\}. \tag{4.1.9}$$

If $s \leq \sqrt{3}$, take

$$f(x) = \begin{cases} \dfrac{1-\epsilon}{2s} & 0 \leq |x| \leq s \\ \dfrac{\epsilon}{2\eta} & s < |x| < s + \eta \end{cases}$$

with η such that $\int x^2 f(x)\mathrm{d}x = 1$. This shows that we can get arbitrary close to

$$\frac{1}{2s} = \frac{1}{2} \min \left\{ \frac{1}{s}, \frac{3}{s^3} \right\} \text{ if } s < \sqrt{3}.$$

If $s \geq \sqrt{3}$, take

$$f(x) = \begin{cases} \dfrac{3-\epsilon}{2s^3} + \dfrac{1}{2\eta}\left(1 - \dfrac{3-\epsilon}{s^2}\right) & 0 \leq |x| \leq \eta \\ \dfrac{3-\epsilon}{2s^3} & \eta < |x| < s \end{cases}$$

with η such that $\int x^2 f(x)\mathrm{d}x = 1$. This shows that we can get arbitrary close to

$$\frac{3}{2s^3} = \frac{1}{2} \min \left\{ \frac{1}{s}, \frac{3}{s^3} \right\} \text{ if } s \geq \sqrt{3}.$$

Therefore, inequality (4.1.9) is sharp and the result is proved. □

The upper bound (4.1.7) leads to the following upper bound on $f(s)/\phi(s)$ (cf. (4.1.4)).

s	1.5	2	2.5	3
$\sup_{f\in\mathcal{F}_1} \frac{f(s)}{\phi(s)}$	2.57	3.47	5.48	12.54

Next we restrict $\mathcal{F}_1$ furthermore by the extra requirement $f(x)/\phi(x) \geq f(s)/\phi(s)$ for all $|x| \geq s$. We call this class $\mathcal{F}_2$.

Let

$$z(x) = \frac{1-\Phi(x)}{\phi(x)}, \tag{4.1.10}$$

then we have that

Lemma 4.1.2 *For all* $s > 0$

$$\sup_{f\in\mathcal{F}_2} f(s) = \frac{1}{2\left(\frac{1}{3}s^3 + s + z(s)\right)}. \tag{4.1.11}$$

Proof. If $f \in \mathcal{F}_2$ and $s > 0$ then

$$f(x) \geq f(s) \text{ for } |x| \leq s$$

and

$$f(x) \geq \frac{f(s)}{\phi(s)}\phi(x) \text{ for } |x| > s.$$

Therefore

$$\begin{aligned} \tfrac{1}{2} &= \int_0^\infty x^2 f(x)\mathrm{d}x \\ &\geq \int_0^s x^2 f(s)\mathrm{d}x + \int_s^\infty \frac{f(s)}{\phi(s)} x^2 \phi(x)\mathrm{d}x \\ &= f(s)\left(\tfrac{1}{3}s^3 + s + z(s)\right), \end{aligned}$$

implying

$$\sup_{f\in\mathcal{F}_2} f(s) \leq \frac{1}{2\left(\frac{1}{3}s^3 + s + z(s)\right)} = u(s), \text{ say.} \tag{4.1.12}$$

Taking

$$f(x) = \begin{cases} u(s)(1-\epsilon) + a(\epsilon,\eta) & |x| \leq \eta \\ u(s)(1-\epsilon) & \eta < |x| \leq s \\ \dfrac{u(s)}{\phi(s)}(1-\epsilon)\phi(x) & |x| > s \end{cases}$$

with $a(\epsilon, \eta)$ such that $\int f(x)\mathrm{d}x = 1$ and η such that $\int x^2 f(x)\mathrm{d}x = 1$, it is seen that we can get arbitrarily close to the upper bound in (4.1.12), thus completing the proof. □

The upper bound (4.1.11) leads to the following upper bound on $f(s)/\phi(s)$ (cf. (4.1.4)).

s	1.5	2	2.5	3
$\sup_{f \in \mathcal{F}_2} \frac{f(s)}{\phi(s)}$	1.23	1.82	3.54	9.18

Dropping the symmetry, unimodality does no longer help. Arbitrarily high densities at s are possible by taking a narrow spike at s and some mass (far) on the negative axis to get $\int x f(x)\mathrm{d}x = 0$ and $\int x^2 f(x)\mathrm{d}x = 1$. Therefore consider the class $\mathcal{F}_3$ of all densities with $f(x)/\phi(x) \geq f(s)/\phi(s)$ for all $x \geq s$. (Hence $\mathcal{F}_2 = \mathcal{F}_1 \cap \mathcal{F}_3$.)

Lemma 4.1.3 *For all* $s > 0$

$$\sup_{f \in \mathcal{F}_3} f(s) = \frac{1}{\frac{1}{2}(s^2+4)^{1/2} + \frac{1}{2}s + z(s)}. \tag{4.1.13}$$

Proof. If $f \in \mathcal{F}_3$ and $s > 0$, then

$$\int_s^\infty f(x)\mathrm{d}x \geq f(s)z(s) > 0,$$

$$\int_s^\infty x f(x)\mathrm{d}x \geq \frac{f(s)}{\phi(s)} \int_s^\infty x\phi(x)\mathrm{d}x = f(s) > 0$$

and

$$\int_s^\infty x^2 f(x)\mathrm{d}x \geq \frac{f(s)}{\phi(s)} \int_s^\infty x^2\phi(x)\mathrm{d}x = f(s)\{s + z(s)\}. \tag{4.1.14}$$

Hence

$$0 < \int_{-\infty}^s f(x)\mathrm{d}x \leq 1 - f(s)z(s) \tag{4.1.15}$$

and

$$\int_{-\infty}^s x f(x)\mathrm{d}x \leq -f(s),$$

which leads to

$$\begin{aligned}\int_{-\infty}^{s} x^2 f(x)\mathrm{d}x &\geq \left\{\int_{-\infty}^{s} x f(x)\mathrm{d}x\right\}^2 \left\{\int_{-\infty}^{s} f(x)\mathrm{d}x\right\}^{-1} \\ &\geq \frac{f^2(s)}{1-f(s)z(s)}\end{aligned}$$

and

$$\int_{s}^{\infty} x^2 f(x)\mathrm{d}x = 1 - \int_{-\infty}^{s} x^2 f(x)\mathrm{d}x \leq 1 - \frac{f(s)^2}{1-f(s)z(s)}.$$

In combination with (4.1.14) we obtain

$$f(s)\{s+z(s)\} \leq 1 - \frac{f(s)^2}{1-f(s)z(s)}$$

and,

$$f^2(s)\{1-z(s)(s+z(s))\} + f(s)\{s+2z(s)\} - 1 \leq 0. \tag{4.1.16}$$

From Kotz and Johnson (1985, p. 505), we obtain

$$z(s) > \frac{2}{s+\sqrt{s^2+4}},$$

from which it follows that $1 - z(s)\{s+z(s)\} < 0$.

Since for the largest root of (4.1.16) $f(s) > z(s)^{-1}$, which is in conflict with (4.1.15), we obtain

$$\sup_{f\in\mathcal{F}_3} f(s) \leq \frac{\sqrt{s^2+4} - \{s+2z(s)\}}{2(1-z(s)\{s+z(s)\})} = \left\{\tfrac{1}{2}(s^2+4)^{1/2} + \tfrac{1}{2}s + z(s)\right\}^{-1}.$$

The sharpness of this bound is shown in a similar way as in lemma 4.1.1 and lemma 4.1.2. We omit the details. □

The upper bound (4.1.13) leads to the following upper bound on $f(s)/\phi(s)$ (cf. (4.1.4)).

s	1.5	2	2.5	3
$\sup_{f\in\mathcal{F}_3} \frac{f(s)}{\phi(s)}$	3.07	6.53	17.80	62.74

To show both the influence of nonnormality and the estimation process, simulation results are presented for some familiar distributions. The normal test limit

$$\hat{t}^N = s - \hat{a}_u(\hat{\mu}_X, \hat{\sigma}_X, \sigma)\cdot\sigma \tag{4.1.17}$$

is computed, with a_u from (3.4.12). Since we are interested in the influence of nonnormality of X on the test limit, estimation of σ is avoided and thus σ is assumed to be known in the simulations. Consequently, the term $(k(a_1)\{2a_1k(a_1)+1-a_1^2\}/(4n)$ in (3.4.12) to correct for estimating σ is omitted, according to '$n \to \infty$'. The estimators $\hat{\mu}_X$ and $\hat{\sigma}_X$ (cf. (3.4.11), however with $\hat{\sigma}_U$ replaced by σ) are based on a sample $X_1+U_1,\ldots,X_m+U_m$ where $X_1,\ldots,X_m,U_1,\ldots,U_m$ are independent, $U_i \sim N(0,\sigma^2)$, but X_i does not necessarily have a normal distribution.

The following distributions of X are considered.

distribution	*density*
beta	$\frac{1}{B(p,q)}\frac{(x-\mu_1)^{p-1}(\mu_2-x)^{q-1}}{(\mu_2-\mu_1)^{p+q-1}}$ $(\mu_1 \le x \le \mu_2)$
normal mixture	$\frac{1}{2}\left\{\frac{1}{\sigma_1}\phi\left(\frac{s-\mu_1}{\sigma_1}\right)+\frac{1}{\sigma_2}\phi\left(\frac{s-\mu_2}{\sigma_2}\right)\right\}$
gamma	$\frac{1}{\nu\Gamma(\alpha)}\left(\frac{x-\mu}{\nu}\right)^{\alpha-1}\exp\left(-\frac{x-\mu}{\nu}\right)$ $(x>\mu)$

where B denotes the beta-function and Γ the gamma-function. In the simulation the parameters are chosen in such a way that $\mathrm{E}X=0$ and $\mathrm{VAR}X=1$. The results are summarized in table 4.1.1 on the next page.

In view of the results in table 4.1.1, additional simulation results not presented here and the preceding theorem we conclude that

- for symmetric and unimodal densities the *CL* based on the normal test limits may differ from γ, but not strongly (note that beta $p=2$, $q=2$, is symmetric and unimodal, while gamma $\alpha=32$ is unimodal and almost symmetric)
- in general, substantial deviations up to a factor 8 are not uncommon, especially if π is small (s is large)
- there is need for robust test limits.

Remark 4.1.1 If the distribution of X is heavy-tailed and the specification limit is not large, the *CL* based on the normal test limit will in general not violate γ, but even may be conservative. This is understood by the following argument. If $f_X(s)/\phi(s)$ is large for large values of s, for instance by heavy-tailedness, usually this is compensated by values $f_X(s)/\phi(s)$ smaller than 1 for

moderate s, resulting in a conservative CL for those values of s. Such an effect is clearly seen in table 4.1.1. □

Table 4.1.1. Simulated mean (standard deviation) of $CL(\hat{t}^N)$

The test limit $\hat{t}^N$ as in (4.1.17) is computed, but $\hat{\mu}_X$ and $\hat{\sigma}_X$ are based on samples, of size m, from various distributions. $\gamma = 100$ ppm, $\sigma = 0.10$. We write $Q(s) = f_X(s)/\phi(s)$. The simulation is carried out with 10000 replications.

		$\pi = 0.10$		$\pi = 0.01$	
distr.	m		$CL(\hat{t}^N)$		$CL(\hat{t}^N)$
$N(0,1)$	100	$s = 1.28$	100.8(15.7)	$s = 2.33$	104.7(53.7)
	400	$f_X(s) = \phi(s)$	100.1(7.3)	$f_X(s) = \phi(s)$	101.1(21.4)
	1600		100.0(3.5)		100.1(10.2)
beta	100	$s = 1.36$	135.6(21.3)	$s = 1.97$	118.6(32.6)
$p = 2$	400	$Q(s) = 1.336$	135.7(10.3)	$Q(s) = 1.305$	120.9(15.9)
$q = 2$	1600		135.6(5.1)		121.4(7.8)
beta	100	$s = 1.33$	92.3(17.7)	$s = 2.63$	187.5(181.9)
$p = 8$	400	$Q(s) = 0.900$	90.8(7.8)	$Q(s) = 1.664$	175.6(55.2)
$q = 32$	1600		90.6(3.9)		172.7(25.9)
beta	100	$s = 1.40$	89.5(21.5)	$s = 2.85$	374.7(540.2)
$p = 2$	400	$Q(s) = 0.852$	87.0(9.0)	$Q(s) = 2.837$	318.4(141.7)
$q = 8$	1600		86.6(4.4)		304.1(61.3)
n.mixt	100	$s = 1.47$	96.4(19.3)	$s = 2.68$	187.1(144.5)
$\mu_1 = 0.53$	400	$Q(s) = 0.932$	94.9(8.7)	$Q(s) = 1.705$	178.4(54.0)
$\sigma_1 = 1.0$	1600		94.6(4.2)		174.4(24.7)
n.mixt	100	$s = 1.55$	116.8(43.3)	$s = 2.89$	542.5(930.8)
$\mu_1 = 0.63$	400	$Q(s) = 1.067$	111.3(15.7)	$Q(s) = 3.562$	406.0(213.2)
$\sigma_1 = 1.1$	1600		109.8(7.3)		380.4(85.5)
gamma	100	$s = 1.34$	74.3(21.6)	$s = 3.28$	1655.7(2532)
$\alpha = 2$	400	$Q(s) = 0.689$	70.5(8.1)	$Q(s) = 6.677$	996.6(928)
	1600		69.8(3.8)		804.8(299)
gamma	100	$s = 1.3$	83.9(18.1)	$s = 2.90$	387.6(607)
$\alpha = 6$	400	$Q(s) = 0.805$	81.7(7.9)	$Q(s) = 2.707$	311.7(148)
	1600		81.3(3.8)		294.8(64.0)
gamma	100	$s = 1.31$	92.1(16.1)	$s = 2.58$	160.9(119.3)
$\alpha = 32$	400	$Q(s) = 0.907$	91.4(7.5)	$Q(s) = 1.464$	153.5(45.8)
	1600		91.2(3.6)		151.3(21.6)

4.2 Second order unbiased test limits based on density estimation

4.2.1 Deconvolution

It is clearly seen in the previous section that nonnormality cannot be ignored in determination of test limits. At the same time the proof of theorem 4.1.1 indicates how to proceed if X has density f_X. In (4.1.5) ϕ should be replaced by f_X. In general the density f_X is unknown and therefore it has to be estimated by available observations. Estimation of the density of X, however, is complicated by the fact that we observe $\tilde{X} = X + U$ instead of X. In chapter 3 this problem did not occur because the density f concerned (being ϕ) is determined by parameters (mean and variance of X) which can be estimated consistently. Estimation of f_X, not assuming a parametric model for the distribution of X, leads to the so called nonparametric deconvolution problem. We are interested in the density of X, having observations from $X + U$. It is well-known that nonparametric deconvolution with normal error is very difficult, i.e. giving extremely slow optimal rates of convergence. This is shown e.g. by Zhang (1990) and Fan (1991a) (cf. also Fan (1991b)), where many further references on the deconvolution problem are found.

Therefore we do not apply deconvolution estimators here, but take into account the extra information that σ is small, implying that the density $f_{\tilde{X}}$ of $X + U$ is close to the density f_X of X. In fact,

$$f_{\tilde{X}}(x) - f_X(x) = \int_{-\infty}^{\infty} f_X(x - \sigma y)\phi(y)\mathrm{d}y - f_X(x) = \frac{1}{2}\sigma^2 f_X''(x) + \ldots$$

and thus the difference will typically be of order σ^2. In that case the deconvolution problem is more tractable (cf. Fan (1992)).

In this section we derive a test limit for which the resulting consumer loss in expectation is equal to the prescribed bound γ, to sufficient precision. It turns out that $f'_{\tilde{X}}$ has to be estimated as well. In chapter 3 estimation of the derivative itself did not come up because it is determined by μ_X and σ_X.

Now, let $\hat{t} = t(\hat{\sigma}, \hat{f}_{\tilde{X}}, \hat{f}'_{\tilde{X}})$ be a test limit based on estimators of σ, $f_{\tilde{X}}$ and $f'_{\tilde{X}}$ and let $\hat{a} = a(\hat{\sigma}, \hat{f}_{\tilde{X}}, \hat{f}'_{\tilde{X}}) = (s - \hat{t})/\hat{\sigma}$. By $CL(\hat{t})$ we denote again the probability of a new item to be nonconforming and accepted, that is $P(X > s,\ \tilde{X} < \hat{t})$, given $\hat{t} = s - \hat{a} \cdot \hat{\sigma}$. We are looking for a test limit $\hat{t}_u = t_u(\hat{\sigma}, \hat{f}_{\tilde{X}}, \hat{f}'_{\tilde{X}})$ for which

$$\mathrm{E}CL(\hat{t}_u) = \gamma \tag{4.2.1}$$

to sufficient precision. More precisely, we are looking for second order unbiasedness.

4.2.2 Selecting estimators

For estimation of σ, $f_{\tilde{X}}$ and $f'_{\tilde{X}}$ we consider the following model. We have a set of n variables which are measured twice and a sample of size m from production, where m typically is much larger than n. In formula,

$$\begin{aligned} \tilde{X}_{ij} &= X_i + U_{ij}, \qquad i = 1, \ldots, n, \ j = 1, 2 \\ \tilde{X}_i &= X_i + U_i, \qquad i = n+1, \ldots, n+m, \end{aligned} \tag{4.2.2}$$

where

$$(\tilde{X}_{11}, \tilde{X}_{12}), \ldots, (\tilde{X}_{n1}, \tilde{X}_{n2})$$

are observations in duplicate (test data) and

$$\tilde{X}_{n+1}, \ldots, \tilde{X}_{n+m}$$

are single observations (production data).

We can estimate σ as before by

$$\hat{\sigma} = \sqrt{\frac{1}{2n} \sum_{i=1}^{n} (\tilde{X}_{i2} - \tilde{X}_{i1})^2}. \tag{4.2.3}$$

In the proof of theorem 3.4.1 we already concluded that

$$\mathrm{E}\left(\frac{\hat{\sigma}}{\sigma} - 1\right)^k = \nu_k + O(n^{-2}), \text{ as } n \to \infty, \ k = 1, 2, 3, 4, \tag{4.2.4}$$

where $\nu_1 = -\frac{1}{4}$, $\nu_2 = \frac{1}{2}$ and $\nu_3, \nu_4 = 0$.

To estimate $f_{\tilde{X}}$ (and $f'_{\tilde{X}}$) we apply the Rosenblatt-kernel estimator. It is generally recognized that the choice of the bandwidth is more important than the choice of the kernel (cf. Silverman (1986, p. 43), Härdle (1991, p. 78)). We therefore simply take the uniform kernel. The conventional bandwidth of order $m^{-1/5}$ turns out to be not the optimal one here. The choice of the bandwidth is treated in sections 4.2.5 and 4.2.6.

Apart from considering these kernel estimators, we have also considered the density estimator proposed by Abramson (1982), especially because of the reported good small-sample behavior of this estimator (cf. Terrell and Scott (1992) p. 1248 and references there). Since there were no important improvements, we do not propose this estimator here.

As it is seen in the definition of a_1 in (4.2.7), in fact we have to estimate $1/f_{\tilde{X}}(s)$. Therefore we have also considered quantile estimators, cf. e.g. Bloch and Gastwirth (1968), Hall and Sheather (1988), Jones (1992). Also these estimators, based on order statistics, give no substantial improvement.

Hence,

$$\hat{f}_{\tilde{X}}(s) = \frac{1}{2mh} \sum_{i=1}^{m} Z_i, \quad \text{and} \quad \hat{f}'_{\tilde{X}}(s) = \frac{1}{m\bar{h}^2} \sum_{i=1}^{m} \bar{Z}_i, \tag{4.2.5}$$

where

$$Z_i = \begin{cases} 1 & \text{if } \tilde{X}_{n+i} \in [s-h, s+h] \\ 0 & \text{otherwise} \end{cases}$$

and

$$\bar{Z}_i = \begin{cases} -1 & \text{if } \tilde{X}_{n+i} \in [s-\bar{h}, s] \\ 1 & \text{if } \tilde{X}_{n+i} \in (s, s+\bar{h}] \\ 0 & \text{otherwise} \end{cases}$$

for $i = 1, \ldots, m$. The estimator of $f_{\tilde{X}}(s)$ boils down to just counting the observations around the specification limit. The choice of the bandwidth h will be considered in section 4.2.5.

Direct calculation gives as $h \to 0$, $m \to \infty$,

$$\begin{aligned}
&\mathrm{E}\left(\frac{\hat{f}_{\tilde{X}}}{f_{\tilde{X}}} - 1\right) = O(h^2) \\
&\mathrm{E}\left(\frac{\hat{f}_{\tilde{X}}}{f_{\tilde{X}}} - 1\right)^2 = \frac{1}{2mhf_{\tilde{X}}(s)} - \frac{1}{m} + O(h^4 + h/m) \\
&\mathrm{E}\left(\frac{\hat{f}_{\tilde{X}}}{f_{\tilde{X}}} - 1\right)^3 = O((mh)^{-2} + h/m + h^6) \\
&\mathrm{E}\left(\frac{\hat{f}_{\tilde{X}}}{f_{\tilde{X}}} - 1\right)^4 = O((mh)^{-2} + (mh)^{-3} + h^8) \qquad (4.2.6) \\
&\mathrm{E}\left(\frac{\hat{f}'_{\tilde{X}}}{f'_{\tilde{X}}} - 1\right) = O(\bar{h}^2) \\
&\mathrm{E}\left(\frac{\hat{f}'_{\tilde{X}}}{f'_{\tilde{X}}} - 1\right)\left(\frac{\hat{f}_{\tilde{X}}}{f_{\tilde{X}}} - 1\right) = O(h/(m\bar{h}^2) + 1/m + (h\bar{h})^2).
\end{aligned}$$

4.2.3 Expansion of the consumer loss

In chapter 3 accurate approximations to and various bounds on the consumer loss have been derived for the case $X \sim N(\mu_X, \sigma_X^2)$. From (3.2.4) it is obvious that by writing F_X instead of Φ and imposing suitable regularity conditions on F_X, many of the results presented in section 3.2 can be generalized. One of the reasons not to do this is because we are actually interested in a test limit in terms of $f_{\tilde{X}}$ and $f'_{\tilde{X}}$ (and consequently in terms of $\hat{f}_{\tilde{X}}$ and $\hat{f}'_{\tilde{X}}$) instead of in f_X and f'_X.

Therefore, let

$$a_1 = g_1^{-1}\left(\frac{\gamma}{\sigma f_{\tilde{X}}(s)}\right), \qquad (4.2.7)$$

analogous to (3.2.9). In view of the results in section 3.4, let a correction term $\hat{d} = d(\hat{\sigma}, \hat{f}_{\tilde{X}}, \hat{f}'_{\tilde{X}})$ be defined by

$$\hat{t}_u = s - (\hat{a}_1 + \hat{d})\hat{\sigma} \tag{4.2.8}$$

where $\hat{a}_1 = a_1(\hat{\sigma}, \hat{f}_{\tilde{X}})$ and with $\hat{t}_u$ aimed to satisfy (4.2.1).

Further we define

$$\bar{a} = (s - \hat{t}_u)/\sigma = (\hat{a}_1 + \hat{d})\hat{\sigma}/\sigma. \tag{4.2.9}$$

The following lemma gives a nonasymptotic upper and lower bound of the relative error in $CL(\hat{t}_u)$ in terms of $\bar{a}$ when $\bar{a} > 0$.

Lemma 4.2.1 *Assume that* $f''_X(x) \geq 0$ *for all* $x \geq s$ *and assume that* $f_X(s) > \frac{3}{4}\sigma \max\{-f'_X(s), 0\}$, *then if* $\bar{a} > 0$

$$0 \leq \frac{CL(\hat{t}_u) - \sigma f_X(s) g_1(\bar{a}) - \frac{1}{2}\sigma^2 f'_X(s) g_2(\bar{a})}{\sigma f_X(s) g_1(\bar{a}) + \frac{1}{2}\sigma^2 f'_X(s) g_2(\bar{a})} \leq \frac{\frac{1}{3}\sigma^2 \max\{f''_X(s); x \geq s\}}{f_X(s) + \frac{3}{4}\sigma \min\{f'_X(s), 0\}}, \tag{4.2.10}$$

with g_1 *and* g_2 *as in (3.2.7).*

Proof. Writing $Y = -U/\sigma$ we have

$$\begin{aligned} & CL(\hat{t}_u) \\ &= P(X > s,\ X + U < \hat{t}_u) = P(Y > \bar{a},\ s < X < s + \sigma(Y - \bar{a})) \\ &= \int_{\bar{a}}^{\infty} \{F_X(s + \sigma(y - \bar{a})) - F_X(s)\}\phi(y)\mathrm{d}y \\ &= \sigma f_X(s) g_1(\bar{a}) + \tfrac{1}{2}\sigma^2 f'_X(s) g_2(\bar{a}) + \tfrac{1}{6}\sigma^3 \int_{\bar{a}}^{\infty} (y - \bar{a})^3 f''_X(\xi_y)\phi(y)\mathrm{d}y, \end{aligned} \tag{4.2.11}$$

for some ξ_y between s and $s + \sigma(y - \bar{a})$. Since $0 \leq f''_X(z) \leq \max\{f''_X(s); x \geq s\}$ for all $z \geq s$ we get

$$0 \leq \int_{\bar{a}}^{\infty} (y - \bar{a})^3 f''_X(\xi_y)\phi(y)\mathrm{d}y \leq \max\{f''_X(s); x \geq s\} g_3(\bar{a}). \tag{4.2.12}$$

Because $\bar{a} > 0$ we can use that $g_3(\bar{a}) \leq 2g_1(\bar{a})$ and $g_2(\bar{a}) \leq \frac{3}{2}g_1(\bar{a})$ (cf. the proof of lemma 3.2.1) and hence

$$\sigma f_X(s) g_1(\bar{a}) + \tfrac{1}{2}\sigma^2 f'_X(s) g_2(\bar{a}) \geq \sigma g_1(\bar{a})(f_X(s) + \tfrac{3}{4}\sigma \min\{f'_X(s), 0\}) > 0. \tag{4.2.13}$$

Combination of (4.2.11)-(4.2.13) completes the proof. □

The condition $f''_X(x) \geq 0$ for all $x \geq s$ is not always met in examples of interest (cf. lemma 3.2.3). It is easily seen in the proof of lemma 4.2.1 that

without this condition the lemma still holds if 0 in the left-hand side of (4.2.10) is replaced by

$$\frac{\frac{1}{3}\sigma^2 \min\{0, \min\{f''_X(s); x \geq s\}\}}{f_X(s) + \frac{3}{4}\sigma \min\{f'_X(s), 0\}}$$

In view of theorem 4.1.1 and the discussion in section 4.2.1 we put the following assumptions.

A1 $\quad |f''_X(x)| \leq M$ for all $x \geq s$,

A2 $\quad f_{\tilde{X}}(s) = f_X(s)\{1 + O(\sigma^2)\} > 0$ and
$f'_{\tilde{X}}(s) = f'_X(s)\{1 + O(\sigma)\}$, as $\sigma \to 0$,

A3 $\quad a_1$ is bounded as $\sigma, \gamma \to 0$.

We prove the following asymptotic result.

Lemma 4.2.2 *Assume* A1 *and* A2, *then*

$$CL(\hat{t}_u) = \{\sigma f_{\tilde{X}}(s) g_1(\bar{a}) - \tfrac{1}{2}\sigma^2 f'_{\tilde{X}}(s) g_2(\bar{a})\}(1 + O(\sigma^2)) \tag{4.2.14}$$

as $\sigma \to 0$, *uniformly in* $\bar{a} \geq 0$.

Proof. As (4.2.10) we have

$$|CL(\hat{t}_u) - \sigma f_X(s) g_1(\bar{a}) - \tfrac{1}{2}\sigma^2 f'_X(s) g_2(\bar{a})| \leq \tfrac{1}{6}\sigma^3 M g_3(\bar{a}) \leq \tfrac{1}{3}\sigma^3 M g_1(\bar{a})$$

and the proof is completed by application of A2 and the fact that $g_2 \leq \frac{3}{2} g_1$. □

4.2.4 The correction term

To determine the correction term $\hat{d}$ defined by (4.2.8) we need to expand $CL(\hat{t}_u)$ in terms of $\hat{\sigma}$, $\hat{f}_{\tilde{X}}$ and $\hat{f}'_{\tilde{X}}$. Therefore we define the set A by

$$A = \left\{ \left| \frac{\hat{\sigma}}{\sigma} - 1 \right| \leq c_1, \left| \frac{\hat{f}_{\tilde{X}}}{f_{\tilde{X}}} - 1 \right| \leq c_2, \left| \frac{\hat{f}'_{\tilde{X}}}{f'_{\tilde{X}}} - 1 \right| \leq c_3 \right\}, \tag{4.2.15}$$

where c_1, c_2 and c_3 are sufficiently small positive constants. Note that $O(P(A^c))$ still depends on the bandwidths h and $\bar{h}$ and hence the order $O(P(A^c))$ cannot yet be specified. On the set A we have (cf. (4.2.8) and (4.2.14) and cf. (3.4.16))

$$\begin{aligned} CL(\hat{t}_u) = \sigma f_{\tilde{X}}(s) &\left\{ g_1\left(\frac{\hat{\sigma}}{\sigma}\hat{a}_1\right) + \frac{\hat{\sigma}}{\sigma}\hat{d} g'_1\left(\frac{\hat{\sigma}}{\sigma}\hat{a}_1\right) + O(\hat{d}^2) \right\} \\ &+ \tfrac{1}{2}\sigma^2 f'_{\tilde{X}}(s) g_2\left(\frac{\hat{\sigma}}{\sigma}\hat{a}_1\right) + O(\sigma^2\hat{d}) + O(\gamma\sigma^2), \end{aligned} \tag{4.2.16}$$

Writing 1_A as the indicator function of the set A, we have, with $k(a)$ as in (3.2.15),

Lemma 4.2.3 *With $\hat{\sigma}$ and $\hat{f}_{\tilde{X}}$ from (4.2.3) and (4.2.5), respectively, we have as $m, n \to \infty$ and $h \to 0$*

$$\begin{aligned}
&\mathrm{E}\left\{g_1\left(\frac{\hat{\sigma}}{\sigma}\hat{a}_1\right) - g_1(a_1)\right\} 1_A \\
&= \frac{\phi(a_1)\{2a_1k(a_1) + 1 - a_1^2\}}{4n} + g_1(a_1)\left(\frac{1}{2mhf_{\tilde{X}}(s)} - \frac{1}{m}\right) \qquad (4.2.17) \\
&\quad + O(n^{-2} + h^2 + (mh)^{-2} + P(A^c)).
\end{aligned}$$

Proof. We expand $m(X,Y) = g_1(Xg_1^{-1}(b/(XY)))$ around $b = m(1,1)$, where

$$b = \frac{\gamma}{\sigma f_{\tilde{X}}(s)} = g_1(a_1),\ X = \frac{\hat{\sigma}}{\sigma},\ Y = \frac{\hat{f}_{\tilde{X}}}{f_{\tilde{X}}}.$$

In obvious notation,

$$\begin{aligned}
&m(x,y) - m(1,1) \\
&= m_{10}(x-1) + m_{01}(y-1) + \tfrac{1}{2}m_{20}(x-1)^2 + \tfrac{1}{2}m_{02}(y-1)^2 \\
&\quad + m_{11}(x-1)(y-1) + \tfrac{1}{6}m_{30}(x-1)^3 + \tfrac{1}{2}m_{21}(x-1)^2(y-1) \\
&\quad + \tfrac{1}{2}m_{12}(x-1)(y-1)^2 + \tfrac{1}{6}m_{03}(y-1)^3 \\
&\quad + O((x-1)^4 + (y-1)^4),
\end{aligned}$$

as $(x,y) \to (1,1)$. By independence of $\hat{\sigma}$ and $\hat{f}_{\tilde{X}}$, (4.2.6) and (4.2.4) we get

$$\begin{aligned}
&\mathrm{E}\{m(X,Y) - m(1,1)\} \cdot 1_A \\
&= m_{10}\nu_1 n^{-1} + \tfrac{1}{2}m_{20}\nu_2 n^{-1} + \tfrac{1}{2}\left(\frac{1}{2mhf_{\tilde{X}}(s)} - \frac{1}{m}\right) m_{02} \\
&\quad + O(n^{-2} + h^2 + (mh)^{-2} + P(A^c)).
\end{aligned}$$

The $P(A^c)$-term on the O-term comes from the replacement of terms like $\mathrm{E}\left(\frac{\hat{\sigma}}{\sigma} - 1\right) \cdot 1_A$ by $\mathrm{E}\left(\frac{\hat{\sigma}}{\sigma} - 1\right)$. The difference of the two can be estimated by

$$|E\left(\frac{\hat{\sigma}}{\sigma} - 1\right) \cdot 1_{A^c}| \le E\left(\frac{\hat{\sigma}}{\sigma} - 1\right)^4 + \mathrm{E}1_{A^c}$$

and so $P(A^c) = \mathrm{E}1_{A^c}$ appears in the O-term. Finally, straightforward calculations give

$$m_{10} = -\phi(a_1),\ m_{20} = a_1\phi(a_1)\{2k(a_1) - a_1\},\ m_{02} = 2g_1(a_1), \qquad (4.2.18)$$

which proves the lemma. □

Remark 4.2.1 In (4.2.17) both σ and γ are considered fixed. Under assumption A3 the lemma also holds when letting $\sigma, \gamma \to 0$. □

The lemma in combination with (4.2.16) and (4.2.1) suggests that we should take (following the notation of chapter 3)

$$d = d(\sigma, f_{\tilde{X}}, f'_{\tilde{X}}) = c + c_u, \tag{4.2.19}$$

where

$$c = -\frac{\sigma}{2}\frac{f'_{\tilde{X}}(s)}{f_{\tilde{X}}(s)}\frac{g_2(a_1)}{g'_1(a_1)} = \frac{\sigma}{2}\frac{f'_{\tilde{X}}(s)}{f_{\tilde{X}}(s)}(a_1^2 + 1 - a_1 k(a_1)), \tag{4.2.20}$$

$$\begin{aligned} c_u = {} & \frac{k(a_1)\{2a_1 k(a_1) + 1 - a_1^2\}}{4n} \\ & + (k(a_1) - a_1)\left(\frac{1}{2mhf_{\tilde{X}}(s)} - \frac{1}{m}\right) \end{aligned} \tag{4.2.21}$$

Comparison with (3.4.12) shows that the term to correct for estimation of σ has not changed, but the different way of estimating the density does result in a different correction.

In view of (4.2.16) and (4.2.1) we prove

Lemma 4.2.4 *Under the assumptions of lemma 4.2.3 in combination with* A3, *with* $\hat{d} = d(\hat{\sigma}, \hat{f}_{\tilde{X}}, \hat{f}'_{\tilde{X}})$ *from (4.2.19) and* $\hat{f}'_{\tilde{X}}$ *from (4.2.5), we have as* $\gamma, \sigma \to 0$, $\bar{h} \to 0$ *and if* $\bar{h} \geq h$ *and* $m \geq n$ *that*

$$\begin{aligned} & \mathrm{E}\left\{\frac{\hat{\sigma}}{\sigma}\hat{d}g'_1\left(\frac{\hat{\sigma}}{\sigma}\hat{a}_1\right) - dg'_1(a_1)\right\} 1_A \\ & = O\left(n^{-2} + h^2 n^{-1} + hm^{-1} + (mh)^{-2} + \sigma n^{-1} + \sigma(mh)^{-1} + \sigma\bar{h}^2 + P(A^c)\right) \end{aligned} \tag{4.2.22}$$

and

$$\mathrm{E}\left\{g_2\left(\frac{\hat{\sigma}}{\sigma}\hat{a}_1\right) - g_2(a_1)\right\} 1_A = O\left(n^{-1} + h^2 + (mh)^{-1} + P(A^c)\right). \tag{4.2.23}$$

Proof. Let b, X and Y be as in the proof of lemma 4.2.3. Then, letting $\hat{a}_1 = a_1(X, Y) = g_1^{-1}(b/(XY))$, we write (on the set A)

$$\begin{aligned} & X\hat{d}g'_1(X\hat{a}_1) - dg'_1(a_1) \\ & = n^{-1}\{f_1(X,Y) - f_1(1,1)\} + (mh)^{-1}\{f_2(X,Y) - f_2(1,1)\} \\ & \quad + \sigma\left\{f_3(X,Y)\left(\frac{\hat{f}'_{\tilde{X}}}{f'_{\tilde{X}}} - 1\right) + f_3(X,Y) - f_3(1,1)\right\} f'_{\tilde{X}} \\ & \quad + m^{-1}\{f_4(X,Y) - f_4(1,1)\} \end{aligned}$$

with, writing $h(X,Y) = Xg'_1(X\hat{a}_1)$,

$$f_1(X,Y) = \frac{k(\hat{a}_1)\{2\hat{a}_1 k(\hat{a}_1) + 1 - \hat{a}_1^2\}}{4} h(X,Y)$$

$$
\begin{aligned}
f_2(X,Y) &= \frac{k(\hat{a}_1)-\hat{a}_1}{2f_{\tilde{X}}(s)Y}h(X,Y)\\
f_3(X,Y) &= -\frac{1}{2}\frac{g_2(\hat{a}_1)}{g_1'(\hat{a}_1)f_{\tilde{X}}(s)\ Y}h(X,Y)\\
f_4(X,Y) &= -(k(\hat{a}_1)-\hat{a}_1)h(X,Y).
\end{aligned}
$$

Combination of (4.2.6) and (4.2.4) and dealing with 1_A as in the proof of lemma 4.2.3 yields

$$
\begin{aligned}
&\mathrm{E}\left\{\frac{\hat{\sigma}}{\sigma}\hat{d}g_1'\left(\frac{\hat{\sigma}}{\sigma}\hat{a}_1\right)-dg_1(a_1)\right\}\cdot 1_A =\\
&\qquad O\left(\{n^{-1}+(mh)^{-1}+\sigma+m^{-1}\}\{n^{-1}+h^2+(mh)^{-1}+P(A^c)\}\right)\\
&\qquad + O\left(\sigma\{\bar{h}^2+m^{-1}+(h\bar{h})^2+h/(m\bar{h}^2)+P(A^c)\}\right)\\
&= O\Big(n^{-2}+n^{-1}h^2+m^{-1}h+(mh)^{-2}+\sigma n^{-1}+\sigma(mh)^{-1}+\sigma\bar{h}^2\Big)\\
&\qquad + O\Big(P(A^c)\Big),
\end{aligned}
$$

using $\bar{h}\geq h$ and $m\geq n$, which completes the proof of (4.2.22). By the same argument (4.2.23) is proved. □

With $\hat{t}_u = s-(\hat{a}_1+\hat{d})\hat{\sigma}$ from (4.2.8), using $\hat{d}=d(\hat{\sigma},\hat{f}_{\tilde{X}},\hat{f}'_{\tilde{X}})$ from (4.2.19), combination of the results in this section yields

$$
\mathrm{E}CL(\hat{t}_u)=\gamma+O\left(\gamma\{n^{-2}+h^2+(mh)^{-2}+\sigma^2+\sigma\bar{h}^2\}+P(A^c)\right), \qquad (4.2.24)
$$

for $\bar{h}\geq h$ and $m\geq n$.

4.2.5 Optimal rates of bandwidths

To find the optimal bandwidths we have to consider $P(A^c)$ since it also depends on h and $\bar{h}$.

For the estimator $\hat{\sigma}$ we have that for each $\epsilon>0$ there exists $c(\epsilon)>0$ such that

$$
P\left(\left|\frac{\hat{\sigma}}{\sigma}-1\right|>\epsilon\right)=O(\mathrm{e}^{-c(\epsilon)n}),\ \text{as } n\to\infty. \qquad (4.2.25)
$$

By Kolmogorov's Exponential Inequality, cf. Shorack and Wellner (1986 p. 855), it follows that for each $\epsilon>0$ there exists constants $c_1(\epsilon),c_2(\epsilon)>0$ such that

$$
P\left(\left|\frac{\hat{f}_{\tilde{X}}}{f_{\tilde{X}}}-1\right|>\epsilon\right)=O(\mathrm{e}^{-c_1(\epsilon)mh}) \qquad (4.2.26)
$$

$$
P\left(\left|\frac{\hat{f}'_{\tilde{X}}}{f'_{\tilde{X}}}-1\right|>\epsilon\right)=O(\mathrm{e}^{-c_2(\epsilon)m\bar{h}^3}). \qquad (4.2.27)
$$

This implies that we should choose h and $\bar{h}$ such that $mh \to \infty$ and $m\bar{h}^3 \to \infty$. It is seen from (4.2.24) that the optimal rate for h equals $m^{-1/2}$, implying $mh \to \infty$ as should be the case to get $P(A^c)$ sufficiently small.

Note that this rate differs from the conventional optimal rates in density estimation implied by minimizing mean squared error. For the optimal bandwidth $\bar{h}$ there is some freedom. Taking $\bar{h} = m^{-\alpha}$ for some $\alpha \in [\frac{1}{4}, \frac{1}{3})$, the order term $O(\sigma\bar{h}^2)$ in (4.2.24) vanishes since $\sigma\bar{h}^2 \leq \sigma^2 + \bar{h}^4 \leq \sigma^2 + 1/m$ by this choice. On the other hand, the error term in (4.2.27) is sufficiently small.

As a result we have now proved the following theorem, which shows that we have established a second order unbiased test limit.

Theorem 4.2.1 *Assume* A1-A3. *Let* $\hat{\sigma}$ *be given by (4.2.3), and* $\hat{f}_{\tilde{X}}$ *and* $\hat{f}'_{\tilde{X}}$ *be given by (4.2.5) with* h *of order* $m^{-1/2}$ *and* $\bar{h}$ *of order* $m^{-\alpha}$ *for some* $\alpha \in [\frac{1}{4}, \frac{1}{3})$. *For the test limit* $\hat{t}_u = s - (\hat{a}_1 + \hat{d})\hat{\sigma}$ *with* $\hat{a}_1 = a_1(\hat{\sigma}, \hat{f}_{\tilde{X}})$ *and* $\hat{d} = d(\hat{\sigma}, \hat{f}_{\tilde{X}}, \hat{f}'_{\tilde{X}})$ *given by (4.2.7) and (4.2.19), respectively,*

$$\mathrm{E}CL(\hat{t}_u) = \gamma + O(\gamma(n^{-2} + m^{-1} + \sigma^2) + \mathrm{e}^{-rn} + \mathrm{e}^{-rm^{1-3\alpha}}), \tag{4.2.28}$$

for some $r > 0$ *as* $\sigma, \gamma \to 0$ *and* $m \geq n \to \infty$

Comparison of theorem 4.2.1 with theorem 3.4.1 shows that an extra term m^{-1} is entered (and a further exponential term). This means that if m is of the order n^2 or larger there is no loss in precision when using test limits based on density estimation, if we have normal observations. It is clear from section 4.1 and theorem 4.1.1 that a substantial gain is obtained in case of nonnormal observations.

4.2.6 Choice of bandwidths

Although theoretically a different constant in the bandwidth does not change the order-term in (4.2.28), from a practical point of view the constant is rather important if m is not extremely large. The remark in Hall and Sheather (1988, p. 382) that the optimal rate is more a qualitative guide than a quantitative assertion applies here as well. To get a concrete quantitative proposal we therefore explore the most important terms of the O-term in (4.2.28).

After some tedious, but straightforward calculation it turns out that we should take

$$h = \frac{3^{1/4}}{m^{1/2}\{f''_{\tilde{X}}(s)f_{\tilde{X}}(s)\}^{1/4}} \left(\frac{1}{2} + \frac{g''_1(a_1)g_1(a_1)}{4g'_1(a_1)^2}\right)^{1/4}. \tag{4.2.29}$$

Direct calculation shows that

$$\frac{g''_1(a)g_1(a)}{g'_1(a)^2} = k(a)\{k(a) - a\}.$$

From Ito and McKean (1974, p. 17) we obtain the inequalities, for $a \geq 0$,

$$\frac{1}{2}\left(a + \sqrt{a^2+2}\right) < k(a) \leq \frac{1}{2}\left(a + \sqrt{a^2+4}\right)$$

and hence

$$0.89 = (\tfrac{1}{2} + \tfrac{1}{4} \cdot \tfrac{1}{2})^{1/4} \leq \left(\frac{1}{2} + \frac{g_1''(a_1)g_1(a_1)}{4g_1'(a_1)^2}\right)^{1/4} < (\tfrac{1}{2} + \tfrac{1}{4})^{1/4} = 0.93.$$

As often the case in density estimation the optimal bandwidth depends on the unknown density. A possible solution is to use a reference distribution. Here we might take the normal distribution with mean μ and variance τ^2, say. For values of s with $(s-\mu)/\tau$ not too far from 2 and taking $f_{\tilde{X}}(x) = \phi((x-\mu)/\tau)/\tau$ we get

$$\frac{3^{1/4}}{(f_{\tilde{X}}''(s)f_{\tilde{X}}(s))^{1/4}} \approx \frac{\tau}{\sqrt{\phi((s-\mu)/\tau)}}.$$

Estimating μ and τ from the data by $\hat{\mu}$ and $\hat{\tau}$ and substituting these estimates leads to the proposal

$$h = \frac{\hat{\tau}}{\sqrt{m\ \phi((s-\hat{\mu})/\hat{\tau})}}. \tag{4.2.30}$$

Similarly, we propose

$$\bar{h} = \frac{\hat{\tau}}{\{m\ \phi((s-\hat{\mu})/\hat{\tau})\}^{1/4}}. \tag{4.2.31}$$

Either the usual sample mean and sample standard deviation or more robust estimators of μ and τ can be used.

4.2.7 Numerical results

In section 4.1 it was shown by theory and simulation results that the normal test limit may violate γ substantially, when observations are nonnormal. The preceding theory shows that a more robust test limit can be obtained using density estimation. It is seen in theorem 4.2.1 that asymptotically the new test limit behaves well under all kind of distributions. What remains is the behavior of the new test limit for moderate sample sizes. Two questions now arise when comparing the new test limit with the normal one for moderate sample sizes:

- how much do we lose when normality holds,
- how much do we gain when normality fails.

To answer the questions Monte Carlo experiments are performed for a broad range of alternatives. A confounding factor is the estimation of σ. Since we are interested in the influence of nonnormality of X on the test limits, both

for the normal test limit (table 4.1.1) as here, σ is assumed to be known in the simulations (implying again that the term $k(a_1)\{2a_1k(a_1)+1-a_1^2\}/(4n)$ in (4.2.21) disappears, according to '$n \to \infty$'). Results of the simulation study are presented in table 4.2.1 below. They should be compared with table 4.1.1.

Table 4.2.1. Simulated mean and standard deviation of $CL(\hat{t}_u)$

The test limit $\hat{t}_u$ as in (4.2.8) is calculated, in which $\hat{d} = d(\sigma, \hat{f}_{\bar{X}}, \hat{f}'_{\bar{X}})$ given in (4.2.19) and h and $\bar{h}$ as in (4.2.30) and (4.2.31). $\gamma = 100$ ppm, $\sigma = 0.10$. The simulation is carried out with 10000 replications.

	$\pi = 0.10$			$\pi = 0.01$		
distr.	m	$CL(\hat{t}_u)$	*freq.* $\hat{f}_{\bar{X}}(s)=0$	m	$CL(\hat{t}_u)$	*freq.* $\hat{f}_{\bar{X}}(s)=0$
$N(0,1)$	100	98.6(39.2)	0.0013	100	67.4(26.8)	0.0127
	400	99.8(27.0)	-	400	92.0(45.7)	0.0006
	1600	99.7(18.9)	-	1600	98.2(33.5)	-
beta	100	100.8(33.8)	0.0003	100	93.2(41.6)	0.0015
$p=2$	400	100.7(24.2)	-	400	101.4(54.1)	-
$q=2$	1600	100.6(16.7)	-	1600	100.4(24.1)	-
beta	100	97.3(40.7)	0.0001	100	65.2(23.6)	0.0043
$p=8$	400	99.6(29.3)	-	400	89.7(38.9)	0.0002
$q=32$	1600	99.6(20.1)	-	1600	97.0(29.9)	-
beta	100	96.7(40.4)	0.0054	100	63.0(22.5)	0.0005
$p=2$	400	99.8(30.9)	-	400	87.6(32.4)	-
$q=8$	1600	99.7(21.2)	-	1600	96.3(24.6)	-
n.mixt	100	98.1(66.9)	0.0002	100	101.2(29.8)	0.0298
$\mu_1=0.53$	400	100.3(28.0)	-	400	86.8(30.5)	0.0013
$\sigma_1=1.0$	1600	100.2(19.6)	-	1600	95.7(25.0)	-
n.mixt	100	98.0(38.5)	0.0032	100	129.8(35.3)	0.0439
$\mu_1=0.63$	400	100.3(28.8)	-	400	92.4(28.4)	0.0013
$\sigma_1=1.1$	1600	100.2(20.2)	-	1600	94.1(21.6)	-
gamma	100	93.8(38.9)	0.0053	100	36.0(8.9)	-
$\alpha=2$	400	99.7(36.2)	0.0003	400	73.6(19.7)	-
	1600	99.5(23.6)	-	1600	91.6(20.3)	-
gamma	100	96.6(40.5)	0.0032	100	56.7(20.0)	0.0006
$\alpha=6$	400	99.5(31.7)	-	400	85.8(32.1)	0.0001
	1600	99.7(21.7)	-	1600	96.2(27.3)	-
gamma	100	96.6(39.9)	0.0001	100	63.7(23.3)	0.0067
$\alpha=32$	400	99.8(29.3)	-	400	90.0(40.7)	0.0006
	1600	99.9(20.5)	-	1600	97.1(31.3)	-

It is seen in tables 4.2.1 and 4.1.1 that the loss of the robust test limit is small if normality holds, except when π and m are both small. Under nonnormality there is a drastic improvement. This is exactly what is expected for a robust procedure, a small loss when the claimed model holds and a large gain if the assumed model does not hold. Note that $n = 40$ is even a rather small value for estimating σ. In view of (4.2.28) one may think on $m \approx n^2$. It is seen that for $m = 1600$ the simulated mean $\mathrm{E}CL(\hat{t}_u)$ is close to γ in all cases in table 4.2.1. This does not hold for the normal test limit.

In the theory it is assumed that $h \to 0$, or $m\phi(s) \to \infty$. That small values of $m\phi(s)$ may indeed lead to larger errors is seen in table 4.2.1 when considering $\pi = 0.01$ and the gamma distribution with $\alpha = 2$. In this situation $m\phi(s) = 0.18, 0.74, 2.95$ for $m = 100, 400, 1600$ respectively. This explains the errors in that case. Using the normal test limits it is seen that much larger errors occur, which remain if m becomes larger (cf. table 4.1.1). Moreover, in most cases the average consumer loss based on density estimation is conservative, in contrast to the average consumer loss based on normal test limits.

4.2.8 No observations in $[s - h, s + h]$

If π and m are small, it may occur that there are no observations in $[s-h, s+h]$, implying $\hat{f}_{\tilde{X}}(s) = 0$. For instance, if the density of X is standard normal, $\pi = 0.01$ $(s = 2.33)$, $\sigma = 0.1$ and the density is estimated with $m = 100$, then $h = 0.615$ and the probability of no observations in $[s - h, s + h]$ equals 0.013. Formally, $\hat{f}_{\tilde{X}}(s) = 0$ would imply to take $\hat{t}_u = \infty$. The idea is that the observations indicate that there is no probability mass on the right hand side of s and therefore there is no reason to reject products. As a consequence, in this situation $CL(\hat{t})$ equals π and hence in the preceding example the contribution of this event to $\mathrm{E}CL(\hat{t})$ equals $0.013 \cdot 0.01$, irrespective of γ, and thus $\gamma = 100$ ppm is already exceeded by this contribution. In table 4.2.1 the number of cases in which there are no observations in $[s - h, s + h]$ is reported. For calculations of the simulated mean and standard deviations these cases are deleted.

Application of other density estimators does not help to solve this problem. This is reasonable, since if we make no assumptions about the relation between the density in the 'middle' and in the 'tail' (as we do with nonparametric density estimation) and if there are no observations in the tail, then the only reasonable nonparametric estimate is $\hat{f}_{\tilde{X}}(s) = 0$.

One way to use information in the 'middle' is to assume a parametric model, somewhere between on one hand the too restrictive normal model and on the other hand the nonparametric approach. On such a intermediate approach will be reported in chapter 5. As it is seen in table 4.2.1 the problem disappears almost for larger m. There is no such problem with $\hat{f}'_{\tilde{X}}(s) = 0$. The corresponding term in $\hat{c}$ (cf. (4.2.20)) simply vanishes.

4.2.9 Conclusions, recommendations

In view of the theoretical and simulations results we conclude the following.

- The second order unbiased test limit using Rosenblatt's kernel estimator for estimating $f_{\tilde{X}}(s)$ and $f'_{\tilde{X}}(s)$ has the required robustness property with respect to the normal test limit:
  - only a small loss under the normal model
  - a large gain in case of nonnormality.

- If one is pretty sure that $\dfrac{f_X(s)}{\phi((s-\mu)/\tau)/\tau}$ is close to 1, where μ and τ^2 are the mean and variance of X, the normal test limit may be preferred. Otherwise, it is (as a rule much) better to use the test limit derived in this chapter.

- If $\hat{f}_{\tilde{X}}(s) = 0$, try to get more observations. If this is impossible, the normal test limit may be applied, but note that for small π and m the normal test limit is unreliable.

- When taking the number n of the test data to estimate σ of order $m^{1/2}$, with m the number of production data, the error terms induced by estimating σ, $f_{\tilde{X}}$ and $f'_{\tilde{X}}$ are of the same order.

4.3 Test limits for which γ is violated with small probability

It is seen in the theory of section 4.1 that the consumer gloss based on normal test limits may be far from γ on the average if X does not have a normal distribution. The simulation results in table 4.1.1 show this rather clearly. Further, it is seen by the standard deviation (also presented in table 4.1.1) that the consumer loss varies widely.

In section 3.4.2 it is shown that the standard deviation of $CL(\hat{t})$ when applying the normal test limit, hence estimating μ_X and σ_X, is of order $m^{-1/2}$. This is clearly seen in table 4.1.1. For the test limit based on density estimation it is seen in table 4.2.1 that if $\pi = 0.10$ the standard deviation for $m = 1600$ is about one half of its value at $m = 100$. Indeed, we will show that the standard deviation is of order $m^{-1/4}$.

From (4.2.16) writing $\hat{d} = \hat{c} + \hat{c}_u$ and from the proof of lemma 4.2.3 we get, on the set A (cf. (4.2.15)),

$$\begin{aligned}\frac{CL(\hat{t}_u) - \gamma}{\gamma} \\ \approx\ -\frac{\phi(a_1)}{g_1(a_1)}\left(\frac{\hat{\sigma}}{\sigma} - 1\right) - \left(\frac{\hat{f}_{\tilde{X}}}{f_{\tilde{X}}} - 1\right) + \left(\frac{\hat{f}_{\tilde{X}}}{f_{\tilde{X}}} - 1\right)^2 + c_u \frac{g_1'(a_1)}{g_1(a_1)}.\end{aligned} \tag{4.3.1}$$

The right-hand side is asymptotically normal $AN(\mu_{CL}, \sigma_{CL}^2)$ with (neglecting terms of order n^{-1} and $(mh)^{-1}$ in μ_{CL}),

$$\mu_{CL} = -\frac{h^2}{6}\frac{f''_{\tilde{X}}(s)}{f_{\tilde{X}}(s)} + c_u\frac{g_1'(a_1)}{g_1(a_1)} \tag{4.3.2}$$

$$\sigma_{CL}^2 = \frac{1}{2n}\left(\frac{\phi(a_1)}{g_1(a_1)}\right)^2 + \frac{1}{2mhf_{\tilde{X}}(s)}. \tag{4.3.3}$$

The choice $h = (m\phi(s))^{-1/2}$ indeed leads to σ_{CL} of order $m^{-1/4}$.

Having explained the order of magnitude of the standard deviation it may be appropriate, when the unbiasedness property is not adequate, to replace $\hat{c}_u$ by a larger $\hat{c}_i$ leading to

$$\hat{t}_i = s - (\hat{a}_1 + \hat{c} + \hat{c}_i)\hat{\sigma}, \tag{4.3.4}$$

for which, with sufficient precision,

$$P(CL(\hat{t}_i) > \gamma) \leq \alpha, \tag{4.3.5}$$

where α is a small positive number.

Considering the asymptotic normality this requirement implies that $\mathrm{E}CL(\hat{t}_i)$ should be smaller than γ. Thus by introducing a negative bias, it is expected that this gives opportunity to reduce the variance at the same time! In view of the new criterion we therefore reconsider the bandwidth h.

We do not absorb the term $\frac{h^2}{6}\frac{f''_{\tilde{X}}(s)}{f_{\tilde{X}}(s)}$ into c_i, since we do not want to estimate $f''_{\tilde{X}}$. As a consequence we put the condition

$$\frac{h^2}{\sigma_{CL}} \to 0, \tag{4.3.6}$$

which is satisfied if

$$h^5 m \to 0. \tag{4.3.7}$$

Assuming (4.3.6) we take, as in the normal case, c_i such that $-\mu_{CL}/\sigma_{CL}$ is the upper α-quantile of the standard normal distribution. We obtain

$$c_i = u_\alpha\sqrt{\frac{k(a_1)^2}{2n} + \frac{(k(a_1)-a_1)^2}{2mhf_{\tilde{X}}(s)}}, \tag{4.3.8}$$

where $u_\alpha = \Phi^{-1}(1-\alpha)$. (Since $(n^{-1}+(mh)^{-1})/\sigma_{CL} \to 0$, indeed terms of order n^{-1} and $(mh)^{-1}$ can be neglected in (4.3.2)).

It is seen from (4.3.8) and (4.2.21) that, ignoring terms of order n^{-1} and $(mh)^{-1}$, c_i indeed is larger than c_u and hence $\hat{t}_i = s - (\hat{a}_1 + \hat{c} + \hat{c}_i)\hat{\sigma}$ is smaller than $\hat{t}_u$.

Reducing the bias as much as possible under the stronger requirement (4.3.5) to maximize the yield $P(\tilde{X} < \hat{t}_i)$, we should minimize the variance σ^2_{CL}. Therefore, h should be chosen as large as possible, but satisfying (4.3.7).

Ignoring at this point the n^{-1} term in (4.3.3), deletion of the h^2 term in μ_{CL} (cf. (4.3.6)) can be translated into

$$\left|\frac{h^2}{6}\frac{f''_{\tilde{X}}(s)}{f_{\tilde{X}}(s)}\right| \leq \frac{1}{K}\left|\frac{1}{2mhf_{\tilde{X}}(s)}\right|^{1/2},$$

for some large K and hence

$$h^5 \leq \frac{18 f_{\tilde{X}}(s)}{K^2 m\{f''_{\tilde{X}}(s)\}^2}.$$

If $f_{\tilde{X}} = \phi$ and s is close to 2, then $18f_{\tilde{X}}(s)/\{f''_{\tilde{X}}(s)\}^2$ is close to 1 and we get $h = \{K^2 m\phi(s)\}^{-1/5}$. Taking for instance $m = 100$, $s = 2.33$ (corresponding to $\pi = 0.01$) we get $\{m\phi(s)\}^{-1/5} = 0.82$ and $\{m\phi(s)\}^{-1/2} = 0.61$, hence for reasonable values of K the choice of h here is even smaller than in (4.2.30). A similar discussion is found in Hall and Sheather (1988).

Resuming we may conclude

- the stronger requirement $P(CL(\hat{t}_i) > \gamma) \leq \alpha$ is met,
- for large values of m the bandwidth h should be taken somewhat larger than as in (4.2.30), while for moderate values of m still (4.2.30) may be applied.

Chapter 5

A comparison between parametric density estimators and Rosenblatt's estimator

In the last chapter we have seen that to determine test limits, it is important to estimate the density of the characteristic at the specification limit accurately. Application of Rosenblatt's estimator yields good results in many situations. Because nonparametric density estimators typically use observations around the specification limit only, problems occur in situations with a small number of observations and where the specification limit is in the tail of the distribution (cf. the discussion in section 4.2.9).

In this chapter we investigate in what situations application of parametric models could be considered a complement to, or even a competitor of the nonparametric approach. We concluded in chapter 4 that the normal density as a density estimator, using just a location and scale parameter, turns out to be no competitor of the nonparametric estimator. But what do we gain if for example parameters of skewness and kurtosis are added? Another extension to a parametric family is to use some quantiles as shape parameters.

A larger parametric model could also be viewed as a compromise between assuming perfect knowledge of the form of the density (normality) and no knowledge at all (nonparametric approach). Instead of making a big step at once, an intermediate approach assuming moderate knowledge of the density may give an improvement in particular situations where the nonparametric approach is less reliable.

In any case, with a parametric model the observations from the center of the distribution are used as well, and improvements are therefore expected especially in those situations where the specification limit is in the tail.

5.1 Overview of the investigations

In five sections, 5.3-5.7, we will discuss five parametric models by looking at two things. By simulation we investigate how well the true density can be estimated if the observations are from the model. Then we investigate in what sense the parametric model can be used in situations where observations are not from the model, but still from a reasonably smooth distribution. We compare these results again with those of Rosenblatt's estimator, but also with those of the normal distribution. As we are especially interested in improvements of Rosenblatt's estimator if the density has to be estimated in the tail with only a relatively small number of observations, we will study these situations in particular.

After introduction of the set-up and the notations in section 5.2, in section 5.3 we reconsider application of the normal density as a density estimator. In the following four sections we will then study some larger parametric models, which are all, in various ways, extensions to the normal distribution.

- exponential power distribution (section 5.4)
- Pearson system (section 5.5)
- Johnson system (section 5.6)
- Box-Cox transformation model (section 5.7)

This chapter is concluded by an appendix containing some technicalities.

It will turn out that these larger parametric models often lead to improvements compared to the normal distribution. However, they do not always compete well against the kernel estimator.

5.2 Set-up and notations

From the proof of theorem 4.1.1 it is seen that if the test limit is based on an estimator $\hat{f}_X$ of f_X (f_X denoting the density of the characteristic), the consumer loss satisfies

$$\frac{\widehat{CL} - \gamma}{\gamma} \approx \frac{f_X(s)}{\hat{f}_X(s)} - 1, \tag{5.2.1}$$

at least, if $f_X(s) > 0$ and provided $\hat{f}_X(s)$ is close to the true density.

In this chapter we pass over the complication that to set test limits we usually have to rely on observations from X which are contaminated by a small random measurement error U. We simply assume that we have observations from X itself. In the previous chapter it is seen that if the test limit is based on Rosenblatt's estimator, there is asymptotically no loss of precision. For

parametric estimators in general it is unduly complicated to investigate the consequences if observations from $X+U$ instead are used, however. Therefore, at first we simply ignore this complication and assume that we have independent observations $X_1, \ldots, X_m$ from X. In the remainder of this chapter we omit the subscript $_X$ of f_X.

Let $\hat{f}_R$ be Rosenblatt's estimator (cf. (4.2.5), however, with $\tilde{X}_{n+i}$ replaced by X_i, $i = 1, \ldots, m$). In section 4.2.5 we have shown that there exists a set A, defined by

$$A = \left\{ \left| \frac{\hat{f}_R}{f} - 1 \right| < \epsilon \right\}, \tag{5.2.2}$$

such that $\mathrm{P}(A^c) = O(\mathrm{e}^{-c_\epsilon mh})$ for some positive constant c_ϵ. Together with the fact that $\widehat{CL}$, being a probability, is bounded, this implies that quantities like mean, variance and mean squared error of $\widehat{CL}$ can be evaluated with respect to 1_A. The error caused is $O(\mathrm{e}^{-c_a mh})$, while this enables us to use (5.2.1) throughout, when applicated to Rosenblatt's estimator. From section 4.2.2 it follows that, as $m \to \infty$,

$$\mathrm{E}\left(\frac{f(s)}{\hat{f}_R(s)} - 1 \right)^2 \cdot 1_A \approx \frac{h^4}{36} \left(\frac{f''(s)}{f(s)} \right)^2 + \frac{1}{2mhf(s)} \tag{5.2.3}$$

leading to the recommendation

$$h = \left\{ m \frac{1}{\sigma_X} \phi\left(\frac{s - \mu_X}{\sigma_X} \right) \right\}^{-1/5},$$

where, if unknown, the expectation μ_X and variance σ_X^2 should be estimated. The rate of convergence now equals $m^{-4/5}$.

Writing $\{f_\theta, \theta \in \Theta\}$ for the parametric family of densities, there should be f_θ for some $\theta \in \Theta$ close to the unknown f and the estimator $f_{\hat{\theta}}$ should in turn be close to f_θ. We assume $f_\theta(s) > 0$. Under weak regularity conditions (cf. appendix A.1 at the end of this chapter) there exists also a set B with $P(B^c)$ exponentially small, such that

$$\mathrm{E}\left(\frac{f(s)}{f_{\hat{\theta}}(s)} - 1 \right)^2 \cdot 1_B = \left(\frac{f(s)}{f_\theta(s)} - 1 \right)^2 + O(m^{-1}), \tag{5.2.4}$$

as $m \to \infty$.

What can we say about the difference between (5.2.3) and (5.2.4) in general? First of all, if $f_\theta(s) = f(s)$ the rate of convergence of the parametric estimator equals m^{-1} and hence from a theoretical point of view faster convergence is obtained using the parametric estimator than when using the nonparametric one.

For small sample sizes the difference between $m^{-4/5}$ and m^{-1} is not really large and the coefficients of the $m^{-4/5}$- and m^{-1}-terms are of importance as

well. The main coefficients in (5.2.3) are $\phi^{1/5}/f$ and $\phi^{-4/5}$ which are large when f is estimated in the tail. The coefficients of the parametric density estimator depend on f_θ and the estimator $\hat{\theta}$. For the sake of illustration, consider the situation of chapter 3 in which $f(s) = f_\theta(s) = \phi((s-\mu)/\sigma)/\sigma$. If we expand the left-hand side of (5.2.4) around $(\hat{\mu}, \hat{\sigma}) = (\mu, \sigma)$ we encounter terms like $\mathrm{E}((\hat{\mu}-\mu)/\sigma)^l((\hat{\sigma}-\sigma)/\sigma)^k$ (for $k, l = 0, 1, 2$, cf. (5.A.5) and (5.A.6) in the appendix). When μ, σ are estimated by the sample mean and the sample standard deviation, respectively, the terms are quite small. Estimation of the parameter σ using sample α_1- and α_2-quantiles $\hat{x}_{(\alpha_1)}$ and $\hat{x}_{(\alpha_2)}$ leads to terms containing, among others,

$$\mathrm{COV}(\hat{x}_{(\alpha_1)}, \hat{x}_{(\alpha_2)}) \approx \frac{\alpha_1(1-\alpha_2)}{mf(x_{(\alpha_1)})f(x_{(\alpha_2)})}, \quad \text{if } \alpha_1 < \alpha_2. \tag{5.2.5}$$

Comparing this expression to (5.2.3) we conclude that, at least for small sample sizes, it is not sensible to use sample quantiles beyond the specification limit s. In that case the coefficients easily become larger than those of Rosenblatt's estimator.

Problems may arise in parametric density estimation when, besides parameters of location and of scale, also parameters of skewness and kurtosis have to be estimated. It is well known that estimation of higher moments causes problems if these moments are large.

In the numerical examples in the following sections the exact mean and mean squared error of $(f/\hat{f}_R) \cdot 1_{\{\hat{f}_R > 0\}}$ as well as

$$p_0 = P(\hat{f}_R(s) = 0) \tag{5.2.6}$$

are calculated. As we deal with standardized densities only, μ_X=0 and σ_X^2=1, yielding $h = \{m\phi(s)\}^{-1/5}$. (Note that $2mh\hat{f}_R(s)$ has a binomial distribution $B(m, p)$, with $p = F(s+h) - F(s-h)$, F being the distribution function of X),

For estimation in the parametric models we rely on simulation. Together with the simulation results we will provide the number of replications, if positive, for which the parametric estimate of the density is less than a certain quantity (as a rule 10^{-5}), or it cannot be determined by numerical problems.

The specification limit s is chosen such that

$$\pi = P(X > s)$$

is 0.10 or 0.01.

5.3 Normal distribution

In this section we reconsider the normal density as a parametric model, that is, we take $f_\theta(s) = \phi((s-\mu)/\sigma)/\sigma = \phi_\theta(s)$ with $\theta = (\mu, \sigma)$. The parameter θ may

be estimated by the maximum likelihood estimators

$$\hat{\theta} = (\hat{\mu}, \hat{\sigma}), \text{ with } \hat{\mu} = \bar{X} = \frac{1}{m}\sum_{i=1}^{m} X_i, \ \hat{\sigma}^2 = \frac{1}{m}\sum_{i=1}^{m}(X_i - \bar{X})^2.$$

In view of the discussion at the end of section 5.2 we will also consider quantile estimators of θ. Let $\hat{x}_{(p)}$ be the sample p-quantile, for instance, $\hat{x}_{(1/2)}$ is the sample median, then an estimator of θ is also defined as follows.

$$\hat{\mu} = \text{sample median}, \ \ \hat{\sigma} = (\hat{x}_{(p)} - \hat{\mu})/\Phi^{-1}(p),$$

where $\Phi^{-1}(p)$ is the p-quantile of the standard normal distribution.

In the table 5.3.1 below $f/f_{\hat{\theta}}$ is compared to $f/\hat{f}_R$ when sampling from the normal family itself, that is, $f(s) = \phi(s)$.

Table 5.3.1 $\phi(s)$ estimated in three ways ($\pi = 0.01$)

The parameters μ and σ are estimated either by the maximum likelihood estimators or by means of the sample median and the 0.975-sample-quantile. The mean and mean squared error of $\phi/\phi_{\hat{\theta}}$ are estimated on the basis of a simulation with 10000 replications.

		Rosenblatt		Maximum likelihood estimators of μ, σ		Quantile estimators of μ, σ	
m	p_0	$E\phi/\hat{f}_R$	$\sqrt{MSE}$	$E\phi/\phi_{\hat{\theta}}$	$\sqrt{MSE}$	$E\phi/\phi_{\hat{\theta}}$	$\sqrt{MSE}$
100	0.001	0.809	0.522	1.160	0.599	1.155	1.074
400	0.000	0.828	0.285	1.033	0.214	1.047	0.354
1600	0.000	0.878	0.180	1.007	0.098	1.010	0.157
6400	0.000	0.921	0.113	1.002	0.049	1.002	0.076

As it is seen, the parametric estimator based on maximum likelihood estimators is the best one. In terms of MSE, Rosenblatt's estimator is almost as good for $m \leq 400$. It performs for $m \leq 400$ better than the parametric estimator based on quantiles. For very large m all three estimators behave very well, the convergence rate of Rosenblatt's estimator being slightly slower.

In section 4.1 we have investigated both theoretically and by means of some examples how much the ratio f/f_θ can differ from 1, with $f_\theta \equiv \phi$ and f an arbitrary but standardized density out of certain classes of densities. If the parameter θ is estimated using sample quantiles these results are of less value, because the mean and variance of ϕ_θ are not necessarily 0 and 1, respectively. We extend the result obtained for $f \in \mathcal{F}_1$ (the class of unimodal and symmetric densities) to the case in which the normal density is such that two quantiles correspond to the unknown f.

Call $\mathcal{F}_q(p)$ the class of unimodal and symmetric densities for which the distribution function F is s.t. $F^{-1}(\frac{1}{2}) = x_{(1/2)}$ (median) and $F^{-1}(p) = x_{(p)}$ with $x_{(1/2)}$ and $x_{(p)}$ given and $p > \frac{1}{2}$.

Proposition 5.3.1 *For $s > x_{(1/2)}$*

$$\sup_{f \in \mathcal{F}_q(\frac{1}{2},p)} f(s) = \frac{p - \frac{1}{2}}{s - x_{(1/2)}} \qquad \text{if } x_{(1/2)} < s \leq x_{(p)},$$

$$\sup_{f \in \mathcal{F}_q(\frac{1}{2},p)} f(s) = \min\left\{ \frac{p - \frac{1}{2}}{x_{(p)} - x_{(1/2)}}, \frac{1-p}{s - x_{(p)}} \right\} \qquad \text{if } x_{(p)} < s.$$

Proof. For $f \in \mathcal{F}_q(p)$ we have that

$$\int_{x_{(1/2)}}^{x_{(p)}} f(x)\mathrm{d}x = p - \frac{1}{2}.$$

Because f is unimodal and symmetric around $x_{(1/2)}$, f is not increasing on $[x_{(1/2)}, \infty)$. It follows that

$$(s - x_{(1/2)}) \cdot f(s) \leq p - \frac{1}{2},$$

for $s \in (x_{(1/2)}, x_{(p)}]$.
Moreover, from

$$\int_{x_{(p)}}^{\infty} f(x)\mathrm{d}x = 1 - p$$

it follows that

$$(s - x_{(p)}) \cdot f(s) \leq 1 - p,$$

for $s > x_{(p)}$. Since $f(s)$ should also be less than or equal to $f(x_{(p)})$ for $s > x_{(p)}$, we have

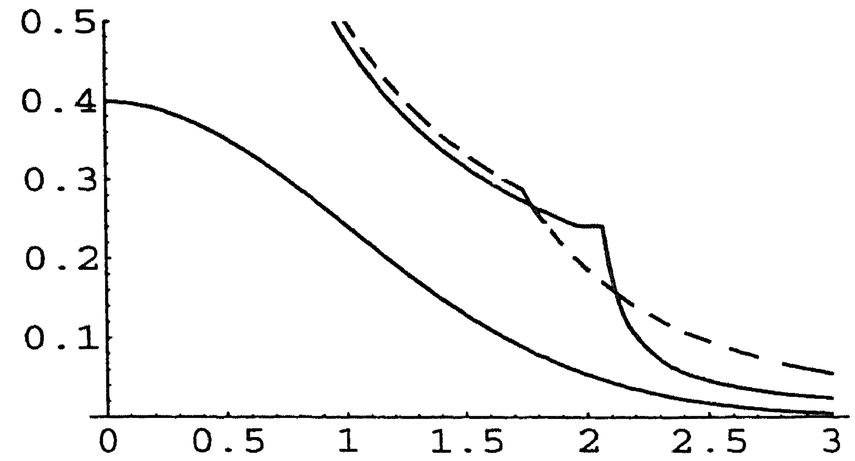
$$f(s) \leq \min\left\{ \frac{p - \frac{1}{2}}{x_{(p)} - x_{(1/2)}}, \frac{1-p}{s - x_{(p)}} \right\},$$

for $s > x_{(p)}$.

In the same way as in the proof of lemma 4.1.1, we can construct densities that come arbitrary close to the upper bound. Hence the proposition follows. □

Figure 5.3.1 on the next page clearly shows the difference with the moment approach.

Figure 5.3.1 Suprema of densities from and $\mathcal{F}_1$ (cf. lemma 4.1.1) and $\mathcal{F}_q(0.975)$



The solid line shows the suprema of $f \in \mathcal{F}_q(0.975)$
The dashed line shows the suprema of $f \in \mathcal{F}_1$
The standard normal density ϕ is drawn as well

Table 5.3.2 and table 5.3.3 (page 101 and 102) provide some examples of mean and mean squared error of the ratios $f/\phi_{\hat{\theta}}$ when sampling from β- and Γ-distributions (cf. section 4.1, p. 62). The parameter θ is estimated using the maximum likelihood estimator (derived in the normal family) and the quantile estimator with $p = 0.975$. Under f, the maximum likelihood estimator converges to (0,1) as $m \to \infty$, since $EX = 0$ and $\sqrt{\text{VAR}X} = 1$ under f. Therefore the limiting value $f(s)/\phi(s)$ is also presented in the tables.

The sample median and sample 0.975-quantile under f converge to the median and 0.975-quantile of X with density f, as $m \to \infty$. Let $\theta = (\mu, \sigma)$ be the parameter value for which the normal distribution with mean μ and standard deviation σ has the same median and 0.975-quantile as the distribution of of X with density f. Now the limiting value of $f/\phi_{\hat{\theta}}$ equals f/ϕ_θ and therefore this quantity is presented in table 5.3.2 and table 5.3.3 as well. The value $p = 0.975$ is chosen afterwards, out of several values. For values below 0.90 the results are a lot worse. The results are again compared to those of Rosenblatt's estimator.

We summarize table 5.3.2 and table 5.3.3 as follows.

- The ratio's f/ϕ_θ in the quantile case are very close to one for various f.

- There is a very large bias and mean squared error of $f/\phi_{\hat{\theta}}$ when θ is estimated by the maximum likelihood estimator and the ratio f/ϕ is large. This is (mainly) due to the large ratios $f(s)/\phi(s)$. (cf. (5.A.7) in appendix A.1).

- For small sample sizes the bias and mean squared error of $f/\phi_{\hat{\theta}}$, using the quantile estimators, are not smaller than the bias and mean squared error obtained with Rosenblatt's estimator (cf. the discussion in section 5.2).

Based on the computations in this section, we conclude that if one chooses to apply the normal density in a nonnormal situation, its two parameters should be estimated using the sample median and a sample p-quantile where p is somewhat

smaller than $1 - \pi$. But even then the normal density estimator is not to be considered a competitor of Rosenblatt's estimator.

5.4 Exponential Power Distribution

The exponential power distribution (*EPD*) looks as a simple and obvious extension to the normal distribution. Its density is given by

$$f_\theta(x) = \frac{1}{2\sigma\Gamma(\beta+1)} \exp(-\left|\frac{x-\mu}{\sigma}\right|^{1/\beta}), \quad \text{for } -\infty < x < \infty,$$

where Γ is the gamma function. The extra parameter β affects the tail of the distribution. The larger β, the more heavy tails, while for $\beta \to 0$ the *EPD* converges to a uniform distribution. For example, $\beta = 2$ corresponds to a kurtosis of 25.2.

Two types of estimators of $\theta = (\mu, \sigma, \beta)$ suggest themselves, moment estimators and maximum likelihood estimators.

The expressions for the moment estimators are simple but not explicit. The equations are

$$\mathrm{E}X = \mu$$

$$\mathrm{E}\,|X - EX|^i = \sigma^i \, \frac{\Gamma((i+1)\cdot\beta)}{\Gamma(\beta)}$$

and β follows from

$$\frac{\mathrm{E}\,|X - \mathrm{E}X|^j}{(\mathrm{E}\,|X - \mathrm{E}X|^i)^{j/i}} = \frac{\Gamma((j+1)\cdot\beta)\ \Gamma(\beta)^{j/i} - 1}{\Gamma((i+1)\cdot\beta)^{j/i}}.$$

It is natural to base the estimators of σ and β on the variance and kurtosis ($i = 2, j = 4$), but simulations showed that good results are obtained as well if the estimators are based on the third absolute moment in combination with either the first absolute moment or the variance.

Table 5.4.1 (page 103) shows simulation results of $f_\theta / f_{\hat{\theta}}$, where θ is estimated using sample mean, sample variance and sample kurtosis, in comparison with Rosenblatt's estimator. The conclusions are

- To estimate the density parametrically at a point in the tail of the distribution, 100 observations do certainly not suffice. In this situation Rosenblatt's estimator performs much better.
- There seems to be no considerable loss when the normal density is estimated by the density of the *EPD*, provided the number of observations is sufficiently large (i.e. larger than 100). Moreover, the bias and mean squared error do not vary much with the parameter β.

Because the small sample behavior of the moment estimators is not really satisfying, we also studied maximum likelihood estimation. The estimators have to be determined iteratively from a system of three equations. As initial estimates we took the moment estimates based on the variance and third absolute moment. In appendix A.2 the estimation procedure is described. It shows that these estimators are not to be recommended for practical purposes.

Table 5.4.2 on page 104 shows some simulation results with the maximum likelihood estimators. The situation with 6400 observations is omitted because of the extremely large amount of computer time involved. We conclude that the results in the situation with 100 observations have improved, but they are not yet satisfying. For larger numbers of observations the results seem to improve only for larger values of β ($\beta \geq 1$), but this improvement is accompanied by an increasing number of failures. These failures are due to the fact that we stopped the estimation procedure after 200 iterations.

Table 5.4.3 and table 5.4.4 on page 105 and page 106 are the analogs of table 5.3.2 and table 5.3.3. Γ- and β-densities are estimated by means of the density of the *EPD*. These tables are interesting because they are examples of the improvement that is obtained if, besides mean and variance, the kurtosis is involved in estimation as well. (It turned out that there is not much difference between estimators based on the variance and absolute third moment and estimators based on the variance and kurtosis. This makes the latter preferable.) We may conclude that compared to the normal density, there is an improvement in the ratio f/f_θ, but for smaller sample sizes this is annulled by the poor estimators.

A question that emerges here is what improvements of the ratio f/f_θ can be achieved by involving the kurtosis. In lemma 4.1.1 we saw that for standardized, unimodal and symmetric densities f,

$$f(s) \leq \min\left\{\frac{1}{2s}, \frac{3}{2s^3}\right\}.$$

If the kurtosis of such a density f should equal d, we also have that

$$d/2 \geq \int_0^s x^4 f(x)\mathrm{d}x \geq f(s)\int_0^s x^4\mathrm{d}x = f(s)\cdot s^5/5 \Rightarrow f(s) \leq \frac{5d}{2s^5}.$$

For instance for $d = 3$, corresponding to the normal distribution, the condition becomes effective for $s \geq \sqrt{5} \approx 2.24$.

We finish this section with the conclusion that involving the kurtosis as an extra parameter to estimate often gives an improvement in the ratio f/f_θ compared to f/ϕ (using only mean and variance). The estimator based on the density of *EPD* does not improve the results of Rosenblatt's estimator.

5.5 Pearson system

Although the Pearson system is very old, still literature appears about the system and it is still used in a lot of applications (cf. Parrish (1983)). As the system is especially meant to approximate unknown densities, we studied its possible application as well. For a detailed description of the system we refer to Johnson & Kotz (1970), here we will briefly summarize the idea.

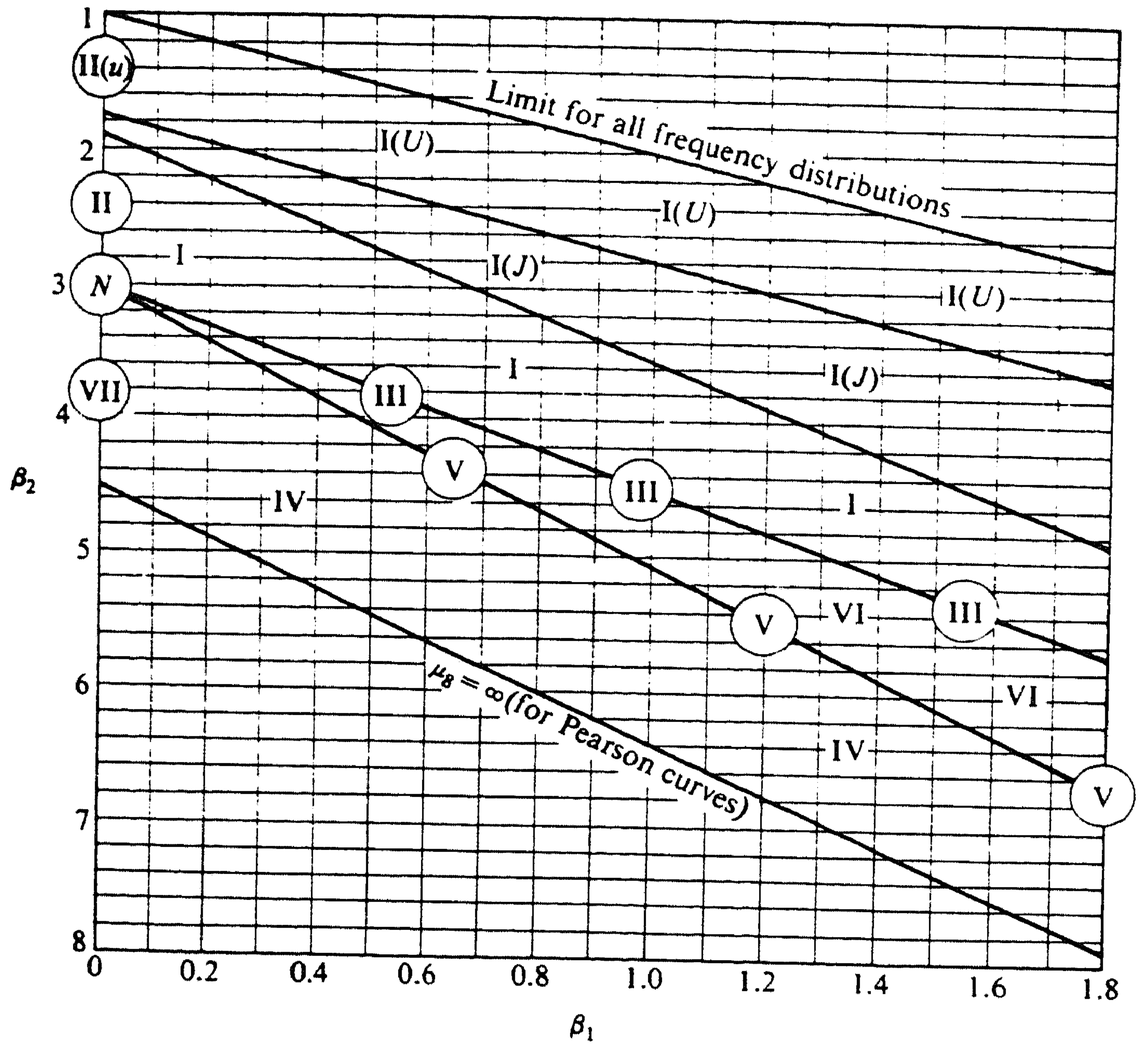
The Pearson system is a system of types of distributions stemming from the following differential equation for the density f:

$$\frac{1}{f(x)}\frac{\mathrm{d}f(x)}{\mathrm{d}x} = -\frac{a+x}{c_0+c_1x+c_2x^2}.$$

Different conditions on the roots of $c_0+c_1x+c_2x^2$ correspond to different types of distributions. For example, the condition that $c_0+c_1x+c_2x^2$ has a positive and a negative root leads to the β-distribution, while if $c_1 = c_2 = 0$, the normal distribution follows.

Figure 5.5.1 below summarizes the Pearson system.

Figure 5.5.1 Pearson system (figure from Johnson&Kotz (1970))



$\beta_1 = (\text{skewness})^2$, $\beta_2 = \text{kurtosis}$

Each combination of skewness and kurtosis corresponds to one (type of) distribution.
The main types of distributions are

I β – distribution, cf. page 62,

IV 'Type IV', with density

$$f(x) = K(s_0 + c_2(x+s_1)^2)^{-1/(2c_2)} \exp(-\frac{a-s_1}{\sqrt{c_2 s_0}} \arctan \frac{x+s_1}{\sqrt{s_0/c_2}})$$

with $-\infty < x < \infty$ and $0 < c_2 < 1/5$,
K to be determined numerically,

VI 'Type VI', with density

$$f(x) = \frac{(x-\mu_1)^{p-1}(x-\mu_2)^{q-1}}{B(-(p+q-1),q)\ (\mu_2-\mu_1)^{p+q-1}},$$

with $0 < \mu_1 < \mu_2 \leq x$ or $x \leq \mu_2 < \mu_1 < 0$,
$q > 1$ and $p + q - 1 < 0$ $(\rightarrow p < 1)$.

The other types are called 'transition types'. For example type N is the normal distribution, type III the gamma distribution and type VII Student's distribution.

The Pearson system thus contains a large variety of densities. The parameters of the system can be estimated by (explicit) moment estimators (cf. Johnson & Kotz (1970)). There are also quantile estimators available, but as they are not explicit we prefer application of the moment estimators. Besides, as we will see in the next section, the quantile approach is more suitable for the Johnson system. Since both systems are much alike, this is also a reason not to apply these quantile estimators here. A disadvantage of the Pearson system is the complexity of type IV.

About the Pearson system we discuss again two things, the behavior of the estimators and the ability to approximate densities which are not from the system. We concentrate again on the particular situations.

Table 5.5.1 on page 107 and page 108 gives a comparison between the behavior of $f_\theta / f_{\hat{\theta}}$ and $f_\theta / \hat{f}_R$. We summarize the results as follows.

- To use the moment estimators, a rather large number of observations seems necessary. Failures in situations with 100 observations partly occur because sometimes s falls outside the estimated support of the β-distribution.

- Estimation of type IV is a problem. The large number of failures, even in situations with 1600 observations, is caused by the fact that the normalizing constant K of type IV can easily be something like 10^{40}, leading to numerical problems.
- Only in the situation with 1600 observations the moment estimators lead to better results than Rosenblatt's estimator.

Remark 5.5.1 The poor results with $\beta_2 = 15$ occur because the 8^{th} central moment is infinite and with this the variance of the sample kurtosis as well. Assumption (5.A.2) (page 95) is not satisfied. □

While we hoped for an improvement of Rosenblatt's estimator, especially for m not too large, the bad behavior of the parametric estimator when sampling from the family itself, is striking. Although we thus conclude that the Pearson system is not at all a competitor of Rosenblatt's estimator, we shortly discuss in what way the system approximates unknown densities. We consider standardized Exponential Power and Contaminated Normal Distributions. The density of a Contaminated Normal Distribution (*CND*) is given by

$$f(x) = (1-\tau)\frac{1}{\sigma_1}\phi\left(\frac{x-\mu_1}{\sigma_1}\right) + \tau\frac{1}{\sigma_2}\phi\left(\frac{x-\mu_2}{\sigma_2}\right), \quad \text{for } -\infty < x < \infty,$$

with $0 < \tau < 1$. If $\mu_1 = \mu_2$, the density is unimodal and symmetric. Any such contaminated normal distribution has heavier tails than the normal distribution. Table 5.5.2 provides some results. It shows that application of the Pearson system is not at all a guarantee for a good approximation to a density in a fixed point. Especially when s corresponds to the 0.99-quantile, the ratio's f/f_θ are not really close to 1.

5.6 Johnson system

The Johnson system is relatively new compared to the Pearson system. The first paper goes back to 1949 (Johnson). Like the Pearson system the Johnson system consists of types of distributions that cover the (β_1, β_2)-plane (β_1 is the squared skewness, β_2 is the kurtosis). The three types of distributions of the Johnson system — U(nbounded), L(ognormal) and B(ounded) — flow from transformations of a normal random variable. For a detailed description we refer again to Johnson & Kotz (1970), here we will briefly point out the main characteristics of the system.

The distributions of a r.v. X from the Johnson system are given by the following relations

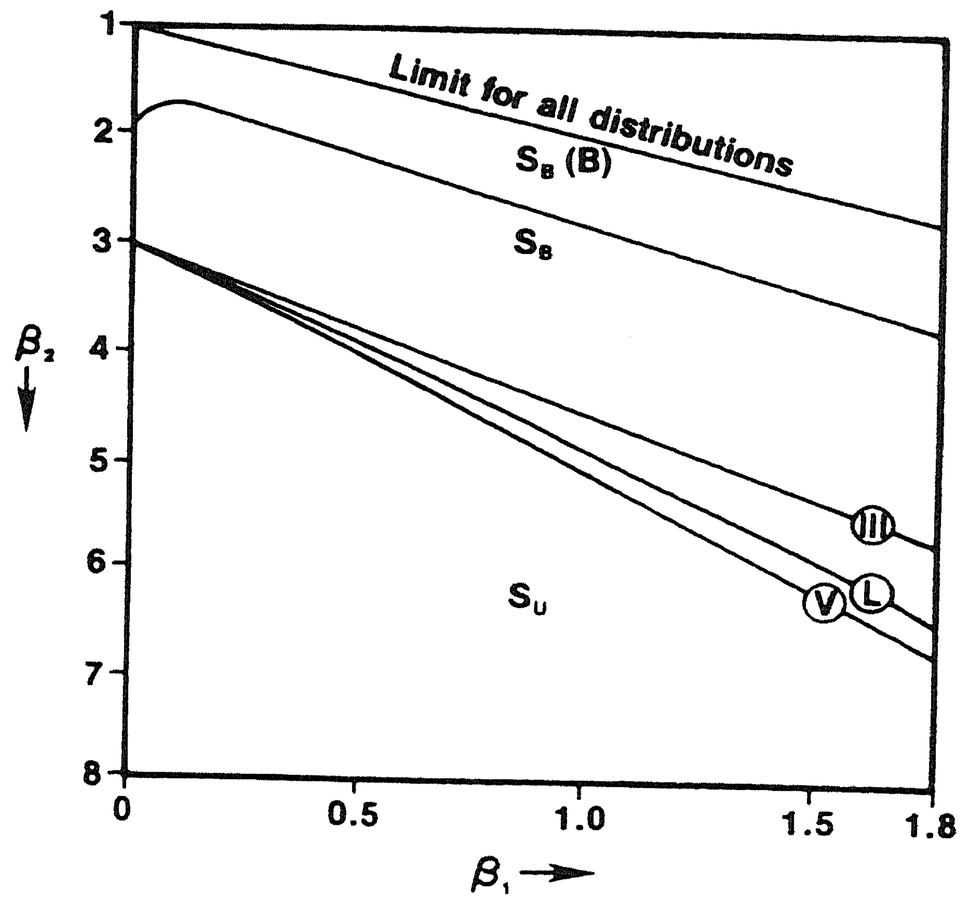
$$Z = \gamma + \delta f_i\left(\frac{X-\xi}{\lambda}\right), \qquad i = (B, L, U),$$

where

$$
\begin{aligned}
Z &\sim N(0,1) \\
f_B(Y) &= \log(Y/(1-Y)) \qquad 0 < Y < 1 \\
f_L(Y) &= \log(Y) \qquad Y > 0 \\
f_U(Y) &= \log(Y + \sqrt{Y^2+1}) \qquad -\infty < Y < \infty.
\end{aligned}
$$

Figure 5.6.1 shows the relation between (β_1, β_2) and the types of distributions.

Figure 5.6.1 Johnson system (figure from Stuart&Ord (1987))



$\beta_1 = (\text{skewness})^2$, $\beta_2 = \text{kurtosis}$

Because the system is a transformation model, estimation based on quantiles is obvious. However, only since 1980 explicit estimators are available. These estimators, derived by Slifker & Shapiro (1980), are based on four quantiles,

$$x_{(i)} = F^{-1}(\Phi(a_i \cdot z)), \quad \text{with} \quad a_4 = -a_1 = 3 \text{ and } a_3 = -a_2 = 1.$$

Slifker & Shapiro (1980) also provide a function of these quantiles which discriminates among the three types. For convenience the estimating equations are given in appendix A.3.

We have to choose a value of z to determine which quantiles we use for our estimates. The authors do not give a specific advice about its value. For a discussion we refer to their paper.

Table 5.6.1 (page 110 and page 111) is the analog of table 5.5.1. It provides a comparison between the use of the quantile estimators of the system and Rosenblatt's estimator. Out of several values the value of z which leads to the smallest mean squared error of $f_\theta / f_{\hat{\theta}}$ is selected. From the simulations we draw the following conclusion.

- Compared to the Pearson system, the estimators of the Johnson system perform much better but in the situation with 100 observations, still worse than Rosenblatt's estimator.

Remark 5.6.1 With respect to the choice of z, we advise to use the following rule of thumb. With 100 observations, take z equal to 0.7, with 400 observations or more, take z equal to 0.8. With the choice z=0.7 or z=0.8 the 0.9821-quantile or the 0.9918-quantile has to be estimated. These quantiles are just left and right of the specification limit if $\pi = 0.01$. This explains (partly) the behavior of the quantile estimators compared to Rosenblatt. □

Like the Pearson system, we tried to fit densities that are not from the system. Table 5.6.2 is the analog of table 5.5.2, we fitted densities of Contaminated Normal and Exponential Power Distributions by the Johnson system. Compared to the Pearson system we see a large improvement with respect to the exponential power densities, while the approximations to the contaminated normal densities remain poor.

5.7 Power transformation

Box&Cox (1964) introduced a parametric model by the following assumption about a observable r.v. X

$$Y = \begin{cases} \dfrac{(X+\lambda_2)^{\lambda_1} - 1}{\lambda_1} & \lambda_1 \neq 0 \\ \log(X+\lambda_2) & \lambda_1 = 0 \end{cases}$$

where Y is assumed to have a known type of distribution.

Originally Box&Cox discussed this model in the context of linear models. The classical theory about a normal homoscedastic linear model could be less restrictive if it were appropriate after a transformation of the X's. As we are interested in finding a parametric model that, among other things, gives improvements with respect to the application of the normal density, we take Y to be $N(\mu_Y, \sigma_Y^2)$ distributed. Hinkley (1975) considered the case where Y has an exponential or a gamma distribution.

If the ratio $\dfrac{\lambda_1 \sigma_Y}{\lambda_1 \mu_Y + 1}$ is positive and sufficiently small (smaller than 1/3), the probability that $\lambda_1 Y + 1$ is negative is sufficiently small and we use the

notation, ignoring that X is not well defined on this part of small probability. For the sake of convenience we say that X has a Box-Cox distribution (BCD).

The BCD is skewed to the right for $\lambda_1 < 1$, normal for $\lambda_1 = 1$ and skewed to the left for $\lambda_1 > 1$. This reveals immediately the main limitation of the model, the only symmetric density within the model is the normal density.

Like the Johnson system, the relation between the parameters and the moments is not explicit. For example, for small σ_Y the mean of X will be something like $(\lambda_1\mu_Y + 1)^{1/\lambda_1} - \lambda_2$ but the variance of X is not clear at all. Cox&Reid (1986) derived an approximately orthogonal and intuitively more appealing parameterization. They write,

$$\frac{(X+\lambda_2)^{\lambda_1} - 1}{\lambda_1} \sim N(\frac{\nu_1^{\lambda_1} - 1}{\lambda_1}, \nu_0\ \nu_1^{2\lambda_1 - 2}),$$

$$\text{with} \qquad \nu_1 = (\lambda_1\mu_Y + 1)^{1/\lambda_1}$$
$$\nu_0 = (\lambda_1\mu_Y + 1)^{2/\lambda_1} \left(\frac{\sigma_Y}{\lambda_1\mu_Y + 1}\right)^2.$$

With this parameterization the probability mass we neglect is $1 - \Phi\left(\frac{\nu_1}{\lambda_1\sqrt{\nu_0}}\right)$, hence $\lambda_1\sqrt{\nu_0}/\nu_1$ must be small.

Remark 5.7.1 The attraction of this parameterization can be shown by a simple Taylor expansion. We obtain

$$\begin{aligned} \mathrm{E}X &= -\lambda_2 + (\lambda_1\mu_Y + 1)^{1/\lambda_1}(1 + O\left(\left(\frac{\sigma_Y}{\lambda_1\mu_Y + 1}\right)^2\right)) = \\ &= -\lambda_2 + \nu_1\ (1 + O\left(\frac{\nu_0}{\nu_1^2}\right)) \end{aligned}$$

and

$$\begin{aligned} \mathrm{VAR}(X) &= (\lambda_1\mu_Y + 1)^{2/\lambda_1} \left(\frac{\sigma_Y}{\lambda_1\mu_Y + 1}\right)^2 (1 + O\left(\left(\frac{\sigma_Y}{\lambda_1\mu_Y + 1}\right)^2\right)) \\ &= \nu_0\ (1 + O\left(\frac{\nu_0}{\nu_1^2}\right)), \end{aligned}$$

as $\frac{\nu_0}{\nu_1^2} = \left(\frac{\lambda_1\sigma_Y}{\lambda_1\mu_Y + 1}\right)^2 \to 0.$

We see that ν_0 corresponds to the scale while ν_1 corresponds to the location of the distribution. The threshold parameter λ_2 corresponds to the location as well. Often λ_2 is omitted, but here we also see that without λ_2 the model of limited use for the present application. □

It is a problem to estimate the four parameters involved. About maximum likelihood estimation Hinkley (1975) remarks that it both involves a great deal of calculation and it can be sensitive to outliers. Together with our experience from maximum likelihood estimation of the parameters of the *EPD* (section 5.4), this was reason for us not to try to apply these estimators. For the case $\lambda_2 = 0$ Hinkley (1975) derives and discusses estimators of λ_1. The estimating equations become rather complicated if λ_2 does not necessarily equal 0, however. In appendix A.4 quantile estimators are derived for a fixed value $c = \lambda_1\sqrt{\nu_0}/\nu_1$. The estimators are based on the sample median and sample p- and $(1-p)$-quantiles.

Table 5.7.1 (page 113) gives ratio's f/f_θ, with f again Γ- and β-densities, for $p = 0.9$. The ratio's are all surprisingly close to one and, in contrast with fitting the normal density (section 5.3 and table 5.3.2)and table 5.3.3), the ratio's f/f_θ here are also for other choices of p very close to one. Figures of the densities show a striking resemblance between the Γ- and β-densities and the fitted BC-densities for various choices of p and over a long range of s (the exception is again the $\beta(2,2)$-density with its very finite support). If the parameters are estimated, the choice of the quantiles is more important. Simulations showed that $p = 0.90$ yields the best results. But from table 5.3.2 and table 5.3.3 it is also seen that for $\pi = 0.01$ with Rosenblatt's estimator much better results are obtained. Because estimation of the 0.90-quantile is involved, neither we expect an improvement for $\pi = 0.10$ (cf. (5.2.5)).

To conclude this section, the possible application of the BC-transformation model is certainly restricted to smooth and skewed densities, but even then there is no guarantee that it gives an improvement compared to Rosenblatt's estimator.

5.8 Conclusions

Having studied several parametric models we come to the following conclusion. In estimating a density at a fixed point, larger parametric models certainly give improvements for a great variety of densities, compared to the application of the normal density. But it turns out that it is very difficult to improve on Rosenblatt's estimator. Perhaps one of the most striking results is that, even with observations that are from the parametric family itself, for small sample sizes up to 400 Rosenblatt's estimator often gives better results.

Although we did not explicitly investigate the relation between a test limit based on these parametric models and the corresponding consumer loss, in view of section 5.2 we conclude that it is very difficult to improve the test limit based on Rosenblatt's estimator.

5.A Appendix

5.A.1 Regularity conditions on the parametric estimator

Given a sample $X_1, \ldots, X_m$ from density f_θ, we assume that an estimator $\hat{\theta}$ converges to $\theta = (\theta_1, \ldots, \theta_k)$ such that, as $m \to \infty$,

$$\mathrm{E}(\hat{\theta}_i - \theta_i) = \frac{\zeta_i}{m} + O(m^{-2}), \text{ for some constant } \zeta_i, \quad i = 1, \ldots, k, \tag{5.A.1}$$

$$\mathrm{E}(\hat{\theta}_i - \theta_i)(\hat{\theta}_j - \theta_j) = \frac{\eta_{ij}}{m} + O(m^{-2}), \quad \text{for some constant } \eta_{ij}, \quad i, j = 1, \ldots, k \tag{5.A.2}$$

$$\mathrm{E} \prod_{i=i_1}^{i_j} (\hat{\theta}_i - \theta_i) = O(m^{-2}), \quad j = 3, 4 \quad i_1, \ldots, i_j = 1, \ldots, k, \tag{5.A.3}$$

there exist positive constants $c_{\theta_i} = c_{\theta_i}(\epsilon)$ such that
$$\mathrm{P}(|\hat{\theta}_i - \theta_i| > \epsilon) = O(\mathrm{e}^{-c_{\theta_i} m}), \quad i = 1, \ldots, k. \tag{5.A.4}$$

We write $f_{\hat{\theta}}$ for the parametric density estimator. For estimators of parameters which are based on sample moments or sample quantiles and for maximum likelihood estimators (5.A.1)-(5.A.4) usually hold.

With a sample $X_1, \ldots, X_m$ from an unknown density f which does not belong to the parametric family, $f_{\hat{\theta}}$ will usually not converge to f. Therefore we write, if $f_{\hat{\theta}} > 0$,

$$\begin{aligned}\left(\frac{f}{f_{\hat{\theta}}} - 1\right)^2 = \left(\frac{f}{f_\theta} - 1\right)^2 - 2\frac{f}{f_\theta}\left(\frac{f}{f_\theta} - 1\right) \cdot \left(\frac{f_{\hat{\theta}}}{f_\theta} - 1\right) \\ + \frac{f}{f_\theta}\left(3\frac{f}{f_\theta} - 2\right) \cdot \left(\frac{f_{\hat{\theta}}}{f_\theta} - 1\right)^2 + \ldots\end{aligned} \tag{5.A.5}$$

Although we assumed that $f_\theta(s) > 0$, $f_{\hat{\theta}}(s)$ is not necessarilly positive. If density f_θ has finite support and s falls outside the estimated support we obtain $f_{\hat{\theta}}(s) = 0$, for example. We cannot determine the probablity of $f_{\hat{\theta}}(s)$ being 0 in general. Moreover, it makes no practical difference whether the estimate equals 0 exactly or whether the estimate is 10^{-6}. Therefore, we introduce the set

$$B = \left\{|\hat{\theta} - \theta| < \epsilon\right\},$$

with the idea that on the set B, where $\hat{\theta}$ is close to θ, $f_{\hat{\theta}}(s)$ is also close to $f_\theta(s)$. We make this idea more precise by the assumption that we can write

$$\frac{f_{\hat{\theta}}}{f_\theta} - 1 = (\hat{\theta} - \theta)^T D_1 + (\hat{\theta} - \theta)^T D_2 (\hat{\theta} - \theta) + R_\theta, \tag{5.A.6}$$

where

$$(D_1)_i = \frac{1}{f_\theta}\frac{\partial f_\theta(s)}{\partial \theta_i},$$

$$(D_2)_{i,j} = \frac{1}{f_\theta}\frac{\partial^2 f_\theta(s)}{\partial \theta_i \partial \theta_j},$$

$(D_1)_i$ and $(D_2)_{i,j}$ both finite,

and where the remainder term R_θ is such that

$$\mathrm{E}R_\theta^2 = O(m^{-2})$$

and

$$\mathrm{P}(|R_\theta| > \epsilon) = O(\mathrm{e}^{-c_r m}), \text{ for some positive constant } c_r = c_r(\epsilon).$$

This assumption now implies that

$$\begin{aligned}
&\mathrm{E}\left(\frac{f}{f_{\hat\theta}} - 1\right)^2 \cdot 1_B \\
&= \left(\frac{f}{f_\theta} - 1\right)^2 - 2\frac{f}{f_\theta}\left(\frac{f}{f_\theta} - 1\right)\mathrm{E}\left(\frac{f_{\hat\theta}}{f_\theta} - 1\right)\cdot 1_B + \\
&\quad \frac{f}{f_\theta}\left(3\frac{f}{f_\theta} - 2\right)\mathrm{E}\left(\frac{f_{\hat\theta}}{f_\theta} - 1\right)^2 \cdot 1_B + O(m^{-2}) \\
&= \left(\frac{f}{f_\theta} - 1\right)^2 + O(m^{-1}).
\end{aligned} \tag{5.A.7}$$

5.A.2 Maximum likelihood estimation in the exponential power distribution

The log-likelihood function of the exponential power distribution is given by

$$\begin{aligned}
&L(\mu, \sigma, \beta; X_1, \ldots, X_m) \\
&= -m\,(\,\ln 2 + \ln\sigma + \ln\beta + \ln\Gamma(\beta)\,) - \sum_{i=1}^{m} \left|\frac{X_i - \mu}{\sigma}\right|^{1/\beta}.
\end{aligned}$$

In order to obtain the maximum likelihood estimators of σ and β we simply put the corresponding partial derivatives equal to 0,

$$\frac{\partial L}{\partial \sigma} = 0 \implies \sigma = \left(\frac{1}{m\beta}\sum_{i=1}^{m}|X_i - \mu|^{1/\beta}\right)^{\beta},$$

$$\frac{\partial L}{\partial \beta} = 0 \implies -m(1 + \beta\,\psi(\beta)) + \frac{1}{\beta}\sum_{i=1}^{m}\left|\frac{X_i - \mu}{\sigma}\right|^{1/\beta}\ln\left|\frac{X_i - \mu}{\sigma}\right| = 0,$$

where $\psi(\beta) = \dfrac{\mathrm{d}\ \ln\Gamma(\beta)}{\mathrm{d}\ \beta}$.

With respect to the parameter μ the things are a little more complicated. We have that if $\mu \neq X_i$ $(i = 1, \ldots, m)$, the second derivative of L is negative for $\beta < 1$, 0 for $\beta = 1$ and positive for $\beta > 1$. In the case $\beta > 1$ we have that

$$\lim_{\mu \uparrow X_i} \frac{\partial L}{\partial \mu} = \infty, \quad i = 1, \ldots, m\ ,$$

$$\lim_{\mu \downarrow X_i} \frac{\partial L}{\partial \mu} = -\infty, \quad i = 1, \ldots, m.$$

Because L is continuous in μ, we conclude that if $\beta > 1$ the maximum likelihood estimator of μ must be equal to one of the observations. In the case $\beta < 1$, L is continuously differentiable (with respect to μ). If $\beta = 1$, the maximum likelihood estimator of μ is equal to the sample median.

For the three parameters we now have a system of three equations from which we iteratively determine the estimates.

5.A.3 Estimators of the Johnson system

The density related to the Johnson-system is given by

$$f_X(x) = \frac{\delta}{\lambda} f_i'\left(\frac{x-\xi}{\lambda}\right) \cdot \phi(\gamma + \delta f_i\left(\frac{x-\xi}{\lambda}\right)), \quad i = (B, L, U).$$

Slifker&Shapiro (1980) presented the following estimators of the four parameters.

Define

$$x_{(i)} = F^{-1}(\Phi(a_i \cdot z)), \text{ with } a_1 = -a_4 = -3, \text{ and } a_2 = -a_3 = -1,$$

and

$$\begin{aligned} m &= x_{(4)} - x_{(3)}, \\ n &= x_{(2)} - x_{(1)}, \\ p &= x_{(3)} - x_{(2)}, \\ q &= x_{(2)} + x_{(3)}. \end{aligned}$$

Then the U-type is characterized by

$$\frac{mn}{p^2} > 1,$$

and

$$\delta = \frac{2z}{\operatorname{arccosh}\left[\frac{1}{2}\left(\frac{m}{p} + \frac{n}{p}\right)\right]},$$

$$\gamma = \delta \cdot \operatorname{arcsinh}\left[\frac{\frac{n}{p} - \frac{m}{p}}{2\sqrt{\frac{n}{p}\frac{m}{p} - 1}}\right],$$

$$\lambda = \frac{2p\sqrt{\frac{n}{p}\frac{m}{p} - 1}}{\left(\frac{m}{p} + \frac{n}{p} - 2\right)\sqrt{\frac{m}{p} + \frac{n}{p} + 2}},$$

$$\xi = \frac{q}{2} + \frac{p\left(\frac{n}{p} - \frac{m}{p}\right)}{2\left(\frac{m}{p} + \frac{n}{p} - 2\right)}.$$

The B-type is characterized by

$$\frac{mn}{p^2} < 1,$$

and

$$\delta = \frac{z}{\operatorname{arccosh}\left[\frac{1}{2}\sqrt{\left(1 + \frac{p}{n}\right)\left(1 + \frac{p}{m}\right)}\right]},$$

$$\gamma = \delta \cdot \operatorname{arcsinh}\left[\frac{\left(\frac{p}{n} - \frac{p}{m}\right)\sqrt{\left(1 + \frac{p}{n}\right)\left(1 + \frac{p}{m}\right) - 4}}{2\left(\frac{p}{m}\frac{p}{n} - 1\right)}\right],$$

$$\lambda = \frac{p\sqrt{\left\{\left(1 + \frac{p}{n}\right)\left(1 + \frac{p}{m}\right) - 2\right\}^2 - 4}}{\frac{p}{m}\frac{p}{n} - 1},$$

$$\xi = \frac{q}{2} - \frac{\lambda}{2} + \frac{p\left(\frac{p}{n} - \frac{p}{m}\right)}{2\left(\frac{p}{m}\frac{p}{n} - 1\right)}.$$

For the L-type,

$$\frac{mn}{p^2} = 1.$$

Hence, based on sample quantiles we will (almost) never choose the L-type as sampled distribution.

5.A.4 Estimators of the Box-Cox model

We will derive approximate but accurate estimating equations, based on quantiles, by fixing the ratio $\lambda_1\sqrt{\nu_0}/\nu_1$. The p-quantile is given by

$$x_{(p)} = \nu_1 \left(\lambda_1 \frac{\sqrt{\nu_0}}{\nu_1}\Phi^{-1}(p) + 1\right)^{1/\lambda_1} - \lambda_2.$$

By fixing $c = \dfrac{\lambda_1\sqrt{\nu_0}}{\nu_1}$ we have that X is not well defined on a part with probability mass $1 - \Phi(1/c)$. For the p-quantile we now write

$$x_{(p)} = \sqrt{\nu_0}\left(\Phi^{-1}(p) + \frac{\lambda_1}{c} + \frac{1}{2}c\ \Phi^{-1}(p)^2\left(\frac{1}{\lambda_1} - 1\right) + O\left(c^2\right)\right),$$

as $c \to 0$.

With this Taylor expansion we obtain the following (approximate) estimating equations

$$\text{median} \;=\; \frac{\lambda_1\sqrt{\nu_0}}{c} - \lambda_2,$$

$$\begin{aligned} x_{(p)} - \text{median} \;\doteq\; & \sqrt{\nu_0}\Phi^{-1}(p)\left\{1 + \frac{1}{2}\left(\frac{1}{\lambda_1} - 1\right)c\ \Phi^{-1}(p) + \right. \\ & \left. + \frac{1}{6}\left(\frac{1}{\lambda_1} - 1\right)\left(\frac{1}{\lambda_1} - 2\right)\left(c\ \Phi^{-1}(p)\right)^2\right\} \end{aligned}$$

$$x_{(p)} - x_{(1-p)} \;\doteq\; 2\sqrt{\nu_0}\Phi^{-1}(p)\left\{1 + \frac{1}{6}\left(\frac{1}{\lambda_1} - 1\right)\left(\frac{1}{\lambda_1} - 2\right)\left(c\ \Phi^{-1}(p)\right)^2\right\}.$$

These approximations make sense if λ_1 is not too close to 0 and for quantiles not too far in the tail. By dividing the second by the third equation we obtain a quadratic expression for λ_1 and with λ_1, we also obtain explicit (approximate) expressions for ν_0 and λ_2.

$$\lambda_1 \;\doteq\; \frac{-B - \sqrt{B^2 - 4AC}}{2A},$$

with

$$\begin{aligned} A &= -\frac{1}{2}c\ \Phi^{-1}(p)\ Q + (1 + \frac{1}{3}(c\ \Phi^{-1}(p))^2)(Q - 2) \\ B &= \frac{1}{2}c\ \Phi^{-1}(p)\ Q - \frac{1}{2}(c\ \Phi^{-1}(p))^2(Q - 2) \\ C &= \frac{1}{6}(c\ \Phi^{-1}(p))^2(Q - 2), \end{aligned}$$

and

$$Q = \frac{x_{(p)} - x_{(1-p)}}{x_{(p)} - med}.$$

We checked the accuracy of the estimating equations numerically. It appears that even with $c = 1/3$, for any $p \in [0.10, 0.90]$ the p-quantile according to the parameters derived by the estimating equations practically equals the true p-quantile.

Table 5.3.2. β-densities, at s (0.99-quantile), estimated in three ways

Shown are simulated bias $\mathrm{E}(f/\phi_{\hat{\theta}} - f/\phi_\theta)$ and the square root of the simulated mean squared error $\mathrm{E}(f/\phi_{\hat{\theta}} - 1)^2$ based on 1000 replications, in comparison with mean and square root of the mean squared error corresponding to the Rosenblatt estimator. The parameters of the normal density are both estimated by the maximum likelihood estimators and by means of the sample median in combination with the sample 0.975-quantile.

		Rosenblatt		Maximum likelihood estimators of μ, σ			Quantiles estimators of μ, σ		
$\beta(2,2)$ $(s = 1.973)$									
m	p_0	$\mathrm{E}\, f/\hat{f}_R$	$\sqrt{\mathrm{MSE}}$	f/ϕ	bias	$\sqrt{\mathrm{MSE}}$	f/ϕ_θ	bias	$\sqrt{\mathrm{MSE}}$
100	0.000	0.948	0.328	1.305	0.102	0.579	1.671	0.031	0.882
400	0.000	0.972	0.173	1.305	0.022	0.369	1.671	0.012	0.726
1600	0.000	1.008	0.102	1.305	0.005	0.321	1.671	0.002	0.683
6400	0.000	1.026	0.065	1.305	0.001	0.309	1.671	0.001	0.674
$\beta(8,32)$ $(s = 2.633)$									
m	p_0	$\mathrm{E}f/\hat{f}_R$	$\sqrt{\mathrm{MSE}}$	f/ϕ	bias	$\sqrt{\mathrm{MSE}}$	f/ϕ_θ	bias	$\sqrt{\mathrm{MSE}}$
100	0.003	0.851	0.567	1.664	0.681	2.554	0.980	0.259	1.710
400	0.000	0.853	0.296	1.664	0.133	0.982	0.980	0.062	0.416
1600	0.000	0.893	0.179	1.664	0.030	0.737	0.980	0.016	0.178
6400	0.000	0.930	0.111	1.664	0.006	0.681	0.980	0.004	0.087
$\beta(2,8)$ $(s = 2.853)$									
m	p_0	$\mathrm{E}f/\hat{f}_R$	$\sqrt{\mathrm{MSE}}$	f/ϕ	bias	$\sqrt{\mathrm{MSE}}$	f/ϕ_θ	bias	$\sqrt{\mathrm{MSE}}$
100	0.002	0.869	0.567	2.837	2.294	10.60	1.003	0.314	2.360
400	0.000	0.865	0.287	2.837	0.343	2.570	1.003	0.064	0.423
1600	0.000	0.901	0.173	2.837	0.090	2.009	1.003	0.013	0.178
6400	0.000	0.935	0.106	2.837	0.021	1.879	1.003	0.003	0.087

Table 5.3.3. Γ-densities, at s (0.99-quantile), estimated in three ways

Shown are simulated bias $\mathrm{E}(f/\phi_{\hat{\theta}} - f/\phi_\theta)$ and the square root of the simulated mean squared error $\mathrm{E}(f/\phi_{\hat{\theta}} - 1)^2$ based on 1000 replications, in comparison with mean and square root of the mean squared error corresponding to the Rosenblatt estimator. The parameters of the normal density are both estimated by the maximum likelihood estimators and by means of the sample median in combination with the sample 0.975-quantile.

		Rosenblatt		Maximum likelihood estimators of μ, σ			Quantiles estimators of μ, σ		
$\Gamma(2)$ $(s = 3.280)$									
m	p_0	$\mathrm{E}f/\hat{f}_R$	$\sqrt{\mathrm{MSE}}$	f/ϕ	bias	$\sqrt{\mathrm{MSE}}$	f/ϕ_θ	bias	$\sqrt{\mathrm{MSE}}$
100	0.005	0.834	0.579	6.676	25.44	95.27(5)	0.977	1.022	9.999
400	0.000	0.848	0.321	6.676	2.741	12.27	0.977	0.146	0.753
1600	0.000	0.889	0.190	6.676	0.560	6.725	0.977	0.028	0.264
6400	0.000	0.927	0.117	6.676	0.148	5.935	0.977	0.009	0.125
$\Gamma(6)$ $(s = 2.902)$									
m	p_0	$\mathrm{E}f/\hat{f}_R$	$\sqrt{\mathrm{MSE}}$	f/ϕ	bias	$\sqrt{\mathrm{MSE}}$	f/ϕ_θ	bias	$\sqrt{\mathrm{MSE}}$
100	0.005	0.859	0.588	2.707	2.762	13.31	0.962	0.484	3.586
400	0.000	0.860	0.315	2.707	0.406	2.554	0.962	0.096	0.526
1600	0.000	0.896	0.185	2.707	0.085	1.888	0.962	0.022	0.218
6400	0.000	0.931	0.114	2.707	0.022	1.752	0.962	0.005	0.106
$\Gamma(32)$ $(s = 2.582)$									
m	p_0	$\mathrm{E}f/\hat{f}_R$	$\sqrt{\mathrm{MSE}}$	f/ϕ	bias	$\sqrt{\mathrm{MSE}}$	f/ϕ_θ	bias	$\sqrt{\mathrm{MSE}}$
100	0.003	0.845	0.568	1.464	0.493	1.856	0.972	0.258	1.755
400	0.000	0.850	0.300	1.464	0.102	0.727	0.972	0.061	0.413
1600	0.000	0.890	0.182	1.464	0.022	0.528	0.972	0.017	0.179
6400	0.000	0.928	0.112	1.464	0.006	0.480	0.972	0.002	0.090

Between brackets the number of 'failures', the number of times that either the parameters could not be determined or that the estimated density was less than 10^{-5}.

Table 5.4.1. Estimation of the *EP*-density at s (0.99-quantile) in two ways

Table shows simulated mean $\mathrm{E}(f_\theta/f_{\hat{\theta}})$ and the square root of the simulated mean squared error $\mathrm{E}(f_\theta/f_{\hat{\theta}}-1)^2$ (based on 1000 replications), when θ is estimated using sample mean, sample variance and sample kurtosis, in comparison with the mean and mean squared error from Rosenblatt's estimator.

		Rosenblatt		**Parametric estimator**	
$EPD(\frac{1}{4})$, $(s=1.913)$					
m	p_0	$\mathrm{E}f/\hat{f}_R$	$\sqrt{\mathrm{MSE}}$	$\mathrm{E}f_\theta/f_{\hat{\theta}}$	$\sqrt{\mathrm{MSE}}$
100	0.000	0.709	0.396	4.253	50.05 (130)
400	0.000	0.764	0.279	1.551	14.99 (12)
1600	0.000	0.831	0.193	1.011	0.098
6400	0.000	0.889	0.126	1.004	0.049
$EPD(\frac{1}{2})$, $(s=2.326)$					
m	p_0	$\mathrm{E}f/\hat{f}_R$	$\sqrt{\mathrm{MSE}}$	$\mathrm{E}f_\theta/f_{\hat{\theta}}$	$\sqrt{\mathrm{MSE}}$
100	0.001	0.809	0.522	3.817	42.04 (5)
400	0.000	0.828	0.285	1.060	0.215
1600	0.000	0.878	0.180	1.013	0.094
6400	0.000	0.921	0.113	1.004	0.045
$EPD(1)$, $(s=2.766)$					
m	p_0	$\mathrm{E}f/\hat{f}_R$	$\sqrt{\mathrm{MSE}}$	$\mathrm{E}f_\theta/f_{\hat{\theta}}$	$\sqrt{\mathrm{MSE}}$
100	0.016	0.925	0.625	1.864	9.312
400	0.000	0.918	0.385	1.065	0.226
1600	0.000	0.924	0.196	1.017	0.097
6400	0.000	0.947	0.116	1.004	0.046
$EPD(2)$, $(s=3.107)$					
m	p_0	$\mathrm{E}f/\hat{f}_R$	$\sqrt{\mathrm{MSE}}$	$\mathrm{E}f_\theta/f_{\hat{\theta}}$	$\sqrt{\mathrm{MSE}}$
100	0.068	0.894	0.549	1.788	5.241
400	0.001	1.023	0.600	1.044	0.286
1600	0.000	0.964	0.243	0.997	0.109
6400	0.000	0.966	0.134	0.990	0.059

Between brackets the number of 'failures': the number of times the estimates of the parameters could not be determined or the estimated density was less than 10^{-5}.

Table 5.4.2. Maximum likelihood estimation of the density of the *EPD*

The table shows simulated mean and the square root of the simulated mean squared error of $f_\theta/f_{\hat{\theta}}$ at s (the 0.99 quantile), based on 1000 replications.

$EPD(\frac{1}{4})$	m	$Ef_\theta/f_{\hat{\theta}}$	$\sqrt{\text{MSE}}$
	100	1.621	7.040 (70)
	400	1.060	0.253
	1600	1.010	0.099

$EPD(\frac{1}{2})$	m	$Ef_\theta/f_{\hat{\theta}}$	$\sqrt{\text{MSE}}$
	100	1.737	4.885
	400	1.046	0.240
	1600	1.012	0.102

$EPD(1)$	m	$Ef_\theta/f_{\hat{\theta}}$	$\sqrt{\text{MSE}}$
	100	1.291	1.315 (1)
	400	1.047	0.205 (1)
	1600	1.013	0.089

$EPD(2)$	m	$Ef_\theta/f_{\hat{\theta}}$	$\sqrt{\text{MSE}}$
	100	1.146	0.595 (18)
	400	1.037	0.181 (1)
	1600	1.010	0.082

Between brackets the number of failures, that is the number of times that either the estimated density is less than 10^{-5} or the estimates could not be determined

Table 5.4.3. β-densities, at s (0.99-quantile), estimated by the normal- and *EP*-density

Estimation of ϕ_θ based on maximum likelihood estimators (derived in the normal family), estimation of the *EP*-density based on sample mean, variance and kurtosis. The table gives the simulated bias $\mathrm{E}(f/f_{\hat{\theta}} - f/f_\theta)$ and the square root of the simulated mean squared error $\mathrm{E}(f/f_{\hat{\theta}} - 1)^2$ based on 1000 replications.

		Normal density			*EPD*		
$\beta(2,2)$	m	f/ϕ_θ	bias	$\sqrt{\mathrm{MSE}}$	f/f_θ	bias	$\sqrt{\mathrm{MSE}}$
(s= 1.973)	100	1.305	0.102	0.579	1.341	3.775	93.59(11)
	400	1.305	0.022	0.369	1.341	0.087	0.580
	1600	1.305	0.005	0.321	1.341	0.019	0.397
	6400	1.305	0.001	0.309	1.341	0.005	0.356
$\beta(8,32)$	m	f/ϕ	bias	$\sqrt{\mathrm{MSE}}$	f/f_θ	bias	$\sqrt{\mathrm{MSE}}$
(s= 2.633)	100	1.664	0.681	2.554	1.527	7.046	60.94(3)
	400	1.664	0.133	0.982	1.527	0.305	1.231
	1600	1.664	0.030	0.737	1.527	0.063	0.651
	6400	1.664	0.006	0.681	1.527	0.016	0.557
$\beta(2,8)$	m	f/ϕ	bias	$\sqrt{\mathrm{MSE}}$	f/f_θ	bias	$\sqrt{\mathrm{MSE}}$
(s= 2.853)	100	2.837	2.294	10.60	2.068	13.427	88.21(11)
	400	2.837	0.343	2.570	2.068	0.484	2.241
	1600	2.837	0.090	2.009	2.068	0.090	1.226
	6400	2.837	0.021	1.879	2.068	0.022	1.105

Between brackets the number of 'failures', the number of times that either the parameters could not be determined or that the estimated density was less than 10^{-5}.

Table 5.4.4. Γ-densities, at s (0.99-quantile), estimated by the normal- and *EP*-density

Estimation of ϕ_θ based on maximum likelihood estimators (derived in the normal family), estimation of the *EP*-density based on sample mean, variance and kurtosis. The table gives the simulated bias $\mathrm{E}(f/f_{\hat{\theta}} - f/f_\theta)$ and the square root of the simulated mean squared error $\mathrm{E}(f/f_{\hat{\theta}} - 1)^2$ based on 1000 replications.

		Normal density			*EPD*		
Γ(2) (s= 3.280)	m	f/ϕ	bias	$\sqrt{\text{MSE}}$	f/f_θ	bias	$\sqrt{\text{MSE}}$
	100	6.676	25.44	95.27(5)	1.797	11.861	65.49(16)
	400	6.676	2.741	12.27	1.797	0.537	2.106
	1600	6.676	0.560	6.725	1.797	0.108	0.968
	6400	6.676	0.148	5.935	1.797	0.029	0.841
Γ(6) (s= 2.902)	m	f/ϕ	bias	$\sqrt{\text{MSE}}$	f/f_θ	bias	$\sqrt{\text{MSE}}$
	100	2.707	2.762	13.31	1.625	12.148	77.94(11)
	400	2.707	0.406	2.554	1.625	0.493	1.781
	1600	2.707	0.085	1.888	1.625	0.094	0.795
	6400	2.707	0.022	1.752	1.625	0.024	0.665
Γ(32) (s= 2.582)	m	f/ϕ	bias	$\sqrt{\text{MSE}}$	f/f_θ	bias	$\sqrt{\text{MSE}}$
	100	1.464	0.493	1.856	1.349	4.605	55.84(2)
	400	1.464	0.102	0.727	1.349	0.213	0.845
	1600	1.464	0.022	0.528	1.349	0.044	0.448
	6400	1.464	0.006	0.480	1.349	0.013	0.375

Between brackets the number of 'failures', the number of times that either the parameters could not be determined or that the estimated density was less than 10^{-5}.

Table 5.5.1. Densities from Pearson system, at 0.99-quantile, estimated in two ways

The table shows simulated mean $E(f_\theta / f_{\hat{\theta}})$ and the square root of the simulated mean squared error $E(f_\theta / f_{\hat{\theta}} - 1)^2$ based on 1000 replications, in comparison with the mean and mean squared error from Rosenblatt's estimator.

		Rosenblatt			Pearson		
(β_1, β_2)= (0,2.5)	m	p_0	$Ef/\hat{f}_R$	$\sqrt{\text{MSE}}$	$Ef_\theta/f_{\hat{\theta}}$	$\sqrt{\text{MSE}}$	
	100	0.000	0.814	0.434	6.700	88.321	(124)
	400	0.000	0.842	0.246	1.116	0.567	(1)
	1600	0.000	0.890	0.157	1.016	0.120	
(β_1, β_2)= (0,3)	m	p_0	$Ef/\hat{f}_R$	$\sqrt{\text{MSE}}$	$Ef_\theta/f_{\hat{\theta}}$	$\sqrt{\text{MSE}}$	
	100	0.001	0.808	0.523	5.143	43.697	(123)
	400	0.000	0.828	0.285	1.116	0.915	(89)
	1600	0.000	0.878	0.180	1.019	0.130	(101)
(β_1, β_2)= (0,6)	m	p_0	$Ef/\hat{f}_R$	$\sqrt{\text{MSE}}$	$Ef_\theta/f_{\hat{\theta}}$	$\sqrt{\text{MSE}}$	
	100	0.012	0.862	0.605	5.441	43.698	(61)
	400	0.000	0.876	0.370	1.431	8.937	(2)
	1600	0.000	0.900	0.201	1.031	0.176	
(β_1, β_2)= (0,15)	m	p_0	$Ef_\theta/\hat{f}_R$	$\sqrt{\text{MSE}}$	$Ef_\theta/f_{\hat{\theta}}$	$\sqrt{\text{MSE}}$	
	100	0.023	0.872	0.609	9.638	73.251	(59)
	400	0.000	0.905	0.425	1.241	0.919	(6)
	1600	0.000	0.913	0.211	1.298	8.421	(5)

Between brackets the number of failures, that is the number of times that either the estimated density is less than 10^{-5} or the estimates of the parameters could not be determined.

Table 5.5.1. (Continued) Densities from Pearson system, at 0.99-quantile, estimated in two ways

(β_1,β_2)=		Rosenblatt			Pearson		
(0.5,3)	m	p_0	$Ef/\hat{f}_R$	$\sqrt{\text{MSE}}$	$Ef_\theta/f_{\hat{\theta}}$	$\sqrt{\text{MSE}}$	
	100	0.001	0.883	0.538	3.424	30.423	(107)
	400	0.000	0.877	0.263	1.083	0.327	(1)
	1600	0.000	0.910	0.159	1.014	0.118	

(β_1,β_2)=							
(0.5,4)	m	p_0	$Ef/\hat{f}_R$	$\sqrt{\text{MSE}}$	$Ef_\theta/f_{\hat{\theta}}$	$\sqrt{\text{MSE}}$	
	100	0.006	0.859	0.593	5.704	42.318	(139)
	400	0.000	0.861	0.323	1.079	0.411	(196)
	1600	0.000	0.896	0.188	1.014	0.123	(250)

(β_1,β_2)=							
(0.5,6)	m	p_0	$Ef/\hat{f}_R$	$\sqrt{\text{MSE}}$	$Ef_\theta/f_{\hat{\theta}}$	$\sqrt{\text{MSE}}$	
	100	0.013	0.866	0.607	2.913	10.446	(80)
	400	0.000	0.881	0.375	1.103	0.415	(13)
	1600	0.000	0.903	0.201	1.010	0.146	

Between brackets the number of failures, that is the number of times that either the estimated density is less than 10^{-5} or the estimates of the parameters could not be determined.

Table 5.5.2. Pearson system applied to fit Contaminated Normal and Exponential Power densities, at their 0.90- and 0.99-quantiles

Tables gives ratios f/f_θ where f_θ is such that skewness and kurtosis of f_θ, from the Pearson system, correspond to the skewness and kurtosis of f.

$CND(\mu_2, \sigma_2, 0.9)$[1]

		$\pi_s = 0.10$		$\pi_s = 0.01$	
μ_2	σ_2	s	f/f_θ	s	f/f_θ
0.0	0.8	1.131	1.137	2.682	1.469
0.0	0.6	0.884	0.998	3.332	3.656
-0.2	0.8	1.306	0.764	2.828	1.160
-0.2	0.6	0.896	0.958	4.078	3.021

$EPD(\beta)$

	$\pi_s = 0.10$		$\pi_s = 0.01$	
β	s	f/f_θ	s	f/f_θ
0.25	1.068	0.763	1.291	2.016
1	0.805	1.273	1.138	2.766
2	0.525	1.228	0.819	3.107

[1] The contamination parameter τ is equal to 0.9 and the parameters μ_1 and σ_1 are such that the mean is equal to 0 and variance equal to 1

Table 5.6.1. Densities from Johnson-system, at 0.99-quantile, estimated in two ways

The table shows simulated mean $\mathrm{E}(f_\theta/f_{\hat\theta})$ and the square root of the simulated mean squared error $\mathrm{E}(f_\theta/f_{\hat\theta}-1)^2$ based on 1000 replications, in comparison with the mean and mean squared error from Rosenblatt's estimator.

$(\beta_1,\beta_2)=$		Rosenblatt			Johnson		
	m	p_0	$\mathrm{E}f/\hat f_R$	$\sqrt{\mathrm{MSE}}$	z	$\mathrm{E}f_\theta/f_{\hat\theta}$	$\sqrt{\mathrm{MSE}}$
(0, 2.5), B	100	0.0001	0.682	0.453	0.8	1.312	1.210(61)
	400	0.0000	0.792	0.273	0.9	1.035	0.397
	1600	0.0000	0.869	0.171	0.9	1.013	0.146
(0, 3), N	m	p_0	$\mathrm{E}f/\hat f_R$	$\sqrt{\mathrm{MSE}}$	z	$\mathrm{E}f_\theta/f_{\hat\theta}$	$\sqrt{\mathrm{MSE}}$
	100	0.0013	0.808	0.523	0.7	1.126	0.872(34)
	400	0.0000	0.828	0.285	0.8	1.059	0.370
	1600	0.0000	0.878	0.180	0.8	1.019	0.151
(0, 6), U	m	p_0	$\mathrm{E}f/\hat f_R$	$\sqrt{\mathrm{MSE}}$	z	$\mathrm{E}f_\theta/f_{\hat\theta}$	$\sqrt{\mathrm{MSE}}$
	100	0.0156	0.575	0.575	0.7	1.093	0.886(12)
	400	0.0000	0.790	0.388	0.8	1.065	0.426
	1600	0.0000	0.875	0.214	0.8	1.016	0.175
(0, 10), U	m	p_0	$\mathrm{E}f/\hat f_R$	$\sqrt{\mathrm{MSE}}$	z	$\mathrm{E}f_\theta/f_{\hat\theta}$	$\sqrt{\mathrm{MSE}}$
	100	0.0296	0.535	0.587	0.7	1.096	0.947(10)
	400	0.0001	0.808	0.434	0.8	1.041	0.402
	1600	0.0000	0.885	0.224	0.8	1.012	0.179

Between brackets the number of failures, that is the number of times that the estimated density is less than 1/10th of the true density.

Table 5.6.1. (Continued) Densities from Johnson-system, at 0.99-quantile, estimated in two ways

		Rosenblatt			Johnson		
(0.49, 3), B	m	p_0	$Ef/\hat{f}_R$	$\sqrt{\text{MSE}}$	z	$Ef_\theta/f_{\hat{\theta}}$	$\sqrt{\text{MSE}}$
	100	0.0009	0.735	0.502	0.8	1.241	0.985(48)
	400	0.0000	0.834	0.272	0.9	1.011	0.279
	1600	0.0000	0.900	0.162	0.9	1.013	0.139
(0.49, 4), U	m	p_0	$Ef/\hat{f}_R$	$\sqrt{\text{MSE}}$	z	$Ef_\theta/f_{\hat{\theta}}$	$\sqrt{\text{MSE}}$
	100	0.0056	0.619	0.558	0.7	1.105	0.860(28)
	400	0.0000	0.782	0.343	0.8	1.043	0.368 (1)
	1600	0.0000	0.868	0.201	0.9	1.000	0.147
(0.49, 6), U	m	p_0	$Ef/\hat{f}_R$	$\sqrt{\text{MSE}}$	z	$Ef_\theta/f_{\hat{\theta}}$	$\sqrt{\text{MSE}}$
	100	0.0142	0.577	0.574	0.7	1.146	0.905(20)
	400	0.0000	0.788	0.384	0.8	1.042	0.412
	1600	0.0000	0.873	0.213	0.8	1.005	0.171

Between brackets the number of failures, that is the number of times that the estimated density is less than 1/10th of the true density.

Table 5.6.2. Johson system applied to fit Contaminated Normal and Exponential Power densities, at their 0.90- and 0.99-quantiles

The densities of the Johnson system are such that the $F(\Phi^{-1}(a_i \cdot z))$-quantiles ($a = (-3, -1, 1, 3)$) correspond with those of the unknown density.

$CND(\mu_2, \sigma_2, 0.9)$[1]

			$\pi = 0.10$		$\pi = 0.01$	
z	μ_2	σ_2	s	f/f_θ	s	f/f_θ
0.7	0.0	0.8	1.131	1.133	2.682	1.529
	0.0	0.6	0.884	1.268	3.332	2.788
	-0.2	0.8	1.306	0.856	2.828	1.095
	-0.2	0.6	0.895	1.092	4.078	2.569
0.8	0.0	0.8	1.131	1.185	2.682	1.268
	0.0	0.6	0.884	1.338	3.332	2.399
	-0.2	0.8	1.306	0.835	2.828	1.250
	-0.2	0.6	0.896	1.098	4.078	3.075

$EPD(\beta)$

		$\pi = 0.10$		$\pi = 0.01$	
z	β	s	f/f_θ	s	f/f_θ
0.7	0.25	1.335	1.017	2.016	0.809
	0.50	1.281	1.000	2.326	1.000
	1.00	1.138	1.032	2.766	1.168
	2.00	0.818	1.088	3.107	1.204
0.8	0.25	1.335	1.044	2.016	0.761
	0.50	1.281	1.000	2.326	1.000
	1.00	1.138	0.992	2.766	1.249
	2.00	0.818	1.026	3.107	1.317

[1] The contamination parameter τ is equal to 0.9 and the parameters μ_1 and σ_1 are such that the mean is equal to 0 and variance equal to 1

Table 5.7.1. Box-Cox model applied to estimate β- and Γ densities, at the 0.9- and 0.99-quantiles

The table gives the simulated bias $\mathrm{E}(f/f_{\hat{\theta}} - f/f_{\theta})$ and the square root of the simulated mean squared error $\mathrm{E}(f/f_{\hat{\theta}} - 1)^2$, based on 1000 replications. The estimators of the parameters are based on the 0.1-, 0.5- and 0.9-quantile with $c = \frac{1}{3}$.

		$\pi = 0.10$			$\pi = 0.01$		
$\beta(2,2)$	m	f/f_θ	bias	$\sqrt{\text{MSE}}$	f/f_θ	bias	$\sqrt{\text{MSE}}$
	100	1.228	0.053	0.386	1.113	0.162	0.636 (28)
	400	1.228	0.006	0.304	1.113	0.023	0.192
	1600	1.228	0.002	0.285	1.113	0.009	0.134
$\beta(8,32)$	m	f/ϕ	bias	$\sqrt{\text{MSE}}$	f/f_θ	bias	$\sqrt{\text{MSE}}$
	100	1.004	0.040	0.185	0.990	1.064	5.300 (9)
	400	1.004	0.008	0.088	0.990	0.099	0.469
	1600	1.004	0.004	0.042	0.990	0.025	0.171
$\beta(2,8)$	m	f/f_θ	bias	$\sqrt{\text{MSE}}$	f/f_θ	bias	$\sqrt{\text{MSE}}$
	100	1.061	0.037	0.202	0.937	0.550	2.784
	400	1.061	0.011	0.109	0.937	0.046	0.296
	1600	1.061	0.001	0.076	0.937	0.013	0.136
$\Gamma(2)$	m	f/f_θ	bias	$\sqrt{\text{MSE}}$	f/f_θ	bias	$\sqrt{\text{MSE}}$
	100	0.997	0.040	0.199	1.008	0.849	3.409 (4)
	400	0.997	0.007	0.083	1.008	0.092	0.441
	1600	0.997	0.002	0.042	1.008	0.027	0.171
$\Gamma(6)$	m	f/f_θ	bias	$\sqrt{\text{MSE}}$	f/f_θ	bias	$\sqrt{\text{MSE}}$
	100	0.983	0.040	0.179	1.032	1.834	9.718 (8)
	400	0.983	0.008	0.086	1.032	0.104	0.515
	1600	0.983	0.004	0.044	1.032	0.035	0.201
$\Gamma(32)$	m	f/f_θ	bias	$\sqrt{\text{MSE}}$	f/f_θ	bias	$\sqrt{\text{MSE}}$
	100	0.987	0.034	0.176	1.020	1.850	11.19 (9)
	400	0.987	0.011	0.087	1.020	0.115	0.592
	1600	0.987	0.001	0.044	1.020	0.029	0.193

Between brackets the number of 'failures': the number of times the estimates of the parameters could not be determined or the estimated density was less than 10^{-5}.

Chapter 6

Nonnormal measurement error

In this final chapter we consider the situation in which a test limit has to be determined when the measurement error does not have a normal distribution. In case of a $N(0, \sigma^2)$ distributed measurement error (chapter 3 and chapter 4) the test limit is based on an estimator of the standard deviation σ. It will turn out that information on the standard deviation of the measurement error will not be sufficient to determine an accurate test limit in a more general situation. While σ can simply be estimated by means of repeated measurements on the characteristic, in this chapter we will have to assume that we have observations on the measurement error itself. These observations are usually obtained by comparison of the standard measurement results for a number of products to corresponding precise measurements. Once we have observations on the measurement error, the assumption that its mean is equal to 0 (there is no systematic measurement error) is no longer necessary.

6.1 Preliminaries.

The measurement error U has mean μ and variance σ^2. We assume that

$$V = -\frac{U - \mu}{\sigma}, \tag{6.1.1}$$

with mean 0 and variance 1, has a continuous density, denoted by g, and we suppose that

(A1) $\mathrm{E}|V|^r < \infty$, for some $r > 6$.

In analogy to the functions g_k (cf. (3.2.5)) we define

$$h_k(a) = \mathrm{E}(V - a)^k \cdot 1_{\{V > a\}} = \int_a^\infty (v - a)^k g(v)\, \mathrm{d}v \quad (k = 0, 1, 2), \tag{6.1.2}$$

where 1_A denotes the indicator function of the set A.
Elementary calculations show that

$$h_k'(a) = -k \cdot h_{k-1}(a) \quad (k = 1, 2). \tag{6.1.3}$$

The value of the inspected characteristic is denoted by X and its measurement by $\tilde{X} = X + U$. We assume that X and U are independent. The distribution function and density of X are denoted by F_X and f_X, respectively. Likewise, $F_{\tilde{X}}$ and $f_{\tilde{X}}$ are the distribution function and density, respectively, of $\tilde{X}$. With respect to the densities f_X and $f_{\tilde{X}}$ we assume

$$\text{(A2)} \quad \begin{cases} f_{\tilde{X}}'' \text{ and } f_X'' \text{ are bounded.} \\ f_{\tilde{X}}(s+\mu) = f_X(s) + O(\sigma^2), \quad f_{\tilde{X}}'(s+\mu) = f_X'(s) + O(\sigma), \end{cases}$$

where s denotes the specification limit. Further we will tacitly assume that $f_X(s) > 0$.

The consumer loss, with test limit t, is defined by

$$CL(t) = P(X > s, X + U < t).$$

Let

$$a = \frac{s + \mu - t}{\sigma} > 0 \tag{6.1.4}$$

analogous to (3.2.1). We have the following result for the consumer loss in terms of a.

Lemma 6.1.1 *Assume* (A1) *and* (A2). *Then*

$$CL(t) = \{\sigma f_{\tilde{X}}(s+\mu)h_1(a) + \frac{1}{2}\sigma^2 f_{\tilde{X}}'(s+\mu)h_2(a)\}(1 + O(\sigma^2)), \tag{6.1.5}$$

as $\sigma \to 0$.

Proof: We write

$$\begin{aligned} &P(X > s, X + U < t) \\ &\quad = P(X > s, X < s + \sigma(V - a)) = \\ &\quad = \int_a^\infty \{F_X(s + \sigma(v-a)) - F_X(s)\} g(v) \mathrm{d}v \\ &\quad = \int_a^\infty \Big\{ \sigma f_X(s)(v-a) + \frac{1}{2}\sigma^2 f_X'(s)(v-a)^2 \\ &\qquad\qquad + \frac{1}{6}\sigma^3 (v-a)^3 f_X''(\xi_v) \Big\} g(v) \mathrm{d}v, \end{aligned}$$

with ξ_v between s and $s + \sigma(v - a)$. Since $\int_a^\infty (v-a)^3 g(v)\mathrm{d}v$ is finite under (A1), and because f_X'' is bounded, we get $CL(t) = \{\sigma f_X(s)h_1(a) + \frac{1}{2}\sigma^2 f_X'(s)h_2(a)\}(1 + O(\sigma^2))$, which is under (A2) equal to the right-hand side of (6.1.5). □

6.2 Consumer loss under incorrect assumption of normality

We will discuss whether a test limit based on the normality assumption of the measurement error can be applied in nonnormal situations. To be able to apply lemma 6.1.1, we assume that (A1) and (A2) are satisfied.

Suppose, in order to get a consumer loss equal to γ, a test limit $t_o = s - a_o\sigma$ is obtained under the assumption of a normally distributed measurement error with mean μ and variance σ^2. By lemma 6.1.1 we have in this situation $\gamma = \sigma f_{\tilde{X}}(s+\mu)g_1(a_o)\{1+O(\sigma)\}$, with $g_1(a) = \mathrm{E}(Y-a)\cdot 1_{\{Y>a\}}$ with $Y \sim N(0,1)$. If U is not normally distributed we have $CL(t_o) = \sigma f_{\tilde{X}}(s+\mu)h_1(a_o)\{1+O(\sigma)\}$ and

$$\frac{CL(t_o)}{\gamma} = \frac{h_1(a_o)}{g_1(a_o)}\{1+O(\sigma)\} \quad (\text{ as } \sigma \to 0).$$

The following examples show that the ratio $h_1(a)/g_1(a)$ can differ much from one, even for symmetric and unimodal densities g of V and choices of a between 0 and 3.

Example 6.2.1 In this example it is shown that there are symmetric unimodal densities g for which $h_1(a)$ is arbitrarily small (for any $a > 0$).
Consider densities of the form

$$g_\epsilon(x) = \begin{cases} c + \frac{1}{\epsilon}\left(\frac{1}{2} - bc\right) & \text{if } 0 \le |x| < \epsilon \\ c & \text{if } \epsilon \le |x| \le b \\ 0 & \text{if } |x| > b \end{cases},$$

with $0 < \epsilon < 1$.
We have that $\int x^2 g_\epsilon(x)\,\mathrm{d}x = 1$ if $c = \dfrac{3-\epsilon^2}{2b(b^2-\epsilon^2)}$ and with this choice of c, the density is well defined if $\epsilon < \sqrt{3} < b$.
Now, for every $a > \epsilon$, $\int_a^\infty (x-a)g_\epsilon(x)\,\mathrm{d}x \to 0$ as $b \to \infty$. □

Example 6.2.2 Consider g_ϵ from the previous example.
If we take $b = 3a$ $(> \sqrt{3})$ we have

$$\lim_{\epsilon \downarrow 0} \int_a^\infty (x-a)g_\epsilon(x)\,\mathrm{d}x = \frac{1}{9a}$$

For $a = 1, 2, 3$, the ratio $\dfrac{1/(9a)}{g_1(a)}$ is 1.33, 6.54 and 96.6, respectively. □

Example 6.2.3 In table 6.2.1 below the ratio $h_1(a)/g_1(a)$ is computed if V has an exponential power distribution or a symmetric β distribution. The location and scale parameter of the distributions are such that the mean is 0 and the variance is equal to 1. For example, $EPD(1/8)$ stands for the exponential power distribution with shape parameter 1/8. (cf. section 5.4). The density of the β-distribution is given on page 49.

Table 6.2.1 Ratio $h_1(a)/g_1(a)$ for β and exponential power distributions.

a	$\beta(2,2)$	$\beta(8,8)$	$\beta(32,32)$	$EPD(1/8)$	$EPD(1/4)$	$EPD(1)$	$EPD(2)$
0	1.05	1.01	1.00	1.08	1.05	0.88	0.69
1	0.99	1.00	1.00	0.93	0.96	1.03	0.96
2	0.08	0.71	0.93	0.01	0.22	2.46	3.88
3	< 0.01	0.11	0.67	< 0.01	< 0.01	13.3	40.9

□

From these three examples we conclude that a test limit based on the assumption of normality of the measurement error indeed can lead to severe violation of the bound γ if normality fails. In view of the introduction of this chapter we conclude also that information on the standard deviation of the measurement error only is not sufficient to determine an accurate test limit.

6.3 Second order unbiased test limits in the nonnormal situation

By $CL(\hat{t})$ we denote the probability that a new product is both nonconforming and accepted, for a given test limit $\hat{t}$. In this section we are looking for a test limit t_u such that

$$\mathrm{E}CL(\hat{t}_u) = \gamma \tag{6.3.1}$$

to sufficient precision.

6.3.1 Definition of the test limit

We introduce

$$d = s - t > 0, \tag{6.3.2}$$

and define, for $k = 0, 1, 2$,

$$r_k(d) \;=\; \mathrm{E}(-U-d)^k \cdot 1_{\{-U-d>0\}}. \tag{6.3.3}$$

The relation between r_k and h_k (cf. (6.1.2)) is

$$\sigma^k h_k\left(\frac{d+\mu}{\sigma}\right) = r_k(d).$$

In terms of the parameter d and the functions r_k the consumer loss $CL(t)$ is then given by (cf. (6.1.5))

$$\begin{aligned} & P(X > s,\ X + U < s - d) \\ &= \{f_{\tilde{X}}(s+\mu)r_1(d) + \frac{1}{2}f'_{\tilde{X}}(s+\mu)r_2(d)\}(1+O(\sigma^2)), \end{aligned} \tag{6.3.4}$$

as $\sigma \to 0$.

Analogous to the functions h_k (cf. (6.1.3)) we have (for $k = 1, 2$)

$$r'_k(d) \ = \ -k \cdot r_{k-1}(d). \tag{6.3.5}$$

As a first order approximation to t_u we define

$$d_1 = r_1^{-1}\left(\frac{\gamma}{f_{\tilde{X}}(s+\mu)}\right), \tag{6.3.6}$$

similar to (3.2.9) and (4.2.7). Note that d_1 is uniquely defined since $r'_1(d) \leq 0$ (cf. (6.3.5)), with equality holding only if $r_1(d) = 0$.

In chapter 3 (cf. (3.2.20) and (3.4.5)) and chapter 4 (cf. (4.2.19)) we have seen that the test limit is of the form $t = s - (a_1 + c + c_u)\sigma$. The main term $a_1\sigma$ is similar to d_1, the correction c is used to correct for the second order term in the expansion of the consumer loss (cf. (6.1.5)) and c_u is used to correct for the bias due to estimation of unknown parameters.

The test limit t_u which we are looking for in the present situation will then be typically of the form

$$t_u = s - (d_1 + c + c_u), \tag{6.3.7}$$

with the correction terms c and c_u such that (6.3.1) is obtained.

To compute the test limit we will have to estimate the functions r_k. Given observations $U_1, \ldots, U_n$ from U we define

$$\hat{r}_k(d) = \frac{1}{n}\sum_{i=1}^{n}(-U_i - d)^k \cdot 1_{\{-U_i - d > 0\}}. \tag{6.3.8}$$

6.3.2 Second order unbiased test limits if $f_{\tilde{X}}$ and μ are known

While the notation in terms of the parameter d and functions $r_k(d)$ is adequate to compute the test limit, for the analysis of the consumer loss we prefer the notation in terms of the parameter a and the functions $h_k(a)$ (cf. (6.1.2) and (6.1.4)) since in this way the scale parameter σ is more visible. This makes it more easy to compare the results from this chapter with the results from the previous chapters.

In this section we assume that $f_{\tilde{X}}$ and μ are known.

Let

$$a_1 \ = \ h_1^{-1}\left(\frac{\gamma}{\sigma f_{\tilde{X}}(s+\mu)}\right) \ = \ \frac{d_1 + \mu}{\sigma} \tag{6.3.9}$$

and, given observations $U_1, \ldots, U_n$ from U, let

$$\hat{a}_1 \;=\; \hat{h}_1^{-1}\left(\frac{\gamma}{\sigma f_{\tilde{X}}(s+\mu)}\right) \;=\; \frac{\hat{d}_1+\mu}{\sigma}, \tag{6.3.10}$$

with

$$\hat{h}_k(a) = \frac{1}{n}\sum_{i=1}^{n}(V_i - a)^k \cdot 1_{\{V_i > a\}}, \quad (\text{with } V_i = -(U_i-\mu)/\sigma) \tag{6.3.11}$$

for $k = 0, 1, 2$.
We assume that a_1 is bounded as $\sigma, \gamma \to 0$.

Because of expansions later, we introduce the set

$$\begin{aligned} A \;=\; \Bigg\{ |\hat{a}_1 - a_1| \le \epsilon_1,\; \left|\frac{\hat{h}_2(\hat{a}_1)}{\hat{h}_1'(\hat{a}_1)} - \frac{h_2(a_1)}{h_1'(a_1)}\right| \le \epsilon_2, \\ \left|\frac{1+\hat{h}_1'(\hat{a}_1)}{(\hat{h}_1'(\hat{a}_1))^2} - \frac{1+h_1'(a_1)}{(h_1'(a_1))^2}\right| \le \epsilon_3 \Bigg\}, \end{aligned} \tag{6.3.12}$$

for sufficiently small constants $\epsilon_1, \epsilon_2, \epsilon_3 > 0$.
We will study the consumer loss on the set A. Since we have that

$$E CL(\hat{t}) = E CL(\hat{t}) \cdot 1_A + O(P(A^c)), \tag{6.3.13}$$

the following lemma shows the error caused by this restriction.

Lemma 6.3.1 *Assume (A1) and assume that $g(a_1) > 0$ with a_1 as in (6.3.9). Then, with $\hat{a}_1$ as in (6.3.10) and $\hat{h}_k$ as in (6.3.11),*

$$P(A^c) = O(n^{-r/4}).$$

Proof. The set A is the intersection of the three parts B_1, B_2 and B_3, with $B_1 = \{|\hat{a}_1 - a_1| \le \epsilon_1\}$ etc. We have that

$$P(A^c) \le P(B_1^c) + P(B_1 \cap B_2^c) + P(B_1 \cap B_3^c). \tag{6.3.14}$$

First we consider $P(B_1^c)$.
Since $\hat{h}_1$ is nonincreasing we have

$$x \le y \;\Leftrightarrow\; \hat{h}_1(x) \ge \hat{h}_1(y). \tag{6.3.15}$$

Consequently, using that $\hat{h}_1(\hat{a}_1) = h_1(a_1)$,

$$\begin{aligned} &\hat{a}_1 - a_1 \ge \epsilon_1 \\ &\quad\Rightarrow h_1(a_1+\epsilon_1) - h_1(a_1+\epsilon_1) \;\ge\; h_1(a_1) - h_1(a_1+\epsilon_1) \\ &\qquad\qquad \ge\; -h_1'(a_1+\epsilon_1) \;>\; 0 \\ &\hat{a}_1 - a_1 \le -\epsilon_1 \\ &\quad\Rightarrow \hat{h}_1(a_1-\epsilon_1) - h_1(a_1-\epsilon_1) \;\le\; h_1(a_1) - h_1(a_1-\epsilon_1) \\ &\qquad\qquad \le\; h_1'(a_1) \;<\; 0. \end{aligned}$$

Under (A1), $\mathrm{E}\,|\hat{h}_1(a) - h_1(a)|^r = O(n^{-r/2})$. Hence,

$$P(|\hat{a}_1 - a_1| \geq \epsilon_1) = O(n^{-r/2}).$$

With respect to $P(B_1 \cap B_2^c)$ we have that if $|\hat{a}_1 - a_1| \leq \epsilon_1$ then $\hat{h}_2(a_1 + \epsilon_1) \leq \hat{h}_2(\hat{a}_1) \leq \hat{h}_2(a_1 - \epsilon_1)$. Hence, for some small $\eta > 0$,

$$\begin{aligned} & P\left(|\hat{a}_1 - a_1| \geq \epsilon_1,\ \hat{h}_2(\hat{a}_1) > (1+\eta)\ h_2(a_1 - \epsilon_1)\right) \\ & \quad \leq\ P\left(\hat{h}_2(a_1 - \epsilon_1) > (1+\eta)\ h_2(a_1 - \epsilon_1)\right) \\ & \quad =\ O(n^{-r/4}). \end{aligned}$$

Analogously,

$$\begin{aligned} & P\left(|\hat{a}_1 - a_1| \leq \epsilon_1,\ \hat{h}_2(\hat{a}_1) < (1-\eta)\ h_2(a_1 + \epsilon_1)\right) \\ & \quad \leq\ P\left(\hat{h}_2(a_1 + \epsilon_1) < (1-\eta)\ h_2(a_1 + \epsilon_1)\right) \\ & \quad =\ O(n^{-r/4}). \end{aligned}$$

Since $\hat{h}_1'(a)$ is a sum of bounded r.v.'s, probabilities like

$$P\left(\hat{h}_1'(a_1 - \epsilon_1) > (1-\eta)\ h_1'(a_1 - \epsilon_1)\right)$$

are even exponentially small. Hence the lemma follows. □

Remark 6.3.1 The assumption $r > 6$ in (A1) is needed to prove asymptotic results further in this section. A remainder term in (6.3.13) which is just of order $n^{-r/4}$ for some $r > 6$, however, does not guarantee that the error we make is indeed sufficiently small for finite sample sizes. Since the consumer loss is very small, it is supposed to range from 1 ppm to 100 ppm, say, large values of n or a large value of r is really needed.

If instead of assumption (A1) we impose

(A1′) $\mathrm{E}e^{tV^2} < \infty$, for $0 < t \leq t_0$ for some $t_0 > 0$,

we obtain $P(A^c) = O\left(\mathrm{e}^{-\epsilon n}\right)$, for some $\epsilon > 0$. The error we make by studying the consumer loss on the set A only is then exponentially small, just as in the situation of a normally distributed measurement error (cf. chapter 3 and chapter 4). □

To find the correction term to $\hat{a}_1$ (actually to $\hat{d}_1$, of course, cf. (6.3.6) and (6.3.10)) in order to obtain a consumer loss which is in expectation equal to γ, to sufficient precision, we write

$$h_1(\hat{a}_1) \ =\ h_1(a_1) + (\hat{a}_1 - a_1)h_1'(a_1) + \tfrac{1}{2}(\hat{a}_1 - a_1)^2 h_1''(a_1) + \ldots\ , \qquad (6.3.16)$$

and we will approximate $\mathrm{E}(\hat{a}_1 - a_1)$ and $\mathrm{E}(\hat{a}_1 - a_1)^2$.

From (6.3.15) it follows that

$$\sqrt{n}(\hat{a}_1 - a_1) \leq y \;\Leftrightarrow\; \hat{h}_1(a_1 + y/\sqrt{n}) \leq h_1(a_1). \tag{6.3.17}$$

As h_1 is a sum of independent random variables, we can approximate its distribution by means of an Edgeworth expansion.
We define the following

$$\begin{aligned} H_n(y) &= P((\hat{a}_1 - a_1)\sqrt{n} \leq y) \;=\; P(\hat{h}_1(a_1 + y/\sqrt{n}) \leq h_1(a_1)) \\ \mu_n(y) &= h_1(a_1 + y/\sqrt{n}) \\ \sigma_n^2(y) &= h_2(a_1 + y/\sqrt{n}) - h_1^2(a_1 + y/\sqrt{n}) \\ \rho_n(y) &= \mathrm{E}\{(V - (a_1 + y/\sqrt{n})) \cdot 1_{\{V > a_1 + y/\sqrt{n}\}} - \mu_n(y)\}^3/\sigma_n^3(y) \\ z_n(y) &= \frac{h_1(a_1) - \mu_n(y)}{\sigma_n(y)}\sqrt{n}. \end{aligned} \tag{6.3.18}$$

Lemma 6.3.2 *With H_n and z_n as in (6.3.18), uniformly for $|y| \leq \sqrt{n}$ we have*

$$\{1 + |z_n(y)|^3\}\; |H_n(y) - H_n^*(y)| = o(n^{-1/2}),$$

with

$$H_n^*(y) = \Phi(z_n(y)) - \frac{\rho_n(y)}{6}\phi(z_n(y))\{z_n^2(y) - 1\}n^{-1/2}. \tag{6.3.19}$$

Proof. Direct application of Theorem 20.6 in Bhattacharya and Rao (1976), taking, in their notation, $s = 3$ and

$$f(y) = \{1 + |x|^3\} \cdot 1_{(-\infty, x]}(y) \qquad \text{if } x < 0$$

and

$$f(y) = \{1 + |x|^3\} \cdot 1_{[x, \infty)}(y) \qquad \text{if } x \geq 0,$$

respectively. □

Lemma 6.3.3 *With H_n^* from (6.3.19) and ϵ_1 as in (6.3.12), under the assumptions of lemma 6.3.1,*

$$\begin{aligned} \mathrm{E}\sqrt{n}(\hat{a}_1 - a_1) \cdot 1_A &= \int_0^{\epsilon_1\sqrt{n}} 1 - H_n^*(y) - H_n^*(-y)\; \mathrm{d}y + o(n^{-1/2}) \\ \mathrm{E}\; n(\hat{a}_1 - a_1)^2 \cdot 1_A &= \int_0^{\epsilon_1\sqrt{n}} 2y\{1 - H_n^*(y) + H_n^*(-y)\}\mathrm{d}y + o(n^{-1/2}), \end{aligned}$$

as $n \to \infty$.

Proof. For the set A (cf. (6.3.12)) we write $A = \{B_1 \cap B_{23}\}$, with $B_1 = \{|\hat{a}_1 - a_1| \leq \epsilon_1\}$ and $B_{23} = B_2 \cap B_3$, with B_2 and B_3 as in the proof of lemma 6.3.1. Then

$$\begin{aligned} \mathrm{E}\, n(\hat{a}_1 - a_1)^2 \cdot 1_A &= \mathrm{E}\, n(\hat{a}_1 - a_1)^2 \cdot 1_{B_1}(1 - 1_{B_{23}^c}) \\ &= \mathrm{E}\, n(\hat{a}_1 - a_1)^2 \cdot 1_{\{|\hat{a}_1 - a_1| \leq \epsilon_1\}} \\ &\quad - \mathrm{E}\, n(\hat{a}_1 - a_1)^2 \cdot 1_{B_1} 1_{B_{23}^c}. \end{aligned} \tag{6.3.20}$$

From the proof of lemma 6.3.1 it follows that $\mathrm{E}\, 1_{B_1} 1_{B_{23}^c} = O(n^{-r/4})$, hence that

$$\mathrm{E}\, n(\hat{a}_1 - a_1)^2 \cdot 1_{B_1} 1_{B_{23}^c} = O(n^{-r/4+1}),$$

which is $o(n^{-1/2})$ for $r > 6$. Likewise we find

$$\mathrm{E}\sqrt{n}(\hat{a}_1 - a_1) \cdot 1_{B_1} 1_{B_{23}^c} = O(n^{-r/4+1/2}).$$

We have that (cf. (6.3.20))

$$\mathrm{E}\, n(\hat{a}_1 - a_1)^2 \cdot 1_{\{|\hat{a}_1 - a_1| \leq \epsilon_1\}} = \int_{-\epsilon_1\sqrt{n}}^{\epsilon_1\sqrt{n}} y^2 \mathrm{d}H_n(y).$$

Partial integration yields

$$\begin{aligned} \int_{-\epsilon_1\sqrt{n}}^{\epsilon_1\sqrt{n}} y^2 \mathrm{d}H_n(y) &= -n\epsilon_1^2\{1 - H_n(\epsilon_1\sqrt{n}) + H_n(-\epsilon_1\sqrt{n})\} \\ &\quad + 2\int_0^{\epsilon_1\sqrt{n}} y\{1 - H_n(y) + H_n(-y)\}\mathrm{d}y. \end{aligned} \tag{6.3.21}$$

The first term on the right-hand side is sufficiently small. To see this we write

$$\begin{aligned} &n\{1 - H_n(\epsilon_1\sqrt{n}) + H_n(-\epsilon_1\sqrt{n})\} \\ &\quad = \epsilon_1^2 n\{1 - H_n^*(\epsilon_1\sqrt{n}) + H_n^*(-\epsilon_1\sqrt{n})\} + \\ &\qquad -\epsilon_1^2 n\{H_n(\epsilon_1\sqrt{n}) - H_n^*(\epsilon_1\sqrt{n})\} - \epsilon_1^2 n\{H_n(-\epsilon_1\sqrt{n}) - H_n^*(-\epsilon_1\sqrt{n})\}. \end{aligned}$$

Since $\sigma_n(y)$ (cf. (6.3.18)) is bounded for $|y| \leq \epsilon_1\sqrt{n}$ and for some $\xi \in (-\epsilon_1, \epsilon_1)$ also $\{h_1(a_1) - \mu_n(y)\}\sqrt{n} = -yh_1'(a_1 + \xi)$, there is some $\epsilon > 0$ such that

$$|z_n(y)| \geq \epsilon\,|y|, \tag{6.3.22}$$

for $|y| \leq \epsilon_1\sqrt{n}$. Hence, for some $\delta > 0$ we have that

$$\begin{aligned} \epsilon_1^2 n\{1 - H_n^*(\epsilon_1\sqrt{n})\} &= O(\mathrm{e}^{-\delta n}) \\ \epsilon_1^2 n\, H_n^*(-\epsilon_1\sqrt{n}) &= O(\mathrm{e}^{-\delta n}). \end{aligned}$$

By lemma 6.3.2 we also get

$$\begin{aligned} \epsilon_1^2 n\{H_n(\epsilon_1\sqrt{n}) - H_n^*(\epsilon_1\sqrt{n})\} &= o(n^{-1}) \\ \epsilon_1^2 n\{H_n(-\epsilon_1\sqrt{n}) - H_n^*(-\epsilon_1\sqrt{n})\} &= o(n^{-1}). \end{aligned}$$

Now, in the integral of (6.3.21) we may replace H_n by H_n^*, because by (6.3.22) we have

$$\begin{aligned}\int_{-\epsilon_1\sqrt{n}}^{\epsilon_1\sqrt{n}} |y|\ |H_n(y) - H_n^*(y)|\ \mathrm{d}y &\leq o(n^{-1/2}) \int_{-\epsilon_1\sqrt{n}}^{\epsilon_1\sqrt{n}} \frac{|y|}{1+\epsilon^3|y|^3}\mathrm{d}y \\ &= o(n^{-1/2}).\end{aligned}$$

The approximation to $\mathrm{E}(\hat{a}_1 - a_1)\sqrt{n} \cdot 1_{\{|\hat{a}_1 - a_1| \leq \epsilon_1\}}$ is proved in the same way.□

Lemma 6.3.4 *Assume $g'(a)$ is bounded on $|a_1 - a| \leq \epsilon_1$. With z_n, ρ_n from (6.3.18), uniformly in $|y| \leq \epsilon_1\sqrt{n}$ we have, as $n \to \infty$*

$$\begin{aligned}z_n(y) &= A_1 y + A_2 y^2/\sqrt{n} + O(|y|^3/n) \\ \rho_n(y) &= A_3 + O(|y|/\sqrt{n}),\end{aligned}$$

with

$$\begin{aligned}A_1 &= -h_1'(a_1)\ b \\ A_2 &= -\frac{1}{2}h_1''(a_1)\ b - h_1'(a_1)\{1 + h_1'(a_1)\}h_1(a_1)\ b^3 \\ A_3 &= \{h_3(a_1) - 3h_2(a_1)h_1(a_1) + 2h_1^3(a_1)\}\ b^3 \\ b &= \{h_2(a_1) - h_1^2(a_1)\}^{-1/2}.\end{aligned} \tag{6.3.23}$$

Proof. By Taylor expansion we have, uniformly in $|y| \leq \epsilon_1\sqrt{n}$,

$$\begin{aligned}\mu_n(y) &= h_1(a_1) + n^{-1/2}h_1'(a_1)y/\sqrt{n} + \frac{1}{2}h_1''(a_1)y^2/n + O(|y|^3/n^{3/2}) \\ \sigma_n^2(y) &= h_2(a_1) - h_1^2(a_1) + \{h_2'(a_1) - 2h_1'(a_1)h_1(a_1)\}y/\sqrt{n} + O(y^2/n),\end{aligned}$$

as $n \to \infty$. The results follows by noting that $h_2' = -2h_1$. □

Lemma 6.3.5 *If $y = o(n^{1/2})$, then for some $\tilde{A} \neq 0$, under the definitions (6.3.18) and (6.3.23)*

$$\begin{aligned}H_n^*(y) &= \Phi(A_1 y) + A_2\phi(A_1 y)y^2/\sqrt{n} - \frac{1}{6}\frac{A_3}{\sqrt{n}}\phi(A_1 y)\{(A_1 y)^2 - 1\}\ + \\ &\quad O\left((|y|^5 + 1)\phi(\tilde{A}y)/n\right),\end{aligned}$$

as $n \to \infty$.

Proof. In view of lemma 6.3.4 Taylor expansion of functions like $\Phi(A_1 y(1+x))$ around $x = 0$ yields the result. □

Lemma 6.3.6 *Under the assumptions of lemma 6.3.3–lemma 6.3.5*

$$\begin{aligned} \mathrm{E}\sqrt{n}(\hat{a}_1 - a_1)\cdot 1_A &= -\frac{A_2}{A_1^3}\frac{1}{\sqrt{n}} + o(n^{-1/2}) \\ \mathrm{E}\; n(\hat{a}_1 - a_1)^2 \cdot 1_A &= \frac{1}{A_1^2} + o(n^{-1/2}), \end{aligned} \tag{6.3.24}$$

as $n \to \infty$.

Proof. Combination of lemma 6.3.3–lemma 6.3.5 and direct calculation yields result. □

The result we have obtained, cf. (6.3.16),

$$\begin{aligned} \mathrm{E}h_1(\hat{a}_1) - h_1(a_1) &\approx -\frac{A_2}{nA_1^3}h_1'(a_1) + \frac{1}{2}\frac{1}{A_1^2\, n}h_1''(a_1) \\ &= -\frac{h_1(a_1)}{n}\frac{1 + h_1'(a_1)}{h_1'(a_1)}, \end{aligned} \tag{6.3.25}$$

indicates that the correction to a_1, to cancel out the first order bias term in the consumer loss (cf. (6.1.5)), should be equal to

$$\frac{h_1(a_1)}{n}\frac{1 + h_1'(a_1)}{(h_1'(a_1))^2}.$$

Indeed, in terms of d_1 etc. (cf. (6.3.7)), let

$$\hat{t}_u = s - (\hat{d}_1 + \hat{c} + \hat{c}_u) \tag{6.3.26}$$

with

$$\begin{aligned} \hat{d}_1 &= \hat{r}_1^{-1}\left(\frac{\gamma}{f_{\tilde{X}}(s+\mu)}\right) \\ \hat{c} &= \frac{1}{2}\frac{f_{\tilde{X}}'(s+\mu)}{f_{\tilde{X}}(s+\mu)}\frac{\hat{r}_2(\hat{d}_1)}{\hat{r}_0(\hat{d}_1)} \\ \hat{c}_u &= \frac{\hat{r}_1(\hat{d}_1)}{n}\frac{\{1 - \hat{r}_0(\hat{d}_1)\}}{\hat{r}_0^2(\hat{d}_1)}, \end{aligned} \tag{6.3.27}$$

then

Theorem 6.3.1 *Under* (A1) *and* (A2) *and the assumption that* $g(a_1) > 0$ *and that* $g'(a)$ *is bounded for* $|a_1 - a| \le \epsilon_1$, *we have with* $\hat{t}_u$ *as in (6.3.26)*

$$\mathrm{E}CL(\hat{t}_u) = \gamma + o(\gamma n^{-1}) + O(\gamma\sigma^2) + O(n^{-r/4}), \tag{6.3.28}$$

as $n \to \infty$ *and* $\sigma, \gamma \to 0$ *such that* a_1 *is bounded.*

Proof. In terms of the parameter a and the functions h_k we have

$$\hat{t}_u = s + \mu - \sigma \hat{a}_u,$$

with (cf. (6.1.2))

$$\begin{aligned}
\hat{a}_u &= \hat{a}_1 + \hat{c}^{(a)} + \hat{c}_u^{(a)} \\
\hat{a}_1 &= \hat{h}_1^{-1}\left(\frac{\gamma}{\sigma f_{\bar{X}}(s+\mu)}\right) \\
\hat{c}^{(a)} &= -\frac{\sigma}{2}\frac{f'_{\bar{X}}(s+\mu)}{f_{\bar{X}}(s+\mu)}\frac{\hat{h}_2(\hat{a}_1)}{\hat{h}'_1(\hat{a}_1)} \\
\hat{c}_u^{(a)} &= \frac{\hat{h}_1(\hat{a}_1)}{n}\frac{\{1+\hat{h}'_1(\hat{a}_1)\}}{(\hat{h}'_1(\hat{a}_1))^2}.
\end{aligned}$$

We restrict our attention to the set A (cf. (6.3.12)), at the cost of an order $n^{-r/4}$ in the remainder of $\mathrm{E}CL(\hat{t}_u)$.
We have, in view of the proof of lemma 6.1.1 and the definition of the set A (cf. (6.3.12)),

$$\mathrm{E}CL(\hat{t}_u)\ 1_A = \mathrm{E}\ 1_A\{\sigma f_{\bar{X}}(s+\mu)h_1(\hat{a}_u) + \frac{1}{2}\sigma^2 f'_{\bar{X}}(s+\mu)h_2(\hat{a}_u)\}(1+O(\sigma^2)).$$

Expansion of the first order term around $\hat{a}_u = a_1$ yields

$$\begin{aligned}
&\mathrm{E}\ 1_A\ \sigma f_{\bar{X}}(s+\mu)h_1(\hat{a}_u) \\
&= \mathrm{E}\ 1_A\Big\{\gamma + \sigma f_{\bar{X}}(s+\mu)\Big((\hat{a}_1-a_1)h'_1(a_1) + (\hat{c}^{(a)}+\hat{c}_u^{(a)})h'_1(a_1) \\
&\qquad + \frac{1}{2}(\hat{a}_1-a_1)^2 h''_1(a_1) + \{(\hat{a}_1-a_1)(\hat{c}^{(a)}+\hat{c}_u^{(a)}) \\
&\qquad + \frac{1}{2}(\hat{c}^{(a)}+\hat{c}_u^{(a)})^2\}h''_1(\xi) + \frac{1}{2}(\hat{a}_1-a_1)^2\{h''_1(\xi)-h''_1(a_1)\}\Big)\Big\},
\end{aligned}$$

with ξ between a_1 and $\hat{a}_1 + \hat{c}^{(a)} + \hat{c}_u^{(a)}$.
Writing $c^{(a)}$ and $c_u^{(a)}$ for the correction terms with a_1 and h_k instead of $\hat{a}_1$ and $\hat{h}_k$, we have that $c^{(a)} = O(\sigma)$ and $c_u^{(a)} = O(n^{-1})$ as $n \to \infty$, $\sigma \to 0$, and

$$\begin{aligned}
\mathrm{E}\ 1_A\ (\hat{a}_1-a_1)(\hat{c}^{(a)}+\hat{c}_u^{(a)}) &= O(\sigma n^{-1} + n^{-2}) \\
\mathrm{E}\ 1_A\ (\hat{c}^{(a)})^2 = O(\sigma^2),&\quad \mathrm{E}\ 1_A(\hat{c}_u^{(a)})^2 = O(n^{-2}) \\
\mathrm{E}\ 1_A\ (\hat{a}_1-a_1)^2\{h''_1(\xi)-h''_1(a_1)\} &= o(n^{-1}).
\end{aligned}$$

Further, using e.g. results on the oscillation modulus of the empirical process (cf. Mason et al. (1983); we omit the technical details)

$$\mathrm{E}\ 1_A\ \hat{c}^{(a)} = c^{(a)} + O(\sigma n^{-3/4}\log n),\quad \mathrm{E}\ 1_A\ \hat{c}_u^{(a)} = c_u^{(a)} + O(n^{-7/4}\log n).$$

Since moreover

$$\mathrm{E}\ 1_A \left\{ (\hat{a}_1 - a_1) h_1'(a_1) + c_u^{(a)} h_1'(a_1) + \frac{1}{2}(\hat{a}_1 - a_1)^2 h_1''(a_1) \right\} = o(n^{-1}),$$

we obtain, using that $\sigma f_{\tilde{X}}(s+\mu) h_1(a_1) = \gamma$ and $h_1(a_1) = O(1)$,

$$\begin{aligned} \mathrm{E}\ 1_A\ \sigma f_{\tilde{X}}(s+\mu) h_1(\hat{a}_u) &= \gamma - \frac{1}{2} f_{\tilde{X}}'(s+\mu)\sigma^2 h_2(a_1) \\ &\quad + o(\gamma n^{-1}) + O(\gamma\sigma^2) + O(\gamma n^{-r/4}). \end{aligned}$$

Finally, because

$$\mathrm{E}\ 1_A\ h_2(\hat{a}_u) = h_2(a_1) + O(\sigma + n^{-1}),$$

we get the result. □

6.3.3 Simulation results

To demonstrate theorem 6.3.1 for finite sample sizes, some simulations have been carried out. Based on 10000 replications, the average and standard deviation of the consumer loss obtained with test limit $\hat{t}_u$ (cf. (6.3.26)) is computed if observations on the measurement error are from a normal or a Γ-distribution. In order to see in what way the bias correction $\hat{c}_u$ (cf. (6.3.27)) contributes, we also computed the 'uncorrected' test limit,

$$\hat{t}_2 = s - \left(\hat{d}_1 + \frac{1}{2} \frac{f_{\tilde{X}}'(s+\mu)}{f_{\tilde{X}}(s+\mu)} \frac{\hat{r}_2(\hat{d}_1)}{\hat{r}_0(\hat{d}_1)} \right). \tag{6.3.29}$$

The results are summarized in the two tables on the next pages.

From these tables we conclude that, for the situations considered, the first order approximation to the bias of the consumer loss when using $\hat{t}_2$ (cf. (6.3.29), (6.1.5) and (6.3.25)) is quite accurate, and that the bias correction works well if $r_0(d_1) = P(U < -d_1)$ is not too small. For example, in table 6.3.1 with $\gamma = 20$ ppm, $P(X > s) = 0.15$ and $\sigma = 0.10$, d_1 is such that $P(U < -d_1) = 0.003$. Since by definition $\hat{r}_0(\hat{d}_1) \geq n^{-1}$, this probability and, consequently, the correction term c_u cannot be accurately estimated if $n = 25$. Vice versa, a test limit based on only one observation of the measurement error is not reliable.

Table 6.3.1 Test limit $\hat{t}_u$ with a $N(0, \sigma^2)$-distributed measurement error

The table shows simulated mean and standard deviation (between brackets) of $CL(\hat{t}_2)$ and of $CL(\hat{t}_u)$, based on 10000 replications. Moreover the first order approximation to the mean of $CL(\hat{t}_2)$ (following from (6.3.25)) is shown. The values of γ are in ppm, π denotes $P(X > s)$, the probability that a product is nonconforming, and n is the number of observations on the measurement error.
The characteristic is standard normally distributed.

	theory	*simulated*	
$(\gamma, \sigma, \pi) = (100, 0.01, 0.01)$, $r_0(d_1) = 0.481$			
n	$\mathrm{E}CL(\hat{t}_2)$	$\mathrm{E}CL(\hat{t}_2)$	$\mathrm{E}CL(\hat{t}_u)$
40	102.7	102.6(23.8)	100.0(23.8)
80	101.4	101.4(16.6)	100.0(16.6)
500	100.2	100.2(6.7)	100.0(6.7)
2000	100.1	100.1(3.3)	100.0(3.3)
$(\gamma, \sigma, \pi) = (100, 0.10, 0.01)$, $r_0(d_1) = 0.082$			
n	$\mathrm{E}CL(\hat{t}_2)$	$\mathrm{E}CL(\hat{t}_2)$	$\mathrm{E}CL(\hat{t}_u)$
40	127.9	123.8(73.8)	100.5(70.2)
80	113.9	112.2(49.4)	99.2(48.2)
500	102.2	101.7(19.0)	99.5(18.9)
2000	100.6	100.3(9.3)	99.8(9.3)
$(\gamma, \sigma, \pi) = (20, 0.01, 0.15)$, $r_0(d_1) = 0.023$			
n	$\mathrm{E}CL(\hat{t}_2)$	$\mathrm{E}CL(\hat{t}_2)$	$\mathrm{E}CL(\hat{t}_u)$
40	41.3	43.6(41.0)	32.4(36.8)
80	30.6	30.9(24.4)	23.5(22.6)
500	21.7	21.6(8.2)	20.0(8.0)
2000	20.4	20.4(4.0)	20.0(3.9)
$(\gamma, \sigma, \pi) = (20, 0.10, 0.15)$, $r_0(d_1) = 0.003$			
n	$\mathrm{E}CL(\hat{t}_2)$	$\mathrm{E}CL(\hat{t}_2)$	$\mathrm{E}CL(\hat{t}_u)$
40	194.9	243.4(271.3)	226.2(256.4)
80	107.4	121.9(136.5)	105.7(123.8)
500	34.0	33.8(28.8)	25.5(26.6)
2000	23.5	23.4(12.1)	20.4(11.8)

Table 6.3.2 Test limit $\hat{t}_u$ with a $\Gamma(8)$-distributed measurement error

This table summarizes the same simulation as in table 6.3.1, however with a measurement error which has a Γ-distribution with shape parameter 8 and with the location and scale parameter such that the mean is 0 and the variance σ^2.

	theory	*simulated*	
$(\gamma, \sigma, \pi) = (100, 0.01, 0.01)$, $r_0(d_1) = 0.533$			
n	$\mathrm{E}CL(\hat{t}_2)$	$\mathrm{E}CL(\hat{t}_2)$	$\mathrm{E}CL(\hat{t}_u)$
40	102.2	102.1(20.7)	99.9(20.5)
80	101.1	101.2(14.6)	100.1(14.6)
500	100.2	100.1(5.7)	99.9(5.7)
2000	100.0	100.0(2.9)	100.0(2.9)
$(\gamma, \sigma, \pi) = (100, 0.10, 0.01)$, $r_0(d_1) = 0.120$			
n	$\mathrm{E}CL(\hat{t}_2)$	$\mathrm{E}CL(\hat{t}_2)$	$\mathrm{E}CL(\hat{t}_u)$
40	118.2	117.2(60.6)	100.2(57.3)
80	109.1	108.0(39.8)	99.2(38.4)
500	101.5	101.6(15.3)	100.2(15.2)
2000	100.4	100.3(7.6)	99.9(7.6)
$(\gamma, \sigma, \pi) = (20, 0.01, 0.15)$, $r_0(d_1) = 0.040$			
n	$\mathrm{E}CL(\hat{t}_2)$	$\mathrm{E}CL(\hat{t}_2)$	$\mathrm{E}CL(\hat{t}_u)$
40	32.1	33.3(29.4)	24.1(26.4)
80	26.0	26.4(17.4)	21.0(16.2)
500	21.0	20.9(5.9)	20.0(5.8)
2000	20.2	20.2(2.9)	20.0(2.9)
$(\gamma, \sigma, \pi) = (20, 0.10, 0.15)$, $r_0(d_1) = 0.006$			
n	$\mathrm{E}CL(\hat{t}_2)$	$\mathrm{E}CL(\hat{t}_2)$	$\mathrm{E}CL(\hat{t}_u)$
40	98.7	144.5(174.5)	127.2(160.6)
80	59.3	70.1(77.5)	55.1(68.2)
500	26.3	26.3(18.0)	20.9(16.9)
2000	21.6	21.5(7.6)	20.0(7.5)

6.3.4 Generalization: estimation of μ

In this section we show that if $\mathrm{E}|V|^r < \infty$ for $r > 8$ (cf. assumption (A1)), no additional correction term is needed to the test limit $\hat{t}_u$ (cf. (6.3.26)) when the mean of the measurement error has to estimated.

Let the estimator of μ be the sample mean,

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} U_i, \tag{6.3.30}$$

and let

$$\hat{a}_1 = \hat{h}_1^{-1}\left(\frac{\gamma}{\sigma f_{\bar{X}}(s+\hat{\mu})}\right). \tag{6.3.31}$$

We introduce the set A_μ similar to the set A (cf. (6.3.12)), defined by

$$\begin{aligned} A_\mu = \Bigg\{ & |\hat{a}_1 - a_1| \le \epsilon_1,\ \left|\frac{\hat{h}_2(\hat{a}_1)}{\hat{h}_1'(\hat{a}_1)} - \frac{h_2(a_1)}{h_1'(a_1)}\right| \le \epsilon_2, \\ & \left|\frac{1+\hat{h}_1'(\hat{a}_1)}{(\hat{h}_1'(\hat{a}_1))^2} - \frac{1+h_1'(a_1)}{(h_1'(a_1))^2}\right| \le \epsilon_3,\ |\hat{\mu}-\mu| \le \epsilon_4 \Bigg\}, \end{aligned} \tag{6.3.32}$$

for small constants $\epsilon_1, \epsilon_2, \epsilon_3, \epsilon_4 > 0$. In the same way as in the proof of lemma 6.3.1 it is easily shown that also $P(A_\mu^c) = O(n^{-r/4})$. (Note that under (A1), $P(|\hat{\mu}-\mu| > \epsilon) = O(n^{-r/2})$, for $\epsilon > 0$.)

We now have the following relation between a_1 and $\hat{h}_1$ (cf. 6.3.17)).

$$\hat{a}_1 \le a_1 \ \Leftrightarrow\ \hat{h}_1(\hat{a}_1) \ge \hat{h}_1(a_1) \ \Leftrightarrow\ \hat{h}_1(a_1) \le \frac{\gamma}{\sigma f_{\bar{X}}(s+\hat{\mu})}, \tag{6.3.33}$$

assuming that $f_{\bar{X}}(s+\hat{\mu}) > 0$.

Since $f_{\bar{X}}''$ is bounded we have that

$$\frac{\gamma}{\sigma f_{\bar{X}}(s+\hat{\mu})} = \frac{\gamma}{\sigma f_{\bar{X}}(s+\mu)}\left(1 - (\hat{\mu}-\mu)\frac{f_{\bar{X}}'(s+\mu)}{f_{\bar{X}}(s+\mu)}\right) + W^2,$$

with $W = (\hat{\mu}-\mu)\tilde{f}(\hat{\mu})$ for some bounded function $\tilde{f}$.
Consequently,

$$\begin{aligned} & P((\hat{a}_1 - a_1)\sqrt{n} \le y) \\ & \quad = P\left(\hat{h}_1(a_1 + y\, n^{-1/2}) + (\hat{\mu}-\mu)h_1(a_1)\frac{f_{\bar{X}}'(s+\mu)}{f_{\bar{X}}(s+\mu)} \le h_1(a_1) + W^2\right). \end{aligned} \tag{6.3.34}$$

Since, under (A1), $P(W^2 > \xi_0\ \sigma^2 n^{-1/2}) = O(n^{-r/4})$, for some $\xi_0 > 0$, we get (cf. (6.3.34))

$$\begin{aligned} & P\left(\hat{h}_1(a_1 + y\, n^{-1/2}) + \ (\hat{\mu}-\mu)h_1(a_1)\frac{f_{\bar{X}}'(s+\mu)}{f_{\bar{X}}(s+\mu)} \le h_1(a_1)\right) \\ & \quad \le\ P((\hat{a}_1 - a_1)\sqrt{n} \le y)\ \le \end{aligned} \tag{6.3.35}$$

$$P\left(\hat{h}_1(a_1 + y\, n^{-1/2}) + \ (\hat{\mu}-\mu)h_1(a_1)\frac{f_{\bar{X}}'(s+\mu)}{f_{\bar{X}}(s+\mu)} \le h_1(a_1) + \xi_0\sigma^2 n^{-1/2}\right) + O(n^{-r/4})$$

From lemma 6.3.2 and the proof of lemma 6.3.3 it is seen that $O(n^{-r/4})$ is sufficiently small if $O(n^{-r/4+3/2}) = o(n^{-1/2})$, hence if $r > 8$.

With the definition of $\hat{h}_1$ (cf. (6.3.11)) and $\hat{\mu}$ (cf. (6.3.30)) it is seen that $\hat{h}_1(a) + (\hat{\mu} - \mu)$ is a sum of independent r.v.'s. Consequently, the probabilities in (6.3.35) can be approximated by means of an Edgeworth expansion, in a similar way as in section 6.3.2.

Let

$$\begin{aligned}
K(a) &= \frac{1}{n}\sum_{i=1}^{n}\left((V_i - a)\cdot 1_{\{V_i>a\}} - \sigma V_i h_1(a_1)\frac{f'_{\tilde{X}}(s+\mu)}{f_{\tilde{X}}(s+\mu)}\right)\\
\mu_n(y) &= EK(a_1 + y/n^{1/2}) = h_1(a_1 + y/n^{1/2})\\
\sigma_n^2(y) &= n\mathrm{VAR}K(a_1 + y/n^{1/2})\\
&= h_2(a_1 + y/n^{1/2}) - h_1^2(a_1 + y/n^{1/2})\\
&\quad - 2\sigma EV_i(V_i - a)\cdot 1_{\{V_i>a\}}h_1(a_1)\frac{f'_{\tilde{X}}(s+\mu)}{f_{\tilde{X}}(s+\mu)}\\
&\quad + \sigma^2 h_1^2(a_1)\left(\frac{f'_{\tilde{X}}(s+\mu)}{f_{\tilde{X}}(s+\mu)}\right)^2,\\
z_n(y) &= \frac{h_1(a_1) - \mu_n(y)}{\sigma_n(y)}\sqrt{n}.
\end{aligned}$$

By Taylor expansion we get (cf. lemma 6.3.4),

$$z_n(y) = (A_0\sigma + A_1)y + A_2y^2/\sqrt{n} + O(|y|^3/n + \sigma^2|y|), \tag{6.3.36}$$

uniformly in $|y| \le \epsilon_1\sqrt{n}$ (ϵ_1 as in (6.3.32)), with

$$A_0 = -h_1'(a_1)\, h_1(a_1)\, b^3\frac{f'_{\tilde{X}}(s+\mu)}{f_{\tilde{X}}(s+\mu)}E\, V_i(V_i - a)\cdot 1_{\{V_i>a\}},$$

and b, A_1 and A_2 as in lemma 6.3.4.
Carrying out steps analogous to those leading to lemma 6.3.5 and lemma 6.3.6, both for the left-hand side of (6.3.35) and the right-hand side, we obtain (the term $\xi_0\sigma^2 n^{-1/2}$ in the right-hand side only gives an additional term of order σ^2 in (6.3.37))

$$\begin{aligned}
E\,\sqrt{n}(\hat{a}_1 - a_1)\cdot 1_A &= \frac{A_2}{(A_1 + \sigma A_0)^3}\frac{1}{\sqrt{n}} + o(n^{-1/2}) + O(\sigma^2)\\
E\, n(\hat{a}_1 - a_1)^2\cdot 1_A &= \frac{1}{(A_1 + \sigma A_0)^2} + o(n^{-1/2}) + O(\sigma^2).
\end{aligned} \tag{6.3.37}$$

Following the lines of the proof of theorem 6.3.1, we find that (6.3.37) leads to additional terms of order $\gamma\sigma n^{-1}$ and $\gamma\sigma^2$ in the expectation of the consumer loss, compared to the case in which μ is known. They contribute only to the remainder.

Simulations as the ones presented in table 6.3.1 and 6.3.2 but with estimation of μ, indeed yield results which are practically the same as the results obtained when μ is known. Numerical examples are therefore omitted.

6.3.5 Generalization: estimation of $f_{\tilde{X}}$ but μ known

When the density of $\tilde{X}$ is estimated, an additional correction term is needed to the test limit which has been derived in section 6.3.2. In chapter 3 for example, we have seen that the correction term c_u (cf (3.4.12)) contains a part which corrects for the bias due to estimation of σ and a part which corrects for estimation of the parameters $\mu_X = \mathrm{E}X$, $\sigma_X^2 = \mathrm{VAR}X$. Likewise, in chapter 4 the correction term c_u (cf. (4.2.21)) contains a part which corrects for estimation of σ (which is equal to that in chapter 3) and a part which corrects for estimation of the density.

It will be seen by a simple heuristic argument that also in the present situation there is no mix-up of errors. As a consequence it is not difficult to derive a correction term which corrects for the bias due to the estimation of the density for the present situation. However, a rigorous proof of a theorem, like theorem 3.4.1 or theorem 4.2.1, which gives second order unbiasedness, requires a great deal of technicalities and a large amount of space. Therefore we have chosen not to pursue.

To derive the additional correction term, let $\hat{f}_{\tilde{X}}(s+\mu)$ be the estimator of $f_{\tilde{X}}(s+\mu)$ and define

$$b = \frac{\gamma}{\sigma f_{\tilde{X}}(s+\mu)}, \quad \hat{b} = \frac{\gamma}{\sigma \hat{f}_{\tilde{X}}(s+\mu)},$$

and

$$a_1 = h_1^{-1}(b), \quad \hat{a}_1 = \hat{h}_1^{-1}(\hat{b}).$$

With test limit $\hat{t}_1 = s+\mu-\hat{a}_1\sigma$ the consumer loss is approximated by (cf. (6.1.5))

$$CL(\hat{t}_1) = \sigma f_{\tilde{X}}(s+\mu) h_1(\hat{a}_1),$$

and its relative error by

$$\begin{aligned} \frac{CL(\hat{t}_1)-\gamma}{\gamma} &= \frac{h_1(\hat{a}_1)-h_1(a_1)}{h_1(a_1)} \\ &= \frac{\hat{h}_1(\hat{a}_1)-h_1(a_1)}{h_1(a_1)} + \frac{h_1(\hat{a}_1)-\hat{h}_1(\hat{a}_1)}{h_1(a_1)}. \end{aligned} \tag{6.3.38}$$

The first term in (6.3.38) is equal to

$$\begin{aligned} \frac{\hat{b}-b}{h_1(a_1)} &= \frac{f_{\tilde{X}}(s+\mu)}{\hat{f}_{\tilde{X}}(s+\mu)} - 1 \\ &= -\left(\frac{\hat{f}_{\tilde{X}}(s+\mu)}{f_{\tilde{X}}(s+\mu)} - 1\right) + \left(\frac{\hat{f}_{\tilde{X}}(s+\mu)}{f_{\tilde{X}}(s+\mu)} - 1\right)^2 + \cdots, \end{aligned} \tag{6.3.39}$$

which is very much in correspondence with (4.1.4).
For the second term in (6.3.38) we write

$$\begin{aligned} & h_1(\hat{a}_1) - \hat{h}_1(\hat{a}_1) \\ = \; & h_1\left(\hat{h}_1^{-1}(\hat{b})\right) - \hat{b} \\ = \; & h_1\left(\hat{h}_1^{-1}(b)\right) - b \; + \; (\hat{b} - b) \; \left(\frac{h_1'\left(\hat{h}_1^{-1}(\xi)\right)}{\hat{h}_1'\left(\hat{h}_1^{-1}(\xi)\right)} - 1\right), \end{aligned} \tag{6.3.40}$$

with ξ between b and $\hat{b}$.
The first term in (6.3.40) is precisely the one we have studied in section 6.3.2 (cf. (6.3.16)). The two factors in the last term in (6.3.40) are not independent, but the second factor will not vary much for small changes of ξ. It is clear that the expectation of this term will be of smaller order than both $\mathrm{E}(\hat{b} - b)$ and $\mathrm{E}\left\{h_1\left(\hat{h}_1^{-1}(b)\right) - b\right\}$.

Combination of (6.3.38) - (6.3.40) then gives approximately

$$\mathrm{E}\,CL(\hat{t}_1) - \gamma \;=\; \gamma\,\mathrm{E}\left(\frac{\hat{b} - b}{h_1(a_1)} + \frac{h_1\left(\hat{h}_1^{-1}(b)\right) - b}{h_1(a_1)}\right).$$

So, the additional correction term we are looking for should cancel

$$\mathrm{E}\,\frac{b - b}{h_1(a_1)},$$

to sufficient precision.

To estimate the density (and its derivative) we assume that we have independent observations $\tilde{X}_1, \ldots, \tilde{X}_m$ from $\tilde{X}$, which are also independent of the observations $U_1, \ldots, U_n$ from the measurement error. If Rosenblatt's estimators, as defined in section 4.2.2, are applied, it follows from section 4.2.4 that the additional correction should equal

$$\frac{h_1(a_1)}{h_0(a_1)}\left(\frac{1}{2mhf_{\tilde{X}}(s+\mu)} - \frac{1}{m}\right).$$

The test limit then becomes (cf. (6.3.26))

$$\hat{t}_u = s - (\hat{d}_1 + \hat{c} + \hat{c}_u), \tag{6.3.41}$$

with

$$\begin{aligned} \hat{d}_1 \;&=\; \hat{r}_1^{-1}\left(\frac{\gamma}{\hat{f}_{\tilde{X}}(s+\mu)}\right) \\ \hat{c} \;&=\; \frac{1}{2}\frac{\hat{f}'_{\tilde{X}}(s+\mu)}{\hat{f}_{\tilde{X}}(s+\mu)}\frac{\hat{r}_2(\hat{d}_1)}{\hat{r}_0(\hat{d}_1)} \\ \hat{c}_u \;&=\; \frac{\hat{r}_1(\hat{d}_1)}{n}\frac{\{1 - \hat{r}_0(\hat{d}_1)\}}{\hat{r}_0^2(\hat{d}_1)} + \frac{\hat{r}_1(\hat{d}_1)}{\hat{r}_0(\hat{d}_1)}\left(\frac{1}{2mh\hat{f}_{\tilde{X}}(s+\mu)} - \frac{1}{m}\right). \end{aligned} \tag{6.3.42}$$

Also in the present situation Rosenblatt's estimators are applied with the bandwidths as in section 4.2.6.

We have carried out several simulations with this test limit. The results for choices of γ, σ and $\pi = P(X > s)$ as in the first three situations of table 6.3.1 and table 6.3.2 are presented in table 6.3.3 and table 6.3.4 below.

Table 6.3.3 Test limit $\hat{t}_u$ with a $N(0, \sigma^2)$-distributed measurement error

The table shows simulated mean and standard deviation (between brackets) of $CL(\hat{t}_u)$ (with $\hat{t}_u$ from (6.3.41)) based on 10000 replications. To estimate the density and its derivative Rosenblatt's estimators are applied, with bandwidths as in section 4.2.6. The number of observations to estimate the density is denoted by m. In the table, nr denotes the number of replications for which the test limit could not be determined (caused by a very small estimate of the density). The values of γ are in ppm, π denotes $P(X > s)$, the probability that a product is nonconforming, and n is the number of observations on the measurement error.
The characteristic is standard normally distributed.

$(\gamma, \sigma, \pi) = (100, 0.01, 0.01), \quad r_0(d_1) = 0.481$

	$(m = 100)$		$(m = 400)$		$(m = 1600)$	
n	$\mathrm{E}CL(\hat{t}_u)$	nr	$\mathrm{E}CL(\hat{t}_u)$	nr	$\mathrm{E}CL(\hat{t}_u)$	nr
40	67.0(28.9)	43	90.5(39.7)	37	98.0(36.7)	1
80	67.0(25.4)	33	90.1(36.2)	27	98.2(33.6)	1
500	66.8(21.9)	34	90.4(33.5)	5	97.8(29.8)	1
2000	66.4(21.4)	21	89.7(32.3)	4	98.1(29.9)	-

$(\gamma, \sigma, \pi) = (100, 0.10, 0.01), \quad r_0(d_1) = 0.082$

	$(m = 100)$		$(m = 400)$		$(m = 1600)$	
n	$\mathrm{E}CL(\hat{t}_u)$	nr	$\mathrm{E}CL(\hat{t}_u)$	nr	$\mathrm{E}CL(\hat{t}_u)$	nr
40	74.6(64.9)	23	96.9(81.7)	4	99.6(77.5)	-
80	69.7(47.8)	32	92.4(63.6)	3	97.9(58.9)	-
500	67.9(30.4)	25	93.1(50.7)	2	97.5(39.1)	-
2000	68.2(27.0)	25	92.7(48.5)	3	97.6(34.5)	-

$(\gamma, \sigma, \pi) = (20, 0.01, 0.15), \quad r_0(d_1) = 0.023$

	$(m = 100)$		$(m = 400)$		$(m = 1600)$	
n	$\mathrm{E}CL(\hat{t}_u)$	nr	$\mathrm{E}CL(\hat{t}_u)$	nr	$\mathrm{E}CL(\hat{t}_u)$	nr
40	32.0(37.4)	-	32.5(37.6)	-	32.7(37.1)	-
80	23.9(23.3)	-	24.1(23.5)	-	24.0(23.0)	-
500	19.9(10.2)	-	20.1(9.3)	-	20.2(8.8)	-
2000	19.8(7.5)	-	20.0(6.1)	-	20.0(5.2)	-

Table 6.3.4 Test limit $\hat{t}_u$ with a $\Gamma(8)$-distributed measurement error

This table summarizes the same simulation as in table 6.3.3, however with a measurement error which has a Γ-distribution with shape parameter 8 and with the location and scale parameter such that the mean is 0 and the variance σ^2.

$(\gamma, \sigma, \pi) = (100, 0.01, 0.01), \quad r_0(d_1) = 0.533$

	$(m = 100)$		$(m = 400)$		$(m = 1600)$	
n	$\mathrm{E}CL(\hat{t}_u)$	nr	$\mathrm{E}CL(\hat{t}_u)$	nr	$\mathrm{E}CL(\hat{t}_u)$	nr
40	66.0(26.5)	35	90.0(37.5)	28	97.8(34.9)	1
80	66.1(23.7)	19	90.4(35.0)	16	97.6(32.5)	1
500	66.0(20.9)	24	89.8(32.8)	4	98.0(29.9)	-
2000	65.7(20.4)	27	90.3(32.1)	4	97.8(28.7)	1

$(\gamma, \sigma, \pi) = (100, 0.10, 0.01), \quad r_0(d_1) = 0.120$

	$(m = 100)$		$(m = 400)$		$(m = 1600)$	
n	$\mathrm{E}CL(\hat{t}_u)$	nr	$\mathrm{E}CL(\hat{t}_u)$	nr	$\mathrm{E}CL(\hat{t}_u)$	nr
40	70.6(53.4)	26	94.6(70.2)	3	99.5(65.3)	-
80	68.2(39.4)	34	93.7(58.8)	9	99.4(51.1)	-
500	68.1(27.3)	24	93.5(45.9)	7	99.4(35.8)	-
2000	68.4(25.4)	40	92.7(42.3)	3	99.2(33.6)	-

$(\gamma, \sigma, \pi) = (20, 0.01, 0.15), \quad r_0(d_1) = 0.040$

	$(m = 100)$		$(m = 400)$		$(m = 1600)$	
n	$\mathrm{E}CL(\hat{t}_u)$	nr	$\mathrm{E}CL(\hat{t}_u)$	nr	$\mathrm{E}CL(\hat{t}_u)$	nr
40	24.1(26.8)	-	24.4(26.9)	-	24.5(27.1)	-
80	21.0(17.5)	-	21.2(17.2)	-	21.0(16.6)	-
500	19.8(8.4)	-	19.9(7.3)	-	20.1(6.6)	-
2000	19.7(6.9)	-	20.0(5.5)	-	20.0(4.3)	-

The results once more indicate that 100 observations are not enough if the density has to be estimated at a point in the tail of the distribution (cf. section 4.2.7 and chapter 5). We conclude that the average consumer losses are quite close to γ if a sufficient large number of observations is available.

6.3.6 Generalization: estimation of $f_{\tilde{X}}$ and μ

In section 6.3.4 we have proved that there is no loss of precision, in terms of the asymptotic bias of the consumer loss, if the test limit $\hat{t}_u$ from section 6.3.2 is based on an estimator of μ instead of on the true value. In the previous section we have seen that when the density has to be estimated but μ is known, it suffices to add essentially the same correction term to the test limit as applied in chapter 4.

While we handled the problems in the latter case by a simple heuristic argument and with theory from chapter 4, this argument is not entirely convincing in the more general case where everything has to be estimated. To begin with, we have to consider $\mathrm{E}\hat{f}_{\tilde{X}}(s+\hat{\mu})$, which is complicated by the fact that $\hat{f}_{\tilde{X}}$ is not differentiable if $\hat{f}_{\tilde{X}}$ is Rosenblatt's estimator. But perhaps more difficult it is to determine the expectation of the remainder term in (6.3.40) if also μ is estimated. We then have to consider

$$\mathrm{E}\left(\frac{f_{\tilde{X}}(s+\mu)}{\hat{f}_{\tilde{X}}(s+\hat{\mu})}-1\right)\left(\frac{h_1'\left(\hat{h}_1^{-1}(\xi)\right)}{\hat{h}_1'\left(\hat{h}_1^{-1}(\xi)\right)}-1\right),$$

where ξ now also depends on $\hat{\mu}$.

On the other hand, when the number of $\tilde{X}$-observations m is large, the estimators $\hat{f}_{\tilde{X}}$ and $\hat{f}'_{\tilde{X}}$ are close to $f_{\tilde{X}}$ and $f'_{\tilde{X}}$, respectively, and we can use the results of section 6.3.4, the situation where $f_{\tilde{X}}$ and $f'_{\tilde{X}}$ are known, but μ is estimated. In this section we will study the number of observations for which the estimators of $f_{\tilde{X}}$ and $f'_{\tilde{X}}$ may be identified with their true values.

Let $\hat{t}_u$ denote the test limit as in (6.3.26), however with the estimator $\hat{\mu}$, as in (6.3.30), instead of the parameter μ. In terms of the functions h_k and the parameter a the test limit is then defined by

$$\hat{t}_u = s+\mu-(\sigma\hat{a}_1+\hat{c}+\hat{c}_u), \tag{6.3.43}$$

with

$$\begin{aligned}\hat{a}_1 &= \hat{h}_1^{-1}\left(\frac{\gamma}{\sigma f_{\tilde{X}}(s+\hat{\mu})}\right)\\ \hat{c} &= \frac{\sigma^2}{2}\frac{f'_{\tilde{X}}(s+\hat{\mu})}{f_{\tilde{X}}(s+\hat{\mu})}\frac{\hat{h}_2(\hat{a}_1)}{\hat{h}_0(\hat{a}_1)}\\ \hat{c}_u &= \frac{\sigma\,\hat{h}_1(\hat{a}_1)}{n}\frac{\{1-\hat{h}_0(\hat{a}_1)\}}{\hat{h}_0^2(\hat{a}_1)}.\end{aligned} \tag{6.3.44}$$

By $\hat{t}_u^*$ we denote the test limit with estimators of μ, $f_{\tilde{X}}$ and $f'_{\tilde{X}}$.

$$\hat{t}_u^* = s+\mu-(\sigma\hat{a}_1^*+\hat{c}^*+\hat{c}_u^*), \tag{6.3.45}$$

with

$$\begin{aligned}\hat{a}_1^* &= \hat{h}_1^{-1}\left(\frac{\gamma}{\sigma\hat{f}_{\tilde{X}}(s+\hat{\mu})}\right)\\ \hat{c}^* &= \frac{\sigma^2}{2}\frac{\hat{f}'_{\tilde{X}}(s+\hat{\mu})}{\hat{f}_{\tilde{X}}(s+\hat{\mu})}\frac{\hat{h}_2(\hat{a}_1^*)}{\hat{h}_0(\hat{a}_1^*)}\\ \hat{c}_u^* &= \frac{\sigma\hat{h}_1(\hat{a}_1^*)}{n}\frac{\{1-\hat{h}_0(\hat{a}_1^*)\}}{\hat{h}_0^2(\hat{a}_1^*)}.\end{aligned} \tag{6.3.46}$$

We will study

$$\mathrm{E}|\,CL(\hat{t}_u^*) - CL(\hat{t}_u)|$$

for the situation in which $\hat{t}_u^*$ is computed with Rosenblatt's estimators of $f_{\tilde{X}}$ and $f'_{\tilde{X}}$ (cf. section 4.2.2). The bandwidths h and $\bar{h}$ are not yet fixed. They will be determined when it has become clear in what way the bandwidths affect $|\,CL(\hat{t}_u^*) - CL(\hat{t}_u)|$.

We assume that we have observations $\tilde{X}_1, \ldots, \tilde{X}_m$ from $\tilde{X}$ and observations $U_1, \ldots, U_n$ from the measurement error, which are all mutually independent. It will be assumed throughout that $f_{\tilde{X}}$ is bounded, that $f_{\tilde{X}}(s+\mu) > 0$ and that $f'_{\tilde{X}}$ is bounded in a neighborhood of $s+\mu$.

First we consider the difference between $\hat{a}_1$ and $\hat{a}_1^*$.

Lemma 6.3.7 *Assume* (A1) *and assume that* $g(a_1) > 0$, *with* a_1 *as in (6.3.9). In the notation as above, there exists a constant* $C_1 > 0$ *such that*

$$|\hat{a}_1^* - \hat{a}_1| \leq C_1\ |\hat{f}_{\tilde{X}}(s+\hat{\mu}) - f_{\tilde{X}}(s+\hat{\mu})|, \tag{6.3.47}$$

on a set D *with* $P(D^c) = O(n^{-r/2})$.

Proof. By definition

$$\hat{h}_1(\hat{a}_1^*) \;=\; \frac{\gamma}{\sigma \hat{f}_{\tilde{X}}(s+\hat{\mu})}.$$

The first requirement for the set D is that $\hat{a}_1$ and $\hat{a}_1^*$ are in a neighborhood of a_1 and that

$$|\hat{h}_1'(\hat{a}_1^* + \epsilon_1)| = \hat{h}_0(\hat{a}_1^* + \epsilon_1) \geq \epsilon_2, \tag{6.3.48}$$

for sufficiently small constants $\epsilon_1, \epsilon_2 > 0$ (Note that $\hat{h}_0$ is nonincreasing). As $P(V > a_1 + \epsilon_1) = h_0(a_1 + \epsilon_1) > 0$, the set where $\hat{a}_1^*$ is in a neighborhood of a_1 and (6.3.48) does not hold has exponentially small probability. The complement of the set where $\hat{a}_1^*$ and $\hat{a}_1$ are in a neighborhood of a_1 has probability of order $n^{-r/2}$. (See the proof of lemma 6.3.1).

Hence, for any $\delta \in (0, \epsilon_1)$,

$$\hat{h}_1(\hat{a}_1^* - \delta) - \hat{h}_1(\hat{a}_1^*) \geq \epsilon_2 \delta \tag{6.3.49}$$

and

$$\hat{h}_1(\hat{a}_1^* + \delta) - \hat{h}_1(\hat{a}_1^*) \leq -\epsilon_2 \delta. \tag{6.3.50}$$

If

$$\left| \frac{\gamma/\sigma}{\hat{f}_{\tilde{X}}(s+\hat{\mu})} - \frac{\gamma/\sigma}{f_{\tilde{X}}(s+\hat{\mu})} \right| \leq \epsilon_1 \epsilon_2$$

then (6.3.47) follows from (6.3.49) and (6.3.50). Otherwise, $|\hat{f}_{\tilde{X}}(s+\hat{\mu}) - f_{\tilde{X}}(s+\hat{\mu})| > \epsilon_3$, for some $\epsilon_3 > 0$ and (6.3.47) follows from the boundedness of $|\hat{a}_1^* - \hat{a}_1|$ on D. □

Lemma 6.3.8 *Under the assumptions of lemma 6.3.7 there exists a constant* $C_2 > 0$ *such that*

$$
\begin{aligned}
&|(\hat{c}^* + \hat{c}_u^*) - (\hat{c} + \hat{c}_u)| \\
&\quad \leq \; C_2\,(\sigma^2 + \sigma n^{-1})\Big\{|\hat{f}_{\hat{X}}(s+\hat{\mu}) - f_{\hat{X}}(s+\hat{\mu})| + |\hat{f}'_{\hat{X}}(s+\hat{\mu}) - f'_{\hat{X}}(s+\hat{\mu})| \\
&\qquad\qquad + \; \sup_x |\hat{h}_0(x) - h_0(x)|\Big\},
\end{aligned}
$$

on a set D *with* $P(D^c) = O(n^{-r/2})$.

Proof. Estimation of $|\hat{h}_2(\hat{a}_1^*) - \hat{h}_2(\hat{a}_1)|$ and $|\hat{h}_1(\hat{a}_1^*) - \hat{h}_1(\hat{a}_1)|$ is straightforward using that $|\hat{a}_1^* - \hat{a}_1|$ is bounded (see the proof of lemma 6.3.7). Direct estimation of $\hat{h}_0(\hat{a}_1^*) - \hat{h}_0(\hat{a}_1)$ is difficult and therefore replaced by

$$
\begin{aligned}
&|\hat{h}_0(\hat{a}_1^*) - \hat{h}_0(\hat{a}_1)| \\
&\quad \leq \; |\hat{h}_0(\hat{a}_1^*) - h_0(\hat{a}_1^*)| + |h_0(\hat{a}_1^*) - h_0(\hat{a}_1)| + |h_0(\hat{a}_1) - \hat{h}_0(\hat{a}_1)| \\
&\quad \leq \; 2\sup_x |\hat{h}_0(x) - h_0(x)| + |h_0(\hat{a}_1^*) - h_0(\hat{a}_1)| \\
&\quad \leq \; 2\sup_x |\hat{h}_0(x) - h_0(x)| + C_3|\hat{a}_1^* - \hat{a}_1|.
\end{aligned}
$$

The quantity $|\hat{a}_1^* - \hat{a}_1|$is estimated by application of lemma 6.3.7. Hence the lemma follows. □

If $\hat{a}_1 > \hat{a}_1^*$, we have

$$
\begin{aligned}
&CL(\hat{t}_u^*) - CL(\hat{t}_u) \\
&= \int_{\hat{a}_1^*}^{\hat{a}_1} \{F_X(s + \sigma(v - \hat{a}_1^*) + \hat{c}^* + \hat{c}_u^*) - F_X(s)\} g(v)\mathrm{d}v \; + \\
&\qquad \int_{\hat{a}_1}^{\infty} \{F_X(s + \sigma(v - \hat{a}_1^*) + \hat{c} + \hat{c}_u) - F_X(s + \sigma(v - \hat{a}_1) + \hat{c} + \hat{c}_u)\} g(v)\mathrm{d}v.
\end{aligned}
\tag{6.3.51}
$$

Except for a set with probability of order $n^{-r/2}$ we have, for some constant $C_4 > 0$, that

$$
|\hat{c}| \; \leq \; C_4\sigma^2 \sup_{x \in N(a_1)} \hat{h}_2(x), \qquad |\hat{c}_u| \; \leq \; C_4\sigma n^{-1}
$$

with $\mathrm{E}\left\{\sup_{x \in N(a_1)} \hat{h}_2(x)\right\} < \infty$ and $N(a_1)$ a small neighborhood of a_1.
Since f'_X is bounded, the first term on the right-hand side of (6.3.51) can be estimated by

$$
C_5\sigma\,|\hat{a}_1^* - \hat{a}_1|\,\{|\hat{a}_1^* - \hat{a}_1| + n^{-1} + \sigma \sup_{x \in N(a_1)} \hat{h}_2(x)\},
$$

while the latter term is at most

$$C_5\{\sigma|\hat{a}_1^* - \hat{a}_1| + |(\hat{c} + \hat{c}_u) - (\hat{c}^* + \hat{c}_u^*)|\}.$$

A similar result holds if $\hat{a}_1^* > \hat{a}_1$.

Lemma 6.3.9 *Under the assumptions of lemma 6.3.7*

$$\begin{aligned} \mathrm{E}|\hat{f}_{\tilde{X}}(z) - f_{\tilde{X}}(z)| &= O((mh)^{-1/2} + h^2) \\ \mathrm{E}|\hat{f}'_{\tilde{X}}(z) - f'_{\tilde{X}}(z)| &= O((m\bar{h}^3)^{-1/2} + \bar{h}^2), \end{aligned}$$

as $m \to \infty$ *and* $h, \bar{h} \to 0$, *for* z *in a neighborhood of* $s + \mu$.

Proof. With $p = F_{\tilde{X}}(z+h) - F_{\tilde{X}}(z-h)$,

$$\begin{aligned} \mathrm{E}|\hat{f}_{\tilde{X}}(z) - f_{\tilde{X}}(z)| &\leq \mathrm{E}|\hat{f}_{\tilde{X}}(z) - p/(2h)| + |p/(2h) - f_{\tilde{X}}(z)| \\ &\leq \left(\mathrm{VAR}\ \hat{f}_{\tilde{X}}(z)\right)^{1/2} + O(h^2) \\ &= O((mh)^{-1/2} + h^2). \end{aligned}$$

Similarly for $\hat{f}'_{\tilde{X}}(z) - f_{\tilde{X}}(z)$. □

We have obtained

$$\begin{aligned} \mathrm{E}\,|CL(\hat{t}_u^*) - CL(\hat{t}_u)| \;\leq\; & C_6\ \sigma\ \mathrm{E}\,|\hat{f}_{\tilde{X}}(s+\hat{\mu}) - f_{\tilde{X}}(s+\hat{\mu})| \;+ \\ & C_6\ \mathrm{E}\,|(\hat{c} + \hat{c}_u) - (\hat{c}^* + \hat{c}_u^*)| + O(n^{-r/2}). \end{aligned}$$

With the choice

$$h \;\sim\; m^{-1/5}, \quad \bar{h} \;\sim\; m^{-1/7}$$

we get

$$\begin{aligned} \mathrm{E}\,|CL(\hat{t}_u^*) - CL(\hat{t}_u)| \;\leq\; & C_7\ \sigma m^{-2/5} + C_7\ (\sigma^2 + \sigma n^{-1})\{m^{-2/7} + n^{-1/2}\} \\ & + O(n^{-r/2}). \end{aligned}$$

If m is such that $m/n^{5/2} \to \infty$ then

$$\mathrm{E}\,|CL(\hat{t}_u^*) - CL(\hat{t}_u)| \;=\; o(\gamma n^{-1}) + O(\gamma\sigma^2) + O(n^{-r/2}),$$

which disappears in the remainder term which was found in theorem 6.3.1. We conclude that, in general, if m is quite large compared to n (which in practice will frequently the case), the effect of estimating $f_{\tilde{X}}$ can indeed be neglected.

Remark 6.3.2 We studied $\mathrm{E}\,|CL(\hat{t}_u^*) - CL(\hat{t}_u)|$, while in case of unbiased estimation the quantity $|\mathrm{E}\,\{CL(\hat{t}_u^*) - CL(\hat{t}_u)\}|$ actually is of interest. It is therefore to be expected that in practice smaller samples sizes will suffice. □

6.4 Test limits for which γ is violated with small probability

It is evident from section 6.3.2 that the consumer loss is asymptotically normal if $f_{\tilde{X}}$ is known. Under the heuristic approach in section 6.3.5 asymptotic normality still holds if $f_{\tilde{X}}$ is estimated. This enables us to find a correction term for which the consumer loss exceeds γ with small probability only.

Let

$$\hat{t}_i = s + \mu - \hat{a}_1\sigma - \hat{c} - \hat{c}_i,$$

with

$$\hat{a}_1 = \hat{h}_1^{-1}\left(\frac{\gamma}{\sigma \hat{f}_{\tilde{X}}(s+\mu)}\right), \tag{6.4.1}$$

the correction term $\hat{c}$ as in (6.3.42) and $\hat{c}_i$ such that

$$P(CL(\hat{t}_i) \geq \gamma) \leq \alpha \tag{6.4.2}$$

is obtained to sufficient precision, for some choice of α.

From (6.3.16) and (6.3.38) in combination with (6.3.39) and (6.3.40) it follows that

$$\begin{aligned}
&\frac{CL(\hat{t}_i) - \gamma}{\gamma} \\
&\approx \left\{\hat{h}_1^{-1}\left(\frac{\gamma}{\sigma f_{\tilde{X}}(s+\mu)}\right) - h_1^{-1}\left(\frac{\gamma}{\sigma f_{\tilde{X}}(s+\mu)}\right)\right\}\frac{h_1'(a_1)}{h_1(a_1)} \\
&\quad - \left(\frac{\hat{f}_{\tilde{X}}(s+\mu)}{f_{\tilde{X}}(s+\mu)} - 1\right) + \left(\frac{\hat{f}_{\tilde{X}}(s+\mu)}{f_{\tilde{X}}(s+\mu)} - 1\right)^2 + c_i\frac{h_1'(a_1)}{h_1(a_1)}.
\end{aligned}$$

When Rosenblatt's estimator is applied to estimate the density (section 4.2.2), it follows from section 4.3 that the right-hand side is asymptotically normal $AN(\mu_{CL}, \sigma_{CL}^2)$ with (neglecting terms of order n^{-1}, h^2 and $(mh)^{-1}$ in μ_{CL})

$$\mu_{CL} = c_i\frac{h_1'(a_1)}{h_1(a_1)}$$

and (cf. (6.3.24) and (6.3.23))

$$\sigma_{CL}^2 = \frac{h_2(a_1) - h_1^2(a_1)}{nh_1^2(a_1)} + \frac{1}{2mhf_{\tilde{X}}(s+\mu)}.$$

With the same arguments as in section 4.3 we take $h \sim m^{-1/5}$ and we find that c_i should be taken

$$c_i = -u_\alpha\frac{h_1(a_1)}{h_1'(a_1)}\sqrt{\frac{h_2(a_1) - h_1^2(a_1)}{nh_1^2(a_1)} + \frac{1}{2mhf_{\tilde{X}}(s+\mu)}},$$

where $u_\alpha = \Phi^{-1}(1-\alpha)$.

The test limit thus becomes

$$\hat{t}_i = s - \hat{d}_1 - \hat{c} - \hat{c}_i, \tag{6.4.3}$$

with

$$\begin{aligned}
\hat{d}_1 &= \hat{r}_1^{-1}\left(\frac{\gamma}{\hat{f}_{\tilde{X}}(s+\mu)}\right) \\
\hat{c} &= \frac{1}{2}\frac{\hat{f}'_{\tilde{X}}(s+\mu)}{\hat{f}_{\tilde{X}}(s+\mu)}\frac{\hat{r}_2(\hat{d}_1)}{\hat{r}_0(\hat{d}_1)} \\
\hat{c}_i &= u_\alpha \frac{\hat{r}_1(\hat{d}_1)}{\hat{r}_0(\hat{d}_1)}\sqrt{\frac{\hat{r}_2(\hat{d}_1)-\hat{r}_1^2(\hat{d}_1)}{n\hat{r}_1^2(\hat{d}_1)} + \frac{1}{2mh\hat{f}_{\tilde{X}}(s+\mu)}}\,.
\end{aligned} \tag{6.4.4}$$

The discussion on the bandwidth h at the end of section 4.3 applies here as well. For computation of the test limit with moderate values of m the bandwidth h derived in section 4.2.6 may still be applied. For larger values of m the bandwidth should be taken somewhat larger. The bandwidth $\bar{h}$ to estimate $f'_{\tilde{X}}(s+\mu)$ should be taken as in section 4.2.6.

6.5 An application in semiconductor industry

As an example of the theory of this chapter we consider a television color decoder TDA9162/N1 manufactured at Philips' consumer IC plant at Nijmegen. From several characteristics specified we choose one that should be below $s = 670.0$. For $n = 44$ products both a standard measurement and a precise laboratory measurement is carried out. From production $m = 2732$ observations are available. Two histograms summarize the data.

Figure 6.5.1 Histogram for production data

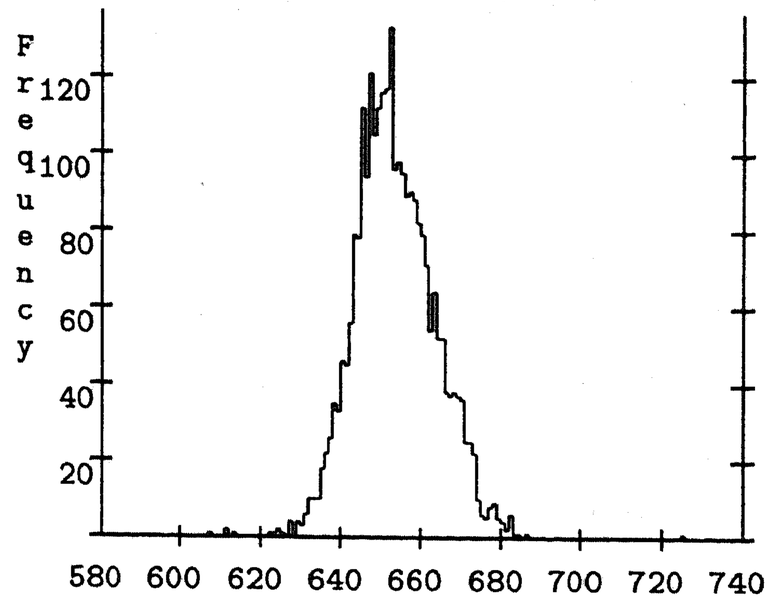
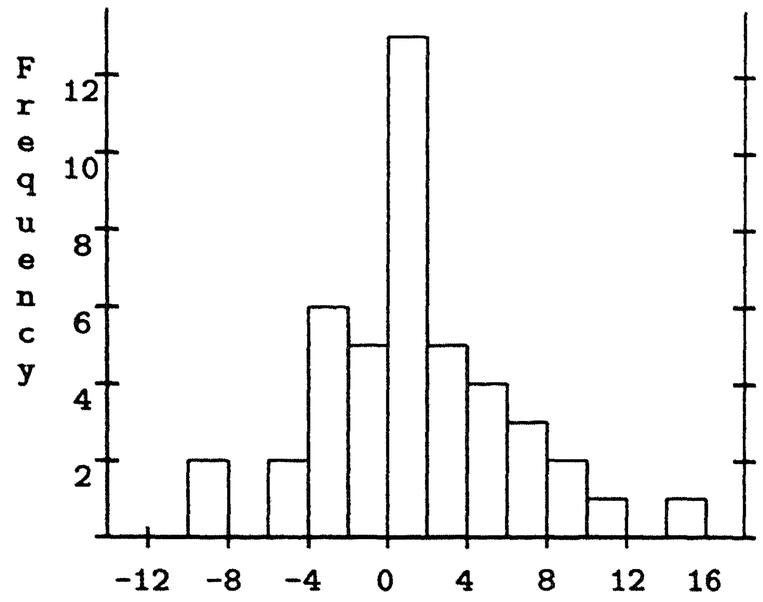


Figure 6.5.2 Histogram for the measurement error



The histograms indicate that the assumption of normality for neither the characteristic nor the measurement error is justified. A plot (not presented here) of the values of X_i against the values of U_i $(i = 1, \ldots, n)$ shows that it is reasonable to assume that the measurement error and the inspected characteristic are independent.

The estimated mean of the measurement error is equal to 1.38, therefore we estimate the density (and its derivative) at $s + \hat{\mu} = 671.38$. To apply Rosenblatt's estimators (section 4.2.4) we determine the bandwidths first, as in section 4.2.6. The sample mean and sample standard deviation of the $\tilde{X}$−observations are 653.6 and 10.02, respectively, leading to $h = 0.666$ and $\bar{h} = 2.584$. We find $\hat{f}_{\tilde{X}}(s + \hat{\mu}) = 0.0096$ and $\hat{f}'_{\tilde{X}}(s + \hat{\mu}) = -0.002$.

For the computation of the test limit itself we refer to section 2.4. Suppose the bound on the consumer loss is $\gamma = 100$ ppm. Then in case of unbiased estimation we find $\hat{t}_u = 661.92$. (cf. (6.3.41)). As a consequence of section 6.3.6, the part of the correction $\hat{c}_u$ (cf. (6.3.42)) which corrects for the estimation of $f_{\tilde{X}}$ is omitted. (Its value is 0.006.) If the consumer loss should exceed γ with probability $\alpha = 0.10$ only, we find $\hat{t}_i = 661.70$ (cf. (6.4.3)). (Again we omit the part of the correction term with m in the denominator.)

About the reliability of the test limit we remark the following. The simulation results (section 6.3.3) show that the accuracy of the test limit depends on the value of $r_0(d_1) = P(U < -d_1)$. In the present situation there are two observations to the left of $-\hat{d}_1 = -7.98$. The simulation results indicate that two observations in expectation to the left of the true value of $-d_1$ is sufficient.

References

Abramson, I.S. (1982). "On bandwidth variation in kernel estimates – a square root law", *Annals of Statistics*, 10, 1217-1223.

Albers, W. and Kallenberg, W.C.M. (1994). "A simple approximation to the bivariate normal distribution with large correlation coefficient", *Journal of Multivariate Analysis*, 49, 87-96.

Albers W., Kallenberg W.C.M. and Otten G.D. (1994a). "Accurate test limits with estimated parameters", *Technometrics*, 36, 92-101.

Albers W., Kallenberg W.C.M. and Otten G.D. (1994b). "Setting test limits under prescribed consumer loss", *Metrika*, 41, 163-181.

Albers W., Kallenberg W.C.M. and Otten, G.D. (1994c). "Robust test limits", Technical Report 1170, University of Twente.

Albers W., Kallenberg W.C.M. and Otten G.D. (1994d). "A comparison between the use of parametric density estimators and a kernel estimator to determine test limits", Technical Report 1211, University of Twente.

Bhattacharya, R.N. and Rao, R.R. (1976). *Normal Approximation and Asymptotic Expansions.* Wiley, New York.

Bloch, D.A. and Gastwirth, J.L. (1968). "On a simple estimate of the reciprocal of the density function", *Annals of Mathematical Statistics*, 39, 1083-1085.

Box, G.E.P., and Cox, D.R. (1964). "An analysis of transformations", *J. Royal Statistical Society, series B*, 26, 211-252.

Cox, D.R. and Reid, N.R. (1986). "Parameter orthogonality and approximate conditional inference", *J. Royal Statistical Society, series B*, 49, 1-39.

Cox, D.R. and Wermuth, N. (1991). "A simple approximation for bivariate and trivariate normal integrals", *International Statistical Review*, 59, 263-269.

Easterling, R.G., Johnson, M.E., Bement, T.R. and Nachtsheim, C.J. (1991). "Statistical tolerancing based on consumer's risk considerations", *Journal of Quality Technology*, 23, 1-11.

Fan, J. (1991a). "On the optimal rates of convergence for nonparametric deconvolution problems", *Annals of Statistics*, 19, 1257-1272.

Fan, J. (1991b). "Global behavior of deconvolution kernel estimators", *Statistica Sinica*, 1, 541-551.

Fan, J. (1992). "Deconvolution with supersmooth distributions", *Canadian J. of Statistics*, 10, 155-169.

Grubbs, F.E. and Coon, H.J. (1954). "On setting test limits relative to specification limits", *Industrial Quality Control*, 10, 15-20.

Hall, P. and Sheather, S.J. (1988). "On the distribution of a studentized quantile", *J. Royal Statistical Society, Ser. B*, 50, 381-391.

Härdle, W. (1991). *Smoothing techniques with implementation in S.* Springer-Verlag, New York.

Hinkley, D.V. (1975). "On power transformation to symmetry", *Biometrika*, 62, 101-111.

Ito, K. and McKean, H.P. (1974). *Diffusion processes and their sample paths.* Springer-Verlag, Berlin, Second Printing, Corrected.

Johnson, N.L. (1949). "Systems of frequency curves generated by methods of translation", *Biometrika*, 36, 149-176.

Johnson, N.L. and Kotz, S. (1970). *Distributions in statistics: Continuous univariate distributions* Volume 1, Wiley, New York.

Jones, M.C. (1992). "Estimating densities, quantiles, quantile densities and density quantiles", *Annals of the Institute of Statistical Mathematics*, 44, 721-727.

Kotz, S. and Johnson, N.L. (1985). *Encyclopedia of statistical sciences*, Volume 5. Wiley, New York.

Lehmann, E.L. (1983). *Theory of point estimation.* Wiley, New York.

Mason, D.M., Shorack, G.R. and Wellner, J.A. (1983). "Strong limit theorems for oscillation moduli of the uniform empirical process", *Z. Wahrscheinlichkeittheorie und verwandte Gebiete*, 65, 83-97.

Mee, R.W., Owen, D.B. and Shyu, J.-C. (1986). "Confidence bounds for misclassification probabilities based on data subject to measurement error", *Journal of Quality Technology*, 18, 29-40.

Mee, R.W. (1990). "An improved procedure for screening based on a correlated, normally distributed variable", *Technometrics*, 32, 331-337.

Mullenix, P. (1990). "The capability of capability indices with an application to guardbanding in a test environment", *1990 International Test Conference IEEE*, 907-915.

Parrish, R.S. (1983). "On an integrated approach to member selection and parameter estimation for Pearson distributions", *Computational Statistics & Data Analysis*, 1, 315-323.

Scheffé, H. (1959). *The analysis of variance.* Wiley, New York.

Shorack, G.R. and Wellner, J.A. (1986). *Empirical Processes with Applications to Statistics.* Wiley, New York .

Slifker, J.F. and Shapiro, S.S. (1980). "The Johnson-system: selection and parameter estimation", *Technometrics*, 22, 239-246.

Silverman, B.W. (1986). *Density estimation for statistics and data analysis.* Chapman and Hall, London.

Stuart, A. and Ord, J.K. (1987). *Kendall's Advanced Theory of Statistics*, Volume 1. Griffin, London.

Terrel, G.R. and Scott, D.W. (1992). "Variable kernel density estimation", *Annals of Statistics*, 20, 1236-1265 .

Zhang, C.H. (1990). "Fourier methods for estimating mixing densities and distributions", *Annals of Statistics*, 18, 806-831.