TU/e technische universiteit eindhoven

Lecture Notes
3TU Course Applied Statistics

A. Di Bucchianico

14th May 2008

Contents

List of Figures

LIST OF FIGURES

List of Tables

LIST OF TABLES

Preface

These notes are the lecture notes for the Applied Statistics course. This course is an elective course in the joint Master's programme of the three Dutch technical universities and is also part of the Dutch National Mathematics Master's Programme. The course Applied Statistics has an alternating theme. In the even years the theme is Statistical Process Control, while in the odd years the theme is Survival Analysis.

Statistical Process Control (SPC) is a name to describe a set of statistical tools that have been widely used in industry since the 1950's and lately also in business (in particular, in financial and health care organisations). Students taking a degree in statistics or applied mathematics should therefore be acquainted with the basics of SPC.

Modern applied statistics is unthinkable without software. In my opinion, statisticians should be able to perform both quick analyses using a graphical (Windows) interface and be able to write scripts for custom analyses. For this course, we use the open source statistical software `R`, which is available from `www.r-project.org`. Please note that `R`-scripts are very similar to scripts in `S` and `S-Plus`. Use of other standard software will be demonstrated during the lecture notes, but is not included in these lecture notes. We try to achieve with this course that students

- learn the basics of practical aspects SPC
- learn the mathematical background of the basic procedures in SPC
- learn to discover the drawbacks of standard practices in SPC
- learn to perform analyses and simulations using `R`.

General information on this course will be made available through the web site

`www.win.tue.nl/` $\sim$ `adibucch/2WS10`.

My personal motivation for writing these lecture notes is that I am unaware of a suitable text book on SPC aimed at students with a background in mathematical statistics. Most text books on SPC aim at an audience with limited mathematical background. Exceptions are Kenett and Zacks (1998) and Montgomery (2000) which are both excellent books (the former addresses much more advanced statistical techniques than the latter), but they do not supply enough mathematical background for the present course. Kotz and Johnson (1993) supplies enough mathematical background, but deals with capability analysis only. Czitrom and Spagon (1997) is a nice collection of challenging case studies in SPC that cannot be solved with standard methods.

Finally, I would like to gratefully acknowledge the extensive help of my student assistant Xiaoting Yu for helping me in making solutions to exercises, R procedures and preparing data sets.

Eindhoven, January 31, 2008
Alessandro Di Bucchianico
www.win.tue.nl/∼adibucch

Chapter 1

Short Historical Introduction to SPC

Contents

Statistical Process Control (SPC) is a name to describe a set of statistical tools that have been widely used in industry since the 1950's and in business (in particular, in financial and health care organisations) since the 1990's. It is an important part of quality improvement programmes in industry. A typical traditional setting for SPC is a production line in a factory. Measurements of important quality characteristics are being taken at fixed time points. These measurements are being used to monitor the production process and to take appropriate action when the process is not functioning well. In such cases we speak of an out-of-control situation.

In this chapter we will give a brief overview of the goals and history of SPC, as well as describe the statistical machinery behind SPC.

1.1 Goal of SPC

The ultimate goal of SPC is to monitor variation in production processes. There is a widely used (but somewhat vaguely defined) terminology to describe the variation in processes. Following the terminology of Shewhart, variation in a production process can have two possible causes:

- common causes
- special causes

Common causes refer to natural, inherent variation that cannot be reduced without making changes to the process such as improving equipment or using other machines. Such variation is often considered to be harmless or it may unfeasible for economical or technical reasons to reduce it. Special causes refer to variation caused by external causes such as a broken part of

a machine. Such causes lead to extra unnecessary variation and must therefore be detected and taken away as soon as possible.

A process that is only subject to common causes is said to be in-control. To be a little bit more precise, an in-control process is a process that

- only has natural random fluctuations (common causes)
- is stable
- is predictable.

An process that is not in-control is said to be out-of-control. An important practical implication is that one should avoid making changes to processes that are not in-control, because such changes may not be lasting.

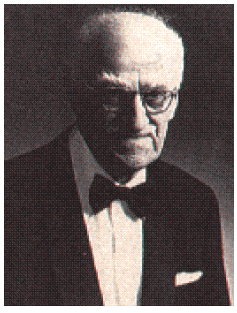
1.2 Brief history of SPC

Deming

Shewhart

Box

Taguchi



Juran

Ishikawa

Figure 1.1: Some famous names in SPC.

Shewhart is usually considered to be the founding father of SPC. As starting point one usually considers the publication of an internal Bell Labs report in 1924. In this note Shewhart described the basic form of what is now called the Shewhart $\overline{X}$ control chart. His subsequent ideas did not catch on with other American companies. One of the few exceptions was Western Electric, where Deming and Juran worked. The famous Western Electric Company 1956 handbook still makes good reading. The breakthrough for SPC techniques came after World

War II when Deming (a former assistant to Shewhart) was hired by the Japanese government to assist in rebuilding the Japanese industry. The enormous successes of the Japanese industry in the second half of the 20^{th} century owes much to the systematic application of SPC, which was advocated with great enthusiasm by Ishikawa. In the early 1950's Box introduced experimental design techniques developed by Fisher and Yates in an agricultural context, in industry, in particular in chemical industries. He made extensive contributions to experimental designs for optimization. Later he moved to the United States and successfully started and led the *Center for Quality and Productivity Improvement* at the University of Wisconsin-Madison. Experimental design was developed in a different way by Taguchi, who successfully introduced it to engineers. It was not until the 1980's that American companies after severe financial losses began thinking of introducing SPC. Motorola became famous by starting the Six Sigma approach, a quality improvement programme that heavily relies on statistics. Both the Taguchi and the Six Sigma approach are being used world-wide on a large scale. The developments in Europe are lagging behind. An important European initiative is ENBIS, the European Network for Business and Industrial Statistics (`www.enbis.org`. This initiative by Bisgaard, a former successor of Box at the University of Wisconsin-Madison, successfully brings together statistical practitioners and trained statisticians from industry and academia.

1.3 Statistical tools in SPC

Since users of SPC often do not have a background in statistics, there is a strong tendency to use simple statistical tools, sometimes denoted by fancy names like "The Magnificent Seven". In many cases this leads to unnecessary oversimplifications and poor statistical analyses. It is curious that several practises like the use of the range instead of the standard deviation are still being advocated, although the original reason (ease of calculation) has long ceased to be relevant. For more information on this topic we refer to Stoumbos et al. (2000), Woodall and Montgomery (1999) and Woodall (2000).

As with all statistical analyses, graphical methods are important and should be used for a first inspection. Scatterplots and histograms are often used. Another simple tool is the Pareto chart. This chart (introduced by Juran) is simple bar chart that shows in an ordered way the most important causes for errors or excessive variation. The rationale is the so-called $80 - 20$ rule (often attributed to the econometrician Pareto), which says that 80% of damage comes from 20% of causes. Histograms are still being used to assess normality of data, although it is widely known that the choice of bins heavily influences the shape of the histogram (see Section 3.8). The use of quantile plots like the normal probability plot or density estimators like kernel density estimators is not widespread (for an exception, see Rodriguez (1992)). In SPC it is often useful to obtain accurate estimate of tail probabilities and quantiles. A typical example are the so-called capability analyses, where process variation is compared with specifications in order to judge whether a production process is capable of meeting specifications (see Chapter 3). This judgement is often based on so-called capability indices, which are often simple parametric estimators of tail probabilities that may not be reliable in practice. Alternatives are tolerance intervals (both parametric and distribution-free; see Section 3.6) that are interval estimators containing a specified part of a distribution or tail estimators based on extreme value theory (although these often require sample sizes that are not available in industrial practice).

Control charts are the most widely known tools in SPC. They are loosely speaking a graph-

ical way to perform repeated hypothesis testing. Shewhart control charts can be interpreted as simple sequential likelihood ration tests depending on the current statistics (a individual observation or a group mean or standard deviation) only, while the Cumulative Sum (CUSUM) charts introduced by Page in the 1950's can be seen as sequential generalized likelihood ratio tests based on all data. These charts also appear change point analysis. The Exponentially Weighted Moving Average (EWMA) charts introduced in Girshick and Rubin (1952), Roberts (1959) and Shiryaev (1963) were inspired by Bayesian statistics, but the procedures have a time series flavour.

In this course we will concentrate on univariate SPC. However, in industrial practice often several quality characteristics are needed to accurately monitor a production processes. These characteristics are often correlated. Hence, techniques from multivariate statistics like principal components are required. For a readable overview of this aspect, we refer to Qin (2003).

General papers with more information about developments of statistical techniques for SPC include Crowder et al. (1997), Hawkins et al. (2003), Lai (1995), Lai (2001), and Palm et al. (1997).

1.4 Exercises

Read the paper Provost and Norman (1990) and answer the following questions.

Exercise 1.1 *Describe the three methods of managing variation mentioned in the text.*

Exercise 1.2 *Describe the different test methods mentioned in this text.*

Exercise 1.3 *Why was variation not a critical issue in the period before 1700?*

Exercise 1.4 *When and why did variation become a critical issue?*

Exercise 1.5 *What was the reason that standards for inspection were formulated at the end of the* 19*th century?*

Exercise 1.6 *Explain the go/no-go principle.*

Exercise 1.7 *What is the goal of setting tolerances and specifications?*

Exercise 1.8 *Explain the term interchangeability and give an example of a "modern" product with interchangeable components*

Exercise 1.9 *Explain the relation between interchangeable components and variation.*

Exercise 1.10 *What is the goal of a Shewhart control chart?*

Exercise 1.11 *In what sense does Shewhart's approach differ from the approach based on setting specifications?*

Exercise 1.12 *What were the reasons for Shewhart to put control limits at a distance of* 3 *standard deviations of the target value?*

Exercise 1.13 *Explain the idea behind Taguchi's quadratic loss function.*

Exercise 1.14 *Explain what is meant by tolerance stack-up (see p. 44, 1st paragraph below Figure 3).*

Chapter 2

A Short Introduction to R

Contents

2.1 The R initiative

There are many commercial statistical softwares available. Well-known examples include SAS, SPSS, S-Plus, Minitab, Statgraphics, GLIM, and Genstat. Usually there is a GUI (graphical user interface). Some softwares allow to perform analyses using the GUI as well as by typing commands on a command line. Larger analyses may be performed by executing scripts.

In the 1970's Chambers of AT&T started to develop a computer language (called `S`) that would be able to perform well-structured analyses. A commercial version of `S` appeared in the early 1990's under the name S-Plus. Ihaka and Gentleman developed a little bit later a free, open source language `R` which is very similar to `S`. Currently `R` is being maintained and continuously improved by a group of world class experts in computational statistics. Hence, `R` has gained enormous popularity among various groups of statisticians, including mathematical statisticians and biostatisticians. The `R`-project has its own web page at `www.r-project.org`. Downloads are available through the CRAN (Comprehensive R Archive Network) at `www.cran.r-project.org`.

2.2 R basics

There are several tutorials available inside R through Help or can be found on the web, *e.g.* through CRAN. The R reference card is very useful. Within R further help can be obtained by typing `help` when one knows the name of a function (*e.g.*, `help(pnorm)`) or `help.search` when one only keywords (*e.g.*, `help.search("normal distribution")`).

2.2.1 Data files

Assignment are read from right to left using the $\leftarrow$ operator:

$$\texttt{a} < -2 + \texttt{sqrt}(5)$$

There are several form of data objects. Vectors can be formed using the `c` operator (concatenation), *e.g.*,

$$\texttt{a} < -\texttt{c}(1, 2, 3, 10)$$

yields a vector consisting of 4 numbers. Vectors may be partitioned into matrices by using the `matrix` command, *e.g.*,

$$\texttt{matrix}(\texttt{c}(1, 2, 3, 4, 5, 6), 2, 3, \texttt{byrow} = \texttt{T})$$

creates a matrix with 2 rows and 3 columns.

The working directory may be set by `setwd` and displayed by `getwd()` (this will return an empty answer if no working directory has been set). Please note that directory names should be written with quotes and that the Unix notation must be used even if R is being used under Windows, *e.g.* `setwd("F:/2WS10")`. A data set may be turned into the default data set by using the command `attach`; the companion command `detach`. Data files on a local file system may be read through the command `scan` when there is only one column or otherwise by

$$\texttt{read.table}(\text{“}\texttt{file.txt}''\texttt{, header} = \texttt{TRUE})$$

Both `read.table` and `scan` can read data files from the WWW (do not forget to put quotes around the complete URL).

Parts of data files may be extracted by using so-called subscripting. The command `d[r,]` yields the rth row of object d, while $d[, c]$ yields the cth column of object d. The entry in row r and column c of object d can be retrieved by using `d[r,c]`. Extracting elements that satisfy a certain condition may also be extracted by subscripting. *E.g.*, `d[d<20]` yields all elements of d that do not exceed 20, while `d["age"]` extracts the column with name "age" (note the double quotes) from object d. The number of elements of an object d is given by `length(d)`.

2.2.2 Probability distributions in R.

Standard probability distributions have short names in R as given by Table 2.1. Several probability functions are available. Their names consists of two parts: the first part is the name of the function (see Table 2.2), while the second part is the name as in Table 2.1. *E.g.*, a sample of size 10 from an exponential distribution with mean 3 is generated in R by `rexp(10,1/3)` (R uses the failure intensity instead of the mean as parameter).
Table 2.3 lists several goodness-of-fit tests that are available in R, either directly or via the package `nortest` (see Subsection 2.2.4).

Distribution	Name in R
normal	norm
(non-central) Student T	t
Weibull	weibull
exponential	exp
(non-central) χ^2	chisq
Gamma	gamma
F	f

Table 2.1: Names of probability distributions in R.

Function	Name in R
d	density
p	probability = cumulative distribution function
q	quantile
r	random numbers

Table 2.2: Names of probability functions in R.

2.2.3 Graphics in R

The standard procedure in R to make 1D and 2D plots is plot. Histogram are available through hist. These commands can be supplied with options to allow for titles, subtitles, and labels on the x-axes:

```
plot(data,main=‘‘Title’’,sub=‘‘Subtitle’’,xlab=‘‘X-axis’’,ylab=‘‘Y-axis’’)
```

Quantile-quantile plots are available through qqplot, while qqnorm yields a plot of the quantiles of a data set against the quantiles of a fitted normal distribution (normal probability plot). A Box-and-Whisker plot is also available for exploratory data analysis through boxplot (if the data set is a data frame like produced by read.table, then multiple Box-and-Whisker plots are produced). The empirical cumulative distribution function is available through ecdf. Kernel density estimators are available through density. Graphics can be saved to files by choosing File and Save as in the menu of the R console.

2.2.4 Libraries in R

Extensions to the basic functions are available through libraries of functions. In the Windows interface of R, these libraries can be loaded or installed by choosing the option Packages in

Test	Name in R	Package
Shapiro-Wilks	shapiro.test	stats
Kolmogorov (1-sample)	ks.test	stats
Smirnov (2-sample)	ks.test	stats
Anderson-Darling	ad.test	nortest
Cramér-von Mises test	cvm.test	nortest
Lilliefors test	lillie.test	nortest

Table 2.3: Names of goodness-of-fit tests in R.

the menu. Libraries may also contain ways to improve exchange of files with other software like Matlab or WinEdt. Examples of useful libraries include:

`survival`: library for survival analysis (Cox proportional hazards etc.)

`qcc`: SPC library

`RWinEdt`: interface with editor WinEdt (recommended for testing scripts, see also Subsection 2.2.7)

`RMatlab`: a bidirectional interface with Matlab

2.2.5 Basic statistics in R

Summary statistics of a data set can be obtained from `summary`, or by using individual commands like `mean`, `sd`, `mad`, and `IQR`. Standard hypothesis tests are also available, *e.g.*, `t.test` yields the standard tests for means of one or two normal samples.

2.2.6 Functions in R

Analyses that have to be performed often can be put in the form of functions, *e.g.*.,

```
simple <- function(data, mean = 0, alpha = 0.05)
          {hist(data), t.test(data, conf.level = alpha, mu = mean)}
```

This means that typing `simple(data,4)` uses the default value $\alpha = 0.05$ and tests the null hypothesis $\mu = 4$.

2.2.7 Editors for R

Instead of pasting commands from an ordinary text editor into the R console, one may also use WinEdt as R editor by using the RWinEdt package. Another choice is Tinn-R, which is a free R editor that helps the user by showing R syntaxis while typing.

2.3 Exercises

Exercise 2.1 *Calculate in* R *the probability that a random variable with a* χ^2 *distribution with* 3 *degrees of freedom is larger than* 5.2.

Exercise 2.2 *Compute the* 0.99 *quantile of the standard normal distribution.*

Exercise 2.3 *Generate* 20 *random numbers from an exponential distribution with mean* 3. *How can you check that you choose the right parametrization?*

Exercise 2.4 *Generate* 40 *random numbers from a normal distribution with mean* 10 *and variance* 2. *Make a histogram and play with the number of bins to convince yourself of the influence on the shape. Check normality through a normal probability plot, a plot of the density and an appropriate goodness-of-fit test. Also test the mean and variance of the sample.*

Exercise 2.5 *A telecommunication company has entered the market for mobile phones in a new country. The company's marketing manager conducts a survey of* 200 *new subscribers for mobile phones. The data of her survey are in the data set* `telephone.txt`*, which contains the first month's bills. Make an appropriate plot of this data set. What information can be extracted from these data? What marketing advice would you give to the marketing manager?*

Exercise 2.6 *Dried eggs are being sold in cans. Each can contains two different types of eggs. As part of a quality control programme, the fat content of eggs is being investigated. The investigation is divided over* 6 *different laboratories. Every laboratory receives the same number of eggs of both types. Testing the fat content of eggs is destructive, so that each egg can only be investigated once. Since measuring the fat content is time consuming, the measurements are divided over* 2 *laboratory assistants within each laboratory.*

A quality manager applies a certain statistical procedure and claims that there is a significant difference between the fat contents measured by the laboratories. This report causes confusion, since the 6 *laboratories all have a good reputation and there is no reason to expect large variation in fat content of the eggs. Find an explanation by making appropriate plots of the data set* `eggs.txt`*.*

Exercise 2.7 *Supermarket chain ATOHIGH has two shops A and B in a certain town. Both stores are similar with respect to size, lay-out, number of customers and spending per customer. The populations of the parts of town of the two shops are quite similar. Management decides to experiment with the lay-out of shop A, including different lighting. After some time a survey is performed on a Saturday afternoon among* 100 *customers in both shops. The survey is restricted to customers which are part of a family of at least* 3 *persons. The data set* `supermarket.txt` *contains the data of this survey. Perform a statistical analysis of this data set, both by producing appropriate plots and by computing appropriate summary statistics. What can you conclude about the spending in both shops?*

Exercise 2.8 *Write an* `R` *function that computes the empirical survivor function given a data set. The empirical survivor function at a point x counts the proportion of observations exceeding x.*

Exercise 2.9 *Write an* `R` *function that produces a confidence interval for the variance from a sample from the normal distribution.*

Exercise 2.10 *Write a function that plots a plot of the values of a data set against the quantiles of a given Weibull distribution. Test your function with a*

1. *a random sample of size* 30 *from a Weibull distribution with shape parameter* 3 *and scale parameter* 2*.*

2. *a random sample of size* 30 *from a Gamma distribution with shape parameter* 7.57 *and rate parameter* 0.235*. Check using formulas that the mean and variance of this Gamma distribution is approximately equal to the mean and variance of the Weibull distribution in 1).*

Exercise 2.11 *In the 19th century French physicists Fizeau and Foucault independently invented ways to measure the speed of light. Foucault's method turned out to be the most accurate*

one. The Foucault method is based on fast rotation mirrors. The American physicists Michelson and Newcomb improved Foucault's method. The data set `light.txt` *contains measurements by Newcomb from 1882. The data are coded times needed by light to travel a distance of* 3721 *metres across a river and back. The coding of these measurements was as follows: from the original times in microseconds measured by Newcomb first* 24.8 *was subtracted, after which the results were multiplied with* 1000*.*

1. *Compute a* 95% *confidence interval for the average speed of light.*
2. *State the assumptions on which the above interval is based. Check your assumptions with a suitable plot. What is your conclusion?*
3. *Make a box plot of the data. What do you observe?*
4. *Plot the data against the observation number. Provide a possible explanation for what you observed in part c).*
5. *Recompute a* 95% *confidence interval for the average speed of light using your findings of part c).*
6. *Use the WWW to find the currently most accurate value for the speed of light. Test whether this new value is consistent with the measurements of Newcomb.*

Exercise 2.12 *To improve rain fall in dry areas, an experiment was carried out with* 52 *clouds. Scientists investigated whether the addition of silver nitrate has a positive effect on rainfall. They chose* 26 *out of a sample of* 52 *clouds and seeded it with silver nitrate. The remaining* 26 *clouds were not treated with silver nitrate. The data set* `clouds.txt` *records the rainfall in feet per acre.*

1. *Apply a t-test to investigate whether the average rainfall increased by adding silver nitrate. Argue whether the data are paired or not.*
2. *The t-test assumes normally distributed data. Check both graphically and by using a formal test whether the data are normally distributed. What conclusion should you draw on the test performed in part a)?*
3. *Because the scientists thought that addition of silver nitrate should have a multiplicative effect, they suggested transforming the data. What transformation should be a logical candidate? What effect does this transformation have on the normality assumption? Apply the t-test of part a) on the transformed data. What is your final conclusion on the addition of silver nitrate?*

Chapter 3

Process Capability Analysis

Contents

In this chapter we present the mathematical background of process capability analysis. We focus in particular on the distributional properties of the capability indices C_p and C_{pk} and related topics like tolerance intervals and density estimation. We refer to Kotz and Johnson (1993) for further information on the mathematical background of capability indices and to Kotz and Lovelace (1998) for further information on applications of capability indices.

3.1 Example of a process capability analysis

In this section we first present the basic ideas behind process capability analyses and then illustrate these ideas with a small case study. In particular, we show how to perform the analysis using the `qcc` package for `R`.

Usually the items produced by a production process have to meet customer requirements. Requirements may also be set by the government through legislation. It is therefore important to know beforehand whether the inherent variation within the production process is such that it can meet these requirements. Requirements are usually defined as specification limits. We denote the upper specification limit by USL and the lower specification limit by LSL. Products that fall outside specification limits are called non-conforming. Within SPC an

investigation whether the process can meet the requirements is called a Process Capability Analysis. Basically one wishes to determine $1 - P(LSL < X < USL)$, the probability of a non-conforming item where X denotes the essential quality characteristic (of course, in practice one usually has more than one important quality characteristic). Such a capability analysis is usually performed at the end of the so-called Phase I of a production process, *e.g.*, the phase in which a pilot study is performed in order to decide whether it is possible to start full production.

A straightforward way to describe process capability would be to use the sample mean and sample standard deviation. As natural bandwidth of a process one usually takes 6σ, which implicitly assumes a normal distribution. For a random variable X which is normally distributed X with parameters μ and σ^2, it holds that $P(X > \mu + 3\sigma) = 0.00135$, and thus $P(\mu - 3\sigma < X < \mu + 3\sigma) = 0.9973$. This is a fairly arbitrary, but widely accepted choice.

Whether a process fits within the 6σ-bandwidth, is often indicated in industry by so-called Process Capability Indices. Several major companies request from their suppliers detailed documentation proving that the production processes of the supplier has certain minimal values for the process capability indices C_p and C_{pk} defined below. These minimal values used to be 1.33 or 1.67, but increasing demands on quality often requires values larger than 2. The simplest capability index is called C_p (in order to avoid confusion with Mallow's regression diagnostic value C_p one sometimes uses P_p) and is defined as

$$C_p = \frac{USL - LSL}{6\sigma}.$$

Note that this quantity has the advantage of being dimensionless. The quantity $1/C_p$ is known as the capability ratio (often abbreviated as CR). It will be convenient to write

$$d = \frac{1}{2}(USL - LSL).$$

If the process is not centred, then the expected proportion of non-conforming items will be higher than the value of C_p seems to indicate. Therefore the following index has been introduced for non-centred processes:

$$C_{pk} = \min\left(\frac{USL - \mu}{3\sigma}, \frac{\mu - LSL}{3\sigma}\right). \tag{3.1}$$

Usually it is technologically relatively easy to shift the mean of a quality characteristic, while reducing the variance requires a lot of effort (usually this involves a major change of the production process).

We now illustrate these concepts in a small case study. The most important quality measure of steel-alloy products is hardness. At a steel factory, a production line for a new product has been tested in a trial run of 27 products (see first column of the data set `steelhardness.txt`). Customers require that hardness of the products is between 65 and 49, with a desired nominal value of 56. The measurements were obtained in rational subgroups of size 3 (more on rational subgroups in Chapter 4). The goal of the capability analysis is to assess whether the process is sufficiently capable to meet the given specifications. A process capability analysis consists of the following steps:

1. general inspection of the data: distribution of the data, outliers (see also Section 3.7)

2. check whether the process was statistically in-control, *i.e.*, are all observations from the same distribution

3. remove subgroups that were not in-control

4. compute confidence intervals for the capability indices of interest (in any case, C_p and C_{pk}; see Sections 3.3 and 3.5 for details)

5. use the confidence intervals to assess whether the process is sufficiently capable (with respect to both actual capability and potential capability)

Since the standard theory requires normal distributions, we start with performing to test for normality using a Box-and-Whisker plot (an informal check for outliers), a normal probability plot, a plot of a kernel density estimate and a goodness-of-fit test see Section 3.7 for more details).

We illustrate these steps using the `qcc` package for `R`. We could perform normality testing outside the `qcc` package, but is more efficient to use as much as possible from this package. Therefore the first step is to create a special `qcc` object, since this required for all calculations in the `qcc` package (which is completely in line with the object orientated philosophy behind `R`).

```
setwd("D:/2WS10")
# point R to the directory where the data is (change this to your directory)

library(qcc)
# load additional library, if not done automatically

steel <- read.table("steelhardness.txt",nrows=27,header=TRUE,fill=TRUE)
# fill is needed since columns have different lengths

boxplot(steel$original,horizontal=TRUE,col="blue",
main="Box-and-whisker plot of original steel data")

plot(density(steel$original),main="Kernel density estimate for original steel data",
col="blue",lwd=3)
# lwd makes lines thicker

qqnorm(steel$original,main="Normal probability plot of original steel data",
pch=19,cex=2,fg="darkgreen")
qqline(steel$original,lwd=3,col="blue",lty="dashed")
# lty is line type, cex = size bullets, pch = choice of points

shapiro.test(steel$original)
# standard normality test

samplenumber <- rep(1:9, each=3) # generate sample numbers
steeldata <- qcc.groups(steel$original,samplenumber)
```

```
steelqcc <- qcc(steeldata,type="xbar",plot=FALSE,confidence.level=0.95,
nsigmas=3)

process.capability.sixpack(steelqcc, spec.limits=c(49,65),target=56,nsigmas=3)

process.capability(steelqcc, spec.limits=c(49,65),target=56,nsigmas=3)
```

Normality checks on the individual data using density kernel estimators, the normal probability plot and the Shapiro-Wilks test indicate no serious problems.

Control charts should be used to check whether the process was in control during the pilot study. They will be treated in more detail in Chapter 4. It suffices here to check that all observation fall within the so-called control limits. Both control charts seem to be in-control so it is allowed to perform the capability study using the standard capability indices which are based on normality. The values for C_p are not impressive but reasonable (note the wide range of the confidence interval, which reflects that we have a limited sample at our disposal to deduce capability). The main problem is with C_{pk}, which indicates that the mean (location) of the hardness is the main problem, rather than the variability.

Steel production is a highly competitive market, so the customer is in a strong position for negotiating prices, specifications and delivery times. One of the customers with whom a good contractual relationship can be established wants to buy steel-alloy products with little variation in hardness. The required target value for the hardness should be 53 and the deviations from this target value should be not larger than 5 hardness units. Changing the specification limits has no effect at all whether the process was in-control or not. We thus may use the findings above with respect to being in control etc. .

```
process.capability(steelqcc, spec.limits=c(48,58),target=53,nsigmas=3)
```

We now see that the values of C_p and C_{pk} are quite close, indicating that the process is well centred with respect to the specification limits. The main problem now is the variability, which needs to be decreased significantly.

The R&D department of the company performs a designed experiment (see second column of the data set `steelhardness.txt`). The department claims that the production process has improved. We perform a statistical test to verify this.

```
var.test(steel$original,steel$improved,alternative="greater",
conf.level=0.95)
```

The test shows that the variance has decreased significantly. So the R&D department has done a good job. This also shows in the capability analysis for the improved production process. Unfortunately, the sample of the improved production process is too small to be confident that the improved process meets the specifications (reflected in the wide ranges of the confidence intervals).

3.2 Basic properties of capability indices

The capability index C_p is useful if the process is centred around the middle of the specification interval. If that is the case, then the proportion of non-conforming items of a normally

distributed characteristic X equals

$$\begin{aligned} 1 - P(LSL < X < USL) &= P(X < LSL) + P(X > USL) \\ &= 2P(X < LSL) \\ &= 2P\left(\frac{X - (USL + LSL)/2}{\sigma} < \frac{-(USL - LSL)/2}{\sigma}\right) \\ &= 2\Phi(-d/\sigma) \\ &= 2\Phi(-3C_p), \end{aligned} \tag{3.2}$$

where Φ denotes the distribution function of the standard normal distribution. Using the identity $\min(a, b) = \frac{1}{2}(|a + b| - |a - b|)$, we obtain the following representations:

$$\begin{aligned} C_{pk} &= \frac{\min(USL - \mu, \mu - LSL)}{3\sigma} \\ &= \frac{d - |\mu - \frac{1}{2}(LSL + USL)|}{3\sigma} \\ &= \left(1 - \frac{|\mu - \frac{1}{2}(LSL + USL)|}{d}\right) C_p. \end{aligned} \tag{3.3}$$

We immediately read off from (3.3) that $C_p \geq C_{pk}$. Moreover, since $C_p = d/(3\sigma)$, we also have that $C_p = C_{pk}$ if and only the process mean equals (LSL+USL)/2. The notation

$$k = \frac{|\mu - \frac{1}{2}(LSL + USL)|}{d} \tag{3.4}$$

is often used. It is also possible to define C_{pk} in terms of a target value T (also called *nominal value* instead of the process mean μ. This is done by simply replacing μ by T in (3.1). In this case one measures everything with respect to a desired value for the process mean rather than the actual process mean.

The expected proportion non-conforming items for a non-centred process with normal distribution can be defined in terms of C_p and C_{pk} as follows (cf. (3.2). The expected proportion equals $\Phi\left(\frac{LSL-\mu}{\sigma}\right)+1-\Phi\left(\frac{USL-\mu}{\sigma}\right)$. Now assume that $\frac{1}{2}(USL+LSL) \leq \mu \leq USL$. Then $C_{pk} = \frac{USL-\mu}{3\sigma}$ and

$$\frac{LSL - \mu}{3\sigma} = \frac{(USL - \mu) - (USL - LSL)}{3\sigma} = C_{pk} - 2C_p \leq -C_{pk},$$

because $C_p \geq C_{pk}$. Hence, the expected proportion non-conforming items can be expressed as

$$1 - P(LSL < X < USL) = \Phi(-3(2C_p - C_{pk})) + \Phi(-3C_{pk}). \tag{3.5}$$

In the Six Sigma approach, a successful quality programme that heavily relies on statistics, additional terminology is used. The fraction of non-conforming items is presented in *dpm* (defects per million) or *dmo* (defects per million opportunities). It is assumed that no production process can be stable. A basic assumption, which is often questioned because no clear motivation is available, is that process means may shift as much as 1.5σ. The capability indices that we defined above correspond to what are called *short-term capability indices*. Long-term capability indices correspond to the case when the process mean that is not at $(USL + LSL)/2$, but has shifted away over a distance of 1.5σ.

3.3 Parametric estimation of capability indices

In this section we first recall some general estimation principles. These principles will be used to obtain (optimal) estimators for capability indices from normal data. We will assume throughout this section that the available data is a sample from a normal distribution.

Definition 3.3.1 (Induced likelihood function) *Let P be a probability distribution depending on parameters $\theta_1, \ldots, \theta_k$, where $\theta = (\theta_1, \ldots, \theta_k)$ ranges over a set $\Theta \subset \mathbb{R}^k$ and let $L(\theta; x_1, \ldots, x_n)$ be the likelihood function of a sample $X_1, \ldots, X_n$ from f. Let τ be an arbitrary function on Θ. The function $M_\tau(\xi; x_1, \ldots, x_n) := \sup_{\theta \in \Theta\,:\,\tau(\theta)=\xi} L(\theta; x_1, \ldots, x_n)$ is the **induced likelihood function by** τ. Any number $\xi \in \Theta$ that maximizes M_τ is said to be an MLE of $\tau(\theta)$.*

The rationale behind this definition is as follows. Estimation of θ is obtained by maximizing the likelihood function $L(\theta; x_1, \ldots, x_n)$ as function of θ for fixed $x_1, \ldots, x_n$, while estimation of $\tau(\theta)$ is obtained by maximizing the induced likelihood function $L(\tau(\theta); x_1, \ldots, x_n)$ as function of $\tau(\theta)$ for fixed $x_1, \ldots, x_n$.

The following theorem from Zehna (1966) describes a useful invariance property of Maximum Likelihood estimators. Note that the theorem does not require any assumption on the function τ.

Theorem 3.3.2 (Invariance Principle) *Let P be a distribution depending on parameters $\theta_1, \ldots, \theta_k$ and let $\widehat{\Theta} = (\widehat{\Theta_1}, \ldots, \widehat{\Theta_k})$ be an MLE of $(\theta_1, \ldots, \theta_k)$. If τ is an arbitrary function with domain Θ, then $\tau(\widehat{\Theta})$ is an MLE of $\tau((\theta_1, \ldots, \theta_k))$. If moreover the MLE $(\widehat{\Theta})$ is unique, then $\tau(\widehat{\Theta})$ is unique too.*

Proof: Define $\tau^{-1}(\xi) := \{\theta \in \Theta \mid \tau(\theta) = \xi\}$ for any $\xi \in \Theta$. Obviously, $\theta \in \tau^{-1}(\tau(\theta))$ for all $\theta \in \Theta$. Hence, we have for any $\xi \in \Theta$ that

$$\begin{aligned} M_\tau(\xi; x_1, \ldots, x_n) &= \sup_{\theta \in \tau^{-1}(\xi)} L(\theta; x_1, \ldots, x_n) \\ &\leq \sup_{\theta \in \Theta} L(\theta; x_1, \ldots, x_n) \\ &= L(\widehat{\Theta}; x_1, \ldots, x_n) \\ &= \sup_{\theta \in \tau^{-1}\left(\tau(\widehat{\Theta})\right)} L(\theta; x_1, \ldots, x_n) \\ &= M_\tau(\tau\left(\widehat{\Theta}\right); x_1, \ldots, x_n). \end{aligned}$$

Thus $\tau(\widehat{\Theta})$ maximizes the induced likelihood function, as required. Inspection of the proof reveals that if $\widehat{\Theta}$ is the unique MLE of $(\theta_1, \ldots, \theta_k)$, then $\tau(\widehat{\Theta})$ is the unique MLE of $\tau((\theta_1, \ldots, \theta_k))$. □

We now give some examples that illustrate how to use this invariance property in order to obtain an MLE of a function of a parameter.

Examples 3.3.3 *Let $X, X_1, X_2, \ldots, X_n$ be independent random variables, each distributed according to the normal distribution with parameters μ and σ^2. Let Z be a standard normal random variable with distribution function Φ. Recall that the ML estimators for μ and σ^2 are $\widehat{\mu} = \overline{X}$ and $\widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X})^2$, respectively.*

a) Suppose we want to estimate σ, where μ is unknown. Theorem 3.3.2 with $\Theta = (0, \infty)$ and $\tau(x) = \sqrt{x}$ yields that the MLE $\widehat{\sigma}$ of σ equals $\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}$.

b) Suppose we want to estimate $1/\sigma$, where μ is unknown. Theorem 3.3.2 with $\Theta = (0, \infty)$ and $\tau(x) = 1/\sqrt{x}$ yields that the MLE of σ equals $\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2\right)^{-1/2}$. The MLE's for C_p and C_{pk} easily follow from the MLE of $1/\sigma$ and are given by $\frac{USL-LSL}{6\hat{\sigma}}$ and $\frac{\min(USL-\overline{X}, \overline{X}-LSL)}{3\hat{\sigma}}$.

c) Let p be an arbitrary number between 0 and 1 and assume that both μ and σ^2 are unknown. Suppose that we want to estimate the p-th quantile of X, that is we want to estimate the unique number x_p such that $P(X \leq x_p) = p$. Since

$$p = P(X \leq x_p) = P\left(Z \leq \frac{x_p - \mu}{\sigma}\right) = \Phi\left(\frac{x_p - \mu}{\sigma}\right),$$

it follows that $x_p = \mu + z_p\,\sigma$, where $z_p := \Phi^{-1}(p)$. Thus Theorem 3.3.2 with $\Theta = \mathbb{R} \times (0, \infty)$ and $\tau(x, y) = x + z_p\sqrt{y}$ yields that the MLE of x_p equals $\overline{X} + z_p\,\widehat{\sigma}$, where $\widehat{\sigma}$ is as in a).

d) Let $a < b$ be arbitrary real numbers and assume that μ is unknown and that σ^2 is known. Suppose we want to estimate $P(a < X < b) = F(b) - F(a)$. Since

$$P(a < X < b) = P\left(\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}\right) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right), \tag{3.6}$$

Theorem 3.3.2 with $\Theta = \mathbb{R}$ and $\tau(x) = \Phi\left(\frac{b-x}{\sigma}\right) - \Phi\left(\frac{a-x}{\sigma}\right)$ yields that the MLE for $P(a < X < b)$ equals

$$\Phi\left(\frac{b-\overline{X}}{\sigma}\right) - \Phi\left(\frac{a-\overline{X}}{\sigma}\right). \tag{3.7}$$

e) Let $a < b$ be arbitrary real numbers and assume that both μ and σ^2 are unknown. Suppose we want to estimate $P(a < X < b) = F(b) - F(a)$. Theorem 3.3.2 with $\Theta = \mathbb{R} \times (0, \infty)$ and $\tau(x, y) = \Phi\left(\frac{b-x}{\sqrt{y}}\right) - \Phi\left(\frac{a-x}{\sqrt{y}}\right)$ yields that the MLE for $P(a < X < b)$ equals

$$\Phi\left(\frac{b-\overline{X}}{\widehat{\sigma}}\right) - \Phi\left(\frac{a-\overline{X}}{\widehat{\sigma}}\right), \tag{3.8}$$

where $\widehat{\sigma}$ is as in a).

Note that the estimators in d) and e) are biased (cf. Exercise 3.4). We now set out to show a systematic way to reduce the variance of unbiased estimators. As an application, we will find an unbiased estimator with minimum variance for all values of μ and σ^2 (UMVU estimator) for the proportion of non-conforming items. We will first use the Rao-Blackwell Theorem to

improve upon a simple unbiased estimator and then apply the Lehmann-Scheffé Theorem to prove optimality of this estimator.

In order to state these theorems, we first need the concepts of sufficiency and completeness. A precise and general statement requires measure theory. Roughly speaking, a statistic is sufficient for a certain parameter if one cannot learn more about this parameter from the sample by knowing only the values of the sufficient statistics rather than knowing all values of the sample.

Definition 3.3.4 (Sufficient statistic) *Let $X_1, \ldots, X_n$ be a sample from a distribution P_θ with density function f_θ where $\theta \in \Theta \subset \mathbb{R}^d$. A statistic S defined on this sample is said to be* sufficient *for $\{P_\theta \mid \theta \in \Theta\}$ if the conditional distribution of $X_1, \ldots, X_n$ given S does not depend on θ.*

It may not be easy to prove sufficiency by verifying Definition 3.3.4 directly. A convenient alternative to prove sufficiency of statistics is provided by the following lemma. The most general form requires measure theory. For convenience, we only state the lemma for absolutely continuous distributions.

Lemma 3.3.5 (Factorization Lemma (Neyman)) *Let $X_1, \ldots, X_n$ be a sample from a distribution P_θ with density function f_θ where $\theta \in \Theta \subset \mathbb{R}^d$. A statistic S defined on this sample is sufficient for $\{P_\theta \mid \theta \in \Theta\}$ if and only if there exists functions $g_\theta : \mathbb{R}^n \to [0, \infty)$ and $h : \mathbb{R}^n \to [0, \infty)$ such that*

$$f_\theta(x_1, \ldots, x_n) = g_\theta(S(x_1, \ldots, x_n))\, h(x_1, \ldots, x_n).$$

Examples 3.3.6 *We give two simple applications of the Factorization Lemma.*

1. *Let $X_1, \ldots, X_n$ be a sample from a normal distribution with unknown mean μ and known variance σ_0^2. By rewriting the joint density of $X_1, \ldots, X_n$ as*

$$\exp\left[\frac{2\mu \sum_{i=1}^n x_i - n\mu^2}{2\sigma_0^2}\right] \left\{ \frac{1}{\left(\sqrt{2\pi\sigma_0^2}\right)^n} \exp\left[-\frac{\sum_{i=1}^n x_i^2}{2\sigma_0^2}\right] \right\},$$

   we obtain from the Factorization Lemma with $\theta = \mu$ that $\sum_{i=1}^n X_i$ is a sufficient statistic for μ.

2. *Let $X_1, \ldots, X_n$ be a sample from a normal distribution with unknown mean μ and unknown variance σ^2. By expanding the exponent of the joint density of $X_1, \ldots, X_n$ as*

$$\frac{1}{\left(\sqrt{2\pi\sigma^2}\right)^n} \exp\left[-\frac{\sum_{i=1}^n x_i^2 - 2\mu \sum_{i=1}^n x_i + n\mu^2}{2\sigma^2}\right],$$

   we obtain from the Factorization Lemma with $\theta = (\mu, \sigma)$ and $h = 1$ that $\sum_{i=1}^n X_i$ and $\sum_{i=1}^n X_i^2$ are joint sufficient statistics for (μ, σ^2).

In many elementary examples $\sum_{i=1}^n X_i$ and/or $\sum_{i=1}^n X_i^2$ are sufficient statistics. For an example of another sufficient statistic, we refer to Exercise 3.13.

It follows from the Factorization Lemma that if T is a sufficient statistic for a parameter θ and r is an injective function on the sample space, then $r(T)$ is also a sufficient statistic for

θ. Hence, it follows from Example 3.3.6 b) that $\left(\overline{X}, S^2\right)$ are also joint sufficient statistics for (μ, σ^2) in a normal sample.

The following theorem shows that conditioning on a sufficient statistic improves estimators in the sense that one may reduce the variance of unbiased estimators. The conditioning idea was hidden in a paper by C.R. Rao; Blackwell deserves credit for showing the important implications for improving estimators.

Theorem 3.3.7 (Rao-Blackwell Theorem) *Let $X_1, \ldots, X_n$ be a sample from a distribution P_θ with density function f_θ where $\theta \in \Theta \subset \mathbb{R}^d$. If S is a sufficient statistic for θ, then for any statistic T we have $E(E(T \mid S)) = E(T)$ and $Var(E(T \mid S)) \leq Var(T)$.*

Proof: The proof can be found in any book on mathematical statistics. □

The equality in the Rao-Blackwell Theorem only occurs in pathological cases. To illustrate the Rao-Blackwell Theorem, we return to our task of improving the ML estimator of the proportion of conforming items, *i.e.*, $P(LSL < X < USL)$ when $X \sim N(\mu, \sigma^2)$. Rather than applying the Rao-Blackwell Theorem directly to (3.7) and (3.8), we follow the typical path of starting with a very simple estimator. We define

$$T = \begin{cases} 1 & \text{if } LSL < X_1 < USL \\ 0 & \text{otherwise.} \end{cases}$$

It is obvious that T is an unbiased estimator of $P(LSL < X < USL)$. Since T depends on X_1 only, it is a very poor estimator in the sense that it has a large variance. Surprisingly, the estimators $E(T \mid \overline{X})$ and $E(T \mid \overline{X}, S^2)$ suggested by the Rao-Blackwell theorem turn out to be the unique unbiased estimator which has minimum variance for all μ and (μ, σ^2), respectively. Before we present some additional theory to show this, we first show how to obtain explicit formulas for these estimators.

We conclude this section by showing how to obtain explicit formulas for the estimators $E(T \mid \overline{X})$ and $E(T \mid \overline{X}, S^2)$. Various authors have obtained different forms of these estimators, sometimes being unaware of contributions by others (see Barton (1961), Bowker and Goode (1952), Folks et al. (1965), Guenther (1971), Liebermann and Resnikoff (1955) and Wheeler (1970)). The form derived in Wheeler (1970) seems to be the only one that allows a computation of the variance of these estimators.

We start with the case where μ is unknown and σ^2 is known. Since $(X_1, \overline{X})$ is a linear transformation of the normal vector $(X_1, \ldots, X_n)$, it follows that $(X_1, \overline{X})$ has a bivariate normal distribution with mean vector (μ, μ) and covariance matrix $\sigma^2 \begin{pmatrix} 1 & 1/n \\ 1/n & 1/n \end{pmatrix}$. Recall that the density of a bivariate normal random vector (X, Y) is given by

$$\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left[-\frac{(x_1-\mu_1)^2/\sigma_1^2 - 2\rho(x_1-\mu_1)(x_2-\mu_2)/(\sigma_1\sigma_2) + (x_2-\mu_2)^2/\sigma_2^2}{2(1-\rho^2)}\right],$$

where ρ is the correlation coefficient of X and Y. Since $\rho(X_1, \overline{X}) = 1/\sqrt{n}$, the conditional density of X_1 given $\overline{X} = s$ is given by

$$\begin{aligned} f_{X_1|\overline{X}=s}(x) = \frac{f_{X_1,\overline{X}}(x,s)}{f_{\overline{X}}(s)} &= \frac{\exp\left[-\frac{n\left((x-\mu)^2-2(s-\mu)(x-\mu)+n(s-\mu)^2\right)}{2(n-1)\sigma^2}\right]}{2\pi\sigma^2\sqrt{\frac{n-1}{n^2}}} \frac{\sqrt{2\pi\sigma^2/n}}{\exp\left[-\frac{n(s-\mu)^2}{2\sigma^2}\right]} \\ &= \frac{1}{\sqrt{2\pi\left(\frac{n-1}{n}\right)\sigma^2}} \exp\left[-\frac{(x-\mu^2)-2(x-\mu)(s-\mu)+(s-\mu)^2}{2\left(\frac{n-1}{n}\right)\sigma^2}\right] \\ &= \frac{1}{\sqrt{2\pi\left(\frac{n-1}{n}\right)\sigma^2}} \exp\left[-\frac{(x-s)^2}{2\left(\frac{n-1}{n}\right)\sigma^2}\right]. \end{aligned}$$

Hence, X_1 given $\overline{X} = s$ has a normal distribution with mean s and variance $\frac{n-1}{n}\,\sigma^2$. We conclude that an explicit formula for the estimator $E(T \mid \overline{X})$ for $P(LSL < X < USL)$ is given by

$$E(T \mid \overline{X}) = P(LSL < X_1 < USL \mid \overline{X}) = \Phi\left(\frac{USL-\overline{X}}{\sqrt{\frac{n-1}{n}}\,\sigma}\right) - \Phi\left(\frac{LSL-\overline{X}}{\sqrt{\frac{n-1}{n}}\,\sigma}\right). \tag{3.9}$$

Now we treat the case when both μ and σ^2 are unknown. Along the same lines of the proof that for a sample from a normal distribution we have that $\overline{X}$ and S^2 are independent, one can prove that $(X_1 - \overline{X})/S$ is independent of both $\overline{X}$ and S. Hence, $\sqrt{n/(n-1)}(X_1 - \overline{X})/S$ has a Student t_{n-1} distribution. This implies that

$$\begin{aligned} E(T \mid \overline{X} = x, S = s) &= P(LSL < X_1 < USL \mid \overline{X} = x, S = s) \\ &= P\left(\sqrt{\frac{n}{n-1}}\,\frac{LSL-\overline{X}}{S} < \frac{X_1-\overline{X}}{S} < \sqrt{\frac{n}{n-1}}\,\frac{USL-\overline{X}}{S} \,\middle|\, \overline{X} = x, S = s\right) \\ &= P\left(\sqrt{\frac{n}{n-1}}\,\frac{LSL-x}{s} < Y_{n-1} < \sqrt{\frac{n}{n-1}}\,\frac{USL-x}{s}\right), \end{aligned} \tag{3.10}$$

where Y_{n-1} denotes a random variable with a Student t-distribution with $n-1$ degrees of freedom. Note that because $(n-1)S^2 > (X_1 - \overline{X})^2$, one has to exclude certain cases in the above formula when the conditional probabilities involved are 0 or 1.

Definition 3.3.8 *A family of probability distributions $\mathcal{P}$ is* complete *if if for an arbitrary integrable function g we have that $E_P\, g = \int g(x)\, dP(x)$ for all $P \in \mathcal{P}$ implies that $P(g = 0) = 1$ for all $P \in \mathcal{P}$. A statistic T is* complete *with respect to a family probability distributions $\mathcal{P}$ if for an arbitrary integrable function g we have that*

$$E_P\, g(T) = \int g(t(x_1, \ldots, x_n))\, dP(x_1) \ldots dP(x_n) = 0$$

for all $P \in \mathcal{P}$ implies that $P(g = 0) = 1$ for all $P \in \mathcal{P}$.

It can be shown as a consequence of the uniqueness of Fourier transforms that the 2-parameter family of normal distributions with $\mu \in \mathbb{R}$ and $\sigma \in (0, \infty)$ are complete. In fact, this holds for the sufficient statistics of any family of distributions that belongs to the exponential family, provided that the parameter space has non-empty interior.

Theorem 3.3.9 (Lehmann-Scheffé) *Let $X_1, \ldots, X_n$ be a sample from a distribution P_θ with density function f_θ where $\theta \in \Theta \subset \mathbb{R}^d$. Let T be a sufficient statistic for θ which is complete with respect to $\{P_\theta \ \theta \in \Theta\}$. If g is a function on $\mathbb{R}^d$ such that there exists an unbiased estimator of $g(\theta)$, then there*

1. *exists a unique UMVU estimator of $g(\theta)$.*
2. *the unique UMVU estimator of $g(\theta)$ is the unique unbiased estimator of $g(\theta)$ which is a function of T.*

The uniqueness part of the Lehmann-Scheffé Theorem yields that the estimators $E(T \mid \overline{X})$ and $E(T \mid \overline{X}, S^2)$ are the UMVU-estimators of the proportion non-conforming items when σ^2 is unknown and both μ and σ^2 are unknown, respectively.

3.4 Exact distribution of capability indices

Now that we have constructed several estimators, we want to study their distribution. For the background on specific distributions we refer to the appendix. It is well-known that MLE's are biased in general. *E.g*, S is a biased estimator of σ for normal samples.

Recall that if $X_1, \ldots, X_n$ are independent random variables each with a $N(\mu, \sigma^2)$ distribution, then the random variable $(n-1)S^2/\sigma^2$ has a χ^2_{n-1}-distribution. The expected value of $\sqrt{n-1}\ S/\sigma$ thus equals

$$\int_0^\infty \frac{\sqrt{t}}{2^{(n-1)/2}\,\Gamma((n-1)/2)}\, t^{(n-1)/2-1} e^{-t/2}\, dt = \sqrt{2}\ \frac{\Gamma(n/2)}{\Gamma((n-1)/2)},$$

where Γ denotes the Gamma function (see Appendix). Hence,

$$\mathrm{E}\,(S) = \frac{\sqrt{2}}{\sqrt{n-1}} \frac{\Gamma(n/2)}{\Gamma((n-1)/2)}\, \sigma. \tag{3.11}$$

An unbiased estimator for σ is thus given by $S/c_4(n)$, where

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

Recall that $\Gamma(x+1) = x\Gamma(x)$ for $x \neq 0, -1, -2, \ldots$, $\Gamma(1) = 1$, and $\Gamma(1/2) = \sqrt{\pi}$. Thus we have the following recursion: $c_4(2) = \sqrt{2}/\sqrt{\pi}$ and $c_4(n+1) = (\sqrt{n-1}/\sqrt{n})(1/c_4(n))$. Instead of the ML estimators for C_p and C_{pk}, one usually uses the estimators

$$\widehat{C}_p = \frac{USL - LSL}{6S}$$

and

$$\widehat{C}_{pk} = \min\left(\frac{USL - \overline{X}}{3S}, \frac{\overline{X} - \mathrm{LSL}}{3S}\right),$$

where $\overline{X}$ denotes the sample mean and S denotes the sample standard deviation. Confidence intervals and hypothesis tests for C_p easily follow from the identity

$$P\left(\frac{\widehat{C}_p}{C_p} > c\right) = P\left(\chi^2_{n-1} < \frac{n-1}{c^2}\right).$$

In particular, it follows that

$$\mathrm{E}\,\widehat{C}_p = \left(\frac{n-1}{2}\right)^{1/2} \frac{\Gamma\left((n-2)/2\right)}{\Gamma\left((n-1)/2\right)}\, C_p.$$

The distribution of $\widehat{C}_{pk}$ is quite complicated, but explicit formulas for the normal case can be given because $\overline{X}$ and S are independent. We refer to Kotz and Johnson (1993) for details.

In order to describe the exact distribution of the ML estimator for quantiles of a normal distribution, we need a generalization of the Student t-distribution.

Theorem 3.4.1 *The MLE $\widehat{X}_p = \overline{X} + z_p\,\widehat{\sigma}$ for the x_p (*i.e.*, the unique number x_p such that $P(X \le x_p) = p$) with an underlying normal distribution is distributed as follows:*

$$P(\overline{X} + z_p\,\widehat{\sigma} \le t) = P\left(T_n\left(\frac{\sqrt{n}(\mu - t)}{\sigma}\right) \le -z_p\,\sqrt{n}\right), \tag{3.12}$$

where $T_\nu(\lambda)$ denotes a random variable distributed according to the noncentral t-distribution with ν degrees of freedom and noncentrality parameter λ.

Proof: Recall that $n\,\widehat{\sigma^2}/\sigma^2$ follows a χ^2-distribution with n degrees of freedom. Combining this with the definition of the noncentral t-distribution (see Definition .0.12), we obtain

$$\begin{aligned} P(\overline{X} + z_p\,\widehat{\sigma} \le t) &= P\left(\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} + z_p\,\frac{\sqrt{n}}{\sigma}\,\widehat{\sigma} \le \frac{\sqrt{n}\,(t-\mu)}{\sigma}\right) \\ &= P\left(Z + \frac{\sqrt{n}\,(\mu - t)}{\sigma} \le -z_p\,\frac{\sqrt{n}}{\sigma}\,\widehat{\sigma}\right) \\ &= P\left(T_n\left(\frac{\sqrt{n}\,(\mu - t)}{\sigma}\right) \le -z_p\,\sqrt{n}\right). \end{aligned}$$

□

3.5 Asymptotic distribution of capability indices

We now turn to the asymptotic distribution of the estimators that we encountered. It is well-known that under suitable assumptions MLE's are asymptotically normal. In concrete situations, one may alternatively prove asymptotic normality by using the following theorem, which is also known under the name *delta method*. It may be useful to recall that the Jacobian of a function is the matrix of partial derivatives of the component functions. For real functions on the real line, the Jacobian reduces to the derivative.

Theorem 3.5.1 (Cramér) *Let g be a function from $\mathbb{R}^m$ to $\mathbb{R}^k$ which is totally differentiable at a point a. If $(X_n)_{n\in\mathbb{N}}$ is a sequence of m-dimensional random vectors such that $c_n\,(X_n - a) \xrightarrow{d} X$ for some random vector X and some sequence of scalars $(c_n)_{n\in\mathbb{N}}$ with $\lim_{n\to\infty} c_n = \infty$, then*

$$c_n\,(g(X_n) - g(a)) \xrightarrow{d} Jg(a)\,X,$$

where Jg is the Jacobian of g.

Proof: See any text book on mathematical statistics. □

With this theorem, we can compute the asymptotic distribution of many estimators, in particular those discussed above. Among other things, this is useful for constructing confidence intervals (especially when the finite sample distribution is intractable).

We start with the asymptotic distribution of the sample variance.

Theorem 3.5.2 *Let $X, X_1, X_2, \ldots$ be independent identically distributed random variables with $\mu_4 = E\,X^4 < \infty$. Then the following asymptotic result holds for the MLE $\widehat{\sigma^2}$ of σ^2:*

$$\sqrt{n}\,\left(\widehat{\sigma^2} - \sigma^2\right) \xrightarrow{d} N(0, \mu_4 - \sigma^4).$$

In particular, if the parent distribution is normal, then

$$\sqrt{n}\,\left(\widehat{\sigma^2} - \sigma^2\right) \xrightarrow{d} N(0, 2\,\sigma^4).$$

Proof: Because the variance does not depend on the mean, we assume without loss of generality that $\mu = 0$. Since we have finite fourth moments, we infer from the Multivariate Central Limit Theorem (Theorem .0.10) that

$$\sqrt{n}\left[\begin{pmatrix} \frac{1}{n}\sum_{i=1}^n X_i \\ \frac{1}{n}\sum_{i=1}^n X_i^2 \end{pmatrix} - \begin{pmatrix} 0 \\ \sigma^2 \end{pmatrix}\right] \xrightarrow{d} N(0, \mathbf{\Sigma}),$$

where $\mathbf{\Sigma}$ is the covariance matrix of X and X^2. Since $\widehat{\sigma^2} = 1/n\,\sum_{i=1}^n X_i^2 - \overline{X}^2$, we apply Theorem 3.5.1 with $g(x, y) = y - x^2$. We compute $Jg(0, \sigma^2) = (0\ 1)$. Now recall that if Y is a random variable with a multinormal distribution $N(\mu, \Sigma)$ and L is a linear map of the right dimensions, then $L\,Y \stackrel{d}{=} N(L\,\mu, L\,\Sigma\,L^T)$. Hence,

$$\begin{aligned} \sqrt{n}\,\left(\widehat{\sigma^2} - \sigma^2\right) &\xrightarrow{d} N\left(0, Jg(0,\sigma^2)\,\mathbf{\Sigma}\,Jg(0,\sigma^2)^T\right) \\ &= N(0, \operatorname{Var} X^2) \\ &= N(0, \mu_4 - \sigma^4). \end{aligned}$$

The last statement follows from the fact that for zero-mean normal distributions $\mu_4 = 3\,\sigma^4$. □

For the asymptotic distribution of $\widehat{\sigma}$ and $1/\widehat{\sigma}$, see Exercise 3.2.

Theorem 3.5.3 *Let $X, X_1, X_2, \ldots$ be independent normal random variables with mean μ and variance σ^2. If μ and σ^2 are unknown, then the following asymptotic result holds for the MLE $\widehat{X_p} = \overline{X} + z_p\,\widehat{\sigma}$ of the p-th quantile of X x_p (cf. Example 3.3.3):*

$$\sqrt{n}\,\left(\widehat{X_p} - (\mu + z_p\,\sigma)\right) \xrightarrow{d} N\left(0, \sigma^2\,\left(1 + \tfrac{1}{2}\,z_p^2\right)\right).$$

Proof: The Central Limit Theorem yields that $\sqrt{n}\,\left(\overline{X} - \mu\right) \xrightarrow{d} N(0, \sigma^2)$. Combining Theorems 3.5.1 and 3.5.2, we have that $\sqrt{n}\,\left(\widehat{\sigma} - \sigma\right) \xrightarrow{d} N(0, \frac{1}{2}\,\sigma^2)$. Now recall that since we have a normal sample, $\overline{X}$ and $\widehat{\sigma}$ are independent. Hence, it follows from a form Slutsky's Lemma (.0.5) that

$$\sqrt{n}\,\left(\overline{X} + z_p\,\widehat{\sigma} - \mu - z_p\,\sigma\right) \xrightarrow{d} N(0, \sigma^2) * N(0, \tfrac{1}{2}\,z_p^2\,\sigma^2),$$

where $*$ denotes convolution. The result now follows from the elementary fact

$$N(0, \sigma_1^2) * N(0, \sigma_2^2) = N(0, \sigma_1^2 + \sigma_2^2).$$

This concludes the proof. □

Theorem 3.5.4 *Let $X, X_1, X_2, \ldots$ be independent normal random variables with mean μ and variance σ^2 and let φ be the standard normal density. If μ and σ^2 are unknown, then the following asymptotic result holds for the MLE of $P(a < X < b) = P((a, b))$:*

$$\sqrt{n}\,\left(\Phi\left(\frac{b - \overline{X}}{\widehat{\sigma}}\right) - \Phi\left(\frac{a - \overline{X}}{\widehat{\sigma}}\right) - P(a < X < b)\right) \xrightarrow{d} N\left((0, \sigma^2\,\left(c_1^2 + 4\,c_1\,c_2\,\mu + 2\,c_2^2\,(2\mu^2 + \sigma^2)\right)\right),$$

where

$$\begin{aligned} c_1 &= \varphi\left(\frac{b-\mu}{\sigma}\right)\,\frac{b\,\mu - (\mu^2 + \sigma^2)}{2\sigma^3} - \varphi\left(\frac{a-\mu}{\sigma}\right)\,\frac{a\,\mu - (\mu^2 + \sigma^2)}{2\sigma^3} \\ c_2 &= \varphi\left(\frac{b-\mu}{\sigma}\right)\,\frac{\mu - b}{2\sigma^3} - \varphi\left(\frac{a-\mu}{\sigma}\right)\,\frac{\mu - a}{2\sigma^3} \qquad . \end{aligned}$$

Proof: We infer from the multivariate Central Limit Theorem (= Theorem .0.10) that

$$\sqrt{n}\left[\begin{pmatrix} 1/n\,\sum_{i=1}^n X_i \\ 1/n\,\sum_{i=1}^n X_i^2 \end{pmatrix} - \begin{pmatrix} \mu \\ \mu^2 + \sigma^2 \end{pmatrix}\right] \xrightarrow{d} N\left(0, \mathbf{\Sigma}\right),$$

where $\mathbf{\Sigma}$ is the covariance matrix of X and X^2. Since $E\,Z^4 = 3$ and $X \stackrel{d}{=} \mu + \sigma\,Z$ where Z is a standard normal random variable, we have

$$\begin{aligned} \mathrm{E}\;X^3 &= \mu^3 + 3\,\mu\,\sigma^2 \\ \mathrm{Var}\,X^2 &= \mu^4 + 6\,\mu^2\,\sigma^2 + 3\,\sigma^4 - (\mu^2 + \sigma^2)^2 = 2\,\sigma^2(2\,\mu^2 + \sigma^2). \end{aligned}$$

Hence,

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma^2 & 2\,\mu\,\sigma^2 \\ 2\,\mu\,\sigma^2 & 2\,\sigma^2(2\,\mu^2 + \sigma^2) \end{pmatrix}.$$

Now we wish to apply Theorem 3.5.1 with

$$g(x, y) = \Phi\left(\frac{b - x}{\sqrt{y - x^2}}\right) - \Phi\left(\frac{a - x}{\sqrt{y - x^2}}\right).$$

This function is totally differentiable, except on the line $y = x^2$. Since we evaluate at $x = \mu$ and $y = \mu^2 + \sigma^2$, we have that $y - x^2 = \sigma^2 > 0$. Hence, there are no differentiability problems. Note that the partial derivatives of $f(x, y) = \dfrac{c - x}{\sqrt{y - x^2}}$ with respect to x and y are given by $\dfrac{c\,x - y}{2(y - x^2)^{3/2}}$, $\dfrac{x - c}{2(y - x^2)^{3/2}}$ respectively. Thus the *transpose* of the Jacobian of g is given by

$$\begin{pmatrix} \varphi\left(\dfrac{b-x}{\sqrt{y-x^2}}\right)\dfrac{b\,x-y}{2(y-x^2)^{3/2}} - \varphi\left(\dfrac{a-x}{\sqrt{y-x^2}}\right)\dfrac{a\,x-y}{2(y-x^2)^{3/2}} \\ \varphi\left(\dfrac{b-x}{\sqrt{y-x^2}}\right)\dfrac{x-b}{2(y-x^2)^{3/2}} - \varphi\left(\dfrac{a-x}{\sqrt{y-x^2}}\right)\dfrac{x-a}{2(y-x^2)^{3/2}} \end{pmatrix},$$

where $\varphi(x)$ is the standard normal density. Evaluating at $x = \mu$ and $y = \mu^2 + \sigma^2$, we see that this reduces to

$$\begin{pmatrix} \varphi\left(\dfrac{b-\mu}{\sigma}\right)\dfrac{b\,\mu-(\mu^2+\sigma^2)}{2\sigma^3} - \varphi\left(\dfrac{a-\mu}{\sigma}\right)\dfrac{a\,\mu-(\mu^2+\sigma^2)}{2\sigma^3} \\ \varphi\left(\dfrac{b-\mu}{\sigma}\right)\dfrac{\mu-b}{2\sigma^3} - \varphi\left(\dfrac{a-\mu}{\sigma}\right)\dfrac{\mu-a}{2\sigma^3} \end{pmatrix}.$$

Putting everything together yields the result. □

3.6 Tolerance intervals

In the previous section we generalized estimation of parameters to estimation of functions of parameters. Two important examples were the p-th quantile and the fraction $P(a < X < b)$ of a distribution. We now present a further generalization. Instead of considering estimators that are real-valued functions of the sample, we will study estimators that are set-valued functions (in particular, functions whose values are intervals).

Many practical situations require knowledge about the location of the complete distribution. *E.g.*, one would like to construct intervals that cover a certain percentage of a distribution. Such intervals are known as tolerance intervals. Although they are of great practical importance, this topic is ignored in many text books. Many practical applications (and theory) can be found in Aitchison and Dunsmore (1975). The monograph Guttman (1970), the review paper Patel (1986) as well as the bibliographies Jílek (1981) and Jílek and Ackermann (1989) are also excellent sources of information on this topic.

In this section we will give an introduction to tolerance intervals based on the normal distribution. It is also possible to construct intervals for other distributions (see *e.g.*, Aitchison and Dunsmore (1975) and Patel (1986)). The idea of using rank statistics to build non-parametric tolerance intervals goes back to Wilks (1941); see Di Bucchianico et al. (2001) for a more recent contribution.

Definition 3.6.1 *Let $X_1, \ldots, X_n$ be a sample from a continuous distribution P with distribution function F. An interval $T(X_1, \ldots, X_n) = (L, U)$ is said to be a* β-content tolerance interval *at confidence level $1 - \alpha$ if*

$$P\left(P(T(X_1, \ldots, X_n)) \geq \beta\right) = P(F(U) - F(L) \geq \beta) = 1 - \alpha. \tag{3.13}$$

The random variable $P(T(X_1, \ldots, X_n))$ is called the coverage *of the tolerance interval.*

This type of tolerance interval is sometimes called a *guaranteed content interval*. There also exists a one-sided version of this type of tolerance interval.

Definition 3.6.2 *Let $X_1, \ldots, X_n$ be a sample from a continuous distribution P with distribution function F. The estimator $U(X_1, \ldots, X_n)$ is said to be a β-content* upper tolerance limit *at confidence level $1-\alpha$ if*

$$P(F(U(X_1, \ldots, X_n)) \geq \beta) = 1 - \alpha. \tag{3.14}$$

Similarly, the estimator $L(X_1, \ldots, X_n)$ is said to be a β-content lower tolerance limit *at confidence level $1-\alpha$ if*

$$P(1 - F(L(X_1, \ldots, X_n)) \geq \beta) = 1 - \alpha. \tag{3.15}$$

Before we give an explicit example of a guaranteed content tolerance interval, we present a useful lemma.

Lemma 3.6.3 *Let φ be the density of the standard normal distribution with corresponding distribution function Φ. Then the following holds:*

1. *For $c > 0$, the function $x \mapsto \Phi(x+c) - \Phi(x-c)$ is increasing for $x < 0$ and decreasing for $x > 0$. In particular, it has a unique maximum at $x = 0$.*
2. *The function $c \mapsto \Phi(x+c) - \Phi(x-c)$ is increasing on $\mathbb{R}$ and has range $[0, 1)$.*
3. *If $F_{\mu,\sigma}$ is the distribution function of a normal distribution with mean μ and variance σ^2, then $F_{\mu,\sigma^2}(x) = \Phi\left((x-\mu)/\sigma\right)$ and $\Phi(x) = F_{\mu,\sigma^2}(\mu + x\sigma)$.*

Proof: These properties follow directly from the specific form of φ by considering derivatives and using the symmetry of φ. □

For a normal sample with unknown mean μ and known variance σ^2, a $1-\alpha$ confidence interval for μ is given by $(\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}})$. We now derive a guaranteed tolerance interval for the same situation.

Proposition 3.6.4 *Let $X_1, \ldots, X_n$ be a sample from a normal distribution with unknown mean μ and known variance σ^2. If k is chosen such that $\Phi\left(\frac{z_{\alpha/2}}{\sqrt{n}} + k\right) - \Phi\left(\frac{z_{\alpha/2}}{\sqrt{n}} - k\right) = \beta$, then the interval $\left(\overline{X} - k\sigma, \overline{X} + k\sigma\right)$ is a β-content tolerance interval at confidence level $1-\alpha$.*

Proof: First note that k exists and is unique by part 2) of Lemma 3.6.3. It follows from part 1) of Lemma 3.6.3 and the choice of k that $|x| \geq \frac{z_{\alpha/2}}{\sqrt{n}}$ holds if and only if $\Phi(x+k) - \Phi(x-k) \leq \beta$. Since $Y = \frac{\overline{X}-\mu}{\sigma} \sim N(0, 1/n)$, we have

$$P\left((\Phi(Y+k) - \Phi(Y-k)) \geq \beta\right) = P\left(|Y| \leq \frac{z_{\alpha/2}}{\sqrt{n}}\right) = P(|Z| \leq z_{\alpha/2}) = 1 - \alpha.$$

□.

Definition 3.6.5 *Let $X_1, \ldots, X_n$ be a sample from a continuous distribution P with distribution function F. An interval $T(X_1, \ldots, X_n) = (L, U)$ is said to be a β-*expectation tolerance interval *if the expected coverage equals β, i.e.*

$$\mathrm{E}\left(P(T(X_1, \ldots, X_n))\right) = \mathrm{E}\left(F(U) - F(L)\right) = \beta. \tag{3.16}$$

There are interesting relations between these concepts and quantiles. Let $X_1, \ldots, X_n$ be a sample from a continuous distribution P with distribution function F. Since

$$P(F(U) \leq \beta) = P(U \leq F^{-1}(\beta)),$$

it follows immediately that an upper (lower) (α, β) tolerance limit is an upper (lower) confidence interval for the quantile $F^{-1}(\beta)$ and vice-versa. From a computation point of view, a more interesting relationship is the one with prediction intervals.

Definition 3.6.6 *Let $X_1, \ldots, X_n$ be a sample from a continuous distribution P with distribution function F. An interval $T(X_1, \ldots, X_n) = (L, U)$ is said to be a $1-\beta$- prediction interval if*

$$P(L < X < U) = 1 - \beta. \tag{3.17}$$

Prediction intervals are usually associated with regression analysis, but also appear in other contexts as we shall see. The following proposition, first proved in Paulson (1943), shows a surprising link between β-expectation tolerance intervals and prediction intervals. It also has interesting corollaries as we shall see later on.

Proposition 3.6.7 (Paulson) *A β-expectation tolerance interval is a β-prediction interval.*

Proof: We use the following well-known property of conditional expectations:

$$\mathrm{E}\left(\mathrm{E}\left(V \mid W\right)\right) = \mathrm{E}\, V.$$

Hence, rewriting the probability in the definition of prediction interval in terms of an expectation, we obtain:

$$\begin{aligned} P(L < X < U) &= \mathrm{E}\left(1_{L<X<U}\right) \\ &= \mathrm{E}\left(\mathrm{E}\left(1_{L<X<U}\right) \mid L, U\right)) \\ &= \mathrm{E}\left(P\left(L < X < U\right) \mid L, U\right)) \\ &= \mathrm{E}\left(P(L, U)\right) \\ &= \mathrm{E}\left(F(U) - F(L)\right), \end{aligned}$$

as required. □

The following proposition is a typical example of using Proposition 3.6.7 to obtain a β-expectation tolerance interval.

Proposition 3.6.8 *If $X_1, \ldots, X_n$ is a sample from a normal distribution with known variance σ^2, then the interval $\left(\overline{X} - \sqrt{1 + \frac{1}{n}}\, z_{\beta/2}\, \sigma, \overline{X} + \sqrt{1 + \frac{1}{n}}\, z_{\beta/2}\, \sigma\right)$ is a $1-\beta$-expectation tolerance interval.*

Proof: Since μ is unknown and σ^2 is known, we try an interval of the form $(\overline{X} - k\sigma, \overline{X} + k\sigma)$, where we need to determine k such that $\mathrm{E}\left(F(\overline{X} + k\sigma) - F(\overline{X} - k\sigma)\right) = 1 - \beta$. By Proposition 3.6.7 we have $\mathrm{E}\left(F(\overline{X} + k\sigma) - F(\overline{X} - k\sigma)\right) = P(\overline{X} - k\sigma < X < \overline{X} + k\sigma) = 1 - \beta$,

where $X \sim N(\mu, \sigma^2)$, $\overline{X} \sim N(\mu, \sigma^2/n)$ and X is independent of $X_1, \ldots, X_n$. Since $X - \overline{X} \sim N(\mu, (1+\frac{1}{n})\sigma^2)$, it follows that k has to satisfy

$$P(\overline{X} - k\sigma < X < \overline{X} + k\sigma) = \Phi\left(\frac{k}{\sqrt{1+\frac{1}{n}}}\right) - \Phi\left(\frac{-k}{\sqrt{1+\frac{1}{n}}}\right) = \beta.$$

Hence, k should be equal to $\sqrt{1+\frac{1}{n}}\; z_{\beta/2}$. □

For normal distributions, it is rather natural to construct tolerance intervals using the jointly sufficient statistics $\overline{X}$ and S^2. In particular, intervals of the form $\left(\overline{X} - k\,S, \overline{X} + k\,S\right)$ are natural candidates. Unfortunately, even the distribution of the coverage of such simple intervals is very complicated if both μ and σ are unknown. However, we may use the Paulson result to compute the first moment, *i.e.*, the *expected* coverage, and thus obtain β-expectation tolerance intervals. To put this result into perspective, recall that for a normal sample with unknown mean μ and unknown variance σ^2, a $1-\alpha$ confidence interval for μ is given by $(\overline{X} - t_{n-1;\alpha/2}\frac{S}{\sqrt{n}}, \overline{X} + t_{n-1;\alpha/2}\frac{S}{\sqrt{n}})$.

Corollary 3.6.9 *Let $X_1, \ldots, X_n$ be a sample from a normal distribution with unknown mean μ and unknown variance σ^2. The expected coverage of the interval $\left(\overline{X} - k\,S, \overline{X} + k\,S\right)$ equals $1-\beta$ if and only if $k = \sqrt{1+\frac{1}{n}}\; t_{n-1;\beta/2}$.*

Proof: We first have to show that the expectation is finite. Note that the coverage may be written in this case as

$$F_{\mu,\sigma^2}(\overline{X} + kS) - F_{\mu,\sigma^2}(\overline{X} - kS) = \Phi\left(\frac{\overline{X}-\mu}{\sigma} + k\,\frac{S}{\sigma}\right) - \Phi\left(\frac{\overline{X}-\mu}{\sigma} - k\,\frac{S}{\sigma}\right).$$

Since $0 \leq \Phi(x) \leq 1$ for all $x \in \mathbb{R}$, it suffices to show that the expectations of $\overline{X}$ and S are finite. The first expectation is trivial, while the second one follows from (3.11).

Now Proposition 3.6.7 yields that it suffices to choose k such that $(\overline{X} - k\,S, \overline{X} + k\,S)$ is a β-prediction interval. In other words, k must be chosen such that

$$P\left(\overline{X} - k\,S < X < \overline{X} + k\,S\right) = 1 - \beta,$$

where X is independent of $X_1, \ldots, X_n$, but follows the same normal distribution. Hence, $X - \overline{X} \stackrel{d}{=} N\left(0, \sigma^2\,(1+\frac{1}{n})\right)$. Thus we have the following equalities:

$$\begin{aligned} 1-\beta &= P\left(\overline{X} - k\,S < X < \overline{X} + k\,S\right) \\ &= P\left(-k < \frac{X-\overline{X}}{S} < k\right) \\ &= P\left(-k < \sqrt{1+\tfrac{1}{n}}\,T_{n-1} < k\right), \end{aligned}$$

from which we see that we must choose $k = \sqrt{1+\frac{1}{n}}\; t_{n-1;\beta/2}$. □

For several explicit tolerance intervals under different assumptions, we refer to the exercises and to Patel (1986). There is no closed solution for (α, β) tolerance intervals for normal distributions when both μ and σ^2 are unknown; numerical procedures for this problem can be found in Eberhardt et al. (1989).

3.7 Normality testing

The estimators and tests discussed in this chapter heavily depend on the normality assumption. It is therefore necessary to investigate whether this assumption is valid before performing a capability analysis. Testing normality (or any other distributional assumption) should consist of two parts:

1. graphical inspection of the data
   (a) normal probability plot
   (b) density estimator plot
2. formal goodness-of-fit test (Shapiro-Wilks or Anderson-Darlin or Cramér-von Mises)

Warning In SPC in general and in capability analyses in particular, one is interested in extreme tails of distributions. Standard procedures for testing normality as described below do not take this into account, since they tend to emphasize the centre of the distribution rather than the tails. So although normality testing as described above is common practice in SPC, it does not necessarily guarantee that estimates are accurate. Another warning comes from the fact that by first applying a goodness-of-fit test (with a certain significance level) before performing any further testing, the significance level of a further test depends on the normality test performed earlier.

The normal probability plot is a plot of theoretical quantiles of a normal distribution with estimated parameters against the sample quantiles. This plot is available in `R` through the command `qqnorm`. Normal probability plots are useful to graphically detect outliers and deviations in tail behaviour (too heavy or too light tails).

The shape of a distribution (unimodality, asymmetries etc.) is difficult to assess from a normal probability plot. For this we need to estimate the density and present it in a plot. A widely used density estimator (although it is not always recognized as such) is the histogram. Let $X_1, \ldots, X_n$ be a random sample from a distribution function F (pertaining to a law P) on $\mathbb{R}$, with continuous derivative $F' = f$. As before, we denote the empirical distribution function by P_n. Let I be a compact interval on $\mathbb{R}$ and suppose that the intervals $I_1, \ldots, I_k$ form a partition of I, i.e.

$$I = I_1 \cup \ldots \cup I_k, \quad I_i \cap I_j = \emptyset \text{ if } i \neq j.$$

The histogram of $X_1, \ldots, X_n$ with respect to the partition $I_1, \ldots, I_k$ is defined as

$$H_n(x) := \sum_{j=1}^{k} \frac{P_n(I_j)\, I_{I_j}(x)}{|I_j|},$$

where $|I_j|$ denotes the length of the interval I_j. It is clear that the histogram is a stepwise constant function. Two major disadvantages of the histogram are

- the stepwise constant nature of the histogram
- the fact that the histogram heavily depends on the choice of the partition

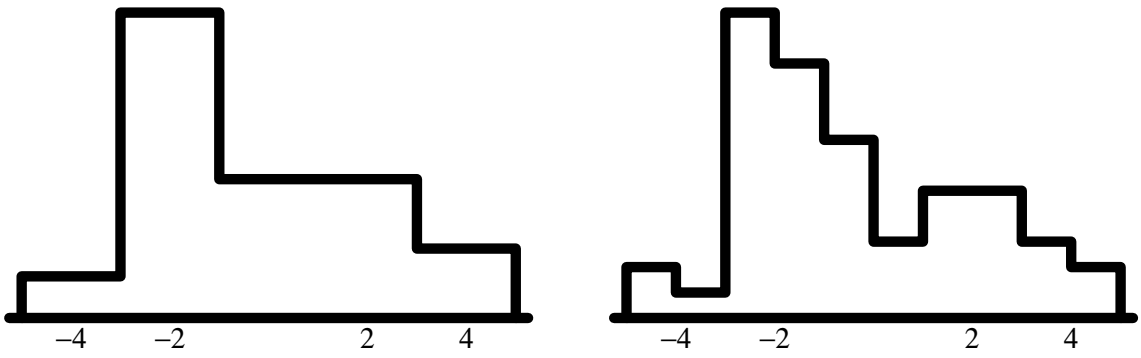


Figure 3.1: Two histograms of the same sample of size 50 from a mixture of 2 normal distributions.

In order to illustrate the last point, consider Figure 3.7 where the two histograms are made from the same data set.
It is because of this phenomenon that histograms are not to be recommended. A natural way to improve on histograms is to get rid of the fixed partition by putting an interval around each point. If $h > 0$ is fixed, then

$$\widehat{N}_n(x) := \frac{P_n((x-h, x+h))}{2h} \tag{3.18}$$

is called the *naive density estimator* and was introduced in 1951 by Fix and Hodges in an unpublished report (reprinted in Fix and Hodges (1989)) dealing with discriminant analysis. The motivation for the naive estimator is that

$$P(x-h < X < x+h) = \int_{x-h}^{x+h} f(t)\,dt \approx 2\,h\,f(x). \tag{3.19}$$

Note that the naive estimator is a local procedure; it uses only the observations close to the point at which one wants to estimate the unknown density. Compare this with the empirical distribution function, which uses all observations to the right of the point at which one is estimating.

It is intuitively clear from (3.19) that the bias of $\widehat{N}_n$ decreases as h tends to 0. However, if h tends to 0, then one is using less and less observations, and hence the variance of $\widehat{N}_n$ increases. This phenomenon occurs often in density estimation. The optimal value of h is a compromise between the bias and the variance. We will return to this topic of great practical importance when we discuss the MSE.

The naive estimator is a special case of the following class of density estimators. Let K be a *kernel function*, that is a nonnegative function such that

$$\int_{-\infty}^{\infty} K(x)\,dx = 1. \tag{3.20}$$

The *kernel estimator* with kernel K and bandwidth h is defined by

$$\widehat{f}_n(x) := \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h} K\left(\frac{x - X_i}{h}\right). \tag{3.21}$$

Thus, the kernel indicates the weight that each observation receives in estimating the unknown density. It is easy to verify that kernel estimators are densities and that the naive estimator is a kernel estimator with kernel

$$K(x) = \begin{cases} \frac{1}{2} & \text{if } |x| < 1 \\ 0 & \text{otherwise.} \end{cases}$$

Remark 3.7.1 *The kernel estimator can also be written in terms of the empirical distribution function F_n:*

$$\widehat{f}_n(x) = \int_{-\infty}^{\infty} \frac{1}{h} K\left(\frac{x - y}{h}\right) dF_n(y),$$

where the integral is a Stieltjes integral.

Examples of other kernels are given in Table 3.1. Kernel density estimators are available in `R` through the command `density`, including an automatic choice of the bandwith h.

name	function
Gaussian	$\frac{1}{\sqrt{2\,\pi}}\, e^{-\frac{1}{2}\,x^2}$
naive/rectangular	$\frac{1}{2}\, 1_{(-1,1)}(x)$
triangular	$(1 - \lvert x\rvert)\, 1_{(-1,1)}(x)$
biweight	$\frac{15}{16}\,(1 - x^2)^2\, 1_{(-1,1)}(x)$
Epanechnikov	$\frac{3}{4}\,(1 - x^2)\, 1_{(-1,1)}(x)$

Table 3.1: Well-known kernels for density estimators.

Formal goodness-of-fit tests should be performed after a graphical inspection, because graphical inspection may cause to reconsider the data (removal of outliers, splitting data sets that come from two sources etc.). In order not to be too strict in rejecting the null hypothesis that the sample comes from a normal distribution, it is recommended to take a small significance value like 0.01 rather than the standard 0.05. The Shapiro-Wilks test is a dedicated goodness-of-fit for normality, based on a tailor-made regression of the normal probability plot. The Anderson-Darling and Cramér-von Mises tests are general tests based on the empirical distribution function. These three tests have comparable power when testing the normality hypothesis against various alternatives. In any case, they are all much better than the Kolmogorov test. These tests are available in `R` as explained in detail in Subsection 2.2.2. An excellent book on goodness-of-fit tests is D'Agostino and Stephens (1986).

3.8 Mathematical background on density estimators

In this section we give an introduction to the mathematical background of the kernel density estimators defined in the previous section. A good impression of kernel estimation is given by the books Silverman (1986) and Wand and Jones (1995). For other types of estimators, we refer to Silverman (1986) and Tapia and Thompson (1978).

3.8.1 Finite sample behaviour of density estimators

In order to assess point estimators, we look at properties like unbiasedness and efficiency. In density estimation, it is very important to know the influence of the bandwidth h (cf. our discussion of the naive estimator). To combine the assessment of these properties, the Mean Square Error (MSE) is used. We now discuss the analogues of these properties for density estimators. The difference is that the estimate is not a single number, but a function. However, we start with pointwise properties.

Theorem 3.8.1 *Let $\widehat{f}_n$ be a kernel estimator with kernel K. Then*

$$\mathrm{E}\ \widehat{f}_n(x) = \frac{1}{h}\int_{-\infty}^{\infty} K\left(\frac{x-y}{h}\right)\ f(y)\,dy = \frac{1}{h}\int_{-\infty}^{\infty} K\left(\frac{y}{h}\right)\ f(x-y)\,dy. \tag{3.22}$$

Proof: This follows from the fact that for a random variable X with density f, we have $\mathrm{E}\, g(X) = \int_{-\infty}^{\infty} g(x)\, f(x)\, dx$. □.

Theorem 3.8.2 *Let $\widehat{f}_n$ be a kernel estimator with kernel K. Then*

$$\mathrm{Var}\ \widehat{f}_n(x) = \frac{1}{n\,h^2}\int_{-\infty}^{\infty} K^2\left(\frac{x-y}{h}\right)\ f(y)\,dy - \frac{1}{n\,h^2}\left\{\int_{-\infty}^{\infty} K\left(\frac{x-y}{h}\right)\ f(y)\,dy\right\}^2. \tag{3.23}$$

Proof: It is easy to see that

$$\big(\hat{f}_n(x)\big)^2 = \frac{1}{n^2}\sum_{i=1}^{n}\frac{1}{h^2}K^2\Big(\frac{x-X_i}{h}\Big) + \frac{1}{n^2}\sum_{\substack{i,j=1\\ i\neq j}}^{n}\frac{1}{h^2}K\Big(\frac{x-X_i}{h}\Big)K\Big(\frac{x-X_j}{h}\Big).$$

Then

$$\mathrm{E}\,\big(\hat{f}_n(x)\big)^2 = \frac{1}{nh^2}\int_{-\infty}^{\infty} K^2\Big(\frac{x-y}{h}\Big)f(y)dy + \frac{n-1}{nh^2}\Big(\int_{-\infty}^{\infty} K\Big(\frac{x-y}{h}\Big)f(y)dy\Big)^2.$$

Next use Theorem 4.2 and the well-known fact that $\mathrm{Var}\, X = \mathrm{E}\, X^2 - (\mathrm{E}\, X)^2$. □

The following general result due to Rosenblatt (see Rosenblatt (1956) for a slightly more general result) shows that we cannot have unbiasedness for all x.

Theorem 3.8.3 (Rosenblatt) *A kernel estimator can not be unbiased for all $x \in \mathbb{R}$.*

Proof: We argue by contradiction. Assume that $\mathrm{E}\,\widehat{f}_n(x) = f(x)$ for all $x \in \mathbb{R}$. Then $\int_a^b\ \widehat{f}_n(x)\,dx$ is an unbiased estimator for $F(b) - F(a)$, since

$$\mathrm{E}\int_a^b\ \widehat{f}_n(x)\,dx = \int_a^b\ \mathrm{E}\,\widehat{f}_n(x)\,dx = \int_a^b\ f(x)\,dx = F(b) - F(a),$$

where the interchange of integrals is allowed since the integrand is positive. Now it can be shown that the only unbiased estimator of $F(b) - F(a)$ symmetric in $X_1, \ldots, X_n$ is $F_n(b) - F_n(a)$. This leads to a contradiction, since it implies that the empirical distribution function is differentiable. □

For point estimators, the MSE is a useful concept. We now generalize this concept to density estimators.

Definition 3.8.4 *The Mean Square Error at x of a density estimator $\widehat{f}$ is defined as*

$$\mathrm{MSE}_x(\widehat{f}) := \mathrm{E}\left(\widehat{f}(x) - f(x)\right)^2. \tag{3.24}$$

The Mean Integrated Square Error of a density estimator $\widehat{f}$ is defined as

$$\mathrm{MISE}(\widehat{f}) := \mathrm{E}\int_{-\infty}^{\infty}\left(\widehat{f}(x) - f(x)\right)^2 dx. \tag{3.25}$$

Theorem 3.8.5 *For a kernel density estimator $\widehat{f}_n$ with kernel K the* MSE *and* MISE *can be expressed as:*

$$\begin{aligned}\mathrm{MSE}_x(\widehat{f}_n) = \frac{1}{n\,h^2}\int_{-\infty}^{\infty} K^2\left(\frac{x-y}{h}\right) f(y)\,dy - \frac{1}{n\,h^2}\left\{\int_{-\infty}^{\infty} K\left(\frac{x-y}{h}\right) f(y)\,dy\right\}^2 + \\ \left(\frac{1}{h}\int_{-\infty}^{\infty} K\left(\frac{x-y}{h}\right) f(y)\,dy - f(x)\right)^2.\end{aligned} \tag{3.26}$$

$$\begin{aligned}\mathrm{MISE}(\widehat{f}_n) = \frac{1}{n\,h^2}\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} K^2\left(\frac{x-y}{h}\right) f(y)\,dy - \left\{\int_{-\infty}^{\infty} K\left(\frac{x-y}{h}\right) f(y)\,dy\right\}^2\right) dx + \\ \int_{-\infty}^{\infty}\left(\frac{1}{h}\int_{-\infty}^{\infty} K\left(\frac{x-y}{h}\right) f(y)\,dy - f(x)\right)^2 dx.\end{aligned} \tag{3.27}$$

Proof: Combination of Exercise 3.20 with formulas (3.22) and (3.23) yields the formula for the MSE. Integrating this formula with respect to x, we obtain the formula for the MISE. □

The above formulas can in general not be evaluated explicitly. When both the kernel and the unknown density are Gaussian, then straightforward but tedious computations yield explicit formulas as shown in Fryer (1976). These formulas were extended in Marron and Wand (1992) to the case of mixtures of normal distributions. Marron and Wand claim in Marron and Wand (1992) that the class of mixture of normal distributions is very rich and that it is thus possible to perform exact calculations for many distributions. These calculations can be used to choose an optimal bandwidth h (see Marron and Wand (1992) for details).

For other examples of explicit MSE calculations, we refer to Deheuvels (1977) and the exercises.

We conclude this section with a note on the use of Fourier analysis. Recall that the convolution of two functions g_1 and g_2 is defined as

$$(g_1 * g_2)(x) := \int_{-\infty}^{\infty} g_1(t)\,g_2(x-t)\,dt.$$

One of the elementary properties of the Fourier transform is that it transforms the complicated convolution operation into the elementary multiplication operation, i.e.

$$\mathcal{F}(g_1 * g_2) = \mathcal{F}(g_1)\,\mathcal{F}(g_2),$$

where $\mathcal{F}(g)$ denotes the Fourier transform of g, defined by

$$(\mathcal{F}(g))(s) = \int_{-\infty}^{\infty} g(t)e^{ist}dt.$$

The formulas (3.22) and (3.23) show that $\mathrm{E}\,\widehat{f}_n(x)$ and $\mathrm{V}ar\,\widehat{f}_n(x)$ can be expressed in terms of convolutions of the kernel with the unknown density. The exercises contain examples in which Fourier transforms yield explicit formulas for the mean and the variance of the kernel estimator.

Another (even more important) use of Fourier transforms is the computation of the kernel estimate itself. Computing density estimates directly from the definition is often very time consuming. Define the function u by

$$u(s) = \frac{1}{n}\sum_{j=1}^{n} e^{i\,s\,X_j}. \tag{3.28}$$

Then the Fourier transform of the kernel estimator is a convolution of u with the Fourier transform of the kernel (see Exercise 3.24). Using Fast Fourier Transform (FFT), one can efficiently compute good approximations to the kernel estimates. For details we refer to (Silverman, 1986, pp. 61-66) and (Wand and Jones, 1995, Appendix D).

3.8.2 Asymptotic behaviour of kernel density estimators

We have seen in the previous section that it is possible to evaluate exactly the important properties of kernel density estimators. However, the unknown density f appears in a complicated way in exact calculations, which limits the applicability. Such calculations are very important for choosing the optimal bandwidth h. Therefore, much effort has been put in obtaining asymptotic results in which the unknown density f appears in a less complicated way. In this section we give an introduction to these results. Many of the presented results can be found in Parzen (1962); Rosenblatt (1956). For an overview of more recent results, we refer to the monographs Silverman (1986); Wand and Jones (1995).

Theorem 3.8.6 (Bochner) *Let K be a bounded kernel function such that $\lim_{|y|\to\infty} y\,K(y) = 0$. Define for any absolutely integrable function g the functions*

$$g_n(x) := \frac{1}{h_n}\int_{-\infty}^{\infty} K\left(\frac{y}{h_n}\right) g(x-y)\,dy,$$

where $(h_n)_{n\in\mathbb{N}}$ is a sequence of positive numbers such that $\lim_{n\to\infty} h_n = 0$. If g is continuous at x, then we have

$$\lim_{n\to\infty} g_n(x) = g(x). \tag{3.29}$$

Proof: Since $\int_{-\infty}^{\infty} \frac{1}{h} K\left(\frac{y}{h}\right)\, dy = \int_{-\infty}^{\infty} K(y)\, dy = 1$, we may write

$$\begin{aligned} |g_n(x) - g(x)| &= \left| g_n(x) - g(x) \int_{-\infty}^{\infty} \frac{1}{h_n} K\left(\frac{y}{h_n}\right) dy \right| \\ &\leq \int_{-\infty}^{\infty} \left| \{g(x-y) - g(x)\} \frac{1}{h_n} K\left(\frac{y}{h_n}\right) \right| dy. \end{aligned}$$

Let $\delta > 0$ be arbitrary. We now split the integration interval into 2 parts: $\{y : |y| \geq \delta\}$ and $\{y : |y| < \delta\}$. The first integral can be bounded from above by

$$\begin{aligned} &\int_{|y|\geq\delta} \frac{|g(x-y)|}{y} \frac{y}{h_n} K\left(\frac{y}{h_n}\right) dy + |g(x)| \int_{|y|\geq\delta} \frac{1}{h_n} K\left(\frac{y}{h_n}\right) dy \quad \leq \\ &\frac{\sup_{|v|\geq\delta/h_n} |v\, K(v)|}{\delta} \int_{|y|\geq\delta} |g(x-y)|\, dy + |g(x)| \int_{|t|\geq\delta/h_n} K(t)\, dt \quad \leq \\ &\frac{\sup_{|v|\geq\delta/h_n} |v\, K(v)|}{\delta} \int_{-\infty}^{\infty} |g(u)|\, du + |g(x)| \int_{|t|\geq\delta/h_n} K(t)\, dt \quad . \end{aligned}$$

Letting $n \to \infty$ and using that K is absolutely integrable, we see that these terms can be made arbitrarily small. The integral over the second region can be bounded from above by

$$\sup_{|y|<\delta} |g(x-y) - g(x)| \int_{|y|<\delta} K(y)\, dy \leq \sup_{|y|<\delta} |g(x-y) - g(x)|.$$

Since this holds for all $\delta > 0$ and g is continuous at x, the above expression can be made arbitrarily small. □.

As a corollary, we obtain the following asymptotic results (taken from Parzen (1962)) for the mean and variance of the kernel estimator at a point x.

Corollary 3.8.7 (Parzen) *Let $\widehat{f}_n$ be a kernel estimator such that its kernel K is bounded and satisfies $\lim_{|y|\to\infty} y\, K(y) = 0$. Then $\widehat{f}_n$ is an asymptotically unbiased estimator for f at all continuity points x if $\lim_{n\to\infty} h_n = 0$.*

Proof: Apply Theorem 3.8.6 to Formula (3.22). □.

In the above corollary, there is no restriction on the rate at which $(h_n)_{n\in\mathbb{N}}$ converges to 0. The next corollaries show that if $(h_n)_{n\in\mathbb{N}}$ converges to 0 slower than n^{-1}, then $\widehat{f}_n(x)$ is consistent in the sense that the MSE converges to 0.

Corollary 3.8.8 (Parzen) *Let $\widehat{f}_n$ be a kernel estimator such that its kernel K is bounded and satisfies $\lim_{|y|\to\infty} y\, K(y) = 0$. If $\lim_{n\to\infty} h_n = 0$ and x is a continuity point of the unknown density f, then*

$$\lim_{n\to\infty} n\, h_n\, \mathrm{Var}\, \widehat{f}_n(x) = f(x) \int_{-\infty}^{\infty} K^2(y)\, dy.$$

Proof: First note that since K is bounded, K^2 also satisfies the conditions of Theorem 3.8.6. Hence, the result follows from applying Theorem 3.8.6 and Exercise 3.27 to Formula (3.23). □.

Corollary 3.8.9 (Parzen) *Let $\widehat{f}_n$ be a kernel estimator such that its kernel K is bounded and satisfies $\lim_{|y|\to\infty} y\,K(y) = 0$. If $\lim_{n\to\infty} h_n = 0$, $\lim_{n\to\infty} n\,h_n = \infty$ and x is a continuity point of the unknown density f, then*

$$\lim_{n\to\infty} \mathrm{MSE}_x(\widehat{f}_n) = 0.$$

Proof: It follows from Corollary 3.8.8 that $\lim_{n\to\infty} \mathrm{V}ar\, \widehat{f}_n(x) = 0$. The result now follows by combining Corollary 3.8.7 and Exercise 3.20. □.

Although the above theorems give insight in the asymptotic behaviour of density estimators, they are not sufficient for practical purposes. Therefore, we now refine them by using Taylor expansions.

Theorem 3.8.10 *Let $\widehat{f}_n$ be a kernel estimator such that its kernel K is bounded and symmetric and such that $\int_{-\infty}^{\infty} |t^3|\,K(t)\,dt$ exists and is finite. If the unknown density f has a bounded third derivative, then we have that*

$$\mathrm{E}\,\widehat{f}_n(x) = f(x) + \frac{1}{2}\,h^2 f''(x) \int_{-\infty}^{\infty} t^2\,K(t)\,dt + o(h^2), h \downarrow 0 \tag{3.30}$$

$$\mathrm{V}ar\,\widehat{f}_n(x) = \frac{1}{n\,h}\,f(x) \int_{-\infty}^{\infty} K^2(t)\,dt + o\left(\frac{1}{n\,h}\right), h \downarrow 0 \text{ and } nh \to \infty \tag{3.31}$$

$$\mathrm{MSE}_x(\widehat{f}_n) = \frac{1}{n\,h}\,f(x) \int_{-\infty}^{\infty} K^2(t)\,dt + \frac{1}{4}\,h^4 \left(f''(x) \int_{-\infty}^{\infty} t^2\,K(t)\,dt\right)^2 + o\left(\frac{1}{n\,h}\right) + o\left(h^4\right), \quad h \downarrow 0 \text{ and } nh \to \infty. \tag{3.32}$$

Proof: By Formula (3.22) and a change of variables, we may write the bias as

$$\mathrm{E}\,\widehat{f}_n(x) - f(x) = \int_{-\infty}^{\infty} K(t)\,\{f(x - th) - f(x)\}\,dt.$$

Now Taylor's Theorem with the Lagrange form of the remainder says that

$$f(x - th) = f(x) - th\,f'(x) + \frac{(th)^2}{2}\,f''(x) - \frac{(th)^3}{3!}\,f'''(\xi),$$

where ξ depends on x, t, and h and is such that $|x - \xi| < |th|$. Since $\int_{-\infty}^{\infty} K(t)\,dt = 1$, it follows that

$$\mathrm{E}\,\widehat{f}_n(x) - f(x) = \int_{-\infty}^{\infty} K(t) \left(-th\,f'(x) + \frac{(th)^2}{2}\,f''(x) - \frac{(th)^3}{3!}\,f'''(\xi)\right) dt,$$

which because of the symmetry of K simplifies to

$$\mathrm{E}\,\widehat{f}_n(x) - f(x) = \int_{-\infty}^{\infty} K(t) \left(\frac{(th)^2}{2}\,f''(x) - \frac{(th)^3}{3!}\,f'''(\xi)\right) dt.$$

If M denotes an upper bound for f''', then the first result follows from

$$\begin{aligned}\left|\mathrm{E}\,\widehat{f}_n(x) - f(x) - \frac{1}{2}\,h^2 f''(x)\int_{-\infty}^{\infty} t^2\,K(t)\,dt\right| &\leq \frac{h^3}{3!}\int_{-\infty}^{\infty}\left|t^3\,K(t)\,f'''(\xi)\right|\,dt \\ &\leq M\,\frac{h^3}{3!}\int_{-\infty}^{\infty}|t^3\,K(t)|\,dt,\end{aligned}$$

where the last term obviously is $o(h^2)$. The asymptotic expansion of the variance follows immediately from Corollary 3.8.8. In order to obtain the asymptotic expansion for the MSE, it suffices to combine Exercise 3.20 with Formulas (3.30) and (3.31). □

These expressions show that the asymptotic expressions are much easier to interpret than the exact expression of the previous section. For example, we can now clearly see that the bias decreases if h is small and that the variance decreases if h is large (cf. our discussion of the naive density estimator).

Theorem 3.8.10 is essential for obtaining optimal choices of the bandwidth. If we assume that f'' is square integrable, then it follows from Formula (3.32) that for $h \downarrow 0$ and $nh \to \infty$:

$$\mathrm{MISE}(\widehat{f}_n) = \frac{1}{nh}\int_{-\infty}^{\infty} K^2(t)\,dt + \frac{1}{4}\,h^4\int_{-\infty}^{\infty}(f'')^2(x)\,dx\left(\int_{-\infty}^{\infty} t^2\,K(t)\,dt\right)^2 + o\left(\frac{1}{nh}\right) + o(h^4). \tag{3.33}$$

The expression

$$\frac{1}{n\,h}\int_{-\infty}^{\infty} K^2(t)\,dt + \frac{1}{4}\,h^4\int_{-\infty}^{\infty}\left(f''\right)^2(x)\,dx\left(\int_{-\infty}^{\infty} t^2\,K(t)\,dt\right)^2 \tag{3.34}$$

is called the *asymptotic MISE*, often abbreviated as AMISE. Note that Formula (3.34) is much easier to understand than Formula (3.27). We now see (cf. Exercise 3.28) how to balance between squared bias and variance in order to obtain a choice of h that minimizes the MISE:

$$h_{\mathrm{AMISE}} = \left(\frac{\int_{-\infty}^{\infty} K^2(t)\,dt}{4n\left(\int_{-\infty}^{\infty} t^2\,K(t)\,dt\right)^2\int_{-\infty}^{\infty}\left(f''\right)^2(x)\,dx}\right)^{1/5}. \tag{3.35}$$

An important drawback of Formula (3.35) is that it depends on $\int_{-\infty}^{\infty}\left(f''\right)^2(x)\,dx$, which is unknown. However, there are good methods for estimating this quantity. For details, we refer to the literature (Silverman (1986); Wand and Jones (1995)). An example of a simple method is given in Exercise 3.29.

Given an optimal choice of the bandwidth h, we may wonder which kernel gives the smallest MISE. It turns out that the Epanechnikov kernel is the optimal kernel. However, the other kernels perform nearly as well, so that the optimality property of the Epanechnikov kernel is not very important in practice. For details, we refer to Silverman (1986) and Wand and Jones (1995).

3.9 Exercises

In all exercises $X, X_1, X_2, \ldots$ are independent identically distributed normal random variables with mean μ and variance σ^2, unless otherwise stated.

Exercise 3.1 *Assume that the main characteristic of a production process follows a normal distribution and that C_p equals* 1.33.

a) What is the percentage non-conforming items if the process is centred (that is, if $\mu = (USL + LSL)/2$)?

b) What is the percentage non-conforming items if $\mu = (2USL + LSL)/3$?

Exercise 3.2 *Find the asymptotic distribution of $\widehat{\sigma}$ and $1/\widehat{\sigma}$, where $\widehat{\sigma}$ is the MLE for σ. What is the asymptotic distribution of $\widehat{C}_p$?*

Exercise 3.3 *Find an explicit formula for a $100\,(1-\alpha)\%$-confidence interval for C_p based on $\widehat{C}_p$ (both exact and asymptotic). Implement your formula in* `R`.

Exercise 3.4 *Assume that σ is known. Show that $\Phi\left(\frac{b-\overline{X}}{\sigma}\right) - \Phi\left(\frac{a-\overline{X}}{\sigma}\right)$ is a biased estimator for $P(a < X < b)$. Hint: use Lemma 3.6.7.*

Exercise 3.5 *Perform a small simulation in* `R` *to compare the performance of the estimators (3.8) and (3.10).*

Exercise 3.6 *A capability study of a process parameter yields that $C_p = 2$ and $C_{pk} = 1.67$. Within the Six Sigma approach to quality control it is assumed that the process mean shifts in the long term. Assuming that this holds, determine which shift of the process mean corresponds to these values of the capability indices.*

Exercise 3.7 *In a certain capability study one wishes to estimate C_p within* 0.1 *accuracy with* 95% *confidence. How many samples does one need in this capability study?*

Exercise 3.8 *Describe the details of a test procedure to test $H_0 : C_p = c_0$ versus $H_0 : C_p > c_0$. Clearly state the test statistic, the assumptions, the distribution of the test statistic, the critical region and an expression for the power function.*

Exercise 3.9 *The file* `ceramic.txt` *contains measurements of electrical characteristics of assemblies during a pilot study of a new production process. The measurements were collected each half hour in rational subgroups of size* 7. *Assess the capability of this process if $USL = 20$, $LSL = 10$ and the nominal value is* 15. *What appears to be the most important problem?*

Exercise 3.10 *A sample has been taken from a new production process. The file* `current.txt` *contains measurements of current, which is a critical quality parameter of the production process. The specification limits of which are* 90 *and* 110 *milli-amperes.*

a) Test with $\alpha = 0.05$ whether this production process has a C_p that exceeds 2.0. *State and check the assumptions that you use.*

b) Compute a 95% β-expectation tolerance interval for the data. State and check the assumptions that you use. Relate your answer with the outcome of a).

Exercise 3.11 *A company makes cigar lighters for cars. It is important that the length of the cigar lighter stays within specifications, since otherwise the cigar lighters do not fit into the dashboard of the cars. The car company has set the following specifications on the lengths of the cigar lighters:* 5.10 ± 0.15 *cm. Production data can be found in the file* `cigar.txt`.

a) Check whether the data is normally distributed. Use $\alpha = 0.02$ *as significance level. How should you group the data in your further analysis?*

b) Check whether the process is in statistical control. From historic data, it is known that the process is in control when $\mu = 5.12$ *and* $\sigma = 0.023$. *If necessary, remove observations that seem to be associated with special causes.*

c) Compute 95% confidence intervals for the capability indices C_p *and* C_{pk}.

d) Estimate the percentage of cigar lighters that do not conform to specifications.

e) How do you assess the capability of the process? Base your assessment on as much as possible information that your statistical software provides.

f) The car company experiences some problems with installing the cigar lighters and decides to change the specifications to 5.11 ± 0.03 *cm. How do you now assess the process capability. What is now the most urgent process improvement issue?*

Exercise 3.12 *An extrusion die is used to produce aluminium rods. The diameter of the rods is a critical quality characteristic. Customers demand that the specifications on the rod diameters are* 32 ± 4 *cm. Samples of each 5 rods are taken. The results can be found in the data set* `rods.txt`. *You are assigned to perform a capability analysis and report back to the plant manager. How do you assess the capability of the process? What are your recommendations to the plant manager?*

Exercise 3.13 *Find a sufficient statistic for the parameter* θ *in a sample of a Beta(θ,1)-distribution. The density of a Beta(θ,1)-distribution is* $\theta\, x^{\theta-1}$ *for* $0 < x < 1$.

Exercise 3.14 *Show that for fixed* n, *the class of binomial distributions with* $0 < p < 1$ *is complete. Hint: use the Fundamental Theorem of Algebra.*

Exercise 3.15 *Let* X *be a Poisson variable. Find an unbiased estimator with minimal variance for* $P(X \neq 0)$.

Exercise 3.16 *Construct a* β*-expectation tolerance interval when* μ *is known and* σ^2 *is unknown.*

Exercise 3.17 *Construct a* β*-content tolerance interval at confidence level* α *when* μ *is known and* σ^2 *is unknown.*

The following optional exercises deal with kernel density estimators.

Exercise 3.18 *Verify that the naive estimator is a kernel estimator.*

Exercise 3.19 *Verify that the kernel estimator is a density.*

Exercise 3.20 *Prove that for any density estimator $\widehat{f}$ we have*

$$MSE_x(\widehat{f}) = \mathrm{V}ar\,\widehat{f}(x) + \Big(\mathrm{E}\,\widehat{f}(x) - f(x)\Big)^2.$$

Exercise 3.21 *Show that formula (3.27) can be rewritten as*

$$\begin{aligned}\mathrm{MISE}(\widehat{f_n}) =& \frac{1}{n\,h}\int_{-\infty}^{\infty} K^2(y)\,dy + \left(1-\frac{1}{n}\right)\int_{-\infty}^{\infty}\frac{1}{h^2}\left(\int_{-\infty}^{\infty} K\left(\frac{x-y}{h}\right) f(y)\,dy\right)^2 dx - \\ & \frac{2}{h}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K\left(\frac{x-y}{h}\right) f(y)\,dy f(x)\,dx + \int_{-\infty}^{\infty} f^2(x)\,dx.\end{aligned}$$

Exercise 3.22 *Calculate the following estimators for μ, σ^2 respectively, where $\widehat{f_n}$ is a kernel estimator for f with a symmetric kernel K (that is, $K(x) = K(-x)$):*

a) $\widehat{\mu} = \displaystyle\int_{-\infty}^{\infty} x\,\widehat{f_n}(x)\,dx.$

b) $\widehat{\sigma^2} = \displaystyle\int_{-\infty}^{\infty} (x-\widehat{\mu})^2\,\widehat{f_n}(x)\,dx.$

Exercise 3.23 *Verify by direct computation that the naive estimator is biased in general.*

Exercise 3.24 *Use (3.28) to find a formula for the Fourier transform of $\widehat{f_n}$.*

Exercise 3.25 *Suppose that K is a symmetric kernel, i.e. $K(x) = K(-x)$. Show that* $\mathrm{MISE}(\widehat{f_n})$ *equals*

$$\frac{1}{2\pi nh}\int_{-\infty}^{\infty}(\mathcal{F}K)^2(t)\,dt + \frac{1}{2\,\pi}\int_{-\infty}^{\infty}\left\{(1-\frac{1}{n})\,(\mathcal{F}K)^2(ht) - 2\,\mathcal{F}K(ht) + 1\right\}|\mathcal{F}f(t)|^2\,dt.$$

Hint: use Parseval's identity

$$\int_{-\infty}^{\infty} g_1(x)\,g_2(x)\,dx = \frac{1}{2\pi}\int_{-\infty}^{\infty}\mathcal{F}g_1(t)\,\overline{\mathcal{F}g_2(t)}\,dt.$$

Exercise 3.26 *The Laplace kernel is defined by $K(x) := \frac{1}{2}\,e^{-|x|}$. Use the results of the previous exercise to derive an expression for the MISE of the kernel estimator with the Laplace kernel when the density is an exponential density.*

Exercise 3.27 *Show that a version of Bochner's Theorem 3.8.6 holds if we relax the conditions $K \geq 0$ and $\int_{-\infty}^{\infty} K(y)\,dy = 1$ to $\int_{-\infty}^{\infty} |K(y)|\,dy < \infty$.*

Exercise 3.28 *Prove Formula (3.35) for the optimal bandwidth based on the AMISE.*

Exercise 3.29 *Prove that if f is a normal density with parameters μ and σ^2, then*

$$h_{\mathrm{AMISE}} = \left(\frac{8\,\sqrt{\pi}\,\int_{-\infty}^{\infty} K^2(t)\,dt}{3\,n\,\left(\int_{-\infty}^{\infty}\ t^2\,K(t)\,dt\right)^2}\right)^{1/5}\sigma.$$

How can this be used to select a bandwidth? What is the rationale behind this bandwidth choice?

Exercise 3.30 *Suppose that we take $h_n = c\,n^{-\gamma}$ where $c > 0$ and $\gamma \in (0,1)$. Which value of γ gives the optimal rate of convergence for the MSE?*

Chapter 4

Control Charts

Contents

In this chapter we present the mathematical background of techniques to detect deviations in a production process that would lead to non-conforming items. Such deviations may lead to increase of production costs or rework costs. The key factor to success is to accurately assess variability in production processes. If we accurately know the variability of a production process that is in control, then we are able to detect observations that indicate that the process has gone out of control. These procedures are known under the name control charts. The simplest and most widely used control chart is the Shewhart $\overline{X}$-chart, which should be used together with an R-chart, an S-chart or an S^2-chart. Shewhart introduced these charts in 1924 as simple tools to be used by workers in production lines. In spite of their simplicity, these charts turned out to be highly effective in practice. When properly set up, these charts quickly detect large changes in process means. In the 1950's CUSUM (Cumulative Sum) control charts were introduced to detect small process changes. Although they are mathematically optimal in a certain sense, it is hard to set them up correctly since they are very sensitive to small changes in the parameters. A convenient alternative is the EWMA chart (Exponentially Weighted Moving Average) which has its roots in time series analysis. The EWMA chart is easy to implement and not very sensitive to both parameter changes and non-normality of data, while it performs almost as good as a CUSUM chart for detecting small process changes.

It is important to note that there are two different uses of control charts in general. In retrospective or Phase I use (sometimes called Initial Study), observations are analyzed after

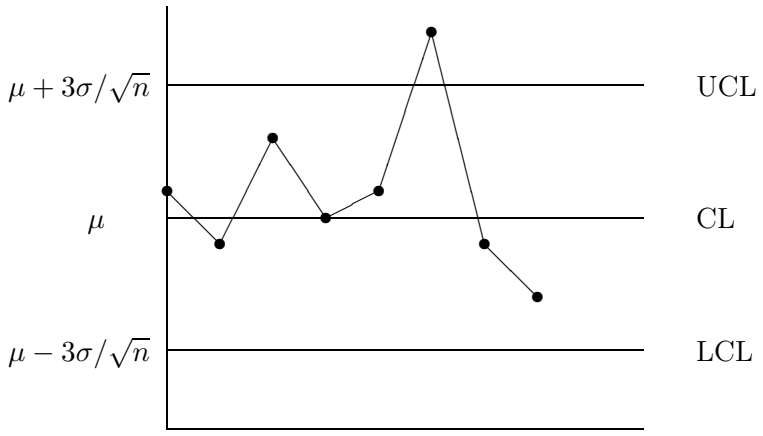


Figure 4.1: Shewhart $\overline{X}$-chart with control lines.

they have all been collected. Usually this is done during a pilot study of a production process, when one needs to estimate in-control process behaviour. This is input to a capability analysis. The other use is on-line or Phase II (sometimes called Control to Standard), in which in-control process parameters are assumed to be known or estimated from Phase I. In this use analysis of data is performed sequentially, that is repeatedly after each observation.

This chapter is organized as follows. In Section 4.1 we discuss Shewhart $\overline{X}$ control charts, including extensions with warning zones and runs rules. Shewhart charts for the variance are discussed in Section 4.2. In Section 4.3 we show a simple way to calculate run length distributions, as well as a general Markov chain approach by Brook and Evans. CUSUM charts and EWMA charts are discussed in Sections 4.4 and 4.5, respectively.

4.1 The Shewhart $\overline{X}$ chart

The basic data collection scheme for a Shewhart chart is as follows. At equidistant points in time one takes a small sample (usually of size 4 or 5) of a product and measures a relevant quality characteristic. Such samples are called rational subgroups. We denote the observations in rational subgroups by X_{ij}, $i = 1, 2, \ldots,$ and $j = 1, \ldots, n$, where n is the size of the rational subgroups. The rational subgroup should be chosen in such a way that the observations are independent and represent the short-term variability. The Shewhart $\overline{X}$ control chart basically is a time sequence plot of the averages $\overline{X_i}$ of the rational subgroups, together with the following 3 horizontal lines that indicate the process location and spread. Assume that μ is the mean $\mathrm{E}(X)$ of the quality characteristic and that σ^2 is the variance $\mathrm{V}ar(X)$ of the quality characteristic. Since $\mathrm{E}(\overline{X}_i) = \mu$ and $\mathrm{V}ar(\overline{X}_i) = \sigma^2/n$ for all i, the standard deviation of $\overline{X}_i$ equals $\sigma/\sqrt{n}$. The centre line (CL) is placed at the process mean μ of the quality characteristic. The other two lines are placed at distance $3\sigma/\sqrt{n}$ of the centre line. These lines are called the upper control limit (UCL) and the lower control limit (LCL), respectively. The control charts signals an alarm if an observation falls outside the region $(\mu - 3\sigma/\sqrt{n}, \mu + 3\sigma/\sqrt{n})$.

If we assume that $\overline{X}_i$ is normally distributed with mean μ and variance σ^2/n, then the

probability of a false alarm equals

$$1 - P(LCL < \overline{X_i} < UCL \mid \mu = \mu_0) = \Phi(-3) + (1 - \Phi(3)) = 2(1 - \Phi(3)) \approx 0.00270 \approx \frac{1}{370}.$$

If however the process mean shifts over a distance $\sigma/\sqrt{n}$, then (see Exercise 4.16)

$$1 - P(LCL < \overline{X_i} < UCL \mid \mu = \mu_0 \pm \frac{\sigma}{\sqrt{n}}) \approx 0.02278 \approx \frac{1}{44}.$$

In Phase I use of control charts, the centre line is placed at the overall mean of the observations. The control limits are computed using (pooled) estimates of the variance from rational subgroups, often in terms of the range rather than the sample standard deviation. The relation between the range and the standard deviation is explored in detail in Section 4.2. In Phase II use of control charts, the values μ and σ are either known from historical data or estimated from a capability study. In the latter case, it is not correct to apply the above calculation, although this is usually being ignored. For the effect of estimated parameters on the false alarm probability we refer to Exercise 4.5. Shewhart based his choice for placing UCL and LCL at distance 3 times the standard deviation of group averages on practical experience. It provides a good compromise to a low false alarm rate and a quick detection of out-of-control situations. Explicit calculations will be discussed in Section 4.3.

4.1.1 Additional stopping rules for the $\overline{X}$ control chart

The decision rule to signal an alarm when a group average falls below the LCL or above the UCL is based on the current rational subgroup and ignores information from previous rational subgroups. In order to increase the detection performance of an $\overline{X}$ control chart, several additional stopping rules have been proposed.

Sometimes warning limits are added at distance $2\sigma/\sqrt{n}$ of centre line. The idea of basing control charts on different regions of observation values is also used in so-called zone control charts; for an example see Exercise 4.1. One of the proposed stopping rules using warning limits is to signal an alarm if either a single group average falls outside the control limits or two successive group averages both fall either above the upper warning limit or below the lower warning limit. Such a stopping rule is known in the SPC literature as a runs rule. Examples of other runs rules include eight successive group averages below or above the centre line or eight increasing or decreasing group averages in a row. Often a combination of such rules is used. It is obvious that such combinations should be used with care, because they lead to an increase in the false alarm rate. Explicit computations can be found in Section 4.3.

4.2 Shewhart charts for the variance

Since the normal distribution is completely specified by the parameters μ (mean) and σ^2 (variance), it does not suffice to monitor the mean of a production process. Therefore it is good practice to set up a control chart for the process variance in addition to the $\overline{X}$ control chart. Both charts use the same rational subgroups. One would expect that the sample variance is the natural choice for the statistic be monitored. However, because control charts were devised by Shewhart in the pre-computer era, a simpler statistic, the range, was and is the standard choice. The range R of a sample $Y_1, \ldots, Y_n$ is defined as

$$R := \max_{1 \le i \le n} Y_i - \min_{1 \le i \le n} Y_i = Y_{(n)} - Y_{(1)}, \tag{4.1}$$

where $Y_{(i)}$ denotes the ith order statistic. It can be shown that for small sample sizes, the range is performing almost as good as the standard deviation (see Exercise 4.8 which requires results from Subsection 4.2.1). Before we define control charts for the variance, we need to study the range and standard deviation.

4.2.1 The mean and variance of the standard deviation and the range

In this subsection we derive expressions for the mean and variance of the standard deviation and the range. These expressions are needed to set up control limits. If the observations X_{ij} are independent and normally distributed with mean μ and variance σ^2, then the statistic $(n-1)S_i^2/\sigma^2$ is χ^2_{n-1} distributed. The expectation of $\sqrt{n-1}\ S_i/\sigma$ is therefore (see also Section 1.3) equal to

$$\int_0^\infty \sqrt{t}\ \frac{1}{2^{(n-1)/2}\ \Gamma((n-1)/2)}\, t^{(n-1)/2-1} e^{-t/2} dt = \sqrt{2}\,\frac{\Gamma(n/2)}{\Gamma((n-1)/2)}.$$

Hence,

$$\mathrm{E}\,(S_i) = \frac{\sqrt{2}}{\sqrt{n-1}}\,\frac{\Gamma(n/2)}{\Gamma((n-1)/2)}\,\sigma.$$

In particular, we have

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

Recall that $\Gamma(1) = 1$, $\Gamma(1/2) = \sqrt{\pi}$ and $\Gamma(x+1) = x\Gamma(x)$ for $x \neq 0, -1, -2, \ldots$. This immediately yields the following recursion for $c_4(n)$:

$$c_4(n+1) = \frac{\sqrt{n-1}}{\sqrt{n}\,c_4(n)},$$

with the initial condition $c_4(2) = \sqrt{2}/\sqrt{\pi}$. Note that the variance of S_i can also be expressed in terms of $c_4(n)$ as defined in Section 3.4), because

$$\mathrm{V}ar\,(S_i) = \mathrm{E}\,(S_i^2) - (\mathrm{E}\,(S_i))^2 = \sigma^2 - c_4^2(n)\sigma^2 = \left(1 - c_4^2(n)\right)\sigma^2. \tag{4.2}$$

We proceed by deriving an integral expression for $\mathrm{E}\,(R_i)$. Let F be the cumulative distribution function of X_{ij}. Then the distribution function of $\max_j(X_{ij})$ at y equals $F(y)^n$ and of $\min_j(X_{ij})$ at y equals $1-(1-F(y))^n$. We thus see that the expectation of $R_i = \max_j(X_{ij}) - \min_j(X_{ij})$ equals

$$\mathrm{E}\,(R_i) = \int_{-\infty}^{\infty} y\frac{\partial}{\partial y}(F(y)^n) - y\frac{\partial}{\partial y}(1-(1-F(y))^n)\,dy = \int_{-\infty}^{\infty} 1 - F(y)^n - (1-F(y))^n\,dy,$$

where the second equality follows from integration by parts. Note that the cumulative distribution function F of a normal distribution with mean μ and variance σ^2 equals $F(y) = \Phi\left(\frac{y-\mu}{\sigma}\right)$. We thus obtain

$$\mathrm{E}\,(R_i) = \int_{-\infty}^{\infty} 1 - F(y)^n - (1-F(y))^n\,dy = \int_{-\infty}^{\infty} 1 - \Phi\left(\frac{y-\mu}{\sigma}\right)^n - \left(1 - \Phi\left(\frac{y-\mu}{\sigma}\right)\right)^n dy.$$

Hence, a simple change of variables yields that for normally distributed samples we have

$$\mathrm{E}\,(R_i) = \sigma \int_{-\infty}^{\infty} 1 - \Phi(y)^n - (1-\Phi(y))^n\,dy.$$

This integral must be computed numerically, since it has been proved that the distribution function of the standard normal distribution has no closed-form primitive. The proportionality constant is usually denoted by $d_2(n)$, *i.e.*, $\mathrm{E}\,(R_i) = d_2(n)\sigma$. We conclude this subsection with the calculation of the variance of the range. It suffices to calculate $\mathrm{E}\,(R^2)$, because $\mathrm{V}ar\,(R_i) = \mathrm{E}\,(R_i^2) - (\mathrm{E}\,(R_i))^2$. Therefore we need to determine the joint distribution G of $(\max_j(X_{ij}), \min_j(X_{ij}))$. As before we denote with $F(x)$ and $f(x)$, the cumulative distribution function and the density, respectively, of the $X_{ij}{}'s$. We thus may write

$$\begin{aligned} G(x,y) &= P\{\max_j(X_{ij}) \le x \text{ and } \min_j(X_{ij}) \le y\} \\ &= P\{\max_j(X_{ij}) \le x\} - P\{\max_j(X_{ij}) \le x \text{ and} \min_j(X_{ij}) > y\} \\ &= F(x)^n - (F(x) - F(y))^n\, I_{\{y<x\}}, \end{aligned} \tag{4.3}$$

where $I_{\{y<x\}}$ denotes the indicator function of the set $\{y < x\}$, which equals 1 if $y < x$ and 0 if $y \ge x$. The joint density thus equals

$$\frac{\partial^2}{\partial x \partial y}\, G(x,y) = n(n-1)(F(x) - F(y))^{n-2} f(x) f(y)\, I_{\{y<x\}}.$$

Hence,

$$\begin{aligned} \mathrm{E}\,(R_i^2) &= \int_{-\infty}^{\infty}\int_{y}^{\infty} (x-y)^2 n(n-1)(F(x) - F(y))^{n-2} f(x) f(y)\, dx\, dy \\ &= -\int_{-\infty}^{\infty}\int_{y}^{\infty} 2(x-y) n[(F(x) - F(y))^{n-1} - (1 - F(y))^{n-1}] f(y)\, dx\, dy \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{x} 2[(F(x) - F(y))^n - (1 - F(y))^n + 1 - F(x)^n]\, dy\, dx \\ &= 2\int_{-\infty}^{\infty}\int_{0}^{\infty} [(F(y+r) - F(y))^n - (1 - F(y))^n + 1 - F(y+r)^n]\, dr\, dy, \end{aligned}$$

where the second and third equalities follow from integration by parts (note the special choice of primitives, which causes some terms to vanish). If we take for F the cumulative distribution function Φ of the standard normal distribution, then we obtain that the expectation of R_i^2 is proportional to σ^2. One usually writes

$$\mathrm{E}\,(R_i^2) = \left(C(n)^2 + d_2(n)^2\right)\sigma^2,$$

so that $\mathrm{V}ar\,(R_i) = C(n)^2\sigma^2$. Numerical calculation of the double integral $\mathrm{E}\,(R_i^2)$ leads to values of $C(n)$.

4.2.2 The R control chart

The R control chart contains 3 lines. The centre line is placed at the expectation R_0 of R, the range of a rational subgroup. If R_0 is unknown, then one uses the estimator $\overline{R}$ of a phase I study. In order to imitate the 3σ limits of the $\overline{X}$ control chart, it is convenient to introduce some notation. We throughout assume that the individual observations are independent and normally distributed with mean μ and variance σ^2. We will prove in this subsection that both the mean and the standard deviation of the range are proportional to the standard deviation

n	c_4	d_2	A_2	D_3	D_4	$D_{.001}$	$D_{.999}$
2	0.7979	1.128	1.880	0.000	3.267		
3	0.8862	1.693	1.023	0.000	2.575		
4	0.9213	2.059	0.729	0.000	2.282	0.199	5.309
5	0.9400	2.326	0.577	0.000	2.115	0.367	5.484
6	0.9515	2.534	0.483	0.000	2.004	0.535	5.619
7	0.9594	2.704	0.419	0.076	1.924	0.691	5.730

Table 4.1: Control chart constants.

of individual observations. The proportionality constants are denoted by $d_2(n)$ and $C(n)$, respectively. *I.e.*,

$$\begin{aligned} \mathrm{E}\,(R) &= d_2(n)\sigma \qquad (4.4)\\ \mathrm{V}ar\,(R) &= C(n)^2\sigma^2. \qquad (4.5) \end{aligned}$$

The upper and lower control limits are placed at 3 times the standard deviation of R. The upper control limit, UCL, is placed at

$$R_0 + 3\sqrt{\mathrm{V}ar\,(R)} = R_0 + 3C(n)\sigma = \left(1 + \frac{3C(n)}{d_2(n)}\right) R_0.$$

One often writes $D_4(n) = 1 + 3C(n)/d_2(n)$. The lower control limit, LCL, is similarly placed at $D_3(n)R_0$ with $D_3(n) = \max(1-3C(n)/d_2(n), 0)$. For an $\overline{X}$ control chart with control limits based on ranges one often defines $A_2(n) = 3/(\sqrt{n}\,d_2(n))$, so the control limits are placed at $\mu \pm A_2(n)\overline{R}$. The range R is not normally distributed, so that alarm probabilities are not equal to 0.0027 as is the case for the $\overline{X}$ chart with known parameters. An alternative way to set up control limits is based on quantiles of the range. These control limit are called probabilistic control limits, defined as

$$\begin{aligned} UCL &= D_{.999}(n)\tfrac{\overline{R}}{d_2(n)} \qquad (4.6)\\ LCL &= D_{.001}(n)\tfrac{\overline{R}}{d_2(n)}. \qquad (4.7) \end{aligned}$$

These limits are set up in such a way that the upper control limit UCL will be exceeded with probability 0.001 The same holds for the lower control limit LCL. There exist tables of values for $D_\gamma(n)$. It would be logical to use the values D_γ and $D_{1-\gamma}$ with $\gamma = 1 - \Phi(3) \approx 0.00135$. The exceedance probabilities for both control limits are then equal. In practice this is not done, because tables for D_γ usually do not include these values. The control limits based on the .001 and .999 quantiles should be compared with 3.09 standard deviation control limits, because $\Phi(3.09) \approx 0.999$. We conclude this subsection with Table 4.1, which gives values of some control chart constants.

4.2.3 The S and S^2 control charts

Alternatively, one may use the statistics S_i^2 or S_i to set up an S^2 control chart or an S control chart, respectively. The control limits for the S^2 control chart follow easily because $(n-1)S^2/\sigma^2$ follows a χ^2_{n-1} distribution. One may set up control limits at distance 3 times the standard deviation of S^2 (see Exercise 4.9). An alternative is to use probabilistic control limits as for the R control chart (see Exercise 4.9).

Control limits for the S control chart may be set up using (4.2). The centre line is placed at the expectation $\tilde{\sigma}$ of S_i, the range of a rational subgroup. If this expectation is unknown, then one uses the estimator $\overline{S}$ of a phase I study. The UCL is placed at $\tilde{\sigma} + 3\sqrt{1 - c_4(n)^2}\frac{\tilde{\sigma}}{c_4(n)} = B_4\tilde{\sigma}$. Likewise the LCL placed at $\tilde{\sigma} - 3\sqrt{1 - c_4(n)^2}\frac{\tilde{\sigma}}{c_4(n)} = B_3\tilde{\sigma}$. Probabilistic control limits may be obtained by taking square roots of the endpoints of two-sided confidence intervals for S_i^2 and replacing σ by $\tilde{\sigma}/c_4(n)$ (see Exercise 4.10).

By setting up appropriate control limits, all three types of control charts for the variance have comparable false alarm probabilities. The difference in performance between these charts is the ability to detect changes in the process variance. This is the topic of the next section.

4.3 Calculation of run lengths for Shewhart charts

The use of control charts in Phase I is retrospective in nature. In this phase it is most important to detect whether out-of-control situations have occurred. A good control chart in Phase II (on-line use), however, should have a low false alarm rate and the ability to quickly detect out-of-control situations. The goal of this section is to put these properties into quantitative terms. We will present two ways to perform calculations.

The way control charts are used is a special way of hypothesis testing. The difference with ordinary one-sample testing is that control charts in phase II involve sequential testing, since each time a new rational subgroup has been observed we again check whether the control chart signals. In that sense the false alarm probability of an individual point on the control chart is related to the type I error in hypothesis testing. The ability to detect process changes is related to the power of the test and thus to the type II error. For detailed discussions on the differences between control charts and hypothesis testing, we refer to Crowder et al. (1997), Stoumbos et al. (2000), Woodall (2000), and Woodall and Montgomery (1999). In order to incorporate the sequential character of control charts in phase II, the notions of type I and II error are replaced by properties of the run length N, defined as the number of rational subgroups observed until the control chart signals an alarm. The most widely studied property of the run length is its mean. In the SPC literature one usually uses the term Average Run Length (ARL). One distinguishes between an in-control ARL (ARL_{in}) and an out-of-control ARL (ARL_{out}). Obviously, ARL_{in} should be large (low false alarm rate or small type I error) and ARL_{out} should be small (fast detection of changes, or high power).

We now show how to compute ARL's for Shewhart $\overline{X}$-charts. It is obvious that if the process is in control (*i.e.*, the distribution of the observations is normal with parameters that correspond to the parameters used to set up the centre line and the control limits), each group average of an $\overline{X}$-chart has equal probability to fall outside the control limits. We denote this probability by p. Obviously $P(N = n) = (1 - p)^{n-1}p$ for $p = 1, 2, \ldots$. Thus in this case N has a geometric distribution. Note that $P(N > n) = (1 - p)^n$. Hence,

$$\mathrm{E}\, X = \sum_{n=0}^{\infty} P(N > n) = \sum_{n=0}^{\infty} (1 - p)^n = \frac{1}{p}.$$

Thus the ARL_{in} of an $\overline{X}$ control chart equals $1/p = (\Phi(-3) + 1 - \Phi(3))^{-1} \approx 370$. A similar calculation yields that $\mathrm{V}ar\, N = (1 - p)/p^2$. Since N is nonnegative, it cannot be symmetric. Moreover, if p is small as is usual the case for control charts, then the standard deviation of N almost equals its mean. The practical implication is that it may be very misleading to

judge the performance of a control chart by considering ARL's only. Run length distributions are usually highly skewed with large variability. The above calculation also holds for the R, S and S^2 control charts if we use the appropriate values for p. Since it is recommended to use the $\overline{X}$ chart in conjunction with one of these three control charts for the variance, so the ARL of interest should really be the ARL of the two used control charts (see Exercise 4.2 for an explicit calculation). Another issue that is often ignored is the fact that the process mean and variance are estimated, which may drastically change the run length distribution (see Exercise 4.16 for a simple example and Albers and Kallenberg (2004a), Albers and Kallenberg (2004b), Chakraborti (2000) and Ghosh et al. (1981) for further discussion).

Before we proceed by discussing run lengths of control charts with runs rules, we present a general lemma, that may be useful for calculating the probability generating function of non-negative lattice distributions.

Lemma 4.3.1 *Let X be a discrete random variable taking values on $1, 2, \ldots$ and define*

$$\widetilde{P}(z) := \sum_{j=1}^{\infty} P(X \geq j)\, z^j.$$

Then we have

a) $P(z) = 1 + \dfrac{z-1}{z}\,\widetilde{P}(z)$ *and* $\widetilde{P}(z) = z\,\dfrac{P(z)-1}{z-1}$,
where $P(z)$ is the probability generating function of X.

b) $\mathrm{E}\, X_{(m)} = P^{(m)}(1) = -\displaystyle\sum_{k=0}^{m-1} \frac{m!}{k!}(-1)^{m-k}\widetilde{P}^{(k)}(1)$,
where $\mathrm{E}\, X_{(m)} = EX(X-1)(X-2)\cdots(X-m+1)$ is the mth factorial moment[1] of X and $P^{(n)}$ denotes the n^{th} derivative of P.
In particular, $\mathrm{E}\, X = P'(1) = \widetilde{P}(1)$ *and* $\mathrm{Var}\,(X) = 2\widetilde{P}'(1) - \widetilde{P}(1) - (\widetilde{P}(1))^2$.

*P*roof.

a) We have

$$\begin{aligned}
\widetilde{P}(z) &= \sum_{j=1}^{\infty}\sum_{k=j}^{\infty} P(X=k)z^j \\
&= \sum_{k=1}^{\infty}\sum_{j=1}^{k} P(X=k)z^j \\
&= \sum_{k=1}^{\infty} P(X=k)\frac{z^{k+1}-z}{z-1} \\
&= \frac{z}{z-1}\left(\sum_{k=1}^{\infty} P(X=k)\left(z^k-1\right)\right) \\
&= \frac{z}{z-1}\,(P(z)-1).
\end{aligned}$$

[1]Here we follow the combinatorial convention, which differs from the convention in analysis and probability theory. In combinatorics, one distinguishes between the lower factorial $x_{(m)} = x(x-1)\ldots(x-m+1)$ and the upper factorial $x_{(m)} = x(x+1)\ldots(x+m-1)$

Rewriting the above formula yields a).

b) The first equality is a basic property of probability generating functions. The second equality follows by applying Leibniz's formula for derivatives of products:

$$(f \cdot g)^{(m)} = \sum_{k=0}^{m} \binom{m}{k} f^{(k)} g^{(m-k)}$$

to a) and evaluating at $z = 1$. □

It has been mentioned in Subsection 4.1.1 that adding runs rules changes the ARL's of a Shewhart control chart. We start with a simple example that can be analyzed directly using a recursion. Suppose we have an $\overline{X}$ chart with warning limits for which we agree that the chart signals if a point is outside the control limits or when there are two successive points beyond either of the warning limits. Let p be the probability that a point of the control charts falls outside the control limits, q be the probability that a point falls between a control limit and the nearby warning limit (the warning region), and r be the probability that a point is between the two warning limits (the "safe" region). Obviously $p + q + r = 1$. Let L denote the remaining ARL after a point in the good region and L' the corresponding ARL after a point in the warning region. If we start with a point in the safe region, then there are three possibilities:

a) the next point is safe again, with probability r, and we expect L more points before an alarm

b) the next point is in the warning region, with probability q, and we expect L' more points before an alarm

c) the next point falls outside the control limits, with probability p, and we stop.

Note that we use that the statistics plotted on a Shewhart control chart are independent. Combining these three possibilities, we obtain

$$L = r(1 + L) + q(1 + L') + p.$$

In a similar way we obtain

$$L' = r(1 + L) + q + p.$$

Combining these two equations, we obtain two linear equations with two unknowns. Hence, we may solve for L and L'. The ARL is then given by $p + q(L + 1) + r(L' + 1)$. Note that since we did not specify p, q and r the above derivation holds for both ARL_{in} and ARL_{out}. A similar way of reasoning yields ARL's for other runs rules.

In order to obtain more detailed results on run length distributions, a different approach is necessary. We present a method from Brook and Evans (1972), which was first presented in the context of run lengths for CUSUM charts. However, the method is very general and may also be applied to run lengths of Shewhart charts, as was first shown in Champ and Woodall (1987). The key in this approach is to represent the control chart as a Markov chain. In order to illustrate the method, we consider an $\overline{X}$ control chart with the runs rule "two successive points outside the same warning limit". We use a discrete-time Markov chain $(Y_k)_{k \in \mathbb{N}}$ with the following four states:

State 1 Current point is in the safe region.

State 2 Current point is in the upper warning region, and the previous point was not.

State 3 Current point is in the lower warning region, and the previous point was not.

State 4 Current point is outside the control limits, or is in the same warning region as the previous point (hence, the control chart signals).

Note that State 4 is an absorbing state. The probabilities of a point to fall into a certain region are denoted as follows: p_1: safe region, p_2: upper warning region, p_3: lower warning region, and $p_4 = 1-p_1-p_2-p_3$ the action region (out-of-control situation). We thus have the following transition matrix, where the rows denote the state before observing a new rational subgroup and the columns after observing a new rational subgroup:

$$P = \begin{pmatrix} p_1 & p_2 & p_3 & p_4 \\ p_1 & 0 & p_3 & p_2+p_4 \\ p_1 & p_2 & 0 & p_3+p_4 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Now let $S_k = (S_{k,1}, S_{k,2}, S_{k,3}, S_{k,4})$ be the distribution of the state space after n observed rational subgroups, *i.e.*, $S_{k,i} = P(Y_k = i)$. By the choice of states, the Markov property holds, so for $k \geq 1$ we have that

$$S_k = S_{k-1}\, P = S_0\, P^k, \tag{4.8}$$

where $S_0 = (p_{01}, p_{02}, p_{03}, p_{04})$ is the initial probability distribution of the states. The default choice for S_0 is $(1, 0, 0, 0)$ (but for other control charts this may not be the most natural choice). The run length distribution of N given the initial state distribution S_0 may now be computed from the relation

$$P(N \leq k) = S_{k,4}. \tag{4.9}$$

This relation can be easily programmed, but a closed form solution is preferable. Therefore we partition the transition matrix P by singling out the absorbing state 4 as follows:

$$P = \begin{pmatrix} R & \mathbf{p} \\ \mathbf{0} & 1 \end{pmatrix},$$

where $\mathbf{0}$ denotes a vector of zeros. Hence, we have $P^k = \begin{pmatrix} R^k & \mathbf{c_k} \\ \mathbf{0} & 1 \end{pmatrix}$. Since the rows of a transition matrix sum to 1, it follows that $\mathbf{c_k} = \mathbf{1} - R^k\mathbf{1} = (I - R^k)\mathbf{1}$. It follows from (4.8) and (4.9) that $\mathbf{c_k}_i$, the ith element of the vector $(I-R^k)\mathbf{1}$, equals $P(N \leq k \mid \text{first point is in state } i)$ if $1 \leq i \leq 3$. Hence, $P(N > k \mid \text{first point is in state } i)$ equals the ith element of the vector $R^k\mathbf{1}$. Now let N be the random vector, such that its ith element is the run length starting from state i with probability one. If we interpret the expression $P(\mathbf{N} > k\mathbf{1})$ as a shorthand for the separate coordinate-wise inequalities, then we can compute $E(\mathbf{N})$ as follows:

$$E(\mathbf{N}) = \sum_{k=0}^{\infty} P(\mathbf{N} > k\mathbf{1}) = \sum_{k=0}^{\infty} R^k\, \mathbf{1} = (I - R)^{-1}\, \mathbf{1}.$$

The above expression not only gives an explicit formula for the ARL, but we also see that the run length distributions in a certain multivariate sense remains geometric (as was the case for

the standard $\overline{X}$ control chart without runs rules). Further note that it is not always necessary to perform matrix inversion (which may be numerically unstable), since in some cases the matrix R is nilpotent, *i.e.*, there is an integer ℓ such that $R^k = 0$ for $k \geq \ell$.

For more calculations of run length distributions for Shewhart charts with runs rules, we refer to Champ and Woodall (1987), Champ and Woodall (1997), Does and Schriever (1992) and Fu et al. (2003) .

4.4 CUSUM procedures

As the basic statistical theory of testing was not developed until the 1950's, Shewhart had no choice but to develop simple monitoring tools. In this section we will show how likelihood ratios can be used to generate a class of sequential tests on which we can base control charts that are more sensitive to small changes. There is a close connection with a branch of statistics called changepoint analysis (see Hawkins et al. (2003)).

In order to better understand the background of sequential tests, we recall the well-known Neyman-Pearson Lemma for simple alternatives. Assume that $X_1, \ldots, X_n$ is a sample from a distribution with density f_θ. In order to test

$$H_0 : \theta = \theta_0$$

versus

$$H_1 : \theta = \theta_1$$

define the likelihood ratio Λ by

$$\Lambda_n = \frac{f_{\theta_1}(x_1) \ldots f_{\theta_1}(x_n)}{f_{\theta_0}(x_1) \ldots f_{\theta_0}(x_n)}. \tag{4.10}$$

The Neyman-Pearson Lemma says that the test which rejects H_0 if $\Lambda > c$ for a certain value k, has optimal power among all other tests of the same size, *i*.e., that have type I error less than or equal to $P_{H_0}(\Lambda > c)$.

A sequential version of this test was developed by Wald during World War II. The idea is to repeatedly test the null and alternative hypotheses when data is being gathered, thereby hoping for a reduced sample size. Again we use the likelihood ratio defined by (4.10), but now n does not denote the fixed sample size, but the current sample size. The decision procedure is based on numbers a and b and reads:

$$\begin{aligned} \text{accept } H_0 &: \text{ if } \Lambda_n \leq a \\ \text{accept } H_1 &: \text{ if } \Lambda_n \geq b \\ \text{continue sampling} &: \text{ if } a < \Lambda_n < b. \end{aligned}$$

The numbers a and b jointly determine the type I and II errors of this procedure, which can be shown to terminate in finite time with probability 1. Wald and Wolfowitz proved that this test, which is called SPRT (Sequential Probability Ratio Test), is optimal with respect to the ASN (Average Sample Number), *i.e.*, the average numbers of samples needed to reach a conclusion, among all tests with type I and II errors that do not exceed those of the SPRT at hand. Calculations of the ASN can be best performed by exploiting that the Λ_n's form a martingale.

Control charts may be interpreted as sequential tests when we interpret the null hypothesis as being in-control and the alternative hypothesis as being out-of-control (see Lai (1995) and Lai (2001)) for extensive discussions on this topic). A crucial difference with the SPRT is that in a control chart setting we never accept the null hypothesis, but continue sampling. This can be done by writing the loglikelihood ratio in a recursive form and resetting it to 0 when the null hypothesis is accepted. To formalize this, we write

$$\log \Lambda_n = \sum_{i=1}^{n} \log \left(\frac{f_{\theta_1}(x_i)}{f_{\theta_0}(x_i)} \right) = \sum_{i=1}^{n} Z_i. \tag{4.11}$$

Then the CUSUM statistic C_n, first introduced in Page (1954), is defined by the recursion

$$C_n = \max(0, C_{n-1} + Z_n), \tag{4.12}$$

where $C_1 = 0$. The CUSUM chart signals if the statistic exceeds a threshold (see example below for details). The CUSUM has an optimality property in terms of an out-of-control average run length.

CUSUM procedures can be developed for a wide range of distributions. For exponential families, the CUSUM procedures admit a convenient general from (see *e.g.*, Chapter 6 of Hawkins and Olwell (1998). Rather than showing these general formulas, we illustrate the derivation of a CUSUM procedure for the case of a normal distribution with known variance σ^2. As null hypotheses we take

$$H_0 : \mu = \mu_0$$

versus

$$H_1 : \mu = \mu_1$$

A straightforward calculation using the elementary fact $\mu_1^2 - \mu_0^2 = (\mu_1 - \mu_0)(\mu_1 + \mu_0)$ shows that

$$Z_i = -\frac{1}{2\sigma^2}\left((x_i - \mu_1)^2 - (x_i - \mu_0)^2\right) = \frac{\mu_1 - \mu_0}{\sigma^2}\left(x_i - \frac{\mu_0 + \mu_1}{2}\right).$$

If $\mu_1 > \mu_0$, then we have a so-called upper CUSUM chart C_n^+ which signals if

$$C_n^+ = \max\left(0, C_{n-1}^+ + X_n - \frac{\mu_0 + \mu_1}{2}\right) > h^+. \tag{4.13}$$

The value $\frac{\mu_0+\mu_1}{2}$ is called the reference value and is often denoted by k. If $\mu_1 < \mu_0$, then we have a so-called lower CUSUM chart C_n^- which signals if

$$C_n^- = \max\left(0, C_{n-1}^- - X_n + \frac{\mu_0 + \mu_1}{2}\right) > h^-. \tag{4.14}$$

One usually simultaneously runs an upper and lower CUSUM chart. A recommended choice for h^+ and h^- is 5σ, where σ is the standard deviation of an individual observation. The value $(\mu_0 + \mu_1)/2$ in 4.13 and (4.14) is called k (warning: one sometimes measures h and k in terms of standard deviations). Unlike Shewhart charts, CUSUM are usually applied to individual data. A refinement to improve the out-of-control ARL is not to restart at 0 but at another value, *e.g.*, at k. This idea was proposed in Lucas and Crosier (1982) and is called FIR (Fast Initial Response). The rationale is that when the process is in-control, the value of

C_n^+ will quickly return to 0, while in out-of-control situations it will take less time to exceed the threshold h^+.

Run length computations for the CUSUM chart are much harder than for the Shewhart charts. The ARL's are best calculated for the upper and lower CUSUM charts separately. In the sequel, we restrict ourselves to the upper CUSUM chart. The recursive form of the CUSUM shows that the conditional distribution of C_n^+ given the complete past equals the conditional distribution given C_{n-1}^+. This fact was used in the seminal paper Brook and Evans (1972) by discretizing the range of C_n^+ into subintervals and setting up a Markov chain where the states are these subintervals (see also William (1984)). Alternatively, one may derive integral equations for the ARL as in Page (1954); see Kemp (2001) for a lucid modern treatment. These integral equations usually can only be solved numerically. It can be shown that the Markov chain method of Brook and Evans corresponds to a simple numerical integration scheme corresponding to discretization at the midpoints of subintervals. Superior integration schemes can be found in Hawkins (1992). Other papers of interest on this topic include Abel and Fuchs (1985), Gold (1989) and Jun and Choi (1993).

An important drawback of CUSUM schemes is that the $ARL_{out-of-control}$ may increase quickly for values that move away from the specific value of the alternative hypothesis that is used to construct the CUSUM. Therefore we now show how to extend the CUSUM procedure to handle composite alternative hypotheses by using generalized likelihood ratios. For a nice comparison of the various principles that may be used to build control charts, we refer to Hawkins et al. (2003). Instead of the simple sequential hypotheses studied above, we now consider hypotheses based on the following model. Assume we have observations (or derived statistics in case of rational subgroups) X_i with distribution function F_θ and density f_θ, where the parameter θ belongs to a certain parameter space Θ.

$$\begin{aligned} &H_{0,i} : X_j \sim F_\theta \ (\theta \in \Theta_0) \text{ for } j \le i \\ &H_{1,i} : \text{for some } k < i \ \begin{cases} X_j \sim F_\theta \ (\theta \in \Theta_0) & \text{for } j \le k \\ X_j \sim F_\theta \ (\theta \in \Theta_1) & \text{for } k < j \le i, \end{cases} \end{aligned} \tag{4.15}$$

where $\Theta_0 \cup \Theta_1 \subset \Theta$ and $\Theta_0 \cap \Theta_1 = \emptyset$. The corresponding Generalized Likelihood Ratio statistic is

$$\Lambda_i = \max_{1 \le k < i} \frac{\prod_{j=1}^{k} \sup_{\theta_j \in \Theta_0} f_{\theta_j}(x_j) \prod_{j=k+1}^{i} \sup_{\eta_j \in \Theta_1} f_{\eta_j}(x_j)}{\prod_{j=1}^{i} \sup_{\theta_j \in \Theta_0} f_{\theta_j}(x_j)} = \max_{1 \le k < i} \frac{\prod_{j=k+1}^{i} \sup_{\eta_j \in \Theta_1} f_{\eta_j}(x_j)}{\prod_{j=k+1}^{i} \sup_{\theta \in \Theta_0} f_{\theta_j}(x_j)}. \tag{4.16}$$

The corresponding control chart is called the GLR control chart. Run length properties of this procedure have to be obtained by simulation.

In order to get more feeling for the flexibility of this setup, we consider some specific examples of common out-of-control situations for normally distributed random variables. For an extensive list of practical out-of-control situations in terms of patterns on means and variances we refer to Chapter 8 of Gitlow et al. (1989) and Chapter 6 of Griffith (1996). A slight variation on this control chart is the so-called cuscore chart. For a comparison of these charts, we refer to Runger and Testik (2003).

Persistent change of mean of given minimal size (σ, μ_0 and $\delta > 0$ known)

$$\begin{aligned} H_{0,i} &: \mu_1 = \ldots \mu_i = \mu_0 \\ H_{1,i} &: \begin{cases} \mu_j = \mu_0, & j = 1, \ldots, k, \\ \mu_j > \mu_0 + \delta, & j = k+1, \ldots, i, \end{cases} \end{aligned} \tag{4.17}$$

In this case (4.16) is evaluated as follows:

$$\begin{aligned} \log \Lambda_i &= \max_{1 \le k < i} \sup_{\mu > \mu_0 + \delta} \sum_{j=k+1}^{i} -(x_j - \mu)^2 + (x_j - \mu_0)^2 \\ &= \max_{1 \le k < i} \sum_{j=k+1}^{i} -(x_j - \max(x_j, \mu_0 + \delta))^2 + (x_j - \mu_0)^2 \end{aligned}$$

Writing $T_i = \log \Lambda_i$ and $W_j = -(x_j - \max(x_j, \mu_0 + \delta))^2 + (x_j - \mu_0)^2$, we have that $T_i = \max_{1 \le k < i} \sum_{j=k+1}^{i} W_j$. Now observe that the terms over which the maximum is taken in T_{i+1} are the same terms as for T_i with the addition of the extra term W_{i+1} resulting from the case $k = i$. The terms in T_{i+1} for $k < i$ differ by W_{i+1} from the corresponding term occurring in the maximum for T_i. This explains the following derivation of a recursion for T_I:

$$\begin{aligned} T_{i+1} &= \max_{1 \le k < i+1} \sum_{j=k+1}^{i+1} W_j \\ &= \max \left(\max_{1 \le k < i} \left(\sum_{j=k+1}^{i+1} W_j \right), \; W_{i+1} \right) \\ &= \max \left(\max_{1 \le k < i} \left(W_{i+1} + \sum_{j=k+1}^{i} W_j \right), \; W_{i+1} \right) \\ &= W_{i+1} + \max \left(\max_{1 \le k < i} \left(\sum_{j=k+1}^{i} W_j \right), \; 0 \right) \\ &= W_{i+1} + \max(T_i, 0). \end{aligned}$$

Epidemic alternative (σ, μ_0 and δ known)

$$\begin{aligned} H_{0,i} &: \mu_1 = \ldots = \mu_i = \mu_0 \\ H_{1,i} &: \begin{cases} \mu_j = \mu_0, & j = 1, \ldots, \ell, \\ \mu_j = \mu_0 + \delta, & j = \ell + 1, \ldots, m, \\ \mu_j = \mu_0, & j = m+1, \ldots, i \end{cases} \end{aligned} \tag{4.18}$$

Persistent non-monotone threshold crossing of mean (σ and δ known)

$$\begin{aligned} H_{0,i} &: \mu_j \le \delta, \quad j = 1, \ldots, i, \\ H_{1,i} &: \begin{cases} \mu_j \le \delta, & j = 1, \ldots, k, \\ \mu_j > \delta, & j = k+1, \ldots, i, \end{cases} \end{aligned} \tag{4.19}$$

A monotone version of this situation has been studied in Chang and Fricker (1999). This situation is a good model for tool wear.

Persistent monotone threshold crossing (σ and δ known)

$$\begin{aligned} H_{0,i} &: \mu_1 \leq \ldots \leq \mu_i \leq \delta \\ H_{1,i} &: \begin{cases} \mu_1 \leq \ldots \leq \mu_k < \delta & j = 1, \ldots, k \\ \delta < \mu_{k+1} \leq \mu_{k+2} \leq \ldots \leq \mu_i & j = k+1, \ldots, i \end{cases} \end{aligned} \tag{4.20}$$

Of course, it also possible to consider alternative hypotheses concerning the variance.

Persistent non-monotone threshold crossing of variance (σ and δ known)

$$\begin{aligned} H_{0,i} &: \sigma_j \leq \delta, \quad j = 1, \ldots, i, \\ H_{1,i} &: \begin{cases} \sigma_j \leq \delta, & j = 1, \ldots, k, \\ \sigma_j > \delta, & j = k+1, \ldots, i, \end{cases} \end{aligned} \tag{4.21}$$

4.5 EWMA control charts

The EWMA (Exponentially Weighted Moving Average) chart is based on ideas from Girshick and Rubin (1952), Roberts (1959) and Shiryaev (1963). The inspiration comes by Bayesian analysis by putting a prior distribution on the changepoint. The chart is defined by the recursion

$$V_i = \lambda X_i + (1-\lambda) V_{i-1}, \quad V_0 = 0 \text{ and } 0 < \lambda < 1. \tag{4.22}$$

It is easy to see that when $\lambda \to 0$, then we obtain a CUSUM procedure, while for $\lambda = 1$ we recover a Shewhart chart. A recommended choice for λ is $0.1 < \lambda < 0.3$. Run length calculations require numerically solving integral equations (**?**, Gan and Chang (2000)), or applying the Brook-Evans method after discretization (see Calzada and Scariano (2003) for the link between these two approaches). The main advantages of the EWMA chart are that its performance in terms of detecting small changes is nearly as good as for CUSUM charts and the robustness against deviations from normality. A disadvantage compared to Shewhart charts is that the control limits change over time, although it must be said that they converge rather quickly.

4.6 Exercises

Exercise 4.1 *Let $L < U$ be given real numbers. Assume that instead of the actual outcomes $X_1, \ldots, X_n$ of a sample from a normal distribution we only have information on the numbers of outcomes that fall below L, fall between U and L, and fall above U. Find Maximum Likelihood estimators for μ and σ^2 based on this information only. This is used in so-called gauge control charts. The idea goes back to Stephens (1948); for modern accounts we refer to Farnum and Stanton (1986) and Farnum and Stanton (1991)).*

Exercise 4.2 *A standard Shewhart $\overline{X}$ chart has an in-control ARL equal to 370. What is the ARL of a combined Shewhart $\overline{X}$ and S^2-chart, when we use probabilistic control limit so that also the S^2-chart has an in-control ARL of 370?*

Exercise 4.3 *The data set* `ceramic.txt` *contains measurements of electrical characteristics of assemblies during a pilot study of a new production process. The measurements were collected each half hour in rational subgroups of size* 7*. The production manager wishes that the average time to detect a shift of the mean of size* 0.7 *should be* 2 *hours. Suggest an appropriate control chart and a corresponding sampling strategy.*

Exercise 4.4 *A quality characteristic of a certain production process is known to be normally distributed with* $\sigma = 4$*. In order to obtain a good quality of the products, it is required to detect changes in the mean of* 4 *units. Currently rational subgroups are taken every* 15 *minutes.*

a) What is the necessary minimum size of the rational subgroups if the mean detection time is 1 *hour?*

b) What should be the sampling frequency if the producer wishes to be 90% *certain of detecting changes in the mean of* 4 *units within* 1 *hour?*

Exercise 4.5 *Assume that a capability study has been performed with* N *observations, subdivided into* m *rational subgroups of size* $n = N/m$*. In the subsequent Phase II, a Shewhart* $\overline{X}$ *chart has been set up based on rational subgroups of size* n*. The control limits are set at set at a distance of* $3\overline{S} = \sqrt{\frac{1}{N}\sum_{i=1}^{m} S_i^2}$*.*

a) Assume that the centre line is set at the known process mean. Compute the probability that a group average falls outside the control limits. Compare your result with the well-known probability 0.0027 *when* N *ranges from* 50 *to* 100*.*

b) Repeat your calculations for the case that the centre line is set at the overall mean from the capability study. Compare your results with a) and the standard 0.0027 *probability for the same ranges of* N*.*

c) The above results are independent of n*. May we therefore conclude that the choice of* n *is unimportant?*

Exercise 4.6 *A laboratory always performs an analysis of a certain process characteristic in duplicate. Discuss whether one should use the individual values or the average value on a Shewhart* $\overline{X}$ *chart.*

Exercise 4.7 *Write a program in* `R` *to produce of table of* $d_2(n)$*. Hint: use the function* `integrate`*.*

Exercise 4.8 *Study the relative efficiency (*i.e.*, the ratio of the variances) of the unbiased estimators* $S/c_4(n)$ *and* $R/d_2(n)$ *for* σ *when we sample from a normal distribution. For which values is* $R/d_2(n)$ *a reasonable alternative for* $S/c_4(n)$*? Hint: use the function* `adapt` *in the* `adapt` *package to evaluate* $\mathrm{Var}\,(R)$*.*

Exercise 4.9 *a) Show that* $\mathrm{Var}\, S^2 = \sigma^4/(n-1)^2$ *when the observations form a sample of size* n *from a normal distribution. Use this to derive control limits placed at a distance of 3 times the standard deviation of the sample variance.*

b) Derive probabilistic control limits for the S^2 *control chart.*

Exercise 4.10 *a) Set up an S control chart by placing control limits at a distance of 3 times the standard deviation of the sample standard deviation.*

b) Derive probabilistic control limits for the S control chart. Hint: use a two-sided confidence interval for the variance and take square roots of the end-points.

Exercise 4.11 *Write a program in* `R` *to produce a table of $D_{0.00135}$ and $D_{0.99875}$.*

Exercise 4.12 *A production process has been running for a long time, during which data has been gathered and analysed. The conclusion is that under stable conditions, the process has a mean of* 64 *and a standard deviation of* 0.02*. Consider the data set* `monthlyreport.txt` *of last month's production. Your task is to report to the Quality Supervisor. What should you report? Give a clear motivation for your report.*

Exercise 4.13 *In an injection moulding process it is important to keep part weight consistent over time. Therefore control charts are applied to monitor part weight. The data collection consists of taking 5 consecutive parts per shift and obtaining the weight of each part. The data set* `injectionmoulding.txt` *contains averages and range of the rational subgroups.*

a) Report through appropriate control charts on the moulding process. Do not forget to check the assumptions on which your control charts are based.

b) Subsequent investigations of the process revealed nothing unusual on the conditions during the data collection. Moreover, subsequent runs with the same unchanged process turned out parts that were consistently acceptable. How can this be explained?

c) Use the sample averages as individual observations and construct a control chart for individual observations. Interpret this new control chart. Compare your results with the previous control charts.

Exercise 4.14 *Give a formal justification of the conditional arguments with average run lengths in the first halve of Section 4.3.*

Exercise 4.15 *Compute ARL_{in} for an $\overline{X}$ control chart with the extra runs rule "two out of three successive point outside either warning limit".*

Exercise 4.16 *Consider the standard assumptions of the Shewhart $\overline{X}$-chart with rational subgroups of size n.*

a) What is the probability that 2 out of 3 consecutive observations fall in the region $(\mu_0 - 2\sigma/\sqrt{n}, \mu_0 + 2\sigma/\sqrt{n})$, if the individual observations are independent and normally distributed with expectation μ_0 and variance σ^2?

b) What is the alarm probability for a single group mean $\overline{X_i}$ if $\overline{X_i}$ is normally distributed with expectation $\mu_0 + \sigma/\sqrt{n}$ and variance σ^2/n?

c) We would like to change the control limits $\mu_0 \pm 3\sigma/\sqrt{n}$ into $\mu_0 \pm k\sigma/\sqrt{n}$. Which value should we choose for k in order to achieve that the probability of a single group mean $\overline{X}_i$ falling outside the control limits equals 0.01*? Assume that $\overline{X}_i$ is normally distributed with expectation μ_0 and variance σ^2/n.*

Exercise 4.17 *The lifetimes of certain parts of a machine need to be replaced often. One wishes to monitor the quality of the parts by monitoring the life times of these parts. Practical experience from the past indicate that lifetimes can be assumed to follow an exponential distribution. A recent sample has been taken and is presented in* `lifetimes.txt`*. The unit is days.*

a) Derive an upper 95%*-mean tolerance interval for the life times.*

b) In order to monitor the replacement frequency of the parts, one wishes to set up a Shewhart $\overline{X}$ *chart with* 3σ *control limits. Compute the centre line and the control limits of such a control chart, if we assume that the observations in the sample correspond to an in-control situation.*

c) Indicate one or two problems in setting up such a Shewhart-like chart for this situation.

d) Derive a one-sided CUSUM procedure based on an SPRT procedure when the in-control mean equals 5 *days and one wishes to detect an out-of-control mean equal to* 3 *days.*

Exercise 4.18 *Suppose that apart from the standard out-of-control criterion ("one observation outside the control limits") we also use the following runs rule on a Shewhart* $\overline{X}$ *chart: two successive points at distance* $2\sigma_{\overline{X}}$ *or more from the centre line. Compute the* $ARL_{\text{in-control}}$ *for this control chart.*

Exercise 4.19 *Compute the in-control Average Run Length of a standard Shewhart* $\overline{X}$*-chart with* 3σ *control limits if we add the runs rule "alarm if 3 consecutive observations on the same side of the centre line".*

Exercise 4.20 *An EWMA control chart is defined in terms of moving averages as follows:* $V_i = \lambda X_i + (1-\lambda)V_{n-1}$*, where* $V_0 = \mu$ *and* $0 < \lambda < 1$*. Prove that*

$$V_3 = (1-\lambda)^3\mu + \lambda(1-\lambda)^2 X_1 + \lambda(1-\lambda)X_2 + \lambda X_3.$$

Assume that the observations X_1*,* X_2*,* X_3 *are independent with expectation* μ *and variance* σ^2*. Calculate the expectation and variance of* Z_3*.*

Chapter 5

Solutions Exercises Historical Introduction to SPC

Solution of Exercise 1.1
The three methods of managing variation mentioned in the text are

Fitness for use In ancient times, products were simple and unique. Usually one craftsman was one craftsman responsible for the product, there were no specifications or designs. The item was inspected and worked until it was assessed by the craftsman to be fit for use.

Specification and tolerance Specifications and tolerances were necessary to communicate between industries and define what should be delivered.

Control chart The control chart is a statistical tool, which is used to distinguish between variation in a process due to common causes and variation due special causes.

Solution of Exercise 1.2
The test methods mentioned in the text are

- inspection by a craftsman
- subjective inspection by an inspector, who disassembles and reassembles the product, and tests for functionality.
- objective inspection through gauges
- use of specifications, dimensioned tolerances and standards.

Solution of Exercise 1.3
The products were simple and unique. The craftsman worked at the crafted item until it was fit for his use.

Solution of Exercise 1.4
Variation became a critical issue through the development of the concept of interchangeable

parts. With the later rise of mass production, variation became even more critical because inspection turned to be too expensive.

Solution of Exercise 1.5
The reason that standards for inspection were formulated was the rise of mass production.

Solution of Exercise 1.6
Gauges are set to specify when a product must be rejected. If the product satisfies the gauges, then the product is a go. It can be shipped and sold. When the product does not satisfy the gauge, the product is a no-go and will be rejected.

Solution of Exercise 1.7
The goal of setting tolerances and specifications is to get a more objective inspection system. It also facilitates the communication between industries. Specifications are necessary to describe quality characteristics.

Solution of Exercise 1.8
Interchangeability means that a part or component of the product can be replaced by another part of component and maintain the same function. The interchangeable parts must satisfy the same specifications. Examples of "modern" products with interchangeable parts include bicycles, televisions and computers.

Solution of Exercise 1.9
It is necessary that there is as little as possible variation in the interchangeable components. If there is too much variation, components will no longer be interchangeable.

Solution of Exercise 1.10
The goal of a control chart is to distinguish between variation in a process due common causes and special causes. Special causes can be identified and should be removed.

Solution of Exercise 1.11
The approach based on setting specifications focuses on quality of individual products. The focus of Shewhart's approach is performance of the process. There is no relationship between the two approaches. A process can be out of statistical control, but all of its outcomes may be within specifications.

Solution of Exercise 1.12
Shewhart determined through practical experience empirical that control limits at a distance of 3 standard deviations perform well. If the distance is smaller than 3σ, then the control chart indicates too often that the process is out of control. This results in an overreaction (too many false alarms). If the distance is bigger than 3σ, then the control chart indicates too late that the process is out of control. This may result in an unnoticed out of control situation.

Solution of Exercise 1.13
Taguchi presented the idea that any variation from a desired target is a loss to society. Taguchi argues that specification limits does not seem to reflect reality, because a product becomes gradually worse when it deviate from the desired target. He therefore thinks that the idea that loss somehow jumps in value at specification limits does not seem to reflect the real world.

Solution of Exercise 1.14
If the variation in critical parameters of parts is too big, the parts will not fit because there is an accumulation of tolerances. This is called tolerance stack-up.

Chapter 6

Solutions Exercises Introduction to R

Solution of Exercise 2.1

```
pchisq(5.2, df=3, lower.tail=FALSE)
```

Solution of Exercise 2.2

```
qnorm(0.99)
```

Solution of Exercise 2.4

```
dataset <- rnorm(100, mean = 10, sd = sqrt(2))
hist(dataset, breaks=10, xlab="10 bins")
hist(dataset, breaks=20, xlab="20 bins")

# A more automated way of playing can be made as follows;

histdemo <- function(data,numberofbins=floor(sqrt(length(data))))
{
hist(dataset, breaks= numberofbins, xlab= paste("number of bins = ",numberofbins))
}

qqnorm(dataset), qqline(dataset) # adds line through 1st and 3rd quartile

shapiro.test(dataset) # Shapiro-Wilks test for normality

t.test(dataset, alternative=c(‘‘two.sided’’),conf.level=0.95,mu=10)

# test for variance does not exist in R, we test using confidence interval:

n <- length(dataset)
```

```
c((n-1)*var(dataset)/qchisq(0.975,df=n-1),(n-1)*var(dataset)/qchisq(0.025,df=n-1))

# alternative: test with critical region

if( (n-1)*var(dataset)/2 >= qchisq(0.975,df=n-1)
    || (n-1)*var(dataset)/2 <= qchisq(0.025,df=n-1))
c("Reject that variance is equal to 2")
else c("Do not reject that variance is equal to 2")
```

Solution of Exercise 2.3

```
dataset <- rexp(10, rate=1/3)
mean(dataset) # check whether parametrization is right
```

Solution of Exercise 2.10

The following implementation is not very efficient, but has the advantage of showing some useful features of `R` like `deparse`, `apply` and the use of anonymous functions. The function `abline` draws lines with a given intercept and slope.

```
Weibullplot <- function(data,shapepar,scalepar,
nameofdata=deparse(substitute(data)))
{
n<-length(data)
orderstat <- sort(data)
quantiles <- sapply(seq(1:n)/(n+1),
(function(x) qweibull(x,shape=shapepar,scale=scalepar)))
plot(orderstat,quantiles,main=paste("Weibull plot of", nameofdata) ,
xlab=paste("Sample quantiles of", nameofdata),
ylab=paste("Quantiles of Weibull (",shapepar,",",scalepar,") distribution"),
abline(0,1) )
}
```

A more efficient solution would be to use the internal function `ppoints` and the fact that `qweibull` is threadable over lists.

```
Weibullplot2 <- function(data,shapepar,scalepar,
nameofdata=deparse(substitute(data)))
{
plot(sort(data),qweibull(ppoints(data),shape=shapepar,scale=scalepar),
main=paste("Weibull plot of", nameofdata) ,
xlab=paste("Sample quantiles of", nameofdata),
ylab=paste("Quantiles of Weibull (",shapepar,",",scalepar,") distribution") ,
abline(0,1))
}
```

The data sets can be generated using

```
data1 <- rweibull(30,shape=3,scale=2)
data2 <- rgamma(30,shape=7.57, rate=0.235)
```

Chapter 7

Solutions Exercises Process Capability Analysis

Solution of Exercise 3.1
We write $d = \frac{1}{2}(USL - LSL)$, so $d/\sigma = 3\,C_p$.

a) Since the process is centred, standardization of X (see Formula (3.2)) yields that the percentage non-conforming items equals

$$P(X < LSL) + P(X > USL) = 2\Phi(-d/\sigma) = 2\Phi(-3\,C_p) = 2\Phi(-3.99) = 0.000066.$$

b) If $\mu = (2USL + LSL)/3$, then in the same way we obtain that

$$P(X < LSL) + P(X > USL) = \Phi(-4C_p) + \Phi(-2C_p) = \Phi(-5.32) + \Phi(-2.66) = 0.0039.$$

Alternatively, we may use (3.3)

$$C_{pk} = \left(1 - \frac{\mid \mu - 1/2(LSL + USL) \mid}{d}\right) C_p = \left(1 - \frac{1/6 \mid USL - LSL \mid}{1/2 \mid USL - LSL \mid}\right) = \frac{2}{3}\, C_p\,.$$

The expected proportion non-conforming items can thus by (3.5) be expressed as

$$P(X < LSL) + P(X > USL) = \Phi(-3(2\,C_p - C_{pk})) + \Phi(-3\,C_{pk}) = \Phi(-5.32) + \Phi(-2.66) = 0.0039.$$

Solution of Exercise 3.2
By Theorem 3.5.2, we have $\sqrt{n}(\widehat{\sigma^2} - \sigma^2) \xrightarrow{d} N\left(0, 2\sigma^4\right)$. Applying Theorem 3.5.1(Cramér) with $g(x) = x^{-\frac{1}{2}}$, we obtain

$$\sqrt{n}\left(\frac{1}{\widehat{\sigma}_n} - \frac{1}{\sigma}\right) \to N\left(0, 2\sigma^4 \left(\frac{-1}{2\sigma^3}\right)^2\right) = N\left(0, \frac{1}{2\sigma^2}\right).$$

Since $\widehat{C_p} = \frac{USL - LSL}{6\widehat{\sigma}}$, it follows that

$$\sqrt{n}(\widehat{C_p} - C_p) \to N\left(0, \frac{(USL - LSL)^2}{72\sigma^2}\right).$$

Solution of Exercise 3.3
We first derive an exact confidence interval for C_p for normal samples. The ML estimator for σ equals $\widehat{\sigma} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2}$. Recall that $\frac{n\widehat{\sigma^2}}{\sigma^2} \sim \chi^2_{n-1}$. Hence,

$$P\left(\chi^2_{n-1,\alpha/2} \leq \frac{n\widehat{\sigma^2}}{\sigma^2} \leq \chi^2_{n-1,1-\alpha/2}\right) = 1 - \alpha$$

By the Invariance Principle (Theorem 3.3.2) we have that $\widehat{C_p}/C_p = \widehat{\sigma}/\sigma$. Substituting this into the above defining relation for a confidence interval for σ^2, we obtain

$$P\left(\chi^2_{n-1,\alpha/2} \leq \frac{nC_p^2}{\widehat{C_p}^2} \leq \chi^2_{n-1,1-\alpha/2}\right) = 1 - \alpha.$$

Hence, an exact $100\,(1-\alpha)\%$-confidence interval for C_p equals

$$\left(\sqrt{\frac{\chi^2_{n-1,\alpha/2}}{n}}\,\widehat{C_p},\ \sqrt{\frac{\chi^2_{n-1,1-\alpha/2}}{n}}\,\widehat{C_p}\right).$$

We now derive an asymptotic confidence interval for C_p. We derived in Exercise 3.2 that

$$\sqrt{n}(\widehat{C_p} - C_p) \to N\left(0, \frac{(USL - LSL)^2}{72\sigma^2}\right).$$

Since $n\widehat{\sigma^2}/\sigma^2 \sim \chi^2_{n-1}$ and the χ^2_{n-1} can be realized as the sum of $n-1$ standard normal random variables, continuity of the function $\sqrt{x}$ and the Strong Law of Large Numbers imply that $\widehat{\sigma}/\sigma$ converges in probability to 1. Hence, Slutsky's Lemma implies that we also have

$$\sqrt{n}(\widehat{C_p} - C_p)\widehat{\sigma} \to N\left(0, \frac{(USL - LSL)^2}{72}\right).$$

Hence, for large n we have

$$P\left(-z_{\alpha/2} \leq \frac{\sqrt{72n\sigma^2}(\widehat{C_p} - C_p)}{USL - LSL} \leq z_{\alpha/2}\right) \approx 1 - \alpha$$

A $100\,(1-\alpha)\%$-confidence asymptotic interval for C_p is thus given by

$$\left[\widehat{C_p} - z_{\alpha/2}\frac{USL - LSL}{\sqrt{72n\sigma^2}},\ \widehat{C_p} + z_{\alpha/2}\frac{USL - LSL}{\sqrt{72n\sigma^2}}\right]$$

These confidence intervals can be implemented in `R` in the following way:

```
CpCIexact <- function(data,LSL,USL,alpha=0.05){
n<-length(data)
Cpest <- (USL-LSL)/(6*sd(data))
c(Cpest*sqrt(qchisq(alpha/2,df=n-1)/(n-1)),
```

```
Cpest*sqrt(qchisq(1-alpha/2,df=n-1)/(n-1)))
}

CpCIasymptotic <- function(data,LSL,USL,alpha=0.05){
n<-length(data)
Cpest <- (USL-LSL)/(6*sd(data))
c("Cp = ",Cpest,100*(1-alpha)," % confidence interval=",
Cpest-qnorm(1-alpha/2)*(USL-LSL)/sqrt(72*n*var(data)),
Cpest+qnorm(1-alpha/2)*(USL-LSL)/sqrt(72*n*var(data)))}
```

Solution of Exercise 3.4

Since the $N(\mu, \sigma^2)$ distribution has cumulative distribution function $F(u) = \Phi\left(\frac{u-\mu}{\sigma}\right)$, it follows that $\Phi\left(\frac{b-\bar{X}}{\sigma}\right) - \Phi\left(\frac{a-\bar{X}}{\sigma}\right) = F(\mu+b-\overline{X}) - F(\mu+a-\overline{X})$. By Proposition 3.6.7, we have

$$\begin{aligned}
\mathrm{E}\left[\Phi\left(\frac{b-\overline{X}}{\sigma}\right) - \Phi\left(\frac{a-\overline{X}}{\sigma}\right)\right] &= P(\mu+a-\overline{X} < X < \mu+b-\overline{X}) \\
&= P\left(\frac{a-\mu}{\sigma\sqrt{1+1/n}} < \frac{X+\overline{X}-2\mu}{\sigma\sqrt{1+1/n}} < \frac{b-\mu}{\sigma\sqrt{1+1/n}}\right) \\
&= \Phi\left(\frac{b-\mu}{\sigma\sqrt{1+1/n}}\right) - \Phi\left(\frac{b-\mu}{\sigma\sqrt{1+1/n}}\right),
\end{aligned}$$

where we because of independence we have that $Var(X - \overline{X}) = (1 + 1/n)\,\sigma^2$. Hence, $\Phi\left(\frac{b-\overline{X}}{\sigma}\right) - \Phi\left(\frac{a-\overline{X}}{\sigma}\right)$ is a biased estimator for $P(a < X < b)$, but the bias vanishes as $n \to \infty$

Solution of Exercise 3.5

Note the use of `apply` in the solution which avoids a loop. It is almost always possible to avoid loops. Loops may be slow because they are memory consuming.

```
MLconf <- function(a,b,data){
n <- length(data)
pt((b-mean(data))/sd(data),df=n-1)-pt((a-mean(data))/sd(data),df=n-1)}

UMVUconf <- function(a,b,data){
n <- length(data)
pt(sqrt(n/(n-1))*(b-mean(data))/sd(data),df=n-1)-
pt(sqrt(n/(n-1))*(a-mean(data))/sd(data),df=n-1)}

simulationloop <- function(iterations=100,size,a,b,mu=0,var=1){
truevalue <- pnorm((b-mu)/sqrt(var))-pnorm((a-mu)/sqrt(var))
setup <- matrix(c(iterations,size,mu,var,a,b,truevalue),nrow=1)
dimnames(setup) <- list("",c("iterations","sample size", "mean",
```

```
"variance","a","b", "P(a<X<b)"))
print(setup,print.gap=2)
cat("\n") # empty new line
sample <-matrix(rnorm(size*iterations,mean=mu,sd=sqrt(var)),nrow=iterations,
byrow=T)
MLestim <-apply(sample,1,function(x) MLconf(a,b,x))
UMVUestim <-apply(sample,1,function(x) UMVUconf(a,b,x))
resultsML <- matrix(c(truevalue,mean(MLestim),sd(MLestim),
truevalue-mean(MLestim),sd(MLestim)+(mean(MLestim)-truevalue)^2),nrow=1,
byrow=T)
resultsUMVU <- matrix(c(truevalue,mean(UMVUestim),sd(UMVUestim),
truevalue-mean(UMVUestim),
sd(UMVUestim)+(mean(UMVUestim)-truevalue)^2),nrow=1,byrow=T)
results <- rbind(resultsML,resultsUMVU)
dimnames(results) <- list(c("ML estimator", "UMVU estimator"),
c("true", "average", "st. dev.", "bias","MSE"))
print(results,digits=3,print.gap=2)
}
```

Solution of Exercise 3.6

Assume without loss of generality that the process has shifted to the right, *i.e*, $\mu = \frac{(USL+LSL)}{2} + \delta$ with $\delta > 0$. Hence, in this case

$$\begin{aligned} C_{pk} &= \min\left(\frac{USL-\mu}{3\sigma}, \frac{\mu-LSL}{3\sigma}\right) = \min\left(\frac{\frac{USL-LSL}{2}-\delta}{3\sigma}, \frac{\frac{USL-LSL}{2}+\delta}{3\sigma}\right) \\ &= \frac{USL-LSL}{6\sigma} - \frac{\delta}{3\sigma} \\ &= C_p - \frac{\delta}{3\sigma} = 2 - \frac{\delta}{3\sigma} = 1.67. \end{aligned}$$

The conclusion is that the long-term shift in the mean equals σ.

Solution of Exercise 3.7

By Exercise 3.3, a $100\,(1-\alpha)\%$-confidence asymptotic interval for C_p equals

$$\left[\widehat{C_p} - z_{\alpha/2}\frac{USL-LSL}{\sqrt{72n\sigma^2}},\ \widehat{C_p} + z_{\alpha/2}\frac{USL-LSL}{\sqrt{72n\sigma^2}}\right],$$

we get that $z_{\alpha/2}\frac{USL-LSL}{\sqrt{72n\sigma^2}} = 0.1$ and $\alpha = 1 - 0.95 = 0.05$. Hence,

$$n = z^2_{0.025}\frac{(USL-LSL)^2}{0.72\sigma^2}\ .$$

Solution of Exercise 3.8

All calculations are based on the assumption that the observations are independent and are

distributed according to the same normal distribution. The statistic $(n-1)\,c_0^2/\widehat{C_p}^2 = (n-1)S^2/\sigma^2$ has a χ^2_{n-1} distribution. Hence,

$$\begin{aligned} P_{H_0}\left(\widehat{C_p} > kc_0\right) &= P_{H_0}\left(\frac{\widehat{C_p}}{c_0} > k\right) \\ &= P_{H_0}\left(\frac{c_0^2}{\widehat{C_p}^2} < 1/k^2\right) \\ &= P_{H_0}\left((n-1)\,\frac{c_0^2}{\widehat{C_p}^2} < \frac{n-1}{k^2}\right) \\ &= P\left(\chi^2_{n-1} < \frac{n-1}{k^2}\right) = \alpha. \end{aligned}$$

Hence, $k = \sqrt{\frac{\chi^2_{n-1;1-\alpha}}{n-1}}$ yields a test with type I error equal to α with corresponding critical region $\left\{\widehat{C_p} < \sqrt{\frac{n-1}{\chi^2_{n-1;1-\alpha}}}\,c_0\right\}$. The power function is given by

$$\begin{aligned} P\left(\widehat{C_p} < \sqrt{\frac{n-1}{\chi^2_{n-1;1-\alpha}}}\,c_0 \mid C_p = c\right) &= P\left(\frac{\widehat{C_p}}{c} < \sqrt{\frac{n-1}{\chi^2_{n-1;1-\alpha}}}\,\frac{c_0}{c}\,\middle|\, C_p = c\right) \\ &= P\left(\frac{(n-1)\widehat{C_p}^2}{c^2} < \chi^2_{n-1;1-\alpha}\left(\frac{c}{c_0}\right)^2\right) \\ &= P\left(\chi^2_{n-1} < \chi^2_{n-1;1-\alpha}\left(\frac{c}{c_0}\right)^2\right). \end{aligned}$$

Solution of Exercise 3.9
To be written.

Solution of Exercise 3.10
To be written.

Solution of Exercise 3.11

a) The individual data are not normally distributed, but the group means are. To assess whether the production data show a process that is in statistical control, we may therefore safely use the group means. To assess capability, we must use individual data, but cannot rely on the standard capability indices, since these indices assume normality.

b) We see from the data set that the subgroups have sample size 4. To assess whether the process in statistical control, we must do an analysis based on control charts with $\mu = 5.12$ and $\sigma = 0.023$. For sake of simplicity, further analysis is based on the standard 3σ -rule, and does not use further runs rules. We first look at the range chart, since the mean chart is based on the assumption that the variance is stable. The process seems

to be in statistical control until group 15. We assume that this has to do with a special cause (in practice this should be investigated, of course). Since we do not know what has happened after subgroup 15, we decide to be on the safe side and restrict further analysis to the first 14 subgroups (= first 56 individual data). Both charts now indicate that the process is in statistical control.

c) The capability indices must be calculated using data from a process that is in statistical control, hence restricted to the first 14 subgroups. With the individual data from subgroup 15 deleted, the individual data are now normal. Hence, we may safely use the standard capability indices that are based on a normality assumption. The 95% confidence intervals are: $C_p = (1.45, 2.12)$, $C_{pk} = (1.37, 2.03)$.

d) 0.000018% is predicted to fall outside the specification limits.

e) Based on the predicted fall-out, the histogram and the point estimates for the capability indices (both around 1.7) capability seems to be in order. However, too much optimism is dangerous since the 95% confidence intervals are wide, a result of not considering enough data. Moreover, a look at the density trace (kernel estimator for the density) reveals that the right-end tail of the distribution seems to be missing. This should be investigated.

f) The process is definitely not capable of meeting these stricter specifications. This can be seen from the dramatically low capability indices (less than 1). The process seems to be on target, but the variance is much too high. An improvement plan for reducing the variance of the process must be set up. Probably drastic changes and/or redesign of the process are necessary.

Solution of Exercise 3.12
A major problem is that the data given only consists of group summary statistics, not the individual data. Thus a reliable capability analysis is not possible. Ignoring this for the moment, we see that the process is not in statistical control. Something is wrong with subgroup 10. If we remove the data from subgroup 10, then both the mean and range charts indicate that the process is in statistical control. Normality tests indicate that there is no reason to doubt that the means come from a normal distribution. Capability analysis shows that both the actual capability and the potential capability are too small. The process is not well centred around the target and the variance is much too high. Although centring the process is relatively easy, the plant manager should first try to reduce variation rather than first centring the process around the target. This is because with such a high variance, it is not assured that the process will stay centred.

Solution of Exercise 3.13
The likelihood equals $\theta^n \left(\prod_{i=1}^n x_i\right)^{\theta-1}$. The factorization Lemma thus yields that $\prod_{i=1}^n X_i$ a sufficient statistic for θ.

Solution of Exercise 3.16
Since μ is known and σ^2 is known, we try an interval of the form $[\mu - kS, \mu + kS]$ with k such that

$$\mathrm{E}\left(F(\mu + kS) - F(\mu - kS)\right) = \beta.$$

By Proposition 3.6.7 we get that

$$\mathrm{E}\left(F(\mu+kS)-F(\mu-kS)\right)=P(\mu-kS<X<\mu+kS)=\beta.$$

where $X\sim N(\mu,\sigma^2)$ and then $\frac{X-\mu}{S}\sim T_{n-1}$

Since

$$P(-k<\frac{X-\mu}{S}<k)=\beta$$

Hence, we get that $k=t_{n-1,\frac{1-\beta}{2}}$.

Solution of Exercise 3.17

Since μ is known, and σ^2 is unknown, we try an interval of the form $[\mu-kS,\mu+kS]$ with k such that

$$P\left((\Phi(kS/\sigma)-\Phi(-kS/\sigma))\geq\beta\right)=\alpha.$$

If $\Phi(kS/\sigma)-\Phi(-kS/\sigma)\geq\beta$, we get that $kS/\sigma\geq z_{\frac{1-\beta}{2}}$. So we have

$$P\left(kS/\sigma\geq z_{\frac{1-\beta}{2}}\right)=\alpha=P\left((n-1)S^2/\sigma^2\geq z^2_{\frac{1-\beta}{2}}(n-1)/k^2\right)$$

Since $(n-1)S^2/\sigma^2\sim\chi^2_{n-1}$, we get that $\frac{(n-1)Z^2_{\frac{1-\beta}{2}}}{k^2}=\chi^2_{n-2,\alpha}$. Hence,

$$k=\frac{z_{\frac{1-\beta}{2}}\sqrt{n-1}}{\sqrt{\chi^2_{n-1,\alpha}}}.$$

Chapter 8

Solutions Exercises Control Charts

Solution of Exercise 4.1
Rather than trying to write down the likelihood equations directly which leads to a cumbersome expression, we invoke the Invariance Principle (Theorem 3.3.2). We first write

$$\begin{aligned} n_L &= |\{i : X_i \leq L\}| \\ n_U &= |\{i : X_i \geq U\}|. \end{aligned}$$

Recall that n_L/n and n_u/n are the Maximum Likelihood estimators for $P(X \leq L)$ and $P(X \geq U)$, respectively. It follows from the Invariance Principle (Theorem 3.3.2) that

$$\frac{n_L}{n} = \widehat{P(X \leq L)} = \Phi\left(\widehat{\frac{L-\mu}{\sigma}}\right) = \Phi\left(\frac{L-\widehat{\mu}}{\widehat{\sigma}}\right).$$

and

$$\frac{n_U}{n} = \Phi\left(\widehat{\frac{\mu-U}{\sigma}}\right) = \Phi\left(\frac{\widehat{\mu}-U}{\widehat{\sigma}}\right),$$

where $\widehat{\mu}$ and $\widehat{\sigma}$ denote the Maximum Likelihood estimators based on n_L and n_U, respectively. Hence,

$$\begin{aligned} \frac{L-\widehat{\mu}}{\widehat{\sigma}} &= \Phi^{-1}(n_L/n) \\ \frac{\widehat{\mu}-U}{\widehat{\sigma}} &= \Phi^{-1}(n_U/n). \end{aligned}$$

If we solve this system of equations with respect to $\widehat{\mu}$ and $\widehat{\sigma}$, then we obtain

$$\begin{aligned} \widehat{\sigma} &= \frac{L-U}{\Phi^{-1}(n_L/n) + \Phi^{-1}(n_U/n)} \\ \widehat{\mu} &= L + \frac{(U-L)\varphi^{-1}(n_L/n)}{\Phi^{-1}(n_L/n) + \Phi^{-1}(n_U/n)}. \end{aligned}$$

Solution of Exercise 4.2
Use the fact that for normal samples, the statistics $\overline{X}$ and S^2 are independent. The probability

that both control charts are in control is for probabilistic control limits equal to $(1-0.0027)^2$. Hence, the probability that at least one of them is out of control at a specific point in time equals $1-(1-0.0027)^2 \approx 1/185$. The ARL of the combined chart is thus approximately equal to 185.

Solution of Exercise 4.3

Solution of Exercise 4.4

Solution of Exercise 4.5

Solution of Exercise 4.6
It is best to put average values on the Shewhart chart, because otherwise one would obtain a biased estimate of the process variance: the variance of individual values also contain contributions to measurement errors.

Solution of Exercise 4.7
Although R is aimed at statistical calculations rather than numerical calculations, it is no problem to produce a reliable table for $d_2(n)$.

```
d2integrand <- function (y,n){1 - pnorm(y)^n - (1-pnorm(y))^n}
d2 <- function(n){integrate(function(y){d2integrand(y,n)},-Inf,Inf)$value}
# integrate returns a vector;
# the component value of that vector is the value of the integral
print(matrix(c(2:10,sapply(2:10,d2)),ncol=2,
dimnames=list(rep("",9),c("n","d2(n)"))),digits=4)
```

Solution of Exercise 4.8

Solution of Exercise 4.9

Solution of Exercise 4.10

Solution of Exercise 4.11

Solution of Exercise 4.12
Although all points on the control charts are within the control limits, this process is definitely out of control. The process variance is much smaller than before (in fact, almost all points fall within 1σ of the centre line). An investigation must be started to find out why this is so, in order to come up with actions that will keep the process at this improved level of performance.

Solution of Exercise 4.13

a) Independence of observations cannot be checked without further information on sampling. Normality of data is OK according to the normal probability plot and the Shapiro-Wilks test.

b) A possible explanation is that the process is indeed in control but that the control limits are too narrow. This may be caused by the sampling procedure. If the time between taking the 5 subsequent items is small, then the in-group variation may be very small, resulting in a too small estimate of the process standard deviation. In other words, the choice of rational subgroups here only picks us the short-term variability, not the true common-cause variability.

c) The control limits are now much wider and there are no out-of-control signals.

Solution of Exercise 4.14

Solution of Exercise 4.15

Solution of Exercise 4.16

Solution of Exercise 4.17

a) The first step is to choose a pivot statistic to form the tolerance interval. Suitable choices are the sufficient statistics $\sum_{i=1}^{40} T_i$ or $\min_{1\leq i\leq 40} T_i$. Since the latter is much easier to handle (it has an exponential distribution, we here show how to deal with $T = \sum_{i=1}^{40} T_i$. We have to find a lower limit L such that $E(P(Y > L)) = 0.95$, where Y has an exponential distribution. Using the parameterisation $\frac{1}{\theta}\, e^{-x/\theta}$ for the density of the exponential distribution, we find that L should satisfy $0.95 = E(1-F(L)) = Ee^{-L/\theta}$. We take $L = cT$ for a suitable constant c. Now $2T/\theta$ has a χ^2_{80}-distribution (one can also express everything in terms of the Gamma distribution). The moment generating function Ee^{sY} of a random variable Y with a χ^2_{80}-distribution equals $(1-2s)^{-40}$, we find that $Ee^{-cT/\theta} = Ee^{-c/2\,2T/\theta} = (1+c)^{-40}$. Hence, $c = \sqrt[40]{1/0.95} - 1 = 0.0013$. Since $T = 186$, a 95% expectation tolerance interval is thus given by $(0.24; \infty)$.

b) The sample mean and sample standard deviation of this sample are 4.65 and 3.95, respectively. The centre line is thus 4.65. The upper control limit is thus $4.65 + 3 * 3.95/\sqrt{k}$, where k is the subgroup size. The lower control limit is $\max(0, 4.65 - 3 * 3.95/\sqrt{k})$. Because for an exponential distribution the parameter θ is both the mean and the standard deviation, an alternative is to base the control limits on the sample mean rather than the sample standard deviation.

c) There are several problems with the approach in b). The exponential distribution is not symmetric, the lower control limit may be negative, and the 3σ limits are not unique. Moreover, the 3σ limits may yield a false alarm probability that is very much different from the usual value 0.0027.

d) The sequential likelihood ratio is given by

$$\log \Lambda_n = n \log \frac{5}{3} + \left(\frac{1}{5} - \frac{1}{3} \right) \sum_{i=1}^{n} X_i.$$

The corresponding lower CUSUM chart is given by

$$C_n^- = \min \left(0, C_{n-1}^- - \left(\frac{1}{5} - \frac{1}{3} \right) X_n - \log \frac{5}{3} \right).$$

Solution of Exercise 4.18

Solution of Exercise 4.19

Useful Results from Probability Theory

In this appendix we present some useful results and notations from probability theory. Proofs of these results can be found in textbooks.

Theorem .0.1 (Markov inequality) *Let X be a non-negative random variable. Then*

$$P(X \geq a) \leq \frac{1}{a}\mathrm{E}\, X,\ a > 0.$$

Definition .0.2 *Let X_n and $Y_n, n \in \mathbb{N}$, be random variables on the same probability space. We write*

$$X_n = O_P(Y_n)$$

to mean that for each $\varepsilon > 0$, there exists an $M \in \mathbb{R}$ such that $P(|X_n| \geq M|Y_n|) < \varepsilon$, for n large enough. We write

$$X_n = o_P(Y_n)$$

to mean that $P(|X_n| \geq \delta|Y_n|) \to 0$, for each $\delta > 0$, as $n \to \infty$. Often these notations are used for deterministic Y_n, or even for $Y_n = 1$.

Lemma .0.3 *Let X_n, Y_n be m-dimensional random vectors and W_n k-dimensional random vectors.*

a) If $X_n \xrightarrow{d} X$ and $f : \mathbb{R}^d \to \mathbb{R}^k$ is continuous, then $f(X_n) \xrightarrow{d} f(X)$.

b) If $X_n \xrightarrow{d} X$ and $X_n - Y_n \xrightarrow{P} 0$, then $Y_n \xrightarrow{d} X$.

c) If $X_n \xrightarrow{d} X$ and $W_n \xrightarrow{d} c$, then

$$\begin{pmatrix} X_n \\ W_n \end{pmatrix} \xrightarrow{d} \begin{pmatrix} X \\ c \end{pmatrix}.$$

Remark .0.4 *Part c) cannot be generalized by assuming $W_n \xrightarrow{d} W$. However, if X_n and W_n are independent, then the result follows from the fact that in this case the joint distribution function factors. We use this in b) of the following corollary.*

Corollary .0.5 *Let X_n and Y_n be m-dimensional random vectors defined on the same probability space.*

a) If $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{d} c$, then $X_n + Y_n \xrightarrow{d} X + c$.

b) If $X_n \xrightarrow{d} X$ *and* $Y_n \xrightarrow{d} Y$ and X_n *and* Y_n *are independent, then* $X_n + Y_n$ *converges in law to the convolution of the distributions* X *and* Y*.*

c) If $X_n \xrightarrow{d} X$ *and* $Y_n \xrightarrow{d} c$*, then* $Y_n^T X_n \xrightarrow{d} c^T X$*.*

Remark .0.6 *Note that it does not necessarily hold in b) that the limit distribution equals* $X + Y$*, since* X *and* Y *need not be independent or worse, need not be defined on the same probability space.*

Lemma .0.7 *Let* X_n *and* Y_n *be d-dimensional random vectors.*

a) If $X_n \xrightarrow{P} X$ *and* $f : \mathbb{R}^d \to \mathbb{R}^k$ *is continuous, then* $f(X_n) \xrightarrow{P} f(X)$*.*

b) If $X_n \xrightarrow{P} X$ *and* $X_n - Y_n \xrightarrow{P} 0$*, then* $Y_n \xrightarrow{P} X$*.*

c) If $X_n \xrightarrow{P} X$ *and* W_n *are k-dimensional random vectors such that* $W_n \xrightarrow{P} W$*, then*

$$\begin{pmatrix} X_n \\ W_n \end{pmatrix} \xrightarrow{P} \begin{pmatrix} X \\ W \end{pmatrix}.$$

Theorem .0.8 (Strong Law of Large Numbers) *Let* $X, X_1, X_2, \ldots$ *be i.i.d. random variables with common distribution function* F *and* $h : \mathbb{R} \to \mathbb{R}$ *a function such that* $\mathrm{E}\,|h(X)| < \infty$*. Then*

$$\frac{1}{n}\sum_{i=1}^{n} h(X_i) \to \mathrm{E}\,h(X) \quad a.s. \quad (n \to \infty).$$

Theorem .0.9 (Lindeberg Central Limit Theorem) *Let* $X_{n,1}, \ldots, X_{n,r_n}$ *be independent random variables for every* $n \in \mathbb{N}$ *with* $r_n \in \mathbb{N}$*. Assume* $\mathrm{E}\,X_{n,k} = 0$ *and* $\sigma_{n,k}^2 = Var\,X_{n,k} < \infty$*. Set* $\sigma_n^2 = Var \sum_{k=1}^{r_n} X_{n,k} = \sum_{k=1}^{r_n} \sigma_{n,k}^2$*. If for all* $\varepsilon > 0$

$$\frac{1}{\sigma_n^2}\sum_{k=1}^{r_n} \mathrm{E}\left(X_{n,k}^2 1_{[|X_{n,k}|>\varepsilon\sigma_n]}\right) \to 0 \ (n \to \infty),$$

then

$$\frac{1}{\sigma_n}\sum_{k=1}^{r_n} X_{n,k} \xrightarrow{d} N(0,1).$$

Theorem .0.10 (Multivariate central limit theorem) *Let* $X_1, \ldots, X_n$ *be i.i.d. random vectors with existing covariance matrix* Σ*. Then, as* $n \to \infty$*,*

$$n^{\frac{1}{2}}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mathrm{E}\,X_1\right) \xrightarrow{d} N(\langle 0, \ldots, 0\rangle, \Sigma).$$

For some distributions it is useful to know the basic properties of the Gamma function . This function is defined for positive x as

$$\Gamma(x) = \int_0^\infty e^{-t}\, t^{x-1}\, dt$$

Useful properties of the Gamma function are

- $\Gamma(n+1) = n!$ for non-negative integer n
- $\Gamma(x+1) = x\,\Gamma(x)$
- $\Gamma(\frac{1}{2}) = \sqrt{\pi}$

The second property also defines the Gamma function for negative, non-integer x.

Definition .0.11 (χ^2-distribution) *The χ^2-distribution is characterised by one parameter, denoted here by n, and known as the "degrees of freedom". Notation: χ^2_n. The name of the χ^2-distribution is derived from its relation to the standard normal distribution: if Z is a standard normal random variable, then its square $X = Z^2$ is χ^2 distributed, with one degree of freedom. If X_i are χ^2 distributed, and mutually independent, then the sum $X = \sum_i X_i$ is χ^2 and the parameter (degrees of freedom) is the sum of the parameters of the individual X_i. If $X_1, \ldots, X_n$ are independent random variables each with a $N(\mu, \sigma^2)$ distribution, then the random variable $(n-1)S^2/\sigma^2$ has a χ^2_{n-1}-distribution. The χ^2-distribution is also a special case of the Gamma distribution, with $\alpha = \nu/2$ and $\lambda = 1/2$. The χ^2-distribution is of great importance in the Analysis of Variance (ANOVA), contingency table tests, and goodness-of-fit tests. The density is given by* $\dfrac{e^{-x/2}\, x^{(\nu-2)/2}}{2^{\nu/2}\,\Gamma(\nu/2)}$.

Definition .0.12 ((Noncentral) t-distribution) *Let Z be a standard normal random variable and let Y be a χ^2-distributed random variable with ν degrees of freedom. If Z and Y are independent, then the distribution of*

$$\frac{Z+\delta}{\sqrt{\dfrac{Y}{\nu}}}$$

is called a noncentral t-distribution with ν degrees of freedom and non-centrality parameter δ. The case $\delta = 0$ is known as the (Student) t-distribution and has density

$$\frac{\Gamma\left(\dfrac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\dfrac{n}{2}\right)\left(1+\dfrac{x^2}{n}\right)^{(n+1)/2}}.$$

For a normal sample $X_1, \ldots, X_n$ with mean μ and variance σ^2, $\left(\overline{X} - \mu\right)/(S/\sqrt{n})$ has a Student t-distribution with $n-1$ degrees of freedom. The Student t-distribution is named after the statistician William Gosset. His employer, the Guinness breweries, prohibited any scientific publication by its employees. Hence, Gosset published using a pen name, Student. For further properties and examples of the use of the noncentral t-distribution, we refer to Kotz and Johnson (1993) and Owen (1968).

Definition .0.13 (F-distribution) *The F-distribution, named after the famous statistician Fisher, is the distribution of a ratio of two independent χ^2 random variables. It has two parameters, denoted by m and n, which are called the degrees of freedom of the numerator and the denominator, respectively. Notation: F^m_n. If X is Student t-distributed with n degrees of freedom, then X^2 is an F^1_n variable. If U is χ^2 distributed with m degrees of freedom, V is χ^2 distributed with n degrees of freedom, and if U and V are independent, then $X = \frac{U/m}{V/n}$ is*

an F_n^m variable. The values $f_{n;\alpha}^m$ are defined by $P\left(F_n^m > f_{n;\alpha}^m\right) = \alpha$ (so they do not follow the customary definition of quantiles). From the definition of F_n^m as a ratio of two χ^2 variables, it follows that $f_{n;1-\alpha}^m = 1/f_{m;\alpha}^n$. The density is given by $\dfrac{\Gamma\left(\dfrac{m+n}{2}\right)}{\Gamma\left(\dfrac{m}{2}\right)\,\Gamma\left(\dfrac{n}{2}\right)}\,\dfrac{m^{m/2}\,n^{n/2}\,x^{(m/2)-1}}{(n+mx)^{(m+n)/2}}$.

Definition .0.14 (Gamma distribution) *Special cases of the Gamma distribution include the χ^2-distribution ($\alpha = \nu/2$ and $\lambda = 1/2$), the Erlang distribution (α positive integer) and the exponential distribution ($\alpha = 1$). Sometimes $\beta = 1/\lambda$ is used as parameter. The density is given by* $\lambda^\alpha\,\dfrac{x^{\alpha-1}\,e^{-\lambda\,x}}{\Gamma(\alpha)}$.

Index

INDEX

Bibliography

V. Abel and E. Fuchs. On the run length of one-sided Cusum procedures in the change-point case. *Meth. Oper. Res.*, 50:133–137, 1985.

J. Aitchison and I. Dunsmore. *Statistical Prediction Analysis*. Cambridge University Press, 1975.

W. Albers and W.C.M. Kallenberg. Estimation in Shewhart control charts: effects and corrections. *Metrika*, 59(3):207–234, 2004a. ISSN 0026-1335.

W. Albers and W.C.M. Kallenberg. Are estimated control charts in control? *Statistics*, 38 (1):67–79, 2004b. ISSN 0233-1888.

D.E. Barton. Unbiased estimation of a set of probabilities. *Biometrika*, 48:227–229, 1961.

A.H. Bowker and H.P. Goode. *Sampling Inspection for Variables*. McGraw-Hill, New York, 1952.

D. Brook and D.A. Evans. An approach to the probability distribution of CUSUM run length. *Biometrika*, 59:539–550, 1972.

M.E. Calzada and S.M. Scariano. Reconciling the integral equation and Markov chain approaches for computing EWMA average run lengths. *Comm. Statistics Simul. Comput.*, 32 (2):591–604, 2003.

S. Chakraborti. Run length, average run length and false alarm rate of Shewhart x-bar chart: Exact derivations by conditioning. *Commun. Statist. Simul.*, 29:61–81, 2000.

C.W. Champ and W.H. Woodall. Exact results for Shewhart control chart with supplementary runs rules. *Technometrics*, 29:393–401, 1987.

C.W. Champ and W.H. Woodall. Signal probabilities of runs rules supplementing a Shewhart control chart. *Comm. Statist. Simul. Comput.*, 26(4):1347–1360, 1997.

J.T. Chang and R.D. Fricker. Detecting when a monotically increasing mean has crossed a threshold. *J. Qual. Technology*, 31:217–234, 1999.

S.V. Crowder, D.M. Hawkins, M.R. Reynolds Jr., and E. Yashchin. Process control and statistical inference. *J. Qual. Technol.*, 29:134–139, 1997.

V. Czitrom and P.D. Spagon, editors. *Statistical Case Studies for Industrial Process Improvement*, 1997. American Statistical Association and SIAM.

BIBLIOGRAPHY

R.B. D'Agostino and M.A. Stephens, editors. *Goodness-of-fit techniques*. Marcel Dekker, New York, 1986.

P. Deheuvels. Estimation nonparametrique de la densité par histogrammes generalisés. *Revue de Statistique Appliquée*, 35:5–42, 1977.

A. Di Bucchianico, J.H.J. Einmahl, and N.A. Mushkudiani. Smallest nonparametric tolerance regions. *Ann. Statist.*, 29(5):1320–1343, 2001.

R.J.M.M. Does and B.F. Schriever. Variables control chart limits and tests for special causes. *Stat. Neerlandica*, 46:229–245, 1992.

K.R. Eberhardt, R.W. Mee, and C.P. Reeve. Computing factors for exact two-sided tolerance limits for a normal distribution. *Communications in Statistics - Simulation and Computation*, 18:397–413, 1989.

N.R. Farnum and L.W. Stanton. Using counts to monitor a process mean. *Journal of Quality Technology*, 18:22–28, 1986.

N.R. Farnum and L.W. Stanton. Narrow limit gauging: Go or No Go? *Quality Engineering*, 3:293–307, 1991.

E. Fix and J.L. Hodges. Discriminatory analysis - nonparametric discrimination: consistency properties. *International Statistical Reviews*, 57:238–247, 1989.

J.L. Folks, D.A. Pierce, and C. Stewart. Estimating the fraction of acceptable product. *Technometrics*, 7:43–50, 1965.

M.J. Fryer. Some errors associated with the non-parametric estimation of density functions. *Journal of the Institute of Mathematics and its Applications*, 18:371–380, 1976.

J.C. Fu, G. Schmueli, and Y.M. Chang. A unified Markov chain approach for computing the run length distribution in control charts with simple or compound rules. *Statist. Probab. Letters*, 65:457–466, 2003.

F.F. Gan and T.C. Chang. Computing average run lengths of exponential EWMA charts. *Journal Of Quality Technology*, 32:183–187, 2000.

B.K. Ghosh, M.R. Reynolds, and Y.V. Hui. Shewhart $\bar{x}$-charts with estimated process variance. *Comm. Statist. Theory Methods*, 10:1797–1822, 1981.

M.A. Girshick and H. Rubin. A Bayes approach to a quality control model. *Ann. Math. Statistics*, 23:114–125, 1952.

H. Gitlow, A. Oppenheim, and R. Oppenheim. *Quality Management: Tools and Methods for Improvement*. Irwin, second edition, 1989.

M.S. Gold. The geometric approximation to the Cusum run length. *Biometrika*, 76:725–733, 1989.

G.K. Griffith. *Statistical Process Control Methods*. ASQC Quality Press, Milwaukee, 1996.

BIBLIOGRAPHY

W.C. Guenther. A note on the Minimum Variance Unbiased estimate of the fraction of a normal distribution below a specification limit. *American Statistician*, 25:18–20, 1971.

I. Guttman. *Statistical Tolerance regions: Classical and Bayesian*. Charles Griffin, London, 1970.

D.M. Hawkins. Evaluation of average run lengths of cumulative sum charts for an arbitrary data distribution. *Communications in Statistics C: Simulation*, 21(4):1001–1020, 1992.

D.M. Hawkins and D.H. Olwell. *Cumulative Sum Charts and Charting for Quality Improvement*. Springer, New York, 1998.

D.M. Hawkins, P. Qiu, and C.W. Kang. The changepoint model for statistical process control. *J. Qual. Technol.*, 35(4):355–366, 2003.

M. Jílek. A bibliography of statistical tolerance regions. *Mathematische Operationsforschung und Statistik - Series Statistics*, 12(441–456), 1981.

M. Jílek and H. Ackermann. A bibliography of statistical tolerance regions II. *Statistics*, 20: 165–172, 1989.

C.-H. Jun and M.S. Choi. Simulating the average run length for CUSUM schemes using variance reduction techniques. *Commun. Statist. Simul.*, 22:877–887, 1993.

K.W. Kemp. *The efficient use of quality control data*. The Institute of Mathematics & its Applications Monograph Series, 4. The Clarendon Press Oxford University Press, Oxford, 2001.

R.S. Kenett and S. Zacks. *Modern Industrial Statistics*. Duxbury Press, 1998.

S. Kotz and N.L. Johnson. *Process Capability Indices*. Chapman & Hall, London, 1993.

S. Kotz and C.R. Lovelace. *Process Capability Indices in Theory and Practice*. Arnold, London, 1998.

T.L. Lai. Sequential changepoint detection in quality control and dynamical systems. *J. Roy. Statist. Soc. Ser. B*, 57:613–658, 1995.

T.L. Lai. Sequential Analysis: Some classical problems and new challenges. *Stat. Sinica*, 11: 303–350, 2001.

G.J. Liebermann and G.J. Resnikoff. Sampling plans for inspection by variables. *J. Amer. Stat. Assoc.*, 50:457–469, 1955.

J.M. Lucas and R.B. Crosier. Fast initial response for CUSUM quality control schemes: Give your CUSUM a head start. *Technometrics*, 24:199–205, 1982.

J.S. Marron and M.P. Wand. Exact mean integrated squared error. *Annals of Statistics*, 20: 712–736, 1992.

D.C. Montgomery. *Introduction to Statistical Quality Control*. Wiley, 4th edition, 2000.

D.B. Owen. A survey of properties and applications of the noncentral t-distribution. *Technometrics*, 10:445–478, 1968.

BIBLIOGRAPHY

E.S. Page. Continuous inspection schemes. *Biometrika*, 41:100–115, 1954.

A.C. Palm, R.N. Rodriguez, F.A. Spiring, and D.J. Wheeler. Some perspectives and challenges for control chart methods. *J. Qual. Technology*, 29:122–127, 1997.

E. Parzen. On estimation of a probability density function and mode. *Annals of Mathematical Statistics*, 33:1065–1076, 1962.

J.K. Patel. Tolerance limits - a review. *Communications in Statistics A - Theory and Methods*, 15:2719–2762, 1986.

E. Paulson. A note on control limits. *Annals of Mathematical Statistics*, 14:90–93, 1943.

L.P. Provost and C.L. Norman. Variation through the ages. *Qual. Progress*, 23(12):39–44, 1990.

S.J. Qin. Statistical process monitoring: basics and beyond. *J. Chemometrics*, 17:480–502, 2003.

S.W. Roberts. Control charts based on geometric moving averages. *Technometrics*, 1:239–250, 1959.

R.N. Rodriguez. Recent developments in process capability analysis. *J. Qual. Technol.*, 24(4): 176–187, 1992.

M. Rosenblatt. Remarks on some nonparametric estimates of a density function. *Annals of Mathematical Statistics*, 27:827–837, 1956.

G.C. Runger and M.C. Testik. Control charts for monitoring fault signatures: Cuscore versus GLR. *Quality Rel. Eng. Int.*, 19(4):387 – 396, 2003.

A.N. Shiryaev. On optimum methods in quickest detection problems. *Theor. Prob. Appl.*, 8: 22–46, 1963.

B.W. Silverman. *Density Estimation for Statistics and Data Analysis*. Chapman & Hall, London, 1986.

W.L. Stephens. Control by gauging. *Journal of the Royal Statistical Society B*, 10:54–108, 1948.

Z. Stoumbos, M.R. Reynolds Jr., T.P. Ryan, and W.H. Woodall. The state of statistical process control as we proceed into the 21st century. *J. Amer. Stat. Assoc.*, 95:992–998, 2000.

R.A. Tapia and J.R. Thompson. *Nonparametric Probability Density Estimation*. The John Hopkins University Press, Baltimore, 1978.

M.P. Wand and M.C. Jones. *Kernel Smoothing*. Chapman & Hall, London, 1995.

Western Electric Company. Statistical Quality Control Handbook, 1956.

D.J. Wheeler. The variance of an estimator in variables sampling. *Technometrics*, 12:751–755, 1970.

BIBLIOGRAPHY

S. S. Wilks. Determination of sample sizes for setting tolerance limits. *Ann. Math. Statistics*, 12:91–96, 1941.

W.H. William. On the Markov chain approach to the two-sided CUSUM procedure. *Technometrics*, 26:41–46, 1984.

W.H. Woodall. Controversies and contradictions in statistical process control. *J. Qual. Technol.*, 32:341–378, 2000.

W.H. Woodall and D.C. Montgomery. Research issues and ideas in statistical process control. *J. Qual. Technol.*, 31:376–386, 1999.

P.W. Zehna. Invariance of maximum likelihood estimation. *Annals of Mathematical Statistics*, 37:744, 1966.