COMPUTER-INTENSIVE STATISTICAL METHODS: SADDLEPOINT APPROXIMATIONS WITH APPLICATIONS IN BOOTSTRAP AND ROBUST INFERENCE

THÈSE N^O 2449 (2001)

PRÉSENTÉE AU DÉPARTEMENT DE MATHÉMATIQUES

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Diego KUONEN

Ingénieur mathématicien diplômé EPF

de nationalité suisse et originaire de Zermatt (VS)

acceptée sur proposition du jury:

Prof. A. C. Davison, directeur de thèse
Prof. T. J. DiCiccio, rapporteur
Prof. S. Morgenthaler, rapporteur
Dr G. A. Young, rapporteur

Lausanne, EPFL
2001

To my mother Augusta

'There is a woman at the beginning of all great things.'

Alphonse de Lamartine (1790–1869)

Acknowledgement

I would like to thank all the people who supported me while I have been working on my thesis, even if I should forget to mention some of their names here.

First of all, I would like to express my deep gratitude to my supervisor, Professor Anthony C. Davison, for his academic guidance, his never-ending patience in answering questions, as well as giving me complete liberty in what I liked to do. His approach to science has been a strong stimulation for me and I also thank him for the knowledge that he has given me — knowledge that one cannot find in books. Without his help I would never have been able to achieve this work.

I am also deeply grateful to Professor Thomas J. DiCiccio (Cornell University, USA), Professor Stephan Morgenthaler (DMA, EPFL) and Dr Alastair G. Young (Cambridge University, UK), who honoured me with their willingness to be members of the jury. They carefully read the thesis and gave very helpful comments and advice. Thanks also to Professor Robert Dalang (DMA, EPFL) for presiding over the jury.

I gratefully acknowledge all current and fellow members and colleagues of the Department of Mathematics and of the Chairs of Statistics at the EPFL. Especially, thanks to Marc-Olivier Boldi, Thomas Gsponer and Dr Sylvain Sardy for proofreading certain parts of this thesis, as well as to Alain Forclaz and Dr Jacques Zuber for their encouragements and advice, especially during the last months of my thesis' work. Special thanks to Reinhard Furrer for his permanent assistance and advice. Last but not least, many thanks to our secretary, Anne-Lise Courvoisier, for her smiles, help, competence, and efficiency.

From the very beginning of my decision to continue the academic life after the diploma, through all phases of doubts, new orientations and frustrating moments until the very end, Christian Glenz has been willing to listen to all my concerns and has given me advice. In the same spirit I would like also to thank Peter Schalbetter, Raoul Schmidhalter, Stephan Zaugg and all of my 'Piggu' friends.

Finally, I am deeply indebted to my mother, Augusta Lauber, and to my sister, Fabienne Kuonen, for all their support during my entire studies, encouraging me to continue my studies, helping me to overcome difficulties when things seemed fuzzy and being patient when I was nervous due to my work. They deserve more thanks that I can ever properly express. Without them and the rest of my family this thesis would not have been completed.

Abstract

The saddlepoint approximation was introduced into statistics in 1954 by Henry E. Daniels. This basic result on approximating the density function of the sample mean has been generalized to many situations. The accuracy of this approximation is very good, particularly in the tails of the distribution and for small sample sizes, compared with normal or Edgeworth approximation methods.

Before applying saddlepoint approximations to the bootstrap, this thesis will focus on saddlepoint approximations for the distribution of quadratic forms in normal variables and for the distribution of the waiting time in the coupon collector's problem. Both developments illustrate the modern art of statistics relying on the computer and embodying both numeric and analytic approximations. Saddlepoint approximations are extremely accurate in both cases. This is underlined in the first development by means of an extensive study and several applications to nonparametric regression, and in the second by several examples, including the exhaustive bootstrap seen from a collector's point of view.

The remaining part of this thesis is devoted to the use of saddlepoint approximations in order to replace the computer-intensive bootstrap. The recent massive increases in computer power have led to an upsurge in interest in computer-intensive statistical methods. The bootstrap is the first computer-intensive method to become widely known. It found an immediate place in statistical theory and, more slowly, in practice. The bootstrap seems to be gaining ground as the method of choice in a number of applied fields, where classical approaches are known to be unreliable, and there is sustained interest from theoreticians in its development. But it is known that, for accurate approximations in the tails, the nonparametric bootstrap requires a large number of replicates of the statistic. As this is time-intensive other methods should be considered. Saddlepoint methods can provide extremely accurate approximations to resampling distributions. As a first step I develop fast saddlepoint approximations to bootstrap distributions that work in the presence of an outlier, using a saddlepoint mixture approximation. Then I look at robust M-estimates of location like Huber's M-estimate of location and its initially MAD scaled version.

One peculiarity of the current literature is that saddlepoint methods are often used to approximate the density or distribution functions of bootstrap estimators, rather than related pivots, whereas it is the latter which are more relevant for inference. Hence the aim of the final part of this thesis is to apply saddlepoint approximations to the construction of studentized confidence intervals based on robust M-estimates. As examples I consider the studentized versions of Huber's M-estimate of location, of its initially

MAD scaled version and of Huber's proposal 2.

In order to make robust inference about a location parameter there are three types of robustness one would like to achieve: robustness of performance for the estimator of location, robustness of validity and robustness of efficiency for the resulting confidence interval method. Hence in the context of studentized bootstrap confidence intervals I investigate these in more detail in order to give recommendations for practical use, underlined by an extensive simulation study.

Version abrégée

a technique de point de selle a été introduite en statistique en 1954 par Henry E. Daniels. A l'origine utilisé pour approcher la fonction de densité de la moyenne, ce résultat a été généralisé à beaucoup de situations. Comparé aux méthodes normales et à l'approximation d'Edgeworth, l'exactitude de cette approximation est très bonne, en particulier dans les queues de la distribution et pour des échantillons de petites tailles.

Avant d'appliquer les approximations de point de selle aux méthodes de bootstrap ou de rééchantillonnage, cette thèse se concentrera sur des approximations de point de selle pour la distribution des formes quadratiques de variables normales et pour la distribution du temps d'attente dans le problème du collecteur de coupons. Les deux développements illustrent l'art moderne de la statistique se fondant sur l'informatique et mélangeant des approximations numériques et analytiques. Les approximations de point de selle sont extrêmement précises dans les deux cas. Ceci est souligné dans le premier développement au moyen d'une étude étendue et plusieurs applications à la régression non paramétrique, et dans le second par plusieurs exemples, y compris le bootstrap exhaustif dans le cas du problème du collectionneur de coupons.

La partie restante de cette thèse est consacrée à l'utilisation des approximations de point de selle afin de substituer le rééchantillonnage. Les récentes augmentations massives de la puissance de calcul des ordinateurs ont mené à une croissance de l'intérêt pour des méthodes statistiques dites 'computer-intensive'. Le bootstrap est la première 'computer-intensive' méthode à devenir largement connue. Il a trouvé un champ d'application immédiat dans la théorie statistique et, plus lentement, dans la pratique. Le bootstrap semble gagner du terrain comme méthode de choix dans un certain nombre de domaines appliqués, où des approches classiques sont connues pour être incertaines. De plus, il y a un intérêt soutenu des théoriciens dans son développement. Mais on sait que pour des approximations précises dans les queues, le bootstrap non paramétrique exige un grand nombre de répliques de la statistique. Comme ceci peut être très lent, d'autres méthodes devraient être considérées. Les méthodes de point de selle peuvent fournir des approximations extrêmement précises aux distributions de rééchantillonnage. Dans un premier temps je développe des approximations de point de selle pour des distributions de rééchantillonnage qui fonctionnent en présence d'une valeur aberrante, en utilisant une approximation de point de selle mélangée. Ensuite, je considère des M-estimateurs robustes de lieu comme le M-estimateur de lieu de Huber et sa version standardisée par

le MAD.

Une particularité de la littérature actuelle est que des méthodes de point de selle sont souvent employées pour approcher la densité ou la fonction de distribution des estimateurs bootstrap, plutôt que celles des pivots associés, tandis que ce sont ces derniers qui sont plus appropriés pour faire de l'inférence. Par conséquent le but de la partie finale de cette thèse est d'appliquer les approximations de point de selle à la construction d'intervalles de confiance bootstrap studentisés basés sur des M-estimateurs robustes. Les exemples que je vais considérer sont les versions studentisées du M-estimateur de Huber, de sa version standardisée par le MAD et du 'proposal 2' de Huber — un M-estimateur robuste de lieu et d'échelle.

Quand on veut faire de l'inférence robuste sur un paramètre de lieu, il y a trois types de robustesse dont on veut s'assurer : la robustesse de performance pour l'estimateur de lieu, la robustesse de validité et la robustesse d'efficacité pour la méthode par intervalles de confiance résultants. Par conséquent dans le contexte des intervalles de confiance bootstrap studentisés, j'étudie ces derniers plus en détails afin de donner des recommandations pour l'usage pratique, lesquelles seront soulignées par une étude étendue de simulations.

Contents

1. Introduction

'Asymptopia. By all accounts, it's a great place to visit. Of course, nobody has ever been there, but it sure looks nice from a distance. Most visitors stop at the outskirts — after all, many tour guides suggest that you really don't need to go too close (as if you could!) since after a certain point, not much changes. Besides, it can be a long, lonely trip. That is not to say that the journey is not worthwhile. ... The promise of saddlepoint methods in particular seems to be that of bringing Asymptotics closer to us, with the promise of excellent asymptotic approximations in even very small sample situations. Asymptopia: How much is a one-way ticket?'

Michael A. Martin (1996, pages 239 and 241)

The term *approximation* refers, in general, to the representation of 'something' by 'something else' that is expected to be a useful replacement for the 'something'. Approximations are sometimes needed because it is not possible to obtain an exact representation of the 'something'; even when an exact representation is possible, approximations may simplify analytical treatments. In scientific work, approximations are in constant use. For example, much scientific argument, and nearly all statistical analysis, is based on mathematical models that are essentially approximations.

Most statistical problems rely in some way on approximations to densities or distributions functions derived from asymptotic theory. Exact answers are rarely available in a form simple enough to be used directly. Typically, the approximations to be used are based on results in the theory of probability. This theory can be usefully combined with asymptotic techniques from analysis and the development of asymptotic expansions. One such example is the *saddlepoint approximation*, which was introduced in statistics by Daniels (1954) for deriving a very accurate approximation to the density of the mean of a sample of independent and identically distributed observations. This pioneering article was reprinted in Springer's 'Breakthroughs in Statistics' series (Daniels, 1997). Saddlepoint approximations can be obtained for any statistic which admits a cumulant generating function. If the cumulant generating function is known the moment generating function may be easily determined and the density and the distribution may be computed by numerically evaluating the convolution formulae or by integrating its

inversion formula. But, in practice the first approach may be time-consuming and inaccurate. The second uses the Fourier inversion formula for the exact density and chooses the contour of integration to pass through the saddlepoint of the integrand on the line of steepest descent. In this form the argument is similar to the term-by-term inversion used to obtain the Edgeworth expansion, but is based on the theory of asymptotic analysis. The saddlepoint approximation to the distribution of a statistic is obtained by integration of its density estimate. Daniels (1954) used the saddlepoint technique of asymptotic analysis, for which general discussions can be found in Courant and Hilbert (1950) or Bleistein and Handelsman (1975).

After Daniels's paper, and up to the mid-1960s, saddlepoint techniques were applied to several types of problem by Daniels and a few authors. The apparent loss of interest in this area after the mid-1960s was followed by a strong revival at the beginning of the 1980s. Following Barndorff-Nielsen and Cox (1979), this method received a dramatic surge in popularity in the literature, and several extensions of Daniels' first formula have been proposed; see for example Lugannani and Rice (1980) for tail probabilities, Field (1982) for M-estimators, Robinson (1982) for permutation tests, Davison and Hinkley (1988) for bootstrap distributions, Booth and Butler (1990) for randomization distributions, and Gatto and Ronchetti (1996) for marginal densities of general statistics. Overall, saddlepoint approximations have been shown to be an important and powerful tool for obtaining accurate expressions for densities and distribution functions. The number of applications of the saddlepoint approximation is quite impressive, as warrants this extremely powerful approximation. Thus, various techniques of accurate approximations, relying on it in one way or another, have been developed since the seminal article by Daniels (1954). Detailed references on the application of saddlepoint approximations in statistics can be found in the review articles by Reid (1988, 1991, 1996) and in the books by Barndorff-Nielsen and Cox (1989), Field and Ronchetti (1990), Kolassa (1994), Jensen (1995), Field and Tingley (1997), and Davison and Hinkley (1997) for bootstrap analysis. More recently, the contributions of saddlepoint and related approximations published in *Biometrika* are reviewed in Davison (2001). To get familiarized with the theory underlying the saddlepoint approximation, I refer to Goutis and Casella (1999), Huzurbazar (1999), or, more mathematically, to Daniels (1987). But, as noted by Ronchetti (1997, page 175),

> *'... it is impossible to enumerate all the applications of saddlepoint techniques but the number, the diversity, and accuracy of the resulting approximations clearly show the great influence of Daniel's paper in statistics.'*

Ronchetti (1997, page 175) also provided an amusing explaination by Henry E. Daniels on the spelling of the word 'saddlepoint'.

> *'The reason is that in the original manuscript of my 1954 paper 'saddlepoint' appeared as a typist's error, but the editor of the Annals of Mathematical Statistics either didn't notice or didn't mind. So I decided to leave it as it matched up rather nicely with the German form Sattelpunkt which also appears as one word.'*

Unlike their direct Edgeworth counterparts, expansions based on the saddlepoint method have relative rather than absolute errors, are nonnegative everywhere (for the density function), and are almost uniformly superior to the former in the tails of the distribution, where most of the statistical interest lies; see, for instance, Easton and Ronchetti (1986), McCullagh (1987), Reid (1988), Barndorff-Nielsen and Cox (1989), Field and Ronchetti (1990) or Pace and Salvan (1997). The relation between them has already been discussed in Hampel (1974a) and in more detail in Field and Hampel (1978, 1982).

Note also that saddlepoint approximations fall into the class of so-called *small sample asymptotics* — an expression coined by Hampel (1974a). As mentioned previously, asymptotic distributions form the basis of most statistical inferences, but it has long been known that the usual large sample approximations — typically normal and chi-squared — can provide poor or misleading inferences from samples of the sizes met in practice. One reason for this is that the central limit theorem and related Edgeworth series corrections give density and distribution function approximations with poor relative error properties and can be inaccurate in the region of most use for statistical purposes, namely the tails. Thus, as illustrated above, the focus of asymptotic research has shifted over the past 25 years to small sample asymptotics based on saddlepoint and related approximations, which can provide very accurate answers even when the sample size n is small. An important feature is that these methods have good relative error properties. In this context, Henry E. Daniels (Whittle, 1993, page 350) noticed that

> '... *the large deviation people, especially the pure mathematicians, are always looking for rates. In this way they miss one of the essential aspects of the saddlepoint approach, which is the extra factor, giving you the next level of approximation. Use of the rate term alone gives you something which is wrong by a term of relative order* $n^{-1/2}$. *It is this second saddlepoint factor which gives you the astonishing accuracy.*'

Small sample inference is one of the central problems of statistics, because samples whose effective size is small or moderate arise very commonly in fields of application. Traditional examples arise in epidemiology, medicine, and social science, but the same issues appear also in established fields such as signal detection and those of emerging interest such as population pharmacokinetics. Their usefulness is summarized by Field (1997, page 39) underlining that

> '... *small sample asymptotics are useful in determining when the asymptotic results are good approximations and providing alternatives in the cases where they are not.*'

However there is a widespread misapprehension that these methods are so complicated that they can only be used by the cognoscenti; see for example the discussion of Reid (1995). It is perhaps for this reason that these methods are little used in applied work, even in contexts where they are known to provide excellent approximations and consequently greatly improved inferences. In this respect the asymptotics have been outstripped by competing approaches such as the bootstrap, Markov chain Monte Carlo,

and Bayesian approaches to inference. Each of these approaches has a role to play, but for parametric inferences in small samples each also has drawbacks that are not shared by small sample asymptotics. This under-use of the parametric methods is made more ironic by the fact that their routine application to the large class of generalized linear models was already discussed in Davison (1988), and that some of the remaining difficulties in applying them in curved exponential family models seem to have been overcome (Skovgaard, 1996), opening the way to a more widespread use in applications. There is also work on their application in general regression problems (DiCiccio *et al.*, 1990; DiCiccio and Field, 1991). Once again, it is impossible to enumerate all applications, but I refer to Brazzale (2000) for further reading and implementation of such so-called higher-order likelihood-based approximations, which, in one way or another, rely on Daniels' (1954) pioneering article.

Parallel to the application of small sample asymptotics in parametric problems has been the realization that they can be applied also in non- and semi-parametric situations. Indeed, one of their earliest modern applications was in robust statistics (Hampel, 1974a), and there is now a large literature on their use in replacing simulation in resampling (see Davison and Hinkley (1997), for a review). One peculiarity of the literature is that these methods are often used to approximate density or distribution functions of estimators (see Strawderman *et al.* (1996), for example), rather than related pivots, whereas it is the latter which are more relevant for inference. This poses problems, because approximation of pivots seems to be intrinsically harder, except in special cases. Two broad ways to estimate densities and distributions of pivots are a direct saddlepoint approach using an approximate cumulant-generating function constructed using the von Mises expansion of the pivot (Gatto, 1994; Gatto and Ronchetti, 1996), and an indirect approach to the density of the pivot that generalizes Daniels and Young (1991) in using Laplace approximation to marginalize the joint density of the estimator and its standard error. Again, saddlepoint methods can provide extremely accurate approximations within these contexts.

Outline

The thesis is organized as follows. Chapter 2 serves as a review of basic saddlepoint approximations. In Chapters 3 and 4 saddlepoint approximations will be successfully applied as a prelude to their application to robust resampling inference presented in Chapters 5, 6 and 7.

More precisely, I will illustrate in Chapter 3 that the saddlepoint approximation is a method of calculating the distribution of quadratic forms in normal variates in an extremely accurate way. The saddlepoint approximation to the distribution of the waiting time in the coupon collector's problem will be given in Chapter 4. Chapter 5 starts the intensive application of saddlepoint approximations to bootstrap distributions. As examples I will consider in Chapter 5 Huber's M-estimate of location and its initial MAD scaled version, whereas in Chapter 6 I consider their studentized versions together with the classical studentized statistic and the studentized version of Huber's proposal 2. These studentized statistics form the basis of the bootstrap-t method discussed in

Possible paths through the thesis

Basic saddlepoint approximations	Chapter 2
Quadratic forms in normal variables	Chapter 3
The coupon collector's problem	Chapter 4
Bootstrap distributions	Chapter 5
Studentized bootstrap distributions	Chapter 6
Studentized bootstrap confidence intervals	Chapter 7
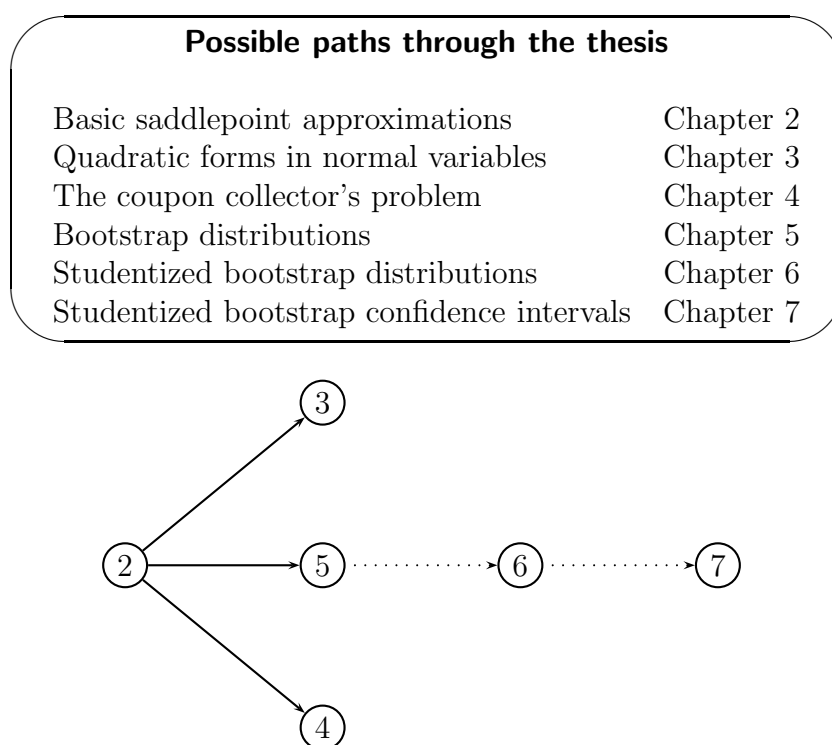
Figure 1.1. *Dependence chart. A solid arrow means that a chapter is pre-requisite for a next chapter. A dotted arrow means a natural continuation. Vertical or horizontal position has no independent meaning.*

Chapter 7. The aim of Chapter 7 is to identify a procedure which is preferable to use in order to perform robust inference. Every chapter ends with a 'Conclusion' section. These are meant to summarize its contents and give an outlook to open questions or serve as bridge for the subsequent chapter. For the reader's convenience Figure 1.1 indicates possible paths through the thesis.

The new contributions of this thesis are given in Chapters 3 and 4, in Section 5.2 and in the examples and comments given in Chapters 5 and 6. Especially, I would like to mention Sections 5.3.2.2, 6.4.1, 6.4.2 and 6.5.2. The simulation study in Chapter 7, together with additional remarks in Chapters 5 and 6, fill several gaps in the bootstrap literature. The recommendation given in Section 7.3, summarising Chapters 5–7, can also be seen as a contribution to robust resampling inference. To my knowledge the current thesis is the only work so far which considered mainly saddlepoint approximations to bootstrap versions of (un)studentized robust M-estimates of location. Much has been done in the non-bootstrap context but not in the bootstrap context. I hope that this thesis will stimulate further discussion and research.

Computing environment

When not mentioned otherwise, calculations were made on a Sun SPARC Ultra 60 workstation with 1Gb RAM using S-Plus or R. S-Plus is a value-added version of S (Becker *et al.*, 1988; Chambers, 1998) sold by Insightful Corporation (formerly

Math Soft, Inc.). The S system was recently recognized with the prestigious 'Association for Computing Machinery (ACM) Software Systems Award'. The ACM citation (`www.acm.org/awards/ss_citations/1998.html`) notes that S '*forever altered the way people analyze, visualize, and manipulate data*'. The S language is often the vehicle of choice for research in statistical methodology and R (Ihaka and Gentleman, 1996) provides a free software route to participation in that activity. R is also known as 'GNU S' and is similar to the ACM award-winning S system. Please consult the R project homepage (`www.R-project.org`) for further information. I have tried where possible to use code that works in all versions of S-Plus and R. Good reference books for S-Plus and R are Venables and Ripley (1999, 2000).

There are various tools available in S-Plus to measure resources, but they differ between versions. The principal resource considered in this thesis is the CPU time. It is clear that more resources, like memory usage, should be considered but I found it sufficient to simply consider the CPU time in order to illustrate the comments made herein. To do so I used the S-Plus function `resources` given in Venables and Ripley (2000, pages 151 and 152).

Typesetting
The thesis was typeset using LaTeX. The figures were generated in PostScript, and were directly incorporated.

Notation and typographic conventions
Besides standard notation and the one introduced throughout the thesis, note that the range of summation is not explicitly stated in the text but is implied by the range of the index. For example, if the index i is known to vary between 1 and n the term $\sum_{i=1}^{n}$ will be simple written as $\sum$. Throughout this thesis S-Plus or R language constructs and commands are set in typewriter font `like this`, or in using the S-Plus or R prompt '>' like

```
>
```

The end of each example or remark is marked with a □.

2. Basic saddlepoint approximations

'Roughly speaking, a higher order approximation around the center of the distribution is replaced by local low order approximations at each point. The unusual characteristic of these approximations is that the first few terms (or even just the leading term) often give very accurate approximations in the far tails of the distributions even for small sample sizes. Besides the theoretical reasons, one empirical reason for the excellent small sample behavior is that the saddlepoint approximations are density-like objects and do not show the polynomial-like waves exhibited for instance by Edgeworth approximations.'

Christopher A. Field and Elvezio M. Ronchetti (1990, page 1)

Saddlepoint approximations to density and distribution functions are the basis of many highly accurate small sample methods. In this chapter the ideas underlying the saddlepoint approximations are informally described for further use in subsequent chapters. I will not sketch detailed derivations of the approximations stated as there are plenty of authors who did this very carefully; see the references given in Chapter 1. Especially, concerning mathematical rigour, I refer to Jensen's (1995) book.

Section 2.1 recalls some statements made in Chapter 1. As saddlepoint approximations can be obtained for any statistic which admits a cumulant generating function I discuss moments and cumulants and their relationship briefly in Section 2.2. Section 2.3 reviews the saddlepoint approximations for the mean stated in the seminal article by Daniels (1954). These basic approximations will be generalised in Section 2.4 and used in subsequent chapters.

2.1. Introduction

Most statistical problems rely in some way on approximations to densities or distributions functions derived from asymptotic theory. Exact answers are rarely available in a form simple enough to be used directly. It has long been known that the usual large sample approximations — typically normal and chi-squared — can provide poor or misleading inferences from samples of the sizes met

in practice. A first alternative is Edgeworth expansion. For a general review of the Edgeworth expansion see Easton and Ronchetti (1986) and the references therein. In general, the Edgeworth expansion provides a good approximation in the centre of the density, but can lead to inaccurate tail probabilities. Moreover, they contain absolute errors, and can be negative. A second alternative is saddlepoint approximations, introduced by Daniels (1954) using the saddlepoint technique of asymptotic analysis. The saddlepoint approximation can be more accurate than Edgeworth expansion, especially in the distribution tails and for small sample sizes. Saddlepoint methods can be derived analytically from an Edgeworth expansion of the density of the statistic of interest. Easton and Ronchetti (1986) presented a technique for converting an Edgeworth expansion into a saddlepoint approximation. Moreover, they mentioned a theoretical advantage of the saddlepoint approximation: the leading term of the saddlepoint approximation is of the same order as the first two terms of the Edgeworth expansion. As already outlined in Chapter 1, saddlepoint approximations contain relative rather than absolute errors, are nonnegative everywhere, and are almost uniformly superior to the former in the tails of the distribution, where most of the statistical interest lies.

As saddlepoint approximations require the entire cumulant generating function, it should be mentioned that Easton and Ronchetti (1986) provide a general method for deriving saddlepoint approximations for very general statistics. The method involves using only the first few terms in a Taylor expansion of the cumulant generating function and then applying the usual saddlepoint technique; see also Gatto (1994) and Gatto and Ronchetti (1996). Since the tail area probability of a distribution critically depends on the behaviour of its cumulant generating function around the origin, truncating the cumulant generating function could have some potentially dramatic effect on the resulting saddlepoint approximation. Furthermore, the saddlepoint equation obtained from the truncated cumulant generating function may not have a unique solution, though there are some suggestions to overcome this difficulty (Wang, 1992).

2.2. Cumulant generating functions

his section deals with the elementary theory of cumulants, or more precisely, gives the definition of the cumulant generating function, which is the basis of the saddlepoint approximation. For a thorough discussion of moments and cumulants and the relationship between them, I refer to McCullagh (1987).

Let $M(\zeta)$ be the *moment generating function* of the random variable X,

$$M(\zeta) = \mathrm{E}\{\exp(\zeta X)\},$$

provided $M(\zeta) < \infty$. For statistical purposes it is often more convenient to work with the *cumulant generating function* of X. If the random variable X has a density function $f(x)$ defined over $-\infty < x < \infty$, then the cumulant generating function is

$$K(\zeta) = \log M(\zeta) = \log \left\{ \int_{-\infty}^{\infty} e^{\zeta x} f(x) dx \right\}.$$

If the distribution of X is discrete, the integral is replaced by a sum over the discrete values. More generally, to take care of both the discrete and the continous case simultaneously, we may use

$$K(\zeta) = \log \left\{ \int_{-\infty}^{\infty} e^{\zeta x} dF(x) \right\},$$

where $F(\cdot)$ is a probability measure. As noticed for instance in McCullagh (1987, page 25) it is often simpler to work with cumulants rather than with moments. The *rth cumulant* is $\kappa_r = K^{(r)}(0)$, so $K(\zeta)$ has power expansion

$$K(\zeta) = \sum_{r=1}^{\infty} \zeta^r \kappa_r / r!,$$

provided all the cumulants exist. The first and second cumulants of X are its mean and variance. The third and fourth cumulants of X, the *skewness*, κ_3, and the *kurtosis*, κ_4, are often used as simple measures of how close a random variable is to normality.

Example 2.1. Consider n independent random variables $X_1, \ldots, X_n$ with respective cumulant generating functions $K_{X_1}(\zeta), \ldots, K_{X_n}(\zeta)$. Then $X_1 + \cdots + X_n$ has cumulant generating function

$$K(\zeta) = \log \{M(\zeta)\} = \log \left\{ \prod_{i=1}^{n} M_{X_i}(\zeta) \right\} = \sum_{i=1}^{n} K_{X_i}(\zeta).$$

It follows that the rth cumulant of a sum of independent random variables is the sum of their rth cumulants. $\square$

Example 2.2 (Binomial distribution). Let $X_1, \ldots, X_n$ be a random sample from a binomial distribution $B(n, p)$. Following Example 2.1, the cumulant generating function of $T = \sum X_i$ is given by $K(\zeta) = nK_X(\zeta)$, where $K_X(\zeta) = \log[1 + p\{\exp(\zeta) - 1\}]$. $\square$

Example 2.3 (Multinomial distribution). Suppose that a vector random variable $X = (X_1, \ldots, X_m)$ has the multinomial distribution with denominator n and probabilities $\pi_1, \ldots, \pi_m$. This distribution arises when n independent observations take values in one of m categories, each falling into the rth category with probability π_r, $\pi_r \geq 0$ and $\sum \pi_r = 1$. Then X_r is the total number falling into the rth category. Its mass function is

$$\Pr(X_1 = x_1, \ldots, X_m = x_m) = \frac{n!}{x_1! \times \cdots \times x_m!} \pi_1^{x_1} \times \cdots \times \pi_m^{x_m}, \qquad \sum_{r=1}^{m} x_r = n.$$

The moment generating function of X is

$$E(e^{\zeta^T X}) = \sum \frac{n!}{x_1! \times \cdots \times x_m!} \pi_1^{x_1} \times \cdots \times \pi_m^{x_m} e^{x_1 \zeta_1 + \cdots + x_n \zeta_n},$$

where the sum is over all integer vectors x such that $x_r \geq 0$ and $\sum x_r = n$. Thus the cumulant generating function of the vector random variable X is

$$K(\zeta) = \log E\{\exp(\zeta^T X)\} = n \log \left\{ \sum_{i=1}^{m} \pi_i \exp(\zeta_i) \right\}.$$

It follows that $\mathrm{E}(X_r) = n\pi_r, \mathrm{var}(X_r) = n\pi_r(1 - \pi_r)$, and $\mathrm{cov}(X_r, X_s) = -n\pi_r\pi_s$ for $r \neq s$. $\hfill\square$

2.3. Saddlepoint approximations for the mean

The first statistical application of the saddlepoint approximation was derived by Daniels (1954). He approached the problem of finding a density approximation through the inversion of a Fourier transform. As shown below, such an approach makes us deal with complex integration.

Let X be a scalar random variable distributed according to the density function $f(x)$. We define the moment generating function,

$$M(\tau) = \int_{-\infty}^{\infty} e^{\tau x} f(x)dx,$$

and the cumulant generating function,

$$K(\tau) = \log \left\{ \int_{-\infty}^{\infty} e^{\tau x} f(x)dx \right\}.$$

To restore $f(x)$ we can apply the Fourier inversion formula,

$$f(x) = \frac{n}{2\pi} \int_{-\infty}^{\infty} M(ir)e^{-irx}dr.$$

Let $X_1, \ldots, X_n$ be a sample of independent and identically distributed scalar variables distributed according to the density function $f(\cdot)$, and let $T = \bar{X}$ denote their average. Following Examples 2.1 and 2.2, its density function $f_T(t)$ is therefore

$$
\begin{aligned}
f_T(t) &= \frac{n}{2\pi} \int_{-\infty}^{\infty} M^n(ir)e^{-nirt}dr \\
&= \frac{n}{2\pi} \int_{-i\infty}^{i\infty} \exp\{n(K(\tau) - \tau t)\}d\tau. \quad\quad (2.3.1)
\end{aligned}
$$

By Cauchy's closed curve theorem of complex variables the value of the integrand is unchanged if we integrate along any line parallel to the imaginary axis,

$$f_T(t) = \frac{n}{2\pi} \int_{-\infty}^{\infty} \exp[n\{K(\zeta + iy) - (\zeta + iy)t\}]dy. \quad\quad (2.3.2)$$

To evaluate the integral we will expand the contents of the exponential as a power series in y about $\tau = \zeta + i0$,

$$K(\zeta + iy) - (\zeta + iy)t = K(\zeta) - \zeta t + i\{K'(\zeta) - t\}y + \sum_{j \geq 2} K^{(j)}(\zeta)(iy)^j/j!, \quad\quad (2.3.3)$$

where $K'(\zeta)$ and $K^{(j)}(\zeta)$ are the first and jth derivatives of $K(\zeta)$ with respect to ζ. Then, combining (2.3.3) and (2.3.2) yields

$$
\begin{aligned}
f_T(t) &= \frac{n}{2\pi} \int_{-\infty}^{\infty} \exp\left(n\left[K(\zeta) - \zeta t + i\{K'(\zeta) - t\}y + \sum_{j\geq 2} K^{(j)}(\zeta)(iy)^j/j!\right]\right) dy \\
&= \frac{n}{2\pi} \exp\left[n\{K(\zeta) - \zeta t\}\right] \int_{-\infty}^{\infty} \exp\left(n\left[i\{K'(\zeta) - t\}y + \sum_{j\geq 2} K^{(j)}(\zeta)(iy)^j/j!\right]\right) dy.
\end{aligned}
$$

Next, we separate the first two terms in the exponential, and then expand the remaining terms as a product of power series:

$$
\begin{aligned}
f_T(t) &= \frac{n}{2\pi} \exp\left[n\{K(\zeta) - \zeta t\}\right] \int_{-\infty}^{\infty} \exp\left(n\left[i\{K'(\zeta) - t\}y - K''(\zeta)y^2/2\right]\right) \\
&\quad \times \prod_{j\geq 3} \exp\left[n\left\{K^{(j)}(\zeta)(iy)^j/j!\right\}\right] dy \\
&= \frac{n}{2\pi} \exp\left[n\{K(\zeta) - \zeta t\}\right] \int_{-\infty}^{\infty} \exp\left(n\left[i\{K'(\zeta) - t\}y - K''(\zeta)y^2/2\right]\right) \\
&\quad \times \prod_{j\geq 3} \left[1 + \sum_{k\geq 1} \left\{nK^{(j)}(\zeta)(iy)^j/j!\right\}^k / k!\right] dy.
\end{aligned}
$$

Multiplying out the product of summations, we get

$$
\begin{aligned}
f_T(t) &= \frac{n}{2\pi} \exp\left[n\{K(\zeta) - \zeta t\}\right] \int_{-\infty}^{\infty} \exp\left(n\left[i\{K'(\zeta) - t\}y - K''(\zeta)y^2/2\right]\right) \\
&\quad \times \left(1 + \sum_{j\geq 1} A_j i^j y^j\right) dy. \tag{2.3.4}
\end{aligned}
$$

The A_j correspond in some extent to the coefficients A_j in Daniels (1954, equation (4.3)), and hence are not given here. We now apply the standard integral, true for $\mathrm{Re}(p) > 0$:

$$
\int_{-\infty}^{\infty} x^n e^{-p^2 x - qx} dx = \left(\frac{i}{2}\right)^n \sqrt{\pi} p^{-(n+1)/2} \exp(q^2/4p) H_n(iq/2\sqrt{p}), \tag{2.3.5}
$$

where $H_j(\cdot)$ is the Hermite polynomial of degree j. Note that $H_{2k}(0) = (-1)^k(2k)!/k!$ and $H_{2k+1}(0) = 0$. Applying (2.3.5) to (2.3.4) yields

$$
\begin{aligned}
f_T(t) &= \left\{\frac{n}{2\pi K''(\zeta)}\right\}^{-1/2} \exp\left(n\left[K(\zeta) - \zeta t - \frac{\{K'(\zeta) - t\}^2}{2K''(\zeta)}\right]\right) \\
&\quad \times \left[1 + \sum_{j\geq 1} A_j(-1)^j \left\{\frac{1}{2\pi K''(\zeta)}\right\}^{j/2} H_j\left\{\frac{K'(\zeta) - t}{\sqrt{2K''(\zeta)/n}}\right\}\right].
\end{aligned}
$$

When $\zeta = 0$ this reduces to the Edgeworth series for $f_T(t)$. The formal proof of convergence of the series requires a number of other results and hence will not be given

here. However, when $K'(\hat{\zeta}) = t$, the arguments of the Hermite functions become zero, thus terms of all odd orders disappear and only the terms of even orders contribute; see also Daniels (1954, page 636). It is certainly reasonable that this should be the best estimate of the integral. This is underlined by the fact that the integrand of (2.3.1) is of the form $\exp\{ng(\cdot)\}$, where $g(x) = K(x) - xt$, and hence the major contribution to this integrand for large n will come from a neighborhood of the saddlepoint $\hat{\zeta}$, a zero of $g(\cdot)$, as $K'(\hat{\zeta}) = t$.

This approach can also be motivated from a purely analytic point of view, by recourse to saddlepoint integration techniques. Pick $\hat{\zeta}$ along the real axis minimizing $K(\zeta) - \zeta t$, and pick the path through $\hat{\zeta}$ along which the real part of $K(\zeta) - \zeta t$ decreases the fastest as one moves away from the real axis. This path is called the path of *steepest descent*; see also Kolassa (1994, Section 4.4) or Field and Ronchetti (1990, Sections 3.2 and 3.3) for more evidence and figures of the resulting steepest descent curves. Another nice application of the method of steepest descent is given in the context of random walk and random flights problems in Hughes (1995, Section 2.7). In addition, note also that $\hat{\zeta} = \hat{\zeta}(t)$, viewed as a point in the complex plane, is neither a maximum nor a minimum but, as stated above, a saddlepoint of $K(\zeta) - \zeta t$, as the function is constant in the imaginary direction and has an extremum in the real direction.

Hence, the so-called *saddlepoint approximation* (Daniels, 1954) to $f_T(t)$ is

$$f_s(t) = \left\{ \frac{n}{2\pi K''(\hat{\zeta})} \right\}^{-1/2} \exp\left\{ n\left[K(\hat{\zeta}) - \hat{\zeta}t \right] \right\}, \qquad (2.3.6)$$

where $\hat{\zeta} = \hat{\zeta}(t)$, known as the *saddlepoint*, is the value of ζ satisfying the *saddlepoint equation* $K'(\hat{\zeta}) = t$, where $K'(\zeta)$ and $K''(\zeta)$ are the first and second derivatives of $K(\zeta)$ with respect to ζ.

Daniels (1954) discussed the existence and properties of the real roots of the saddlepoint equation $K'(\zeta) = t$, upon which the saddlepoint approximations depend. He noticed that the saddlepoint approximation can be used whenever t lies within a restricted range of values assumed by $K'(\zeta)$ in order to have a unique real root. For values of t near the bounds of its admissible range, where the approximation might be expected to fail, he noticed that the approximation is very accurate, whereas the accuracy of the normal approximation, and of the Edgeworth expansion generally, deteriorate as t approaches the ends of its range.

Note that the expression (2.3.6) is always positive, but will not usually integrate to exactly one, so in practice it is renormalized. The renormalized version improves the relative order of the approximation to $O(n^{-3/2})$, it has been $O(n^{-1})$ without renormalization. Calculation of these error terms often requires detailed technical arguments, and I do not include them; see Field and Ronchetti (1990) or Kolassa (1994). Note that the notation $O(n^{-1})$ denotes a function that satisfies $\lim_{n\to\infty} nO(n^{-1}) = $ constant, where for a random sample of size n standard large sample techniques typically give approximations of absolute errors of order $O(n^{-1/2})$. The property of having a relative error of order at worst $O(n^{-1})$ for approximating the density is a major advantage compared to Edgeworth expansions. And as the error is relative the ratio of the true distribution

to its saddlepoint approximation is bounded over the likely range of t. Daniels (1954, Section 7) showed that the coefficient of the term of order $O(n^{-1})$ does not depend on t for a wide class of underlying densities. Thus, in such cases the relative error is of order $O(n^{-1})$ even uniformly (Jensen, 1988, 1995, Chapter 6). The approximation (2.3.6) can also be derived from the Edgeworth expansion; see Barndorff-Nielsen and Cox (1979) or Barndorff-Nielsen and Cox (1989, Section 4.3). In the latter, (2.3.6) is called the *tilted Edgeworth expansion.*

It turns out that the saddlepoint approximation (2.3.6) is exact or exact up to renormalization for the normal, the gamma and the inverse normal distributions (Daniels, 1954, 1980). An analogous approximation for the distribution of the sample mean of discrete variables has the same form as (2.3.6) with n in the left-hand term replaced by $1/n$; see Daniels (1983, 1987).

Finally, Daniels (1954, Section 4) showed that saddlepoint approximations can also be viewed probabilistically by means of the method of conjugate densities. Following Reid (1988, page 214) this leads to a *'more statistical version of the derivation'* of the saddlepoint approximation given in (2.3.6). This technique had been used by Esscher (1932), but is attributed to Cramér (1938), whereas Efron (1981) and others call this operation *exponential tilting.* Although not shown in this thesis, Ronchetti (1997, page 173) noted that:

> *'By proving that the method of steepest descent and the technique based on conjugate density lead to the same approximation, Daniels gave new insight into the properties and in particular the accuracy of the approximation. Whereas Daniels (and this author) shows a clear preference for the derivation by means of the method of steepest descent, we can argue that it is the combination of both which really leads to a deep understanding of the properties of the approximation.'*

So far, we only have seen the saddlepoint approximation to the density of the mean, derived in Daniels (1954) and reviewed in Reid (1988). There are essentially three methods of calculating saddlepoint approximations to the cumulative distribution function of $T = \bar{X}$ in common use: the numerically integrated saddlepoint density, usually renormalized for additional accuracy; the indirect Edgeworth expansion; and the Lugannani–Rice formula. The latter two were compared in Daniels (1987) in the case of the mean, for exponential and inverse normal distributions, examples where the density approximation (2.3.6) is exact. He found that the indirect Edgeworth expansion, not given herein (see Daniels (1987, equation (3.4)), for instance), performs slightly better than the Lugannani–Rice formula (2.3.8), shown below. However, the Lugannani–Rice formula is much easier and simpler to apply than the indirect Edgeworth expansion, which was studied intensively in Robinson *et al.* (1990) or in Jing and Robinson (1994, Section 2). See also, Robinson (1982), who showed that the indirect Edgeworth expansion as given in equation (3.4) in Daniels (1987) is actually the Laplace approximation to the integrated saddlepoint density.

The approximation to the cumulative distribution function used throughout this thesis, originally due to Barndorff-Nielsen (1986, 1990), is an alternative to the Lugannani–

Rice approximation, and is defined as

$$F_s(t) = \Phi \left\{ w + \frac{1}{w} \log \left(\frac{v}{w} \right) \right\}, \qquad (2.3.7)$$

where $\Phi(\cdot)$ denotes the standard normal cumulative distribution function, and

$$w = \text{sign}(\hat{\zeta}) \left\{ 2 \left(\hat{\zeta} t - K(\hat{\zeta}) \right) \right\}^{1/2}, \qquad v = \hat{\zeta} \left\{ K''(\hat{\zeta}) \right\}^{1/2}.$$

Both w and v are functions of t. Calculation of (2.3.7) and (2.3.6) requires knowledge of the cumulant generating function $K(\zeta)$ and computation of $\hat{\zeta}$, the saddlepoint, for each t of interest. The relative error in (2.3.7) is also $O(n^{-3/2})$. The derivation of (2.3.7) is reviewed in Jensen (1995), and uniformity properties of the approximation are discussed in Jensen (1988).

As mentioned, (2.3.7) is an alternative to the Lugannani–Rice approximation (Lugannani and Rice, 1980),

$$F_s^{LR}(t) = \Phi(w) + \phi(w) \left(\frac{1}{w} - \frac{1}{v} \right), \qquad (2.3.8)$$

where $\phi(\cdot)$ is the standard normal density function. For a detailed review and discussion see Daniels (1987). More recently, Routledge and Tsao (1997a) proved that the Lugannani–Rice approximation may be differentiated to obtain Daniel's expansion for the corresponding density function.

Approximations (2.3.7) and (2.3.8) are usually indistinguishable in practice. More precisely, (2.3.8) has similar, often identical, asymptotic properties to (2.3.7); see, for example, Barndorff-Nielsen (1991, Section 5). Jensen (1995, Theorem 5.1.1) proved that (2.3.7) is equivalent to (2.3.8). The main advantage of the approximations (2.3.7) and (2.3.8) over the numerically integrated saddlepoint density is that they are much simpler to use; they only require the evaluation of one saddlepoint, and no numerical integration. The first could be avoided in the integrated saddlepoint approach by means of the transformation given in Daniels (1983, Section 6), but numerical integration still would be needed to get the distribution approximation. See Robert and Casella (1999, Example 6.3.4) for an illustration of the use of the Metropolis–Hastings algorithm to do such a calculation. Routledge and Tsao (1997b) introduced two numerical quadrature methods for computing distributions, being complements to asymptotics. They concluded that the saddlepoint approximation will in general be uniformly more accurate, especially in the extreme tail.

In rewriting approximations (2.3.7) and (2.3.8) as $1 - F_s(t)$, one notices the appearence of the term $1 - \Phi(\cdot)$. As noted by Daniels (1987, page 45), the calculation of very small tail probabilities may lead to numerical instability and he found it helpful to express the formulae in terms of Mill's ratio $R(\cdot) = \{1 - \Phi(\cdot)\}/\phi(\cdot)$ rather than $1 - \Phi(\cdot)$ itself. Fortunately, I did not encounter such instabilities for the applications in this thesis.

Both approximations (2.3.7) and (2.3.8) are often surprisingly accurate throughout the range of $T = \bar{X}$ (see Daniels (1983, 1987) or Davison and Hinkley (1988), for example), except when t is close to $\mathrm{E}(T)$, or similarly $\hat{\zeta} \approx 0$, where the approximations can become numerically unstable, and are best avoided. The approximation should then be replaced by its limit as $\hat{\zeta} \to 0$ (Daniels, 1987, equation (4.10)),

$$F_s\{\mathrm{E}(T)\} = \frac{1}{2} - \frac{1}{6}(2\pi n)^{-1/2}\lambda_3(0),$$

with $\lambda_3(0) = K'''(0)/\{K''(0)\}^{3/2}$, where $K''(\cdot)$ and $K'''(\cdot)$ are the second and the third derivatives of the cumulant generating function with respect to ζ.

With slight modifications similar formulae are available when the $X_1, \dots, X_n$ are a sample of independent and identically distributed discrete variables, taking values on a lattice with probabilities $\Pr(X_i = j) = p_j$, $i = 1, \dots, n$. Following Daniels (1987, Section 6) the coefficient v used in the approximations (2.3.7) and (2.3.8) needs to be changed and a continuity correction to the saddlepoint equation is needed. Indeed, $\hat{\zeta}^*$ denotes the value of ζ satisfying $K'(\zeta) = t - 1/2$, and v is defined as

$$v = 2\sinh\left[\hat{\zeta}^*\left\{K''(\hat{\zeta}^*)\right\}^{1/2}\right].$$

See also Kolassa (1994, Section 5.5).

Finally, note that the above approximations are based on the normal distribution and are the most common saddlepoint approximations. However, derivations based on other distributions are also possible using bases such as the gamma and the inverse normal distributions. The resulting generalized approximations are presented and discussed in Wood *et al.* (1993) for the Lugannani–Rice approximation (2.3.8), and in Booth and Wood (1995) for (2.3.7).

2.4. Generalisation

The saddlepoint approximations for the mean discussed in the previous section will be generalised in this section in order to enable their use in subsequent chapters. First of all, Daniel's (1954) saddlepoint approximation to the density function, (2.3.6), contains already the basic elements needed to go beyond the case of the mean. Indeed (2.3.1) holds for an arbitrary statistic, T, having cumulant generating function $K_T(\cdot)$, provided $K(\tau)$ is replaced by $K_T(n\tau)/n$. As an application, the saddlepoint approximations for sums of independent and identically distributed variates, for instance, is reviewed in Davison and Hinkley (1988, Section 2). Although typically unknown, an approximation for $K_T(\cdot)$, for example, by Taylor expansion could be used in (2.3.1). Easton and Ronchetti (1986) showed that the other steps remain the same.

The basic approximations could be generalized in various ways. But, for the purpose of this thesis, the most useful situation is as follows. Let $X_1, \dots, X_n$ be a sample of independent identically distributed scalar random variables. Similar to Example 2.1, the cumulant generating function of a $d \times 1$ statistic T, which may be expressed as

a linear combination of independent random variables, say $T = \sum a_i X_i$, is $K_T(\zeta) = K(\zeta) = \sum K_{X_i}(\zeta^T a_i)$, where the a_i are constant $d \times 1$ vectors, and ζ is a $d \times 1$ vector. For additional invariance properties of the cumulant generating function see McCullagh (1987, Section 6.2.4). Another way to express T is as $A^T X$, where A^T is the $d \times n$ matrix whose ith column is a_i, and X is the $n \times 1$ vector with ith element X_i. The saddlepoint approximation to the density of T at t becomes

$$f_s(t) = \left\{ (2\pi)^d |K''(\hat{\zeta})| \right\}^{-1/2} \exp \left\{ K(\hat{\zeta}) - \hat{\zeta}^T t \right\}, \qquad (2.4.1)$$

where $\hat{\zeta} = \hat{\zeta}(t)$, the saddlepoint, satisfies the $d \times 1$ saddlepoint equation $\partial K(\zeta)/\partial \zeta = t$ and $K''(\hat{\zeta})$ is the $d \times d$ matrix with elements $\partial^2 K(\hat{\zeta})/\partial \zeta \partial \zeta^T$, and $|\cdot|$ denotes determinant; see also Davison and Hinkley (1997, Section 9.5.2).

When $d = 1$, *i.e.* T is scalar, an approximation to its cumulative distribution function at t is

$$F_s(t) = \Phi \left\{ w + \frac{1}{w} \log \left(\frac{v}{w} \right) \right\}, \qquad (2.4.2)$$

where

$$w = \text{sign}(\hat{\zeta}) \left\{ 2 \left(\hat{\zeta} t - K(\hat{\zeta}) \right) \right\}^{1/2}, \qquad v = \hat{\zeta} \left\{ K''(\hat{\zeta}) \right\}^{1/2}.$$

See Barndorff-Nielsen and Cox (1994, Section 6.7). When t is close to $\text{E}(T)$, approximation (2.4.2) can become numerically unstable, and is best avoided. For the general case the explicit expression is rather complicated. Fortunately such values of t are not of interest in practice.

As we will see in Section 5.3, this approximation generalizes easily when the quantity of interest is the scalar solution of estimating equations as in the context of robust inference.

2.5. Conclusion

In this chapter, saddlepoint approximations were reviewed. It was stated that the saddlepoint method yields extremely accurate approximations to density and distribution functions. This accuracy will be underlined with the applications shown in subsequent chapters. Saddlepoint approximations have widespread applicability, far beyond that illustrated in this chapter; see the review articles by Reid (1988, 1991, 1996), and more recently Davison (2001).

One of the limitations of saddlepoint approximation is the requirement that the cumulant generating function exists. Although this may appear to be rather limiting, the list for commonly used distributions with cumulant functions is already large. Furthermore this limitation can be bypassed in some instances by using the empirical cumulant generating function. Examples include applications to resampling methods such as the bootstrap (see Davison and Hinkley (1988) and Chapters 5–7 of this thesis) as well as applications to density estimation (Feuerverger, 1989).

As a final note, concerning the writing style of Daniels' pioneering article, Ronchetti (1997, page 172) wrote the following:

> *'Daniels writes the first part in the style of a good applied mathematician and this allows him to focus on the ideas. ... Fortunately, the paper was written in 1954 for I suspect such a style would not be acceptable nowadays in the 'Annals of Statistics'.'*

But, fortunately, it was accepted for publication.

3. Quadratic forms in normal variables

'... there is also this interesting fact about saddlepoint approximations: they always seem to be more accurate than one should expect. In fact, they seem to embody the truth of the matter far better than the exact formula, which is itself often difficult to comprehend. The saddlepoint approximation somehow seems to capture the essence, which is why I pursue it. One does feel that there is some underlying reason why the approximation is so much better than expected.'

Henry E. Daniels (Whittle, 1993, page 350)

n this chapter a saddlepoint approximation is used for the distribution of quadratic forms in normal variates. It is comparable in speed with exact methods, almost as accurate and is much easier to program. After stating the problem in Section 3.1, a review of exact and approximate methods for computing the distribution of quadratic forms in normal variates, presented in Section 3.2, is given in Section 3.3. The exact cumulant generating function is easily derived. Extensive study in Section 3.4 shows that this approximates the distribution extremely accurately. This is underlined in Section 3.5 by means of several applications to nonparametric regression.

The main results of this chapter appeared in *Biometrika* (Kuonen, 1999). The application to nonparametric regression in Section 3.5.2 was presented at Prague Stochastics '98 (Kuonen, 1998c), and a review was given at the SRSS meeting (Kuonen, 2000e).

3.1. Introduction

et $X = (X_1, \ldots, X_n)^T$ be a random vector following a multivariate normal distribution with mean vector $\mu = (\mu_1, \ldots, \mu_n)^T$ and variance-covariance matrix Ω. The quadratic form associated with the $n \times n$ matrix A is

$$Q(X) = X^T A X = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} X_i X_j. \tag{3.1.1}$$

Without loss of generality we can assume the matrix A to be symmetric: if not, then one may take the symmetric matrix $(A + A^T)/2$ as the matrix of the quadratic form since $X^T A X = X^T A^T X = X^T \{(A + A^T)/2\} X$.

Quadratic forms enter into many statistics associated with normally distributed random variables, so we may want to calculate the distribution of $Q(X)$ or the probability

$$\Pr\{Q(X) > q\}, \tag{3.1.2}$$

where q is a given scalar. In the simplest case, $A = \Omega = I_n$ and $Q(X) = \sum X_i^2$ is a noncentral chi-squared variable with n degrees of freedom and noncentrality parameter $\xi^2 = \sum \mu_i^2$ (Scheffé, 1959, Appendix IV).

As in general the matrix A is neither assumed to be idempotent (*i.e.* $AA = A$) nor to be positive definite (*i.e.* the quadratic form $X^T A X$ is positive definite, $X^T A X > 0$ for all $X \neq 0$), classical results such as Cochran's theorem (Scheffé, 1959, Appendix VI) implying a chi-squared distribution for the quadratic form do not apply, and another approach to the calculation of (3.1.2) is needed.

Imhof (1961) gives exact methods to compute this distribution using real arithmetic, whereas Ruben (1962, 1963) restricted his considerations to non-negative quadratic forms in providing a simplified expression for (3.1.1) as an infinite linear combination of (noncentral) chi-square distribution functions. An exhaustive discussion of representations of the exact distributions of quadratic forms is given in Johnson and Kotz (1970, Chapter 29). Imhof's method has been programmed in Fortran by Koerts and Abrahamse (1969) and in Pascal by Farebrother (1990). Moreover, as any quadratic form in independent normal variables can be reduced to a linear combination of chi-squared random variables, the Algol algorithm of Davies (1980) can also be used; a C version is available.

3.2.   The general quadratic form

We first consider a central quadratic form in which $\mu_1 = \cdots = \mu_n = 0$. Since the variance-covariance matrix Ω is positive definite and symmetric, it can be factored by Choleski decomposition as $\Gamma\Gamma^T$, where Γ is a non-singular lower triangular matrix. It follows that $\Gamma^T A \Gamma$ is symmetric and therefore its eigenvalues are all real. Let Λ be the diagonal matrix of eigenvalues of $\Gamma^T A \Gamma$ and V the associated matrix of eigenvectors. It may be interesting to make a non-singular orthogonal linear transformation of the variables such that (3.1.1) is particularly simple when expressed in the new variables. Suppose we transform $X = (X_1, \ldots, X_n)^T$ to $Y = (Y_1, \ldots, Y_n)^T$ by the transformation $Y = P^{-1}X$, so $X = PY$, where $P = \Gamma V$. Then $X^T = Y^T P^T$, hence $X^T A X = Y^T P^T A P Y$. As $P^T A P = V^T \Gamma^T A \Gamma V = \Lambda$, the distribution of $X^T A X$ is the same as that of $Y^T \Lambda Y = \sum \lambda_i Y_i^2$, where the Y_i are independent standard normal variables and $\lambda_1 \geq \cdots \geq \lambda_n$ denote the eigenvalues of $\Gamma^T A \Gamma$ (Scheffé, 1959, Appendix II). Hence in the central case our attention can be focused on the distribution

of

$$Q(Y) = \sum_{i=1}^{n} \lambda_i Y_i^2. \tag{3.2.1}$$

For the general noncentral case ($\mu \neq 0$) a similar reduction is possible. Imhof (1961) expressed the general quadratic form (3.1.1) as

$$Q(Y) = \sum_{i=1}^{n} \lambda_i \chi_{h_i;\sigma_i^2}^2, \tag{3.2.2}$$

where the h_i are the orders of multiplicity of the λ_i, the $\chi_{h_i;\sigma_i^2}^2$ are independent chi-squared variables with h_i degrees of freedom and noncentrality parameter σ_i^2, and they are defined by the relation

$$\chi_{h_i;\sigma_i^2}^2 = (Y_1 + \sigma_i)^2 + \sum_{r=2}^{h_i} Y_r^2,$$

where the Y_i are independent standard normal deviates. One remarks that the $\chi_{h_i;\sigma_i^2}^2$ are functions of Y_i and that the distribution of (3.2.2) depends on $\mu_1, \ldots, \mu_n$ only through σ_i^2. Moreover, in setting all h_i to one and all σ_i^2 to zero, we obtain equation (3.2.1). Imhof (1961, equation (2.3)) gives the exact characteristic function of (3.2.2),

$$\mathrm{E}[\exp\{i\zeta Q(Y)\}] = \prod_{j=1}^{n}(1 - 2i\zeta\lambda_j)^{-h_j/2} \exp\left\{ i \sum_{j=1}^{n} \frac{\sigma_j^2 \lambda_j}{1 - 2i\zeta\lambda_j} \right\}. \tag{3.2.3}$$

Johnson and Kotz (1970, page 152) give an equivalent definition. The corresponding cumulant generating function is

$$K(\zeta) = \log \mathrm{E}[\exp\{\zeta Q(Y)\}] = -\frac{1}{2}\sum_{i=1}^{n} h_i \log(1 - 2\zeta\lambda_i) + \sum_{i=1}^{n} \frac{\sigma_i^2 \lambda_i}{1 - 2\zeta\lambda_i}, \tag{3.2.4}$$

assuming $1 - 2\zeta\lambda_i$ to be positive, *i.e.*

$$\zeta < \frac{1}{2} \min_i \lambda_i^{-1}.$$

Note that for the central case (3.2.4) reduces to

$$K(\zeta) = -\frac{1}{2}\sum_{i=1}^{n} \log(1 - 2\zeta\lambda_i). \tag{3.2.5}$$

One remarks that the explicit determination of the eigenvalues $\lambda_1, \ldots, \lambda_n$ is needed.

3.3. Computing methods

n this section the problem of calculating the distribution function of the general quadratic form (3.2.2) is examined. For a detailed account on the methods discussed in the Sections 3.3.1 and 3.3.2, I refer to Kotz and Johnson (1970, Chapter 29) and more recently to Mathai and Provost (1992) who give an extensive mathematical discussion of quadratic forms in real normal random vectors and matrices.

3.3.1. Numerical methods

The existing numerical methods are mainly concerned with the problem of calculating the distribution of (3.2.2) by using numerical integration to invert the characteristic function (3.2.3), which is the Fourier transform of the density function.

Imhof (1961, Section 3) showed that the distribution of the general quadratic form (3.2.2) can be obtained by numerical integration of a modified inversion formula and gives detailed information on the resulting degree of accuracy. Imhof's technique gives excellent results in comparison with other methods, and it can almost be regarded as exact. This method has been programmed in Fortran by Koerts and Abrahamse (1969, pages 155–160); see also Koerts and Abrahamse (1969, pages 76–87) for details. It has been translated by Farebrother (1990) into Pascal using Simpson's rule for the numerical integration. Farebrother's translation (AS 256) can be downloaded from the 'Applied Statistics algorithms' section of the Statlib web site (`lib.stat.cmu.edu/apstat/`).

Two additional numerical integration methods for inverting the characteristic function (3.2.3) have been proposed by Rice (1980). Both make use of paths of integration that pass through or near a suitable saddlepoint. The first method is used when the integrand decreases rapidly, using a trapezoidal rule for the numerical integration. The second method is used when the integrand decreases slowly, as when q is small. Examples illustrating these two methods and information regarding the computation are given in Rice (1980). But the first method is known to not work well for general quadratic forms when q is small, *i.e.* in the center of the distribution, and is very costly, and the second method becomes similar to the one proposed by Davies (1973). In the context of the ratio of quadratic forms in normal variables, Lugannani and Rice (1984) expressed the probability density function of the ratio as a multiple integral and proposed a numerical integration method for its evaluation, which has similar drawbacks to the method of Rice (1980).

As any quadratic form in independent normal variables can be reduced to a linear combination of chi-squared random variables, the algorithm of Davies (1980), based on the method of Davies (1973), can be used. The original program was written in Algol, but I received an updated C version from Robert B. Davies.

Simulations have shown that Imhof's method (as implemented by Farebrother) and Davies's methods do not differ much. This result is very satisfying and was expected as both calculate the distribution exactly. Therefore, only Davies's method will be considered in what follows.

3.3.2. Existing approximations

The various methods for computing the exact distribution of quadratic forms in normal variables require in general extensive numerical computations. Numerical integration methods, though sufficiently accurate in solving the general problem, require a considerable amount of computer time. Several approximation methods have been proposed to reduce those difficulties.

The elementary approximation first presented is of historical interest. The simplest approximation proposed by Patnaik (1949) consists of replacing the distribution of (3.2.1) by that of $a\chi_b^2$, where a and b are determined to match the first two moments.

For the distribution of (3.2.2) an improvement might be expected using Pearson's (1959) three-moment central chi-squared approximation, which basically corresponds to use $a\chi_b^2 + c$ instead of $a\chi_b^2$, matching the first three moments. Pearson (1959) noticed that such distribution will never differ greatly from that of a chi-squared distribution having appropriately chosen parameters. Johnson (1959) showed that this approximation can be remarkably accurate in the tails of the distribution. Imhof (1961) extended Pearson's (1959) three-moment central chi-squared approximation to the distribution of non-central chi-squared variables. As can be seen in Imhof (1961), Pearson's approximation, which requires little more work, gives a much better fit than is achieved with Patnaik's approximation, particularly in the upper tail. As a special case of the generalisation due to Imhof (1961), we obtain the following approximation

$$\Pr\{Q(X) > q\} \doteq \Pr\{\chi_b^2 > r\},$$

where χ_b^2 denotes a chi-squared variable with $b = c_2^3/c_3^2$ degrees of freedom, $r = (q - c_1)(b/c_2)^{1/2} + b$ and $c_s = \sum \lambda_i^s(h_i + s\sigma_i^2)$, $s = 1, 2, 3$. For the case where the quadratic form is non-positive one has to assume that $c_3 > 0$. Otherwise, one must approximate the distribution of $-Q(X)$. As $\Gamma^T A\Gamma$ and ΩA have the same eigenvalues, for the central case, where $\sigma_1^2 = \cdots = \sigma_n^2 = 0$ and $h_1 = \cdots = h_n = 1$, we obtain

$$c_s = \sum_{i=1}^n \lambda_i^s = \text{tr}\{(\Gamma^T A\Gamma)^s\} = \text{tr}\{(\Omega A)^s\}.$$

Azzalini and Bowman (1993) fitted the Johnson family of frequency curves to the distribution of the central quadratic form, matching the first four moments. Fitting by moments is not always a desirable procedure, therefore in using the exact expression for the cumulants (Johnson and Kotz, 1970, page 153),

$$\kappa_s = 2^{s-1}(s-1)!\text{tr}\{(\Omega A)^s\} = 2^{s-1}(s-1)!c_s,$$

Bowman and Azzalini (1997, page 88) rewrite the Pearson approximation in the central case in terms of the first three cumulants. They fit a distribution of the type $a\chi_b^2 + c$, where $a = |\kappa_3|/(4\kappa_2)$, $b = (8\kappa_2^3)/\kappa_3^2$ and $c = \kappa_1 - ab$. The absolute value for κ_3 in a should prevent from the case where the quadratic form is non-positive, assuming that $\kappa_3 = 8c_3 > 0$. The advantage in the central case is that the eigenvalues only

enter through $\mathrm{tr}\{(\Omega A)^s\}$ and therefore need not be calculated explicitly, whereas in the general noncentral case explicit calculation is required.

Additional methods to approximate noncentral distributions using central distributions are given in Cox and Reid (1987); see also Jensen and Solomon (1972) and Solomon and Stephens (1977) for other approximations via variable transformation. Jensen and Solomon (1994) noticed that such transformations give superior accuracy throughout the range of each distribution in comparison with other approximations, including Edgeworth series (Buckley and Eagleson, 1988).

Nevertheless, given the importance of statistics involving quadratic forms associated with normally distributed random variables, including their frequent use in applied statistics, there is undoubtedly a need for tractable, fast and accurate approximations to their distribution function.

3.3.3. Saddlepoint approximation

Saddlepoint methods give highly accurate approximations to density and distribution functions. By contrast with Pearson's approximation they use the entire cumulant generating function (3.2.4). As illustrated in Section 2.4 the saddlepoint approximation to the distribution of (3.2.2) can easily be obtained by means of (2.4.2). Setting $d = 1$, $T = Q(Y)$, $a_i = \lambda_i$ and $X_i = \chi^2_{h_i;\sigma_i^2}$, the saddlepoint approximation to the distribution of $Q(X)$ or $Q(Y)$ at q is

$$\Pr\{Q(X) > q\} \doteq 1 - F_s(q) = 1 - \Phi\left\{ w + \frac{1}{w}\log\left(\frac{v}{w}\right) \right\}, \tag{3.3.1}$$

where

$$w = \mathrm{sign}(\hat{\zeta})\left[2\left\{ \hat{\zeta}q - K(\hat{\zeta}) \right\} \right]^{1/2}, \qquad v = \hat{\zeta}\left\{ K''(\hat{\zeta}) \right\}^{1/2}$$

and $\hat{\zeta} = \hat{\zeta}(q)$, the saddlepoint, is the value of ζ satisfying the equation $K'(\hat{\zeta}) = q$, where $K'(\zeta)$ and $K''(\zeta)$ are the first and second derivatives of $K(\zeta)$ with respect to ζ,

$$K'(\zeta) = \sum_{i=1}^{n} \frac{h_i\lambda_i}{1 - 2\zeta\lambda_i} + 2\sum_{i=1}^{n} \frac{\sigma_i^2\lambda_i^2}{(1 - 2\zeta\lambda_i)^2}$$

and

$$K''(\zeta) = 2\left\{ \sum_{i=1}^{n} \frac{h_i\lambda_i^2}{(1 - 2\zeta\lambda_i)^2} + \sum_{i=1}^{n} \frac{\sigma_i^2(\lambda_i^3 - 2\zeta\lambda_i^2 + \lambda_i)}{(1 - 2\zeta\lambda_i)^3} \right\},$$

which in the central case reduce to

$$K'(\zeta) = \sum_{i=1}^{n} \frac{\lambda_i}{1 - 2\zeta\lambda_i}, \qquad K''(\zeta) = 2\sum_{i=1}^{n} \frac{\lambda_i^2}{(1 - 2\zeta\lambda_i)^2}.$$

The existence and uniqueness of $\hat{\zeta}$ follow from Daniels (1954). Calculation of (3.3.1) requires computation of $\hat{\zeta}$ for each q of interest. Even as $Q(X)$ is not in general a sum

Table 3.1. *Mean CPU time in seconds (over 100 attempts) of the calculation of the eigenvalues of the covariance matrix of a random multivariate standard Gaussian variate of size n.*

n	Time
10	0.008
50	0.034
100	0.158
200	0.935
300	3.291
400	6.752

of n independent identically distributed variables, it can be shown that under additional conditions (3.3.1) gives a relative error of $O(n^{-3/2})$ for the distribution (Jensen, 1995, Section 6.5). Remark also, though not of interest in the present chapter, that the saddlepoint approximation to the density of $Q(X)$ or $Q(Y)$ at q is a straightforward application of (2.4.1).

For the central case a similar result was obtained by Lieberman (1994) by means of a Lugannani–Rice approximation to the tail probability of a ratio of central quadratic forms in normal variables, which is usually indistinguishable in practice for this special case. The saddlepoint method's merits and its superiority over Edgeworth series have been extensively studied; see also Chapter 1. Therefore, there will be made no comparison between Buckley and Eagleson's (1988) Edgeworth expansion and the stated saddlepoint approximation.

Recall that (3.3.1) requires the explicit determination of all eigenvalues $\lambda_1, \ldots, \lambda_n$ of the $n \times n$ matrix ΩA or, equivalently, $\Gamma^T A \Gamma$. One may think that this step is numerically burdensome, and this already for sample sizes $n = 200$ (quite an ordinary situation) as all $n = 200$ eigenvalues must be evaluated. I agree that there is some additional cost in computational time due to the evaluation of all eigenvalues. But as Table 3.1 illustrates this additional cost in CPU time is negligible for small n. Simulations were made using S-PLUS 3.4 Release 1. Moreover, I would like to emphasise on the fact that the proposed saddlepoint approach is very easy to implement: in the statistical packages S-PLUS and R it takes only several lines of commands.

3.4. Comparison

n this section a numerical comparison between several methods discussed previously is given: Davies's method, Pearson's three-moment central chi-squared approximation and the saddlepoint approximation. Note that from the methods given in Section 3.3.2 I restrict consideration to Pearson's three-moment

central chi-squared approximation as this is the most widely used approximation in the statistical literature.

In an intensive numerical comparison study, I considered 39 different quadratic forms. For the 18 most interesting of them, Tables 3.2–3.4 record the values of (3.1.2) found by Davies's method. For the approximations the relative error compared to this method is given. The quadratic forms Q_1 to Q_8 are represented in the first column of Tables 3.2 and 3.3 by the triplet $\lambda_i, h_i, \sigma_i^2$, whereas Q_9 to Q_{18} in Tables 3.3 and 3.4 are represented by the triplet A, μ, Ω. These triplets determine the quadratic forms explicitly. F_n is the symmetric banded $n \times n$ matrix with bandwidth one, the diagonal is $(1, 2, \ldots, 2, 1)$ and the sub-diagonal consists of -1's (Farebrother, 1990),

$$
F_n = \begin{pmatrix}
1 & -1 & & & \\
-1 & 2 & -1 & & \\
& \ddots & \ddots & \ddots & \\
& & -1 & 2 & -1 \\
& & & -1 & 1
\end{pmatrix}.
$$

The $n \times n$ matrix S_n is a positive definite symmetric banded $n \times n$ matrix with bandwidth two, 7's on the diagonal, 2's in the first band and 1's in the second band,

$$
S_n = \begin{pmatrix}
7 & 2 & 1 & & \\
2 & 7 & 2 & \ddots & \\
1 & 2 & \ddots & \ddots & 1 \\
& \ddots & \ddots & 7 & 2 \\
& & 1 & 2 & 7
\end{pmatrix}.
$$

Finally, D_n is a positive definite symmetric banded $n \times n$ matrix with bandwidth one: 10's on the diagonal, -1's in the sub-diagonal, in the n-th line and in the n-th column,

$$
D_n = \begin{pmatrix}
10 & -1 & & & -1 \\
-1 & 10 & -1 & & \vdots \\
& \ddots & \ddots & \ddots & -1 \\
& & -1 & 10 & -1 \\
-1 & \cdots & -1 & -1 & 10
\end{pmatrix}.
$$

The quadratic forms Q_1 to Q_4 were already used by Imhof (1961, Table 1) and Q_5 and Q_6 were used by Farebrother (1990, Table 1).

The values of (3.1.2) given in Tables 3.2–3.4 are rounded to six decimal places. Pearson's three-moment central chi-squared approximation (ii) encounters problems for negative definite or indefinite quadratic forms, such as Q_7, Q_{10}, Q_{12}, Q_{14}, Q_{16} and Q_{18}, and its relative error can be large. The saddlepoint approximation (iii) yields very accurate approximations, even for non-positive quadratic forms. Even with n large, as in Q_{13}, Q_{14}, Q_{17} and Q_{18}, the saddlepoint approximation (iii) is more precise than (ii). In

recapitulating one remarks that the saddlepoint approximation (iii) outperforms Pearson's three-moment central chi-squared approximation (ii) and nears Davies's method (i) quite well, and this over the entire range of the distribution.

A summary of Tables 3.2–3.4, including the quadratic forms Q_2 to Q_{10}, Q_{12}, Q_{14} and Q_{18} for chosen values of q, is given in Table 1 of Kuonen (1999).

As mentioned previously, the saddlepoint approximation (3.3.1) is very easy to compute: in the statistical package S-Plus it takes only about 20 lines of commands, whereas Davies's method requires more than 200 lines of C code, and to have an estimate of desired accuracy one has to adjust all the input variables. Thus the saddlepoint approximation can be applied more easily, though there is some additional cost in CPU time because of the use of an interpreted language like S-Plus, and to the evaluation of all the eigenvalues; this is of the order of a few seconds for small n; see Table 3.1.

3.5. Application to nonparametric regression

uppose that n data points $(x_1, y_1), \ldots, (x_n, y_n)$ of dimension $(d+1)$ have been collected. Their regression relationship can be modelled as

$$y_i = m(x_i) + \epsilon_i, \quad i = 1, \ldots, n, \tag{3.5.1}$$

where $m(\cdot)$ is an unknown function and the ϵ_i are the observation errors with mean zero and variance σ^2. The idea of local regression is to fit locally a low-order polynomial at grid points of interest, with observations receiving different weights. The resulting local least squares estimator of $m(x)$ is

$$\hat{m}_H(x) = e_1^T (X_x^T W_x X_x)^{-1} X_x^T W_x y, \tag{3.5.2}$$

where H is a $d \times d$ symmetric positive matrix ($H^{1/2}$ is known as the bandwidth matrix), e_1 is the $(d+1) \times 1$ vector having one in the first entry and zeros otherwise, $y = (y_1, \ldots, y_n)^T$, $W_x = \text{diag}\{K_H(x_1 - x), \ldots, K_H(x_n - x)\}$ is the kernel weight matrix, where $K_H(u) = |H|^{-1/2} K(H^{-1/2} u)$ is a d-variate normalised kernel, and

$$X_x = \begin{pmatrix} 1 & (x_1 - x)^T \\ \vdots & \vdots \\ 1 & (x_n - x)^T \end{pmatrix}.$$

Note that $X_x^T W_x X_x$ is assumed to be non-singular. Equation (3.5.2) can be rewritten as $\hat{m}_H(x) = S_H y$, where S_H is the $n \times n$ smoothing matrix. Multivariate local regression estimators have proved to be very useful in modelling real data (Cleveland and Devlin, 1988). General expressions for the bias and the variance of $\hat{m}_H(\cdot)$ can be found in Ruppert and Wand (1994). Useful overviews of different approaches to the construction of nonparametric regression estimators are given by Hastie and Tibshirani (1990, Chapter 2) and Fan and Gijbels (1996, Chapter 2). For the local regression approach in particular, recent discussions are provided by Hastie and Loader (1993) and Cleveland and Loader (1995). An introductory account of these methods is given in Bowman and Azzalini (1997).

Table 3.2. *Probability that the quadratic form exceeds q. (i) Davies's method with accuracy 0.0001; (ii) relative error (%) for Pearson's three-moment central chi-squared approximation; (iii) relative error (%) for the saddlepoint approximation. The relative error (%) is* $100[\{\text{Approximation}-(\text{i})\}/(\text{i})]$.

Quadratic form	q	(i)	(ii)	(iii)
$Q_1 = 0.6, 1, 0; 0.3, 1, 0; 0.1, 1, 0$	0.01	0.99804	0.1967	−0.0028
	0.1	0.94579	5.7321	−0.0915
	0.7	0.50644	−3.1917	−1.3575
	1.4	0.23445	−2.2070	−0.5923
	2	0.12397	1.9861	−1.8681
	2.8	0.05469	3.4496	−0.5205
$Q_2 = 0.6, 2, 0; 0.3, 2, 0; 0.1, 2, 0$	0.2	0.99356	0.6477	−0.0072
	1	0.76039	−0.9564	−0.2445
	2.1	0.37177	−0.7521	−0.6872
	3	0.18697	0.8949	−0.7273
	4.5	0.05561	2.2458	−0.1369
	6	0.01612	0.7627	0.7094
$Q_3 = 0.6, 6, 0; 0.3, 4, 0; 0.1, 2, 0$	1	0.99732	0.1106	0.0010
	2	0.95056	0.1138	−0.0096
	3	0.81562	−0.1934	−0.0314
	6	0.28154	0.1684	−0.0548
	8	0.10051	0.6378	0.0080
	10	0.03112	0.3858	0.1319
$Q_4 = 0.6, 2, 0; 0.3, 4, 0; 0.1, 6, 0$	0.5	0.99829	0.1716	0.0001
	1	0.96664	1.0412	−0.0258
	2	0.72331	−1.2231	−0.2742
	4	0.21154	0.9924	−1.2604
	7	0.01986	1.3567	−0.6259
	8	0.00872	−1.8485	−0.0419
$Q_5 = 30, 1, 0; 1, 10, 0$	5	0.98461	1.5633	−0.0576
	15	0.70813	−4.5825	−2.8621
	25	0.48918	−1.7642	−1.0212
	50	0.25007	0.2658	0.1553
	75	0.14181	0.6307	0.1861
	100	0.08366	0.6854	0.2657
$Q_6 = 30, 1, 0; 1, 20, 0$	10	0.99508	0.4949	−0.0098
	25	0.71362	−6.2191	−1.3132
	40	0.42675	−1.9818	−0.9419
	75	0.17789	1.0506	0.9676
	100	0.10357	1.3484	1.1522
	125	0.06192	1.2832	1.0701

Table 3.3. *Probability that the quadratic form exceeds q. (i) Davies's method with accuracy 0.0001; (ii) relative error (%) for Pearson's three-moment central chi-squared approximation; (iii) relative error (%) for the saddlepoint approximation. The relative error (%) is $100[\{\text{Approximation} - (i)\}/(i)]$.*

Quadratic form	q	(i)	(ii)	(iii)
$Q_7 = -0.6, 2, 0; 0.3, 4, 0; 0.1, 6, 0$	-4	0.99002	1.0084	-0.0262
	-2	0.94714	4.1574	-0.1305
	-1	0.87836	-1.2578	-0.3127
	1	0.40199	-12.3782	3.1172
	2	0.14065	20.0263	3.6625
	4	0.00979	168.4783	1.8132
$Q_8 = 0.6, 1, 0.1; 0.3, 2, 0.2; 0.1, 1, 0.2$	0.01	0.99983	0.0169	0.0038
	0.1	0.98834	1.1796	0.0006
	0.4	0.86944	1.1434	0.3374
	0.7	0.71109	-1.7122	1.1808
	3.2	0.07751	3.3471	1.3629
	4	0.03751	3.9922	24.9562
$Q_9 = F_5; 0_5; I_5$	0.1	0.99945	0.0553	0.0004
	0.5	0.98811	1.2037	-0.0171
	5	0.59487	-1.7681	-0.5186
	15	0.12843	1.3698	-1.6413
	25	0.02692	2.1819	-1.8041
	50	0.00061	-11.7038	1.5783
$Q_{10} = -F_5; 0_5; I_5$	-25	0.97308	2.7667	0.0499
	-15	0.87157	13.6302	0.2418
	-10	0.71949	-29.7161	0.4482
	-5	0.40513	-40.7814	0.7615
	-1	0.04101	217.4345	1.3391
	-0.5	0.01189	913.4563	1.4228
$Q_{11} = S_7; 0_7; I_7$	1.5	0.99998	0.0021	-0.0022
	15	0.94309	0.7716	-0.0231
	25	0.80876	-0.4857	-0.1226
	50	0.40241	-0.8474	-0.7777
	100	0.05669	2.6029	-2.3044
	200	0.00071	-11.2303	2.8844
$Q_{12} = -S_7; 0_7; I_7$	-200	0.99932	0.0702	-0.0021
	-100	0.94331	6.0094	0.1385
	-50	0.59759	-28.684	0.5229
	-25	0.19124	-9.2208	0.5186
	-15	0.05691	105.7646	0.3833
	-5	0.00226	$3,341.0644$	-0.7139

Table 3.4. *Probability that the quadratic form exceeds q.   (i) Davies's method with accuracy 0.0001; (ii) relative error (%) for Pearson's three-moment central chi-squared approximation; (iii) relative error (%) for the saddlepoint approximation. The relative error (%) is $100[\{\text{Approximation}-(\text{i})\}/(\text{i})]$.*

Quadratic form	q	(i)	(ii)	(iii)
$Q_{13} = S_{15}; 0_{15}; I_{15}$	15	0.99993	0.0068	-0.0009
	35	0.98913	0.2732	-0.0010
	55	0.91428	0.0467	-0.0258
	80	0.69765	-0.5486	-0.1200
	100	0.48937	-0.4479	-0.2407
	200	0.02778	0.8413	-0.7076
$Q_{14} = -S_{15}; 0_{15}; I_{15}$	-300	0.99924	0.0761	0.0022
	-200	0.97222	2.8572	0.0259
	-150	0.86284	1.9381	0.0951
	-100	0.51063	-22.9427	0.2311
	-50	0.05801	77.4904	0.2815
	-20	0.00045	$8,624.5016$	2.0988
$Q_{15} = D_5; 0_5; I_5$	1	0.99988	0.0202	0.0032
	15	0.91098	0.0867	-0.0039
	30	0.69616	-0.2007	0.0003
	100	0.07701	0.4541	-0.0539
	130	0.02482	0.3833	-0.0752
	200	0.00152	-2.7739	-0.803
$Q_{16} = -D_5; 0_5; I_5$	-300	0.99998	0.0020	-0.0003
	-200	0.99848	0.1523	0.0012
	-100	0.92299	8.3439	0.0045
	-30	0.30390	-27.2578	-0.0007
	-10	0.03852	188.2950	0.1021
	-1	0.00020	$39,624.7886$	-15.8028
$Q_{17} = D_{10}; 0_{10}; I_{10}$	10	0.99983	0.0073	-0.0013
	50	0.88761	-0.0118	-0.0002
	75	0.67286	-0.1223	-0.0023
	125	0.25378	0.0810	-0.0263
	150	0.13450	0.2427	-0.0497
	200	0.03134	0.2260	-0.1196
$Q_{18} = -D_{10}; 0_{10}; I_{10}$	-300	0.99891	0.1087	-0.0001
	-200	0.96866	3.2351	0.0039
	-150	0.86550	2.5433	0.0077
	-50	0.11239	19.9697	0.0020
	-30	0.01958	294.6736	0.0708
	-10	0.00017	$24,457.3681$	0.0078

3.5.1. Testing for no effect

Suppose that one wants to compare the hypotheses

$$
\begin{aligned}
H_0 &: \quad \mathrm{E}(y_i) = \mu, \\
H_1 &: \quad \mathrm{E}(y_i) = m(x_i),
\end{aligned}
$$

known as usual hypotheses for 'testing for no effect', *i.e.* H_0 posits no effect. The standard approach from classical linear models was extended by Azzalini *et al.* (1989) to the nonparametric setting. The residual sums of squares offer a natural means of quantifying the extent to which these models explain the data. These are, under the hypothesis H_0,

$$
\mathrm{RSS}_0 = \sum_{i=1}^{n}(y_i - \bar{y})^2,
$$

and, under the hypothesis H_1,

$$
\mathrm{RSS}_1 = \sum_{i=1}^{n}\{y_i - \hat{m}_H(x_i)\}^2.
$$

Azzalini *et al.* (1989) were then led to the so-called *pseudolikelihood ratio test statistic*,

$$
T = \frac{\mathrm{RSS}_0 - \mathrm{RSS}_1}{\mathrm{RSS}_1}, \tag{3.5.3}
$$

which is proportional to the usual F statistic, but the fact that the test statistic is derived from a likelihood argument does not imply that its distribution is approximately chi-squared. A selected review on the pseudolikelihood ratio test approach is given in Azzalini (1998). It is useful to express the structure of T in terms of quadratic forms as

$$
\begin{aligned}
\mathrm{RSS}_0 &= y^T(I_n - L)^T(I_n - L)y = y^T(I_n - L)y, \\
\mathrm{RSS}_1 &= y^T(I_n - S_H)^T(I_n - S_H)y,
\end{aligned}
$$

where L is the $n \times n$ matrix with all its entries equal to one (Bowman and Azzalini, 1997, Section 5.2). An explicit form of (3.5.3) in terms of quadratic forms is

$$
T = \frac{y^T B y}{y^T C y}, \tag{3.5.4}
$$

where $C = (I_n - S_H)^T(I_n - S_H)$ and B is the matrix $I_n - n^{-1}L - C$. Unfortunately, standard results from linear models do not apply because the matrices C and B do not have the necessary properties, such as positive definiteness. The corresponding p-value can be written as

$$
\Pr(T > t \mid H_0) = \Pr\left(y^T U y > 0\right) = \Pr\left(\epsilon^T U \epsilon > 0\right), \tag{3.5.5}
$$

where $U = B - tC$, t is the observed value of T and $\epsilon = (\epsilon_1, \ldots, \epsilon_n)^T$ is the error vector. If the ϵ_i are normally distributed one can set $\sigma^2 = 1$ without loss of generality as T
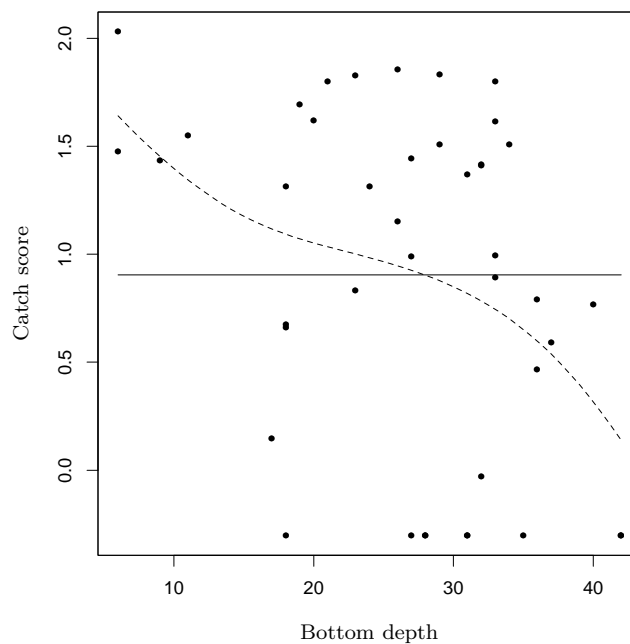
Figure 3.1. *A plot of the relationship between bottom depth and catch score in a subsample of the Great Barrier Reef data. The nonparametric regression curve with $h = 8$, dashed line, and the curve of no effect, solid line, have been superimposed.*

is scale-invariant. Hence, setting $A = U$ and $\Omega = I_n$, the covariance matrix of ϵ, the p-value (3.5.5) can be calculated easily in using one of the methods discussed in Section 3.3.

In order to compare their performances, consider as an application a subsample of the Great Barrier Reef data (Bowman and Azzalini, 1997, page 52). These data refer to a survey of the fauna on the sea bed lying between the coast of northern Queensland and the Great Barrier Reef. The sampling region covered a zone which was closed to commercial fishing in 1993. Two of its variables are the bottom depth, x, and the catch score, y. For details of the survey and an analysis of the data see the references given in Bowman and Azzalini (1997). For these data we have $d = 1$, therefore H becomes the scalar smoothing parameter h. The following calculations were simplified in using the S-Plus library `sm` associated with Bowman and Azzalini (1997). A plot of the relationship between these two variables is given in Figure 3.1. The figure also displays the nonparametric regression curve with $h = 8$ and the curve of no effect. Note that Kuonen (1999, Figure 1) used $h = 5$. Table 3.5 gives the p-value (3.5.5) as function of the smoothing parameter h. The values in column (ii) correspond to the 'significance trace' given in Bowman and Azzalini (1997, Figure 5.1), reproduced in Figure 3.2 by superimposing Davies's method and the saddlepoint approximation. A plot of the type shown in Figure 3.2 was introduced by Azzalini and Bowman (1991). Its usefulness depends on the essential stability of the p-value as a function of h. Bowman and Azzalini (1997) stated that their approach is sufficiently accurate to approximate the p-value. But, Table 3.5 and Figure 3.2 indicate that Pearson's three-moment central

Table 3.5. *The p-value as a function of the smoothing parameter h for testing a relationship between catch score and bottom depth in the Great Barrier Reef data. (i) Davies's method with accuracy 0.0001; (ii) relative error (%) for Pearson's three-moment central chi-squared approximation; (iii) relative error (%) for the saddlepoint approximation. The relative error (%) is* $100[\{\text{Approximation} - \text{(i)}\}/\text{(i)}]$.

h	(i)	(ii)	(iii)
3	0.0604	5.433	0.028
5	0.0633	7.737	0.135
7	0.0516	7.545	0.019
9	0.0407	6.765	0.738
11	0.0334	5.890	1.192
13	0.0289	4.867	1.288
15	0.0258	4.164	0.240
17	0.0234	3.460	1.431

PSfrag replacements
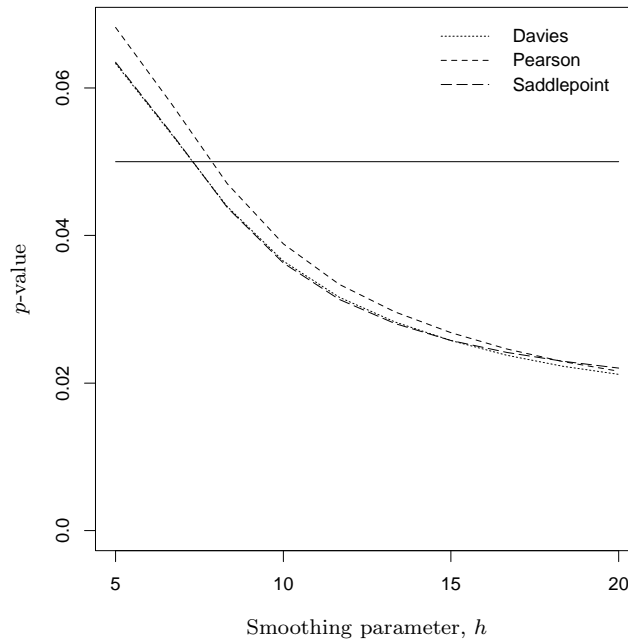
Figure 3.2. *The significance trace to assess the evidence between bottom depth and catch score in the Great Barrier Reef data.*

chi-squared approximation overestimates the exact p-value and therefore the rejection region of H_0 decreases compared to the exact computations given by Davies's exact method. The saddlepoint approximation to the p-value is extremely accurate. In fixing a significance level of 5%, the p-values recorded in Table 3.5 suggest that there exists a relation between bottom depth and catch score for $h > 7$, whereas for $h \leq 7$ there is
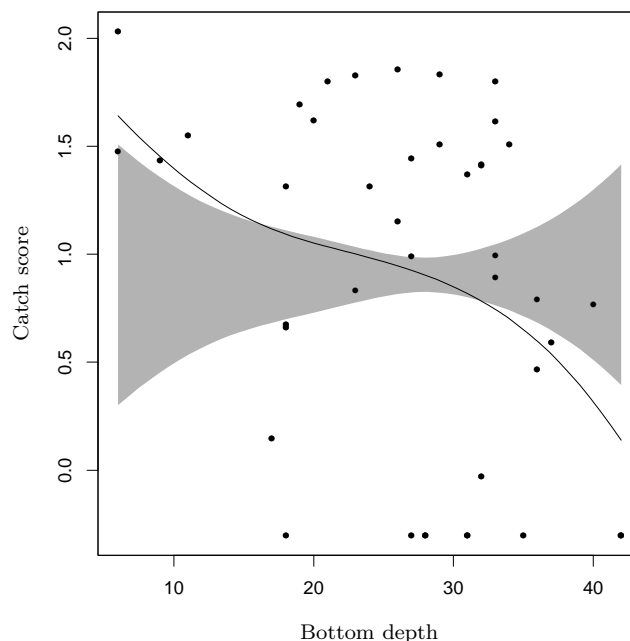
Figure 3.3. *Reference band for the relationship between bottom depth and catch score in a subsample of the Great Barrier Reef data. The nonparametric regression curve with $h = 8$ has been superimposed.*

no significance for a relation. This may result from the fact that when h increases the nonparametric local regression curve becomes smoother and differs from the fitted linear regression curve significantly. In addition to this formal model comparison, providing a useful global assessment, variability bands as described in Bowman and Azzalini (1997, pages 89 and 90) can be used in this context to provide a helpful graphical follow-up. Figure 3.3 displays a reference band for the relationship between bottom depth and catch score. One remarks that the nonparametric regression curve with $h = 8$ exceeds the band at each end of the bottom depth range, illustrating that there is some evidence for a relationship between the two variables.

In practice it is useful to have some idea of the power of a test. Hence, it may be interesting to carry out a small power study to compare the performances of the approximation in more details. To do so the observables y_i, $i = 1, \ldots, n$, are assumed to be generated from $y_i = \hat{m}_h(x_i) + e_i$, where $\hat{m}_h(x_i)$ is the local least squares estimate (3.5.2) evaluated at x_i, the e_i's are drawn independently from the normal distribution $N(0, \eta^2)$. Various combinations of h and η were considered and in each case 500 samples were generated. Tables 3.6 and 3.7 present the percentages of times H_0 was rejected for significance level 5% and some selected values of h and η. A summary of these tables is given in Table 3 of Kuonen (1999). First of all one remarks that the pseudo-likelihood ratio test is very powerful in the context presented in this section and did not have any difficulties to check the hypotheses — even in generating under H_1. This was already suggested by Table 3.5, where H_0 has been rejected by a significant factor. Moreover, one notices that the power decreases when η gets large. This may result from the fact that large η's imply relevant noise and therefore difficulties in distinguishing the two

Table 3.6. *The simulated power as a function of h and η for testing a relationship between catch score and bottom depth in the Great Barrier Reef data. (i) Davies's method with accuracy 0.0001; (ii) Pearson's three-moment central chi-squared approximation; (iii) saddlepoint approximation.*

h	η	(i)	(ii)	(iii)
0.1	2	0.958	0.958	0.958
	1	0.946	0.940	0.946
	0.5	0.932	0.928	0.932
	0.25	0.940	0.936	0.940
	0.1	0.942	0.942	0.942
0.5	2	0.958	0.956	0.958
	1	0.940	0.940	0.940
	0.5	0.960	0.960	0.960
	0.25	0.940	0.936	0.940
	0.1	0.966	0.964	0.966
1	2	0.946	0.944	0.946
	1	0.934	0.934	0.934
	0.5	0.946	0.944	0.946
	0.25	0.954	0.950	0.954
	0.1	0.954	0.954	0.954
5	4	0.956	0.954	0.956
	2	0.934	0.930	0.934
	1	0.940	0.938	0.940
	0.5	0.934	0.934	0.934
	0.25	0.948	0.944	0.948
	0.1	0.950	0.950	0.950
	0.01	0.930	0.926	0.930
	0.0001	0.944	0.942	0.944

models. The saddlepoint approximation (iii) gives exactly the same power as for the exact approach (i), whereas Pearson's three-moment central chi-squared approximation (ii) gives slightly smaller estimated power.

3.5.2. Testing for a linear relationship

The objective of nonparametric smoothing is to relax assumptions on the form of an unknown function of interest and to 'let data speak for themselves'. These approaches can be combined with parametric methods to yield a sensible data analysis as in the problem of testing a linear relationship between a single covariate $x = (x_1, \ldots, x_n)^T$ and

Table 3.7. *The simulated power as a function of h and η for testing a relationship between catch score and bottom depth in the Great Barrier Reef data. (i) Davies's method with accuracy 0.0001; (ii) Pearson's three-moment central chi-squared approximation; (iii) saddlepoint approximation.*

h	η	(i)	(ii)	(iii)
7	2	0.940	0.938	0.940
	1	0.946	0.946	0.946
	0.5	0.934	0.934	0.934
	0.25	0.932	0.928	0.932
	0.1	0.954	0.954	0.954
9	2	0.948	0.948	0.948
	1	0.946	0.946	0.946
	0.5	0.950	0.948	0.950
	0.25	0.956	0.952	0.956
	0.1	0.928	0.924	0.928
11	2	0.944	0.938	0.944
	1	0.950	0.942	0.950
	0.5	0.960	0.958	0.960
	0.25	0.942	0.940	0.942
	0.1	0.932	0.930	0.932
15	2	0.950	0.944	0.948
	1	0.944	0.940	0.944
	0.5	0.942	0.940	0.940
	0.25	0.958	0.956	0.958
	0.1	0.934	0.930	0.934

a response variable $y = (y_1, \ldots, y_n)^T$, related by (3.5.1). Eubank and Spiegelman (1990) discuss some general issues of using nonparametric smoothing in the context of checking a linear relationship. Suppose that one wants to compare the hypotheses

$$H_0: \quad m(x_i) = \alpha + \beta x_i,$$
$$H_1: \quad m(x_i) \text{ is a smooth function.}$$

The pseudolikelihood ratio test statistic (3.5.3) becomes

$$T = \frac{y^T V y - y^T C y}{y^T C y},$$

where $C = (I_n - S)^T (I_n - S)$ and $V = I_n - P$, where P denotes the least squares projection matrix. Since $(I_n - S)^T 1_n$ is the null vector, the distribution of T is free from α, but this does not hold for β. Therefore the dependence of the distribution of

T on the unknown value of β makes T unsuitable for hypothesis testing. In denoting by $e = (e_1, \ldots, e_n)^T$ the residual vector from the fitted simple linear regression model, Azzalini and Bowman (1993) represent the data by the residual vector $e = Vy$. The formal hypotheses can be written as

$$H_0 : \ \ \mathrm{E}(e) = 0,$$
$$H_1 : \ \ \mathrm{E}(e) \text{ is a smooth function of } x.$$

Hence

$$T = \frac{e^T B e}{e^T C e},$$

where $B = I_n - C$ and C as before. The corresponding p-value becomes

$$\Pr(T > t \mid H_0) = \Pr\left(e^T U e > 0\right),$$

where $U = B - tC = I_n - (1 + t)C$ and t is the observed value of the statistic T. In the present setting the e_i are correlated, whereas in the previous section independence was assumed. But, the methods discussed still apply as one remarks that the problem is again reduced to the computation of the distribution of a central quadratic form in normal variates.

As a first application, consider another example from Bowman and Azzalini (1997), dealing with the relationship between longitude and catch score for the subsample of the Great Barrier Reef data, where longitude is an additional variable denoting the longitude of the sampling position. A scatter plot of longitude against catch score is shown in Figure 3.4, together with the nonparametric regression curve using $h = 0.1$ and a fitted linear regression curve. The corresponding p-values are recorded in Table 3.8. Once again one remarks that the values obtained by the saddlepoint approximation are closer to the exact ones, whereas Pearson's approximation underestimates the exact p-value; see also the corresponding significance trace in Figure 3.5. Apart from very small values of the smoothing parameter 0.02 we have strong evidence against the linear model for all four methods. This seems to underline the feeling one gets in looking at Figure 3.4.

To conclude this section, consider the radiocarbon data (Bowman and Azzalini, 1997). These data record high precision measurements for the range of ages between 3000 and 5000 years of radiocarbon on Irish oak, used to construct a calibration curve. As covariate, x, consider the true calendar age and as response variable, y, the age predicted from the radiocarbon dating process. Once again, the aim is to test a linear relationship between the two variables. A plot of their relationship is given in Figure 3.6. Superposed are the fitted regression line and a nonparametric smoother with $h = 6$. From the significance trace in Figure 3.7 one notes that in considering the exact and the saddlepoint approach there is a convincing evidence of non-linearity at all values of h. But, Pearson's three-moment approximation produces insignificance for small values of h.
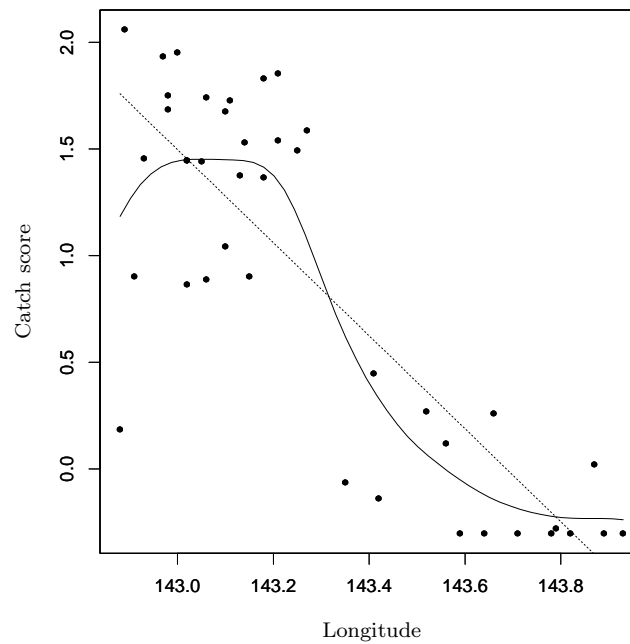
Figure 3.4. *A plot of the relationship between longitude and catch score in a subsample of the Great Barrier Reef data. The nonparametric regression curve with $h = 0.1$, solid line, and a fitted linear regression curve, dotted line, have been superimposed.*

Table 3.8. *The p-value as a function of the smoothing parameter h for testing a linear relationship between longitude and catch score in the Great Barrier Reef data. (i) Davies's method with accuracy 0.0001; (ii) Pearson's three-moment central chi-squared approximation; (iii) saddlepoint approximation.*

h	(i)	(ii)	(iii)
0.02	0.11289	0.11838	0.11288
0.04	0.00905	0.01756	0.00905
0.06	0.00351	0.00633	0.00351
0.08	0.00305	0.00374	0.00305
0.10	0.00427	0.00284	0.00427
0.12	0.00612	0.00225	0.00613
0.14	0.00824	0.00178	0.00826
0.16	0.01096	0.00144	0.01099
0.18	0.01474	0.00119	0.01476
0.20	0.02006	0.00102	0.02006

3.5.3. Other applications

The approach presented in Sections 3.5.1 and 3.5.2 can be applied easily to a number of other cases; see Azzalini (1998, Section 2.3 and Chapter 3). A further application is a
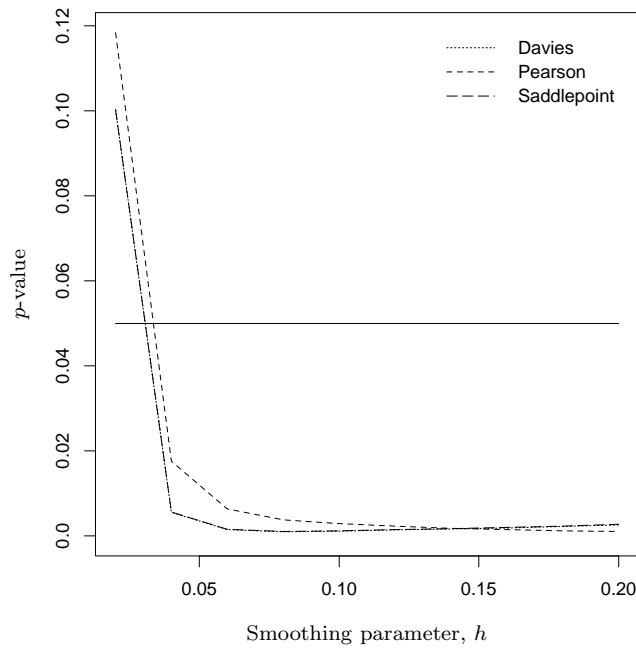
Figure 3.5. *The significance trace to assess a linear relationship between longitude and catch score in the Great Barrier Reef data.*
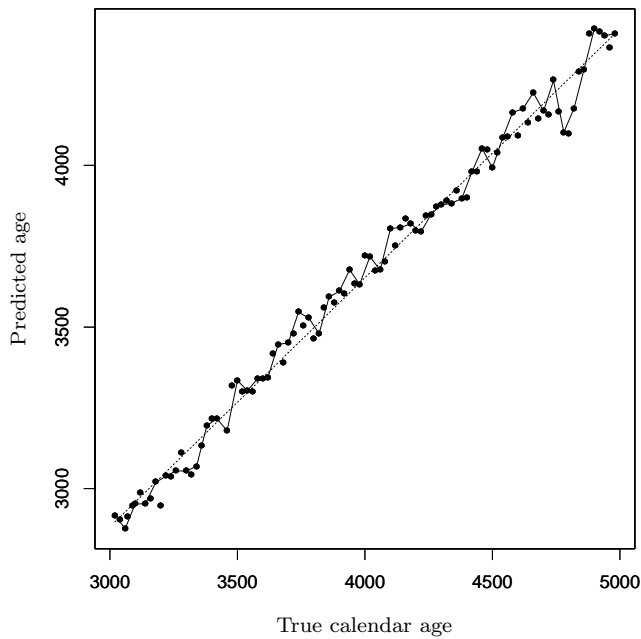

Figure 3.6. *A plot of the relationship between the true calendar age and the age predicted from the radiocarbon dating process in the radiocarbon data. The nonparametric regression curve with $h = 6$, solid line, and a fitted linear regression curve, dotted line, have been superimposed.*

test for constant residual variance in a linear regression scheme (Diblasi and Bowman, 1997).
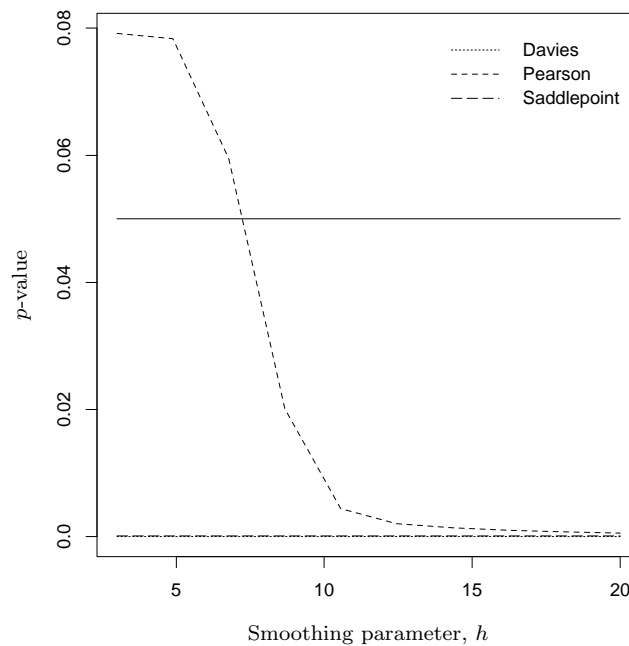
Figure 3.7. *The significance trace to assess a linear relationship between the true calendar age and the age predicted from the radiocarbon dating process in the radiocarbon data.*

3.6.   Conclusion

t has been illustrated in this chapter that the saddlepoint approximation is a method of calculating the distribution of quadratic forms in normal variates in an extremely accurate way; see also Kuonen (1999). This statement has been underlined in considering as example the distribution of the pseudolikelihood test statistic in nonparametric regression analysis. Moreover, the saddlepoint approximation is very easy to implement compared to the exact methods and outperforms existing approximations significantly.

4. The coupon collector's problem

'There was a free gift in my breakfast cereal this morning. I just called the manufacturers and they told me that the gift comes in ten different colors, and they encouraged me to collect all ten, and so eat lots of their cereal. Assuming there is an equal chance of getting any one of the colors, how many boxes I must consume to be 95% sure to get all ten gifts?'

In this chapter I show that a saddlepoint approximation to the distribution of the waiting time in the coupon collector's problem is very accurate. This nice application of the saddlepoint approximation illustrates the modern art of statistics relying on the computer and embodying both numeric and analytic approximations.

The main results derived in this chapter appeared in *The American Statistician* (Kuonen, 2000a), and a review was given at the SRSS meeting (Kuonen, 2000e). The example on gene transfer in Section 4.5 results from a collaboration with the Laboratory of Cellular Biotechnology, EPFL.

4.1. Introduction

A standard combinatorial problem is to estimate the number of coupons, drawn at random with replacement, needed to complete a collection. This problem, known as the coupon collector's problem or the classical occupancy problem, has been investigated intensively; see Feller (1970), Johnson *et al.* (1994), Port (1994) or Read (1998). Attention is given to the special case where all sampling probabilities are equal. For the general case, some results and ideas will be provided as well.

Section 4.2 summaries the underlying basic theory, including existing approximations to the distribution of the waiting time until all coupons are sampled. A saddlepoint approximation to its distribution is derived in Section 4.3, and in Section 4.4 several examples are considered, including the exhaustive bootstrap for symmetric statistics seen from a collector's point of view. An additional example in the context of gene transfer in order to introduce several genes into cells is given in Section 4.5.

4.2. Basic theory

uppose there are n distinct coupons in a collection, and a series of independent draws is made from these. At each draw a coupon is drawn with replacement with probability n^{-1} from the collection. Because of repetitions a random sample of size n will in general contain fewer than n distinct coupons. As the sample size increases, new elements will enter the sample more and more rarely. The interest is in the number of draws, W, needed for all n coupons to have been drawn at least once.

For $w \geq n$, as before it is not possible to have all n coupons, the distribution function of the waiting time W is (Feller, 1970, pages 59 and 61; Port, 1994, pages 301–304; Read, 1998, equation (5))

$$\Pr(W \leq w) = \sum_{i=0}^{n} \binom{n}{i} (-1)^i \left(1 - \frac{i}{n}\right)^w. \tag{4.2.1}$$

Good (1957, 1961) provided a recursion formula for (4.2.1). A direct numerical evaluation of (4.2.1) is limited to the case of relatively small n and w, due to the successive additions and subtractions in (4.2.1), and due to computational limits. But (4.2.1) is an immediate consequence of Bonferroni's inequalities (Feller, 1970, pages 110 and 111) and therefore enjoys the property that the error incurred by truncating the sum after any number of terms has the sign of the first omitted term and is smaller in absolute value. Nevertheless, to overcome such difficulties, approximate methods have to be considered. Baum and Billingsley (1965) observed that W has the same distribution as a sum of independent geometric random variables. By adapting an argument used by Baum and Billingsley (1965), Dawkins (1991) showed that asymptotically

$$\Pr(W \leq w) \approx \exp[-n \exp(-w/n)]. \tag{4.2.2}$$

Read (1998) gives a log-normal approximation to the distribution of W, matching the first three moments of W.

Another method is given by the multinomial occupancy probability

$$\Pr(W \leq w) = \Pr(c_1 \geq 1, \ldots, c_n \geq 1), \tag{4.2.3}$$

where $(c_1, \ldots, c_n)$ has a multinomial distribution with denominator w and probability vector $(n^{-1}, \ldots, n^{-1})$, and c_i denotes the number of times coupon i appears in the sample of size w. Levin (1981) discussed how to calculate multinomial cumulative distribution functions in general and showed that for (4.2.3) in the equiprobable case,

$$\Pr(c_1 \geq 1, \ldots, c_n \geq 1) = \frac{w![1 - \exp(-\nu/n)]^n}{\nu^w \exp(-\nu)} \Pr(S = w), \tag{4.2.4}$$

where the parameter ν is a tuning parameter which may be chosen for convenience and stable computation, and S is a sum of n independent left-truncated Poisson variables

each taking values $j = 1, 2, \ldots$ with probability $\{\exp(-\nu/n)/[1-\exp(-\nu/n)]\}(\nu/n)^j/j!$. The untruncated Poisson variables have mean equal to ν/n, and the truncated variables have mean $\mu = (\nu/n)/[1 - \exp(-\nu/n)]$. Levin (1981) found that the choice $\nu = w$ to be generally satisfactory and very natural in this context as w/n are the expected cell frequencies. Moreover, he noticed that the exact value of (4.2.4) can be stably computed by obtaining the convolution $\Pr(S = w)$, and Levin (1981, 1992) gave for $\nu = w$ the Edgeworth approximation to order n^{-1} for $\Pr(S = w)$,

$$\Pr(S = w) \approx l\left(\frac{w - n\mu}{\sqrt{n\sigma^2}}\right)\frac{1}{\sqrt{n\sigma^2}}, \tag{4.2.5}$$

where

$$l(x) = (2\pi)^{-1/2}\exp[-1/(2x^2)][1 \;\; + \;\; (\gamma_1/6)(x^3 - 3x) + (\gamma_2/24)(x^4 - 6x^2 + 3)$$
$$+ \;\; (\gamma_1^2/72)(x^6 - 15x^4 + 45x^2 - 15)],$$

and where the skewness, $\gamma_1 = n^{-1/2}\mu_3/\sigma^3$, and the kurtosis, $\gamma_2 = n^{-1}[(\mu_4/\sigma^4) - 3]$, are defined in terms of the moments of the truncated Poisson distribution: $\mu = \mu_{(1)} = (w/n)/[1 - \exp(-w/n)]$, $\sigma^2 = \mu_{(2)} + (\mu - \mu^2)$, $\mu_3 = \mu_{(3)} + \mu_{(2)}(3 - 3\mu) + (\mu - 3\mu^2 + 2\mu^3)$, and $\mu_4 = \mu_{(4)} + \mu_{(3)}(6 - 4\mu) + \mu_{(2)}(7 - 12\mu + 6\mu^2) + (\mu - 4\mu^2 + 6\mu^3 - 3\mu^4)$, using the rth factorial moments $\mu_{(r)} = (-w/n)^r/[1 - \exp(-w/n)]$. Combining (4.2.5) and (4.2.4) yields Levin's Edgeworth approximation for $\nu = w$,

$$\Pr(W \leq w) \approx \frac{w![1 - \exp(-w/n)]^n}{w^w\exp(-w)} \;\; l\left(\frac{w - n\mu}{\sqrt{n\sigma^2}}\right)\frac{1}{\sqrt{n\sigma^2}}, \tag{4.2.6}$$

where $l(\cdot)$, μ and σ^2 as before. Levin (1992) made a comparison of (4.2.6) with (4.2.1) and (4.2.2).

4.3. Saddlepoint approximation

The waiting time W has the same distribution as a sum of n independent geometric random variables with probabilities $(n - i + 1)/n$, $i = 1, \ldots, n$. Hence the exact cumulant generating function of W is (Read, 1998, equation (8))

$$K(\zeta) = -\sum_{i=1}^{n} \log\left\{1 - \frac{n[1 - \exp(-\zeta)]}{i}\right\}. \tag{4.3.1}$$

The saddlepoint approximation to the distribution of W can easily be obtained by means of the results illustrated in Section 2.4. Setting $d = 1$, $T = W$, $a_i = 1$ and X_i being the geometric random variates it yields that the saddlepoint approximation to the distribution of W at w, $\Pr(W \leq w)$, is given by (2.4.2). The saddlepoint $\hat{\zeta} = \hat{\zeta}(w)$ is the unique value of ζ satisfying the equation $K'(\hat{\zeta}) = w$, where prime denotes differentiation with respect to ζ. The first two derivatives of $K(\zeta)$ are

$$K'(\zeta) = n\exp(-\zeta)\sum_{i=1}^{n}\frac{1}{i - n[1 - \exp(-\zeta)]}$$

and

$$K''(\zeta) = n \exp(-\zeta) \sum_{i=1}^{n} \frac{n-i}{\{i - n[1 - \exp(-\zeta)]\}^2}.$$

This approximation has a relative error up to $O(n^{-3/2})$ and as the error is relative the ratio of the true distribution to its saddlepoint approximation is bounded over the likely range of w. If w is close to $\mathrm{E}\{W\} = n \sum 1/i$ (Read, 1998, equation (6)), or similarly $\hat{\zeta} \approx 0$, the approximation (2.4.2) is replaced by its limit as $\hat{\zeta} \to 0$, which as a straightforward application of the results in Jensen (1995, Section 3.3) yields

$$\Pr(W \le w) \approx \Phi\left\{\lambda_3/6\right\},$$

where

$$\lambda_3 = \frac{K'''(0)}{\{K''(0)\}^{3/2}} = \frac{\displaystyle\sum_{i=1}^{n}(2n^2 - 3ni + i^2)/i^3}{\sqrt{n}\left\{\displaystyle\sum_{i=1}^{n}(n-i)/i^2\right\}^{3/2}}.$$

The existence of a solution to $K'(\hat{\zeta}) = w$ has been demonstrated by Daniels (1954), and the uniqueness of $\hat{\zeta}$ will be demonstrated in what follows. There exists an unique solution to $K'(\zeta) = w$, or equivalently an unique minimum to the function $g(\zeta) = K(\zeta) - w\zeta$, as the function $g(\cdot)$ is strictly convex: $g''(\zeta) = K''(\zeta) > 0$. This proves the uniqueness of a solution, $\hat{\zeta}$, to $K'(\hat{\zeta}) = w$; provided that the cumulant generating function exists. The cumulant generating function (4.3.1) can be written as

$$K(\zeta) = n\zeta - \sum_{i=1}^{n-1} \log\left\{\frac{n - i\exp(\zeta)}{n - i}\right\},$$

which is defined only if $[n - i\exp(\zeta)]/(n-i) > 0$, $i = 1, \ldots, n-1$, or similarly $\exp(\zeta) < n/i$, $i = 1, \ldots, n-1$. This last condition yields that $\zeta < \log\{n/(n-1)\}$, which for example with $n = 3$ becomes $\zeta < 0.4054$, and tends to 0 as $n \to \infty$. Moreover, when $\zeta < \log\{n/(n-1)\}$, the moment generating function (Read, 1998, page 176) is also defined,

$$M(\zeta) = (n-1)!\exp(n\zeta)\left\{\prod_{i=1}^{n-1}(n - i\exp(\zeta))\right\}^{-1}.$$

Hence, one must only consider values of ζ for which the cumulant generating function exists. In considering only the values where $K(\zeta)$ exists, $\zeta < \log\{n/(n-1)\}$, uniqueness and existence of $\hat{\zeta}$ are guaranteed. For example, using $n = 3$ and $w = 4.5$, the equation $K'(\hat{\zeta}) = w$ is solved for $\hat{\zeta}_1 = -0.2253$ and $\hat{\zeta}_2 = 0.8821$, but as $\hat{\zeta}_2 > 0.4054$ only $\hat{\zeta}_1$ is an admissible solution.

Since the distribution of W, (4.2.1), is discrete, and the saddlepoint approximation is continuous, a better approximation is given by writing $\Pr(W = w)$ as $\Pr(w-1/2 < W < w+1/2)$. This continuity correction implies that $\Pr(W \le w) = \Pr(W < w+1/2)$. Read

(1998) made the same correction for his log-normal approximation, as well as Dawkins (1991) for (4.2.2).

The calculation of (4.2.1) requires computation of $\hat{\zeta}$ for each w of interest. As noted previously, solving the saddlepoint equation $K'(\hat{\zeta}) = w$ is equivalent to finding the minimum of the function $K(\hat{\zeta}) - w\hat{\zeta}$. This can be done using packaged routines such as the S-Plus function `nlmin`, which finds a local minimum using a general quasi-Newton optimizer, with starting value near zero and smaller than $\log\{n/(n-1)\}$. The resulting function `saddle.collector`, which takes only a few lines of code in the statistical package S-Plus, is given in Table 4.1. Consider the first example of Section 4.4: $n = 365$ and $w = 1900$. For $\Pr(W \leq 1900)$ we get

```
> saddle.collector(n=365, w=1900+.5)
[1] 0.1338624
```

This computation may appear intensive and perhaps it was some decades ago, but this obstacle is eliminated by the computing power available today. This fact illustrates the inelegant drawback of saddlepoint approximations: they embody both numeric and analytic approximations. As an example, the mean CPU time in seconds (over 20 attempts) of the calculation of $\Pr(W \leq 1,088,812)$ with $n = 92,378$ was 5.92 seconds using S-Plus 3.4 Release 1. For $\Pr(W \leq 4)$ with $n = 3$ the mean CPU time in seconds (over 20 attempts) was 0.02 seconds.

In the next section, the saddlepoint approximation (2.4.2) will be applied to several examples, and compared to (4.2.1), (4.2.2), to the log-normal approximation (Read, 1998) and to Levin's Edgeworth expansion (4.2.6).

4.4. Examples

As a first example, consider the problem proposed by Feller (1970, page 105): '*in a village of 1900 people, what is the probability of finding no days of the year which are not birthdays?*' In assuming the year as consisting of 365 days and that the birthdays occur randomly and independently, we can take $n = 365$ and $w = 1900$, implying $\Pr(W \leq 1900)$ to be the desired probability. The exact distribution (4.2.1) needed to be computed by means of Bonferroni's inequalities and delivers 0.1323, Feller's (1970, page 106) Poisson approximation gives 0.1353, Levin's approximation (4.2.6) yields 0.1323, and the saddlepoint approximation (2.4.2) delivers 0.1339. For this example Levin's Edgeworth expansion is more accurate than the saddlepoint approximation, whereas the Poisson approximation overestimates the exact value considerably.

Read (1998) observed that his log-normal approximation overestimates probabilities in the tails and underestimates those in the center of the distribution. Read's (1998) Table 2 for $n = 2, 3$ is shown in Table 4.2. For $n = 2$ the eight tabulated values of his log-normal approximation do not agree for three quantile points $(5\%, 50\%, 75\%)$, whereas the saddlepoint approximations completely agree with the exact values. As illustrated in Table 4.2 the same holds for $n = 3$ and the two quantile points $(95\%, 99\%)$.

Table 4.1. *The* S-PLUS *function* `saddle.collector` *(Kuonen, 2000a, Appendix).*

```
saddle.collector <- function(n, w, init=0) {
 if(w<n) stop("w < n")
 if(init>=log(n/(n-1))) stop("init > log(n/(n-1))")
 if(w==(n*sum(1/(1:n)))) {
 lambda3 <- sum((2*n^2-3*n*(1:n)
               +(1:n)^2)/(1:n)^3)/((sum((n-(1:n))
           /(1:n)^2))^(3/2)*sqrt(n))
 tmp.res <- lambda3/6
 }
 else {
  assign("para", list(n, w), frame = 1)
  fct2min <- function(zeta) {
   K <- -sum(log(1-(para[[1]]
            *(1-exp(-zeta)))/(1:para[[1]])))
   K - para[[2]]*zeta
  }
  tmp.min <- nlmin(fct2min, init,
                    max.iter=100, max.fcal=200)
  if(!tmp.min$converged) stop("nlmin not converged")
  saddlepoint <- tmp.min$x
  tmp.fct <- fct2min(saddlepoint)
  tmp.num <- (1:n)-n*(1-exp(-saddlepoint))
  K.prime <- n*exp(-saddlepoint)*sum(1/tmp.num)
  K.primeprime <- n*exp(-saddlepoint)
                   *sum((n-(1:n))/(tmp.num^2))
  u <- sign(saddlepoint)*sqrt(-2*tmp.fct)
  v <- saddlepoint*sqrt(K.primeprime)
  tmp.res <- u+log(v/u)/u
  }
 pnorm(tmp.res)
}
```

Table 4.2. *Comparison of exact and approximate α quantiles for $n = 2, 3$.*
The row labelled Read *is taken from Read's (1998) Table 2.*

		α (%)							
n		5	10	25	50	75	90	95	99
2	Exact	2	2	2	2	3	5	6	8
	Read	1	2	2	3	4	5	6	8
	Saddlepoint	2	2	2	2	3	5	6	8
3	Exact	3	3	4	5	7	9	11	15
	Read	3	3	4	5	7	9	10	14
	Saddlepoint	3	3	4	5	7	9	11	15

PSfrag replacements

$n = 6$ $n = 13$

Figure 4.1. *Exact and approximate distribution functions of W for $n = 6, 13$. Three of the lines are indistinguishable to plotting accuracy.*

As another example consider the case $n = 6$, throws of a die until all six sides have appeared, and $n = 13$, draws of cards from a complete set of cards until all 13 ranks have appeared. Figure 4.1 shows the exact distribution function of W (solid), Dawkins's approximation (dotted), Levin's approximation (long-dashed) and the saddlepoint approximation (dashed). Dawkins's approximation is inaccurate for small n, whereas Levin's approximation and the saddlepoint approximation are extremely accurate, matching the theoretical distribution.

An additional example corresponds to $n = 10$, random generation of digits 0–9 until all have occurred. The exact and approximate distribution functions are given in Figure
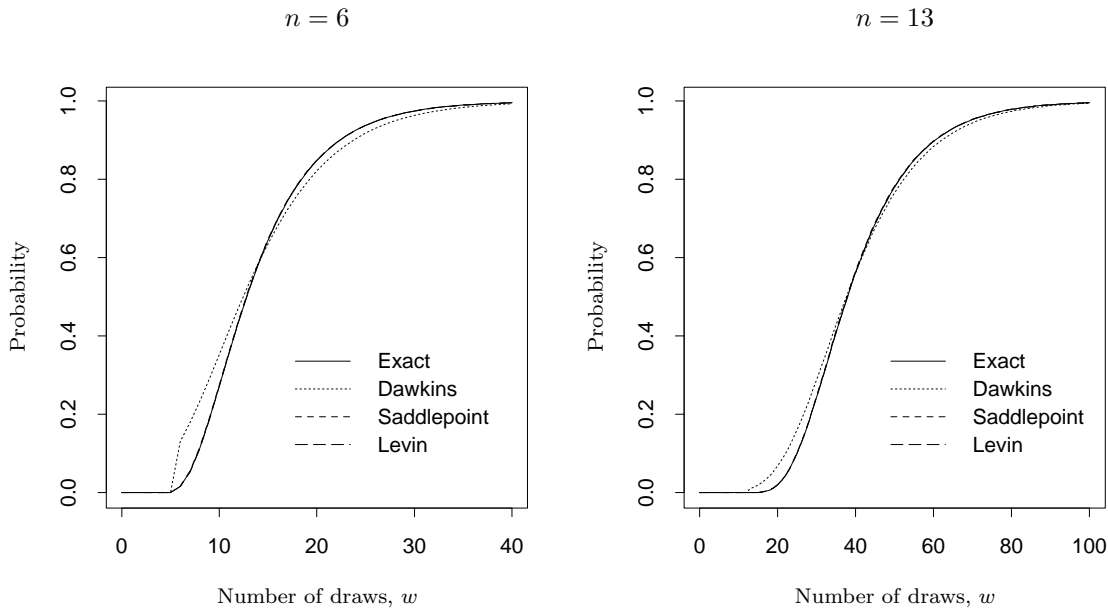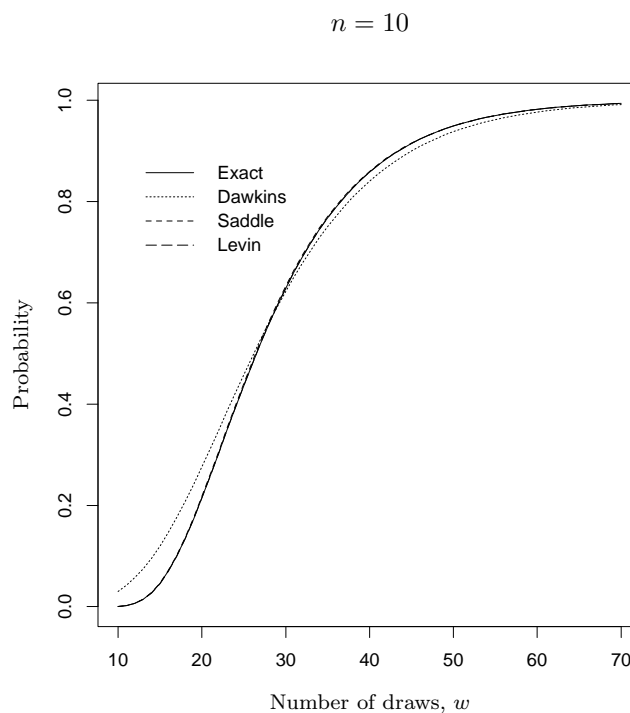
$$n = 10$$

Figure 4.2. *Exact and approximate distribution functions of W for n = 10. Three of the lines are indistinguishable to plotting accuracy.*

Table 4.3. *Exact and approximate probabilities for n = 10 and selected values of w, computed to four decimal places. The column labelled* Levin *is taken from Levin's (1992) Table 1. The relative error (%) is* Re = 100[(Approximation − Exact)/Exact].

w	Exact	Dawkins	Re (%)	Levin	Re (%)	Saddlepoint	Re (%)
10–19	0.1732	0.2411	−39.20	0.1729	0.17	0.1754	1.27
20–29	0.4216	0.3515	16.63	0.4223	−0.17	0.4237	0.49
30–39	0.2483	0.2324	6.40	0.2481	0.08	0.2458	−1.01
40–49	0.1004	0.1068	−6.37	0.1003	0.10	0.0989	−1.49
50–59	0.0366	0.0427	−16.67	0.0365	0.27	0.0363	−0.84
≥60	0.0199	0.0257	−29.15	0.0199	0.00	0.0200	0.50

4.2. Except Dawkins's approximation (dotted) all approximations are indistinguishable graphically from the exact distribution function of W (solid). Using the saddlepoint approximation (2.4.2), Levin's (1992) Table 1 can be augmented. The resulting Table 4.3 indicates that both saddlepoint and Edgeworth approximations are very accurate, outperforming Dawkins's by a significant factor.

In order to know how these approximations behave in the lower tail of W consider the case $n = 26$, generating letters of the alphabet until all have occurred, and $n =$
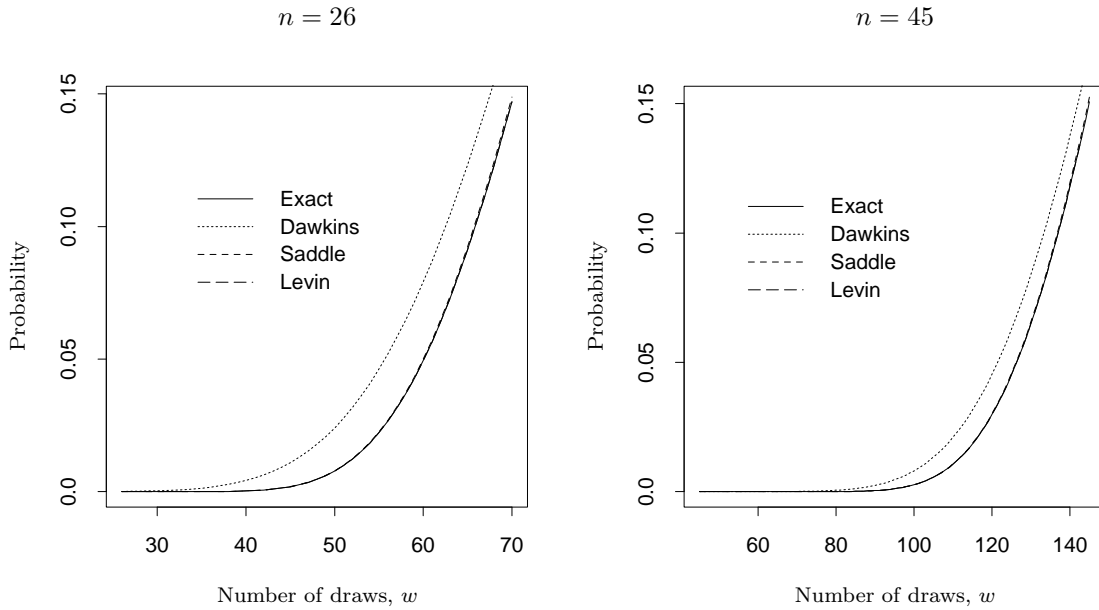
$$n = 26 \qquad\qquad\qquad n = 45$$

Figure 4.3. *Exact and approximate distribution functions for the lower tail of W for $n = 26, 45$. Three of the lines are indistinguishable to plotting accuracy.*

45, draws of numbers from a total of 45 numbers in the Swiss lottery until all have appeared. The lower tails of W are shown in Figure 4.3. One notices the accuracy of Levin's approximation (long-dashed) and the saddlepoint approximation (dashed), and the failure of Dawkins's approximation (dotted) in the lower tails.

In what follows, I try to explore Levin's Edgeworth approximation (4.2.6) and the proposed saddlepoint approximation (2.4.2) in more detail. The saddlepoint method's superiority over the Edgeworth series was briefly discussed in Chapter 1. Levin's Edgeworth expansion (4.2.6) has absolute error of order $O(n^{-1})$, whereas the proposed saddlepoint approximation (2.4.2) achieves a relative error of order $O(n^{-3/2})$. But, as seen in Table 4.3, we can expect Levin's approximation to be more accurate than $O(n^{-1})$ in the center of the distribution of W. Indeed, as mentioned in Section 4.2, Levin's approximation depends on a free parameter ν, which can be chosen for convenience and computational stability. Levin (1981) recommended the value $\nu = w$, which led to the approximation (4.2.6). But Levin (1983) suggested centering the distribution of S, the sum of independent and identically distributed left-truncated Poisson variables, and noted that his approximation could be improved by choosing $\nu > n$. Note that whenever ν/n is large, the truncation to positive integers is negligible, in which case $\nu = w$ is already near the center of the distribution of S. Furthermore, Butler and Sutton (1998) showed that Levin's simple choice of ν leads to a different saddlepoint approximation, illustrating the connection between Edgeworth and saddlepoint approximations. An additional saddlepoint approximation could be obtained by taking Levin's (1981) result given in (4.2.4) and using a saddlepoint approximation to the distribution of S. The latter variable, S, is a sum of independent and identically distributed truncated Poisson

Table 4.4. *Exact, computed by means of Bonferroni's inequalities, and approximate α quantiles of the distribution of the number of resamples needed to obtain the exact bootstrap distribution. The relative error (%) is abbreviated to* Re = 100[(Approximation − Exact)/Exact].

		α (%)					
		90			95		
m	n	Exact	Saddlepoint	Re (%)	Exact	Saddlepoint	Re (%)
7	1,716	16,638	16,612	−0.1563	17,873	17,856	−0.0951
8	6,435	70,909	70,811	−0.1382	75,540	75,476	−0.0847
9	24,310	300,199	299,830	−0.1229	317,698	317,459	−0.0752
10	92,378	1,264,096	1,262,691	−0.1111	1,330,592	1,329,679	−0.0686

variables, which may be already approximately normal for large ν/n, whereas W is a sum of independent but non-identically distributed, skewed, geometric variables.

As last example, consider the exhaustive bootstrap for symmetric statistics. The bootstrap is a general technique for estimating sampling distributions. Davison and Hinkley (1997) give an accessible account of bootstrap and its applications with extensive references. To approximate the ideal situation of sampling from the underlying unknown population distribution, resampling with replacement from the data, say of size m, is performed by putting equal probabilities m^{-1} at each data value. Repeating this procedure R times provides R nonparametric bootstrap samples, and hence R observations of the statistic of interest, T. When T is symmetric in the data values, there are up to $n = \binom{2m-1}{m-1}$ possible values for the statistic, each having probability n^{-1} to be sampled (Davison and Hinkley, 1997, page 27). The values for n corresponding to $m = 7, 8, 9$ and 10 are recorded in the second column of Tables 4.4 and 4.5. One remarks that n increases very fast as m increases. Knowing all n values, exact computation of the bootstrap distribution of T could be carried out. As one can only hope to enumerate completely for moderate sample sizes, one may want to know the distribution of the waiting time of such a procedure. To do so, the exhaustive bootstrap can be seen as a coupon (bootstrap statistic sample) collector's problem, where W is the number of samples needed to obtain all n different possible values of T. For a given value of n and α, the approximate α quantile is defined as the smallest value of w such that $\Pr(W < w + 1/2) \geq \alpha$. The exact and approximate 90%, 95%, 97.5% and 99% quantiles are tabulated in Tables 4.4 and 4.5 for different n. A summary is given in Table 3 of Kuonen (2000a). For $n =$1,716 ($m = 7$) one would need $17,873$ bootstrap samples to be 95% sure to have the exact bootstrap distribution, whereas for $n =$24,310 ($m = 9$) 317,698 are needed. The relative error of the saddlepoint approximation decreases when we go further into the tail of the distribution of W. This example reflects a feature of

Table 4.5. *Exact, computed by means of Bonferroni's inequalities, and approximate α quantiles of the distribution of the number of resamples needed to obtain the exact bootstrap distribution. The relative error (%) is abbreviated to* Re = 100[(Approximation − Exact)/Exact].

		α (%)					
		97.5			99		
m	n	Exact	Saddlepoint	Re (%)	Exact	Saddlepoint	Re (%)
7	1,716	19,093	19,078	−0.0786	20,669	20,676	0.0339
8	6,435	80,106	80,059	−0.0587	86,028	86,054	0.0302
9	24,310	334,983	334,770	−0.0636	357,321	357,420	0.0277
10	92,378	1,396,194	1,395,467	−0.0521	1,481,162	1,481,536	0.0253

the bootstrap distribution of a statistic, the total numbers of replicates increases rapidly with increasing sample size. Hence, for small values of m it is often feasible to calculate a bootstrap estimate exactly, by computing all bootstrap replicates, but as in practical applications data sets are usually larger than $m = 10$, it seems to be impossible to compute the exact bootstrap distribution by random sampling within a reasonable amount of time and effort, though other methods could be used (Diaconis and Holmes, 1994).

4.5. Gene transfer into cells

The example in this section results from a collaboration with the Laboratory of Cellular Biotechnology, DC-IGC, EPFL. They use gene transfer to introduce several genes into cells. Gene transfer is an essential tool in modern biology touching all domains from basic research to industrial application. From earlier studies they know that per cell w genes are randomly added to the genome. Different genes are present in the transfer mixture. The number of each gene present in the transfer mixture is the same, and each gene has identical probability to be added to the genome. One of their interests was to know the probability that a modified cell has at least one copy of each of the n genes ($n < w$). This can be seen as a coupon collector's problem, *i.e.* estimating the number of coupons (genes), drawn at random with replacement, needed to complete a collection (at least one copy of each gene is in the modified cell). In their experiments n is generally between 2 and 10, and w is between 10 and 100. In a concrete experiment, $w = 20$ and $n = 7$, the probabilities are 0.7038 for the exact distribution (4.2.1), 0.6877 for Dawkins's approximation (4.2.2), 0.7045 for Levin's approximation (4.2.6), and the saddlepoint approximation (2.4.2) delivers 0.7069. Dawkin's approximation is not so accurate as the others are as n is small,

whereas the others are very accurate.

4.6. Conclusion

I n this chapter several examples were used to show that a saddlepoint approximation to the distribution of the waiting time in the coupon collector's problem is very accurate. The proposed saddlepoint approximation only covers the case of equal sampling probabilities, whereas the approach proposed by Levin (1981, 1983, 1992) also covers the case of unequal sampling probabilities. For example, it would be interesting to know what the median number of poker hands is to be dealt before one has at least one of each kind of poker hand? A referee of Kuonen (2000a) kindly calculated the median number to be 451,617.

Finally, note that the consumation of 51 breakfast cereal boxes would be needed to be 95% sure to get all ten gifts.

5. Bootstrap distributions

'The bootstrap allows the data analyst to assess the statistical accuracy of complicated procedures, by exploiting the power of the computer. The use of the bootstrap either relieves the analyst from having to do complex mathematical derivations, or in some instances provides an answer where no analytical answer can be obtained. The bootstrap can be used either nonparametrically, or parametrically. In nonparametric mode, it avoids restrictive and sometimes dangerous parametric assumptions about the form of the underlying populations. In parametric mode, it can provide more accurate estimates of error than traditional Fisher information-based methods.'

Bradley Efron and Robert J. Tibshirani (1993, page 394)

The bootstrap (Efron, 1979) found an immediate place in statistical theory and, more slowly, in practice; see Young (1994). But it is known that, for accurate approximations in the tails, the nonparametric bootstrap requires a large number of replicates of the statistic. As this can be time-intensive other methods should be considered. A first alternative is Edgeworth expansion. But although Edgeworth expansion provide good approximations in the centre of the density, it can lead to inaccurate tail probabilities; see the references given in Chapter 1. A second alternative is saddlepoint approximation as reviewed in Chapter 2. Indeed, bootstrap analysis presents an important application of saddlepoint approximation. Davison and Hinkley (1988) used saddlepoint approximations to replace Monte Carlo simulation for bootstrap means and related statistics. The method of Davison and Hinkley (1988) requires the statistic of interest to be linear in the observations. Extensions to nonlinear statistics are discussed in Daniels and Young (1991) for the studentized bootstrap mean, and in DiCiccio *et al.* (1992a, 1992b, 1994) using the approximation developed in DiCiccio and Martin (1991). Davison and Hinkley's technique is justified in Wang (1990a) and extended to certain nonlinear statistics by a different method in Wang (1990b). In the latter the approximation established is similar to that of Skovgaard (1987). Further results in the general context of so-called *empirical saddlepoint approximation*, when the required cumulant generating function is obtained empirically (as for the bootstrap approximation, for instance), are given in Feuerverger (1989) and Jing *et al.* (1994). More recently, Robinson and Skovgaard (1998) derived error rates for two cases where

the empirical saddlepoint approximation is just the saddlepoint approximation to the bootstrap approximation: the univariate standardized mean and in the general case of tests based on smooth functions of means. The empirical saddlepoint approximation also provides an alternative to empiricial likelihood techniques. Their relationship was discussed by Monti and Ronchetti (1993) for multivariate M-estimators using the empirical saddlepoint approximation given by Ronchetti and Welsh (1993). The general case of this empirical saddlepoint approximation will not be considered within this thesis but I refer to the references given above.

The implementation of saddlepoint approximations for bootstrap distributions, as used in this chapter, is discussed in Davison and Hinkley (1997, Section 9.5), Canty and Davison (1996, 1999) or Kuonen (1998a). The importance of such implementations is underlined by Rudolf Beran's (1994, page 149) comment on the practical use of approximations in bootstrap computations:

> *'... Edgeworth expansions suffer from relative inaccuracy in their tails as well as algebraic cumbersomeness. Saddlepoint approximations to bootstrap distributions ... appear to be more accurate, but currently lack convenient implementation outside the simplest cases.'*

I recall the basic notions of the nonparametric bootstrap, of the nonparametric delta method and of the saddlepoint approximation as applied to resampling in Section 5.1. Daniels and Young (1991, Section 6) noticed that saddlepoint approximations to bootstrap distributions may fail if data contain outliers. Therefore, I propose in Section 5.2 a saddlepoint mixture approximation for the case of the average. In Section 5.3 I look at the application of saddlepoint approximations for bootstrap statistics defined by estimating equations. As applications I consider Huber's M-estimate for location and the standardized M-estimator of location with initial MAD scaling.

5.1. Introduction

n this section I review the basic notions of the nonparametric bootstrap, of the nonparametric delta method (von Mises expansion) and of the saddlepoint approximations to bootstrap statistics. A more complete survey on bootstrap methods and their applications can be found in Shao and Tu (1995) or Davison and Hinkley (1997).

5.1.1. Nonparametric bootstrap

The bootstrap method is a simulation procedure to obtain an approximation to the distribution of a statistic. Consider a sample $x_1, \ldots, x_n$, thought of as the outcome of n independent and identically distributed random variables $X_1, \ldots, X_n$ whose probability density function (PDF) and cumulative distribution function (CDF) are denoted by f and F. Let θ be the parameter of interest which is estimated by a statistic T whose value is t. Since the distribution F is unknown in the context of the *nonparametric*

bootstrap, so is the distribution of T. Thus, if we are to use T to estimate θ, then we need to have some idea of the variability of T. Therefore we can use resampling of the data set $x_1, \ldots, x_n$ to estimate the distribution of an estimator. This is the essence of the bootstrap. Let

$$\hat{F}(x) = \frac{1}{n} \sum_{i=1}^{n} I(x_i \leq x)$$

be the *empirical distribution function* (EDF) based on the sample, where $I(\cdot)$ denotes the indicator random variable. The EDF can be used to estimate the unknown cumulative distribution function F. Let $\{p_i : i = 1, \ldots, n\}$ denote an array of weights (or resampling probabilities) for $x_1, \ldots, x_n$, which satisfy $p_i \geq 0$ and $\sum p_i = 1$ for $i = 1, \ldots, n$. The *generalized empirical distribution function* (GEDF) is

$$\hat{F}_G(x) = \sum_{i=1}^{n} p_i I(x_i \leq x).$$

Note that $\mathrm{E}\{\hat{F}_G(x) \mid x_1, \ldots, x_n\} = \hat{F}(x)$ only if $\mathrm{E}(p_i) = n^{-1}$, and that in taking $p_i \equiv n^{-1}$, $\hat{F}_G(x)$ equals $\hat{F}(x)$. In the generalized bootstrap process the GEDF is used to estimate F; consistency is shown in Mason and Newton (1992).

Denote by $x_1^\star, \ldots, x_n^\star$ a random sample of size n from the GEDF. This is equivalent to simple random sampling with replacement from the original sample, $x_1, \ldots, x_n$, and applying the same statistic to the resampled data as to the original data. The resampled data are called the *bootstrap sample*, and the conditional distribution of the statistic applied to the bootstrap sample (given $x_1, \ldots, x_n$) is referred to as the *bootstrap distribution*. Each bootstrap sample, $x_1^\star, \ldots, x_n^\star$, consists of elements from the original data set sampled with probability vector $(p_1, \ldots, p_n)$. Let

$$f_k^\star = \sum_{i=1}^{n} I(x_i^\star = x_k), \quad k = 1, \ldots, n,$$

denote the number of x_k's in the bootstrap sample. Then $(f_1^\star, \ldots, f_n^\star)$ has a multinomial distribution with denominator n and probabilities $p_1, \ldots, p_n$. The marginal distribution of $f_i^\star$ is $B(n, p_i)$, so $\mathrm{E}^\star(f_i^\star) = np_i$ and $\mathrm{var}^\star(f_i^\star) = np_i(1 - p_i)$, $i = 1, \ldots, n$. Following Example 2.3 the cumulant generating function of $(f_1^\star, \ldots, f_n^\star)$ is

$$K(\zeta) = n \log \left\{ \sum_{i=1}^{n} p_i \exp(\zeta_i) \right\}. \tag{5.1.1}$$

Remark 5.1. To apply simulation with the EDF is very straightforward. Because the EDF puts equal probabilities on the original data values $x_1, \ldots, x_n$, each $T^\star$ consists of a independent copy of T applied to the bootstrap sample $x_1^\star, \ldots, x_n^\star$. Therefore we obtain a simulated sample $T_1^\star, \ldots, T_R^\star$. The observed values $t_1^\star, \ldots, t_R^\star$ are called the *bootstrap replicates*. Note that the letter R is reserved for the number of replicate simulations. $\quad\square$

Remark 5.2. The *basic bootstrap confidence interval*, of approximate coverage $(1 - 2\alpha)$, is defined to be the central $(1 - 2\alpha)$ range of the ordered bootstrap replications. In the usual order-statistic notation, this interval is $[t^\star_{(\alpha(R+1))}, t^\star_{((1-\alpha)(R+1))}]$. As mentioned in Remark 5.1, each bootstrap sample gives a bootstrap replication of T, namely $t^\star_r$, $r = 1, \ldots, R$. We assume that R, the number of bootstrap replicates, is chosen so that $\alpha(R + 1)$ and $(1 - \alpha)(R + 1)$ are integers. $\qquad\square$

5.1.2. Nonparametric delta method

The delta method is one of the oldest techniques for assessing the standard errors of complicated statistical estimators. We consider real-valued statistics T as estimators of the parameter of interest θ, and consider estimators which are functionals, so T can be represented as $T = t(\hat{F})$. We assume that the functionals under study are Fisher consistent, so the estimator asymptotically equals $\theta = t(F)$. The key idea of the nonparametric delta method, also known as *von Mises expansion*, is to extend series extension to statistical functions. If some distribution G is 'near' F, then the linear form of the first-order von Mises expansion of T at F evaluated in G is

$$t(G) \doteq t(F) + \int L_t(x; F) dG(x) + \text{remainder}, \qquad (5.1.2)$$

where $L_t(\cdot; F)$, the first derivative of $t(\cdot)$ at F, also known as the *influence function*, see Hampel (1974b) and Hampel *et al.* (1986), is defined by

$$L_t(x; F) = \lim_{\epsilon \searrow 0} \frac{t\{(1 - \epsilon)F + \epsilon \triangle_x\} - t(F)}{\epsilon},$$

where $\triangle_x$ is the point mass 1 at x. The influence function describes the effect of an infinitesimal contamination at the point x on the estimate, standardised by the mass of the contamination. It is a measure of the asymptotic bias caused by contamination in the observations. From (5.1.2) we see that the derivative satisfies $\int L_t(x; F) dF(x) = 0$. We call the empirical approximation $l(x) = L_t(x; \hat{F})$ the *empirical influence function*, and the particular values $l_i = l(x_i)$ the *empirical influence values*.

Let us now look at the relationship between the influence function and the asymptotic variance. Evaluating (5.1.2) at $G = \hat{F}$ yields the *linear nonparametric delta method approximation*:

$$\begin{aligned} t(\hat{F}) &\doteq t(F) + \int L_t(x; F) d\hat{F}(x) + \text{remainder} \\ &= t(F) + \frac{1}{n} \sum_{i=1}^{n} L_t(X_i; F). \end{aligned} \qquad (5.1.3)$$

By the central limit theorem the second term on the right-hand side of equation (5.1.3) is asymptotically normal. In most cases the remainder becomes negligible for $n \to \infty$,

so $t(\hat{F})$ itself is asymptotically normal. Hence, $(T - \theta)/v_L(F)^{1/2}$ tends to $N(0,1)$, where the asymptotic variance equals

$$v_L(F) = \frac{1}{n} \int L_t^2(x; F) dF(x). \tag{5.1.4}$$

In practice $v_L(F)$ is approximated by substituting $\hat{F}$ for F in (5.1.4), that is by using the sample version

$$v_L = v_L(\hat{F}) = \frac{1}{n^2} \sum_{i=1}^{n} l_i^2, \tag{5.1.5}$$

which is known as the *nonparametric delta method variance estimate*. Note also that (5.1.3) implies that $\sum l_i = 0$.

Example 5.1 (Bootstrap estimate). The linear delta approximation (5.1.3) can be applied to the bootstrap statistic $T^\star$ in the following way: let $\hat{F}^\star$ be the EDF of the bootstrap sample. Then (5.1.3) shows that

$$t(\hat{F}^\star) = t(\hat{F}) + \frac{1}{n} \sum_{i=1}^{n} L_t(x_i^\star; \hat{F})$$

or

$$t^\star \doteq t + \frac{1}{n} \sum_{i=1}^{n} f_i^\star l_i,$$

which is the linear approximation to $T^\star$, say $T_L^\star$. In the case of the average, where the empirical influence values l_i are given by $x_i - \bar{x}$, it follows that the linear approximation is

$$T_L^\star = t + \frac{1}{n} \sum_{i=1}^{n} f_i^\star l_i = \frac{1}{n} \sum_{i=1}^{n} f_i^\star x_i$$

as $\sum f_i^\star = n$. In the context of the generalized (weighted) bootstrap, where the unknown distribution F is estimated by the GEDF, the linear delta approximation becomes $t(\hat{F}_G^\star) = t(\hat{F}_G) + \sum p_i L_t(x_i^\star; \hat{F}_G)$. $\square$

Remark 5.3. If the estimator T is defined by an estimating equation of the form $\sum \psi(x_i, t) = 0$ such that $\int \psi(x, t) dF(x) = 0$, its influence function is

$$L_t(x) = \frac{\psi(x, \theta)}{\mathrm{E}\{-\psi'(x, \theta)\}}$$

under the assumption that the denominator is nonzero, where prime denotes differentiation with respect to θ. The nonparametric delta method variance estimate is

$$v_L = \sum_{i=1}^{n} \psi^2(x_i, t) \left\{ \sum_{i=1}^{n} \psi'(x_i, t) \right\}^{-2}.$$

An example is the classical M-estimator, where ψ represents Huber's function. $\square$

5.1.3. Saddlepoint approximations

Suppose that the statistic T can be expressed as a linear combination of $X_1, \ldots, X_n$, say $T = \sum b_i X_i$, where the b_i are scalars. The bootstrap statistic is $T^\star = \sum b_i X_i^\star = \sum f_i^\star a_i$, where $a_i = b_i x_i$ and $f_i^\star$ is the bootstrap frequency of X_i. The $f_i^\star$ have a joint multinomial distribution with denominator n and probabilities $(p_1, \ldots, p_n)$, so $T^\star$ has cumulant generating function

$$K(\zeta) = n \log \left\{ \sum_{i=1}^n p_i \exp(\zeta a_i) \right\}. \tag{5.1.6}$$

As illustrated in Section 2.4, setting $d = 1$, $a_i = b_i x_i$ and $X_i = f_i^\star$, the saddlepoint approximation to the PDF and CDF of $T^\star$ can easily be obtained by means of (2.4.1) and (2.4.2). Here, the saddlepoint $\hat{\zeta}$ is the solution of

$$K'(\zeta) = n \sum_{i=1}^n a_i p_i \exp(\zeta a_i) \left\{ \sum_{i=1}^n p_i \exp(\zeta a_i) \right\}^{-1} = t. \tag{5.1.7}$$

No solution exists for $t \leq \min(a_i)$ and $t \geq \max(a_i)$.

Although the nonparametric bootstrap statistic is discrete, there are generally no problems in approximating it in a continuous way, because usually $T^\star$ takes so many values that continuity corrections do not make a big difference. Indeed, in considering a data set of size n, the maximal number of distinct points is $\binom{2n-1}{n}$; see Hall (1992, Appendix I). For example, $n = 5$ and 10 yield 126 and 11,628 points.

If an estimate of the entire distribution of $T^\star$ is needed, the values of $F_s(t)$, given in (2.4.2), for m evenly spaced values of t between $\min(a_i)$ and $\max(a_i)$ are calculated and a cubic B-spline is fitted to the values of $\Phi^{-1}\{F_s(t)\}$ by means of the S-Plus or R function `smooth.spline`. Hesterberg (1994) recommended the use of a cubic spline with some boundary conditions (using the S-Plus or R function `spline`), but in practice they have little effect here.

Saddlepoint approximations are not easy to implement in a efficient way. Two main problems may arise: the calculation of the cumulant generating function and the computation of the saddlepoint. Daniels and Young (1991, Section 6) remarked another limitation of the saddlepoint approximation itself when applied to an empirical distribution and hence to the bootstrap. They considered a data set containing an outlier and noticed that the usual saddlepoint approximation to the density of the studentized mean breaks down. In the next section I propose a saddlepoint mixture approximation for the average in this situation.

5.2. Saddlepoint mixture approximations

O utliers are observations far removed from the pattern set by the majority of the data. They are mainly due to gross errors or legitimate extreme observations. Outliers typically inflate classical estimates of the error variance, and the decision of a classical hypothesis test may completely change on removing

a single distant point. The proportion of contamination can affect the choice of statistical techniques. In order to make them more stable, the need to bound the influence of outliers on estimates in the bootstrap context may lead us to ideas like the weighted bootstrap. Barbe and Bertail (1995) investigated the weighted bootstrap of statistical functionals. They showed that from the point of view of the adequacy of the whole distribution, we can do as well as the classical bootstrap. But, it seems to be logical that even in putting a small weight on a outlier, the saddlepoint approximation of a classical statistic may break down, as the bootstrap distribution in the presence of an outlier may be multi-modal, whereas the classical saddlepoint approximation is unimodal. In this section, I propose a saddlepoint mixture approximation giving a multi-modal distribution estimate. A major problem is to know the critical value of an outlier for which the saddlepoint approximation may fail. A classical method of outlier detection consists of the use of a boxplot. But, practice showed that the outlying values in the usual boxplot do not correspond to values for which the saddlepoint approximation breaks down. This results from the fact that the aim is to approximate a bootstrap statistic for which the breakdown point of the original estimator is not preserved. Stromberg (1997) pointed out that the bootstrapped sample covariance matrix can have a breakdown point of nearly zero regardless of the breakdown point of the estimator itself. Singh (1998) studied breakdown properties of bootstrap quantiles, and Hu and Hu (2000) applied them to some second-order accurate bootstrap methods. In Chapter 6, some remarks on this will be made in the context of studentized bootstrap statistics. In this section I propose a practical method for detecting an outlier in the context of the saddlepoint approximation for the average.

5.2.1. The average

Let $x_1, \ldots, x_n$ be a sample of size n. Suppose that the parameter of interest θ is the true mean of the unknown distribution F. The statistic used to estimate θ is $T = \bar{X} = n^{-1} \sum X_i$. Hence the bootstrap statistic of the average is $T^\star = \bar{X}^\star = \sum f_i^\star a_i$ with $a_i = x_i/n$ and $f_i^\star$ as before. To apply the saddlepoint approximations (2.4.1) and (2.4.2) we need only the cumulant generating function, which is given by (5.1.6), using $p_i \equiv n^{-1}$.

Example 5.2 (Daniels and Young's data). To illustrate the saddlepoint approximations of the average I consider Daniels and Young's data (1991, Section 6):

$$-0.4621, -0.4608, -0.4492, -0.3830, -0.3109, -0.2282, -0.0896, 0.0927, 0.2183, 2.0723.$$

S-PLUS functions to calculate the needed saddlepoint approximations for standard bootstrap statistics are included in the library `boot` written by Angelo J. Canty (Davison and Hinkley, 1997); the corresponding R library is available as well. It seems plausible that the extreme outlier at 2.0723 may cause problems for the saddlepoint approximation. This belief is justified by Figure 5.1. In the left panel the 49,999 replicates of the bootstrapped statistic and the corresponding saddlepoint approximation to the density of the average are shown. As the saddlepoint approximation is generally unimodal the
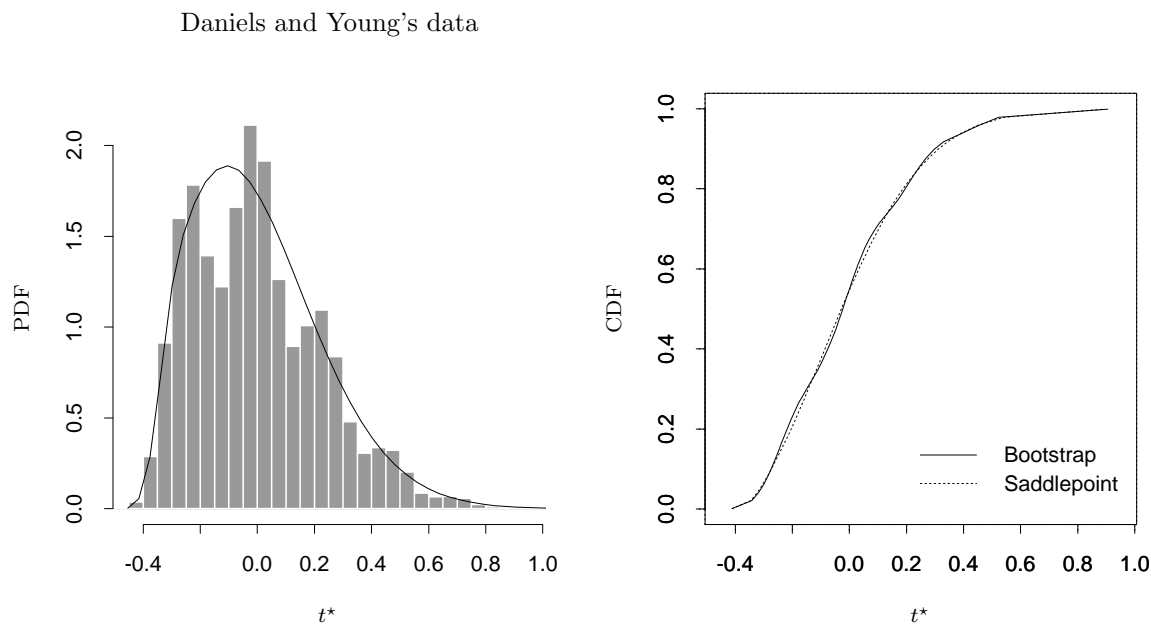
Daniels and Young's data

Figure 5.1. *Saddlepoint approximations for Daniels and Young's data. Left panel: histogram of $R = 49,999$ replicates of the average. The solid line represents the saddlepoint approximation to the bootstrap density of the average. Right panel: bootstrap and saddlepoint approximation estimates of the CDF.*

multi-modality of the replicates is not captured. The outlier at 2.0723 generates variation in the density of the average. This clearly reflects a limitation of the saddlepoint approximation. In the right panel the estimated CDF of the bootstrap approach and the saddlepoint approximation are shown. One could think falsely from this figure that the saddlepoint approximation works quite well when compared to the bootstrap distribution estimate. It is interesting to know if there is a structure in the modes shown in the histogram. Figure 5.2 shows the histogram of the bootstrap estimates with superposed density estimates corresponding to the frequency of the outlying value, found from the bootstrap output. The resampling weights are $p_i = n^{-1}$ for $i = 1, \ldots, n$ and r represents the number of times the outlier is resampled. One remarks that the multi-modality has been captured very well. As mentioned before, a weighted bootstrap may not help as the average breaks down in taking a single outlier. The weighted bootstrap may reduce the variability and hence the number of modes, but saddlepoint approximations would still break down. Hence, the idea of computing a saddlepoint mixture approximation comes to mind. In considering each mode separately the approximation should be more accurate.

I also tried to change the extreme value in order to find out the critical value for which the saddlepoint approximation to the bootstrap density of the average fails for the first time. In taking as criterion the standard boxplot with span equal to 1.5 times the interquartile range (IQR), the upper extreme of the boxplot is 0.2183, and in considering a value larger than 0.2183 the saddlepoint approximation should break down. In practice we do not see multi-modality of the histogram until the value is not larger than 1.4048,
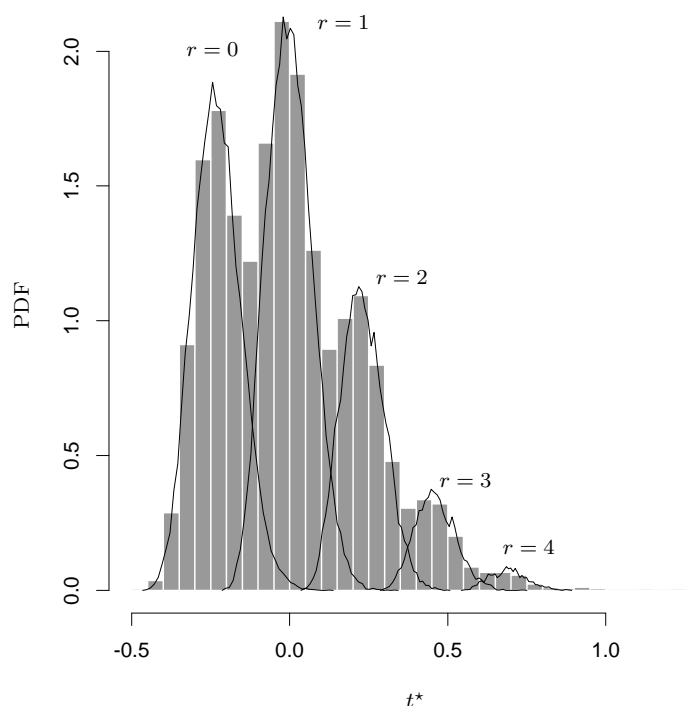
Daniels and Young's data

Figure 5.2. *Mode checking for Daniels and Young's data. Histogram of $R = 49,999$ replicates of the average with superposed density estimates corresponding to the frequency of the outlying value, r.*

and therefore the saddlepoint approximation works well. This value corresponds to a span of $c\,\mathrm{IQR}$ with $c = 2.83$. Therefore, I conclude that for Daniels and Young's data values larger than 1.4048 indicate a failure for the saddlepoint approximation. Hence, the span of the boxplot with $c = 3$ would be more appropriate for our purposes. Moreover, as a boxplot displays the variability of the median such a rule seems to be very stable to outliers.   □

In the previous example the idea of saddlepoint mixture approximations came up. Let us now explore this idea and test their accuracy in the context of the average. Note that the calculations below are valid only for the simple case where a single outlying value is considered, say the jth observation. Let us split the bootstrap statistic as

$$\bar{X}^\star = \sum_{i=1}^{n} f_i^\star a_i = f_j^\star a_j + \sum_{\substack{i=1 \\ i \neq j}}^{n} f_i^\star a_i,$$

where $a_i = x_i/n$ for $i = 1, \ldots, n$. Let r be the number of times the outlier has been resampled. Given $f_j^\star = r$, we have

$$\bar{X}^\star = r a_j + \sum_{\substack{i=1 \\ i \neq j}}^{n} f_i^\star a_i.$$

The CDF of the average may be written as

$$\Pr(\bar{X}^\star \le x) = \sum_{r=0}^{n} \Pr(\bar{X}^\star \le x \mid f_j^\star = r)\Pr(f_j^\star = r). \qquad (5.2.1)$$

Since the marginal distribution of $f_j^\star$ is $B(n, p_j)$, we have

$$\Pr(f_j^\star = r) = \binom{n}{r} p_j^r (1 - p_j)^{n-r}.$$

Hence to approximate (5.2.1) only an estimate of $\Pr(\bar{X}^\star \le x \mid f_j^\star = r)$ is needed. We use the saddlepoint approximation, and define the *saddlepoint mixture approximation* to the bootstrap CDF of the average as

$$\Pr(\bar{X}^\star \le x) \doteq \sum_{r=0}^{n} F_s(\bar{X}^\star \le x \mid f_j^\star = r)\Pr(f_j^\star = r), \qquad (5.2.2)$$

where $F_s(\cdot \mid f_j^\star = r)$ is the saddlepoint approximation to the conditional bootstrap CDF given $f_j^\star = r$. To calculate the approximation, the values of $F_s(t \mid f_j^\star = r)$ for m evenly spaced points are calculated using (2.4.2) and a spline is fitted to the values of $\Phi^{-1}\{F_s(t \mid f_j^\star = r)\}$. In practice, $m = 50$ seems to be a good choice. To obtain the approximations one has to know the cumulant generating function, $K(\zeta)$, of the average given $f_j^\star = r$, and furthermore we need to know the range of values for which the conditional bootstrap density is non-zero. To find a local minimum of $K(\zeta) - t\zeta$ I used a general quasi-Newton optimiser, which is already implemented in S-PLUS (function `nlmin`) or R (function `nlm`). The saddlepoint mixture approximation to the PDF of the average is calculated in the same way. To improve the accuracy of the approximations the saddlepoint approximation for the density is integrated and rescaled so that the density integrates to one. To integrate the approximation, whose value is known at equally spaced points, I use the classical trapezoidal rule. This quadrature method is based on the device of adding up the value of the integrand at a sequence of abscissas within the range of integration.

Remark 5.4. For n large and p_j small, the binomial distribution of $f_j^\star$ may be approximated by a Poisson distribution with parameter np_j. To have an idea of $\Pr(f_j^\star = r)$ we can calculate $(np_j)^r/r! \exp(-np_j)$. On setting $p_j = n^{-1}$ we get

r	0	1	2	3	4	5
$\Pr(f_j^\star = r)$	0.3678	0.3678	0.1839	0.0613	0.0153	0.0030
$\Pr(f_j^\star \le r)$	0.3678	0.7356	0.9195	0.9808	0.9961	0.9991

From this table we see that the probability that the outlier will be resampled five times is small. This suggests that the sum in (5.2.2) needs only be taken up to $r = 4$. □

Let us now calculate the cumulant generating function of the average given $f_j^\star = r$. It is defined as

$$K(\zeta) = \log \mathrm{E}\left\{\exp(\zeta \bar{X}^\star) \mid f_j^\star = r\right\} = \zeta r a_j + (n-r) \log \left\{\sum_{\substack{i=1 \\ i \neq j}}^{n} q_i \exp(\zeta a_i)\right\},$$

where $q_i = \Pr(f_i^\star \mid f_j^\star = r) = p_i/(1-p_j)$ for $i \neq j$ and $a_i = x_i/n$. Note that if $p_i \equiv n^{-1}$, we have $q_i = 1/(n-1)$. Furthermore,

$$
\begin{aligned}
K(\zeta) &= (n-r) \log\left[\exp\left\{\zeta r a_j/(n-r)\right\}\right] + (n-r) \log\left\{\sum_{\substack{i=1 \\ i \neq j}}^{n} q_i \exp(\zeta a_i)\right\} \\
&= (n-r) \log\left(\sum_{\substack{i=1 \\ i \neq j}}^{n} q_i \exp\left[\zeta\left\{a_j r/(n-r) + a_i\right\}\right]\right) \\
&= \tilde{n} \log\left\{\sum_{i=1}^{n} \tilde{p}_i \exp(\zeta \tilde{a}_i)\right\},
\end{aligned}
\tag{5.2.3}
$$

where $\tilde{n} = n - r$, $\tilde{p}_i$ equals $q_i = p_i/(1-p_j)$ for $i \neq j$ and otherwise $\tilde{p}_j = 0$, and $\tilde{a}_i = a_j r/(n-r) + a_i$. We remark that formula (5.2.3) is of the same form as (5.1.6). To apply the saddlepoint approximation we also need the second derivative of $K(\zeta)$ with respect to ζ,

$$K''(\zeta) = \tilde{n} \frac{\displaystyle\sum_{i=1}^{n} \tilde{a}_i^2 \tilde{p}_i \exp(\zeta \tilde{a}_i) \sum_{i=1}^{n} \tilde{p}_i \exp(\zeta \tilde{a}_i) - \left\{\displaystyle\sum_{i=1}^{n} \tilde{a}_i \tilde{p}_i \exp(\zeta \tilde{a}_i)\right\}^2}{\left\{\displaystyle\sum_{i=1}^{n} \tilde{p}_i \exp(\zeta \tilde{a}_i)\right\}^2}.$$

Hence (5.1.7) yields

$$K''(\hat{\zeta}) = \tilde{n} \sum_{i=1}^{n} \tilde{a}_i^2 \tilde{p}_i \exp(\hat{\zeta} \tilde{a}_i) \left\{\sum_{i=1}^{n} \tilde{p}_i \exp(\hat{\zeta} \tilde{a}_i)\right\}^{-1} - \tilde{n}^{-1} t^2.$$

Finally, let me define the grids on which the saddlepoints should be computed. This clearly depends on r, and therefore the idea is the following: for a fixed r I calculate a sort of confidence interval for the bootstrap statistic $T^\star$ given $f_j^\star = r$. I work out the expectation and the variance of $T_L^\star - t$ given $f_j^\star = r$. In Example 5.1 I applied a linear delta approximation to such a statistic in the following way:

$$T_L^\star - t \doteq \frac{1}{n} \sum_{i=1}^{n} f_i^\star l_i,$$

where l_i are the empirical influence values. As t is the observed value of the statistic the variance of $T_L^\star$ equals that of $T_L^\star - t$, and for the expectation one only has to add t. For the conditional expectation one gets

$$\mathrm{E}(T_L^\star - t \mid f_j^\star = r) = \frac{r}{n}l_j + \frac{1}{n}\sum_{\substack{i=1 \\ i \neq j}}^{n} \mathrm{E}^\star(f_i^\star)l_i = \frac{r}{n}l_j + \frac{n-r}{n(1-p_j)}\sum_{\substack{i=1 \\ i \neq j}}^{n} p_i l_i \qquad (5.2.4)$$

as $\mathrm{E}^\star(f_i^\star) = (n-r)\tilde{p}_i = (n-r)p_i/(1-p_j)$ for $i \neq j$. Moreover the conditional variance is

$$\begin{aligned}
\mathrm{var}(T_L^\star - t \mid f_j^\star = r) &= \frac{1}{n^2}\sum_{\substack{i=1 \\ i \neq j}}^{n}\sum_{\substack{k=1 \\ k \neq j}}^{n} \mathrm{cov}^\star(f_i^\star, f_k^\star)l_i l_k \\
&= \frac{n-r}{n^2}\sum_{\substack{i=1 \\ i \neq j}}^{n}\tilde{p}_i(1-\tilde{p}_i)l_i^2 - \frac{n-r}{n^2}\sum_{\substack{i=1 \\ i \neq j}}^{n}\sum_{\substack{k=1 \\ k \neq j}}^{n}\tilde{p}_i\tilde{p}_k l_i l_k \\
&= \frac{n-r}{n^2(1-p_j)^2}\Big[(1-p_j)\sum_{\substack{i=1 \\ i \neq j}}^{n}p_i l_i^2 - \sum_{\substack{i=1 \\ i \neq j}}^{n}p_i^2 l_i^2 - \Big\{\sum_{\substack{i=1 \\ i \neq j}}^{n}p_i l_i\Big\}^2\Big].
\end{aligned}$$
$$(5.2.5)$$

When $p_i \equiv n^{-1}$, (5.2.4) and (5.2.5) reduce to

$$\mathrm{E}(T_L^\star \mid f_j^\star = r) = \mathrm{E}(T_L^\star - t \mid f_j^\star = r) + t = \frac{r}{n}l_j - \frac{n-r}{n(n-1)}l_j + t = \frac{r-1}{n-1}l_j + t$$

as $\sum l_i = 0$, and

$$\begin{aligned}
\mathrm{var}(T_L^\star \mid f_j^\star = r) &= \frac{(n-r)}{(n-1)^2 n^2}\Big[(n-1)\sum_{\substack{i=1 \\ i \neq j}}^{n}l_i^2 - \sum_{\substack{i=1 \\ i \neq j}}^{n}l_i^2 - \Big\{\sum_{\substack{i=1 \\ i \neq j}}^{n}l_i\Big\}^2\Big] \\
&= \frac{(n-r)}{(n-1)^2 n^2}\Big[(n-2)\sum_{\substack{i=1 \\ i \neq j}}^{n}l_i^2 - l_j^2\Big].
\end{aligned}$$

Let us denote $\mathrm{E}(T_L^\star \mid f_j^\star = r)$ by μ_r and $\mathrm{var}(T_L^\star \mid f_j^\star = r)$ by σ_r^2. For a fixed r we calculate the saddlepoint approximations at m evenly spaced points from $\mu_r - c\sigma_r$ to $\mu_r + c\sigma_r$, where c is constant. In practice we took c equal to the 97.5% quantile of the standard normal distribution.

Example 5.3 (Daniels and Young's data). In Example 5.2 I remarked the limitation of the classical saddlepoint approximation in the presence of an outlier, as in Daniels and Young's data. In this example I compare the saddlepoint mixture approximation to the classical saddlepoint approximation and to the nonparametric bootstrap. In the left panel of Figure 5.3 the saddlepoint mixture approximation to the density of
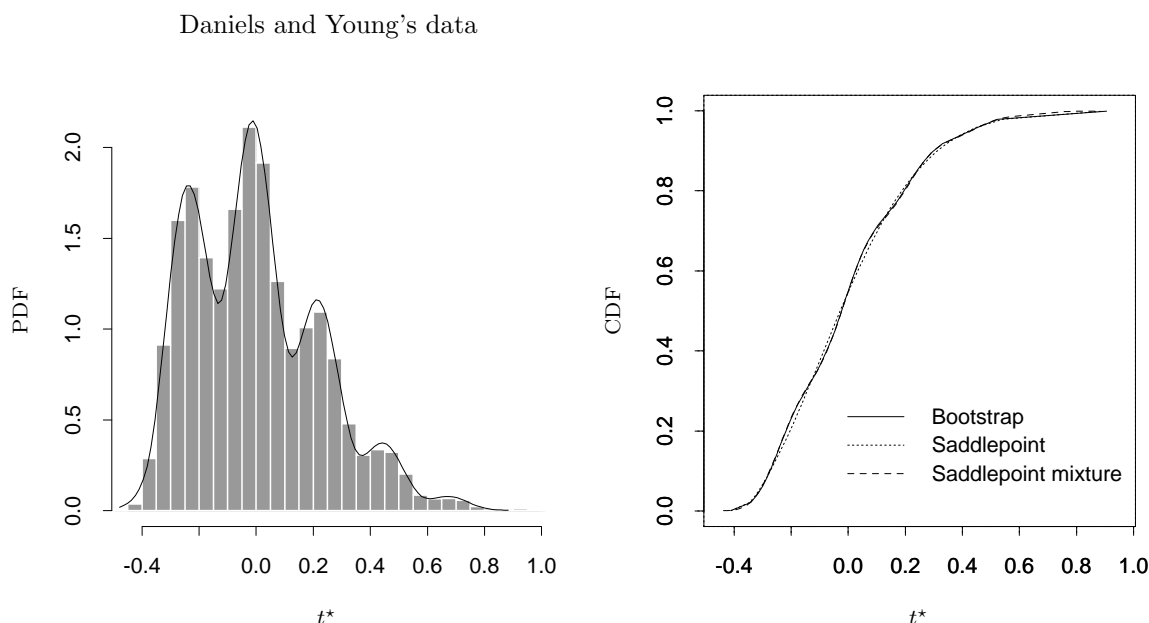
Daniels and Young's data

Figure 5.3. *Saddlepoint mixture approximations for Daniels and Young's data. Left panel: histogram of $R = 49,999$ replicates of the average. The solid line represents the saddlepoint mixture approximation to the density of the average. Right panel: bootstrap, saddlepoint and saddlepoint mixture approximation to the CDF.*

the average is shown. The multi-modal structure has been captured very well. In the right panel of Figure 5.3 approximations to the CDF are shown. The saddlepoint mixture approximation (dashed) is closer to the bootstrap approximation than the simple saddlepoint approximation (dotted) is. □

Example 5.4 (Rosner's data). These data are given in Rousseeuw and Leroy (1987, page 165) and represent a sample containing ten monthly diastolic blood pressure measurements:

$$40, 75, 80, 83, 86, 88, 90, 92, 93, 95.$$

A glance at these measurements, which originally appeared in Rosner (1977), reveals that 40 stands out compared to the other values. The corresponding saddlepoint approximations are shown in in the left panel of Figure 5.4. The usual saddlepoint approximation (solid) fails in the centre, whereas the saddlepoint mixture approximation (dotted) to the density of the average works very well over the entire range of possible values. For the distribution approximation, the right panel of Figure 5.4 suggests that both approximations work well. □

Example 5.5 (Artificial data). Let us consider a data set with one extreme value. We consider a data set containing 50 observations: 49 from an uniform distribution $U(0, 1)$ and an outlier at 10. In the left panel of Figure 5.5 the saddlepoint approximations to the density are superposed on the histogram of $R = 49,999$ bootstrap replicates of the average. Once again the saddlepoint mixture approximation (dotted) to the
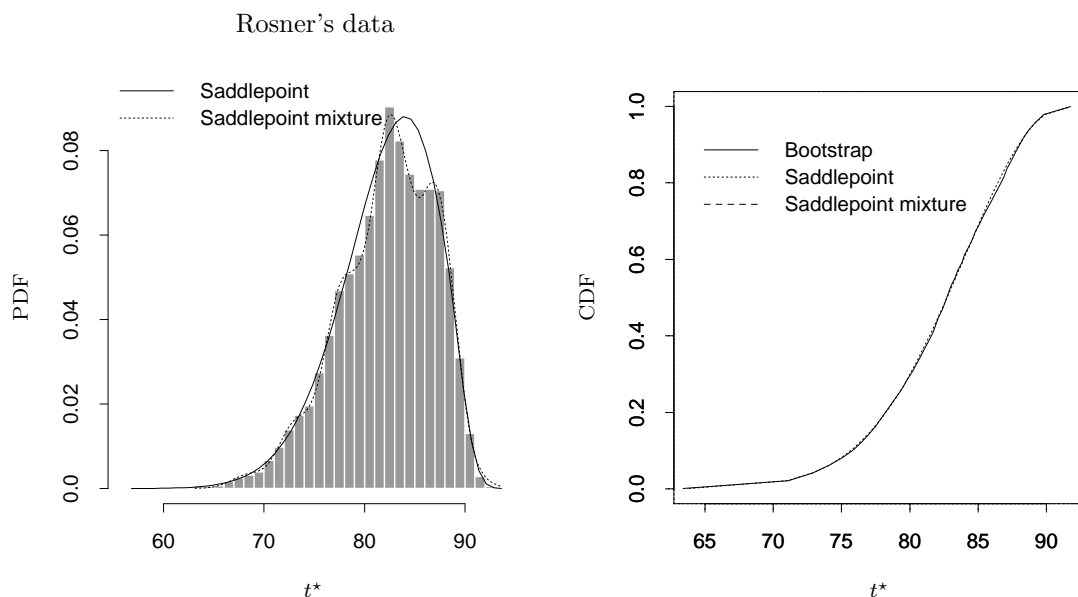
Rosner's data

Figure 5.4. *Saddlepoint mixture approximations for Rosner's data. Left panel: histogram of $R = 49,999$ replicates of the average. The solid line represents the saddlepoint mixture approximation to the density of the average. Right panel: bootstrap, saddlepoint and saddlepoint mixture approximation to the CDF.*

Artificial data

Figure 5.5. *Saddlepoint mixture approximations for the artificial data of Example 5.5. Left panel: histogram of $R = 49,999$ replicates of the average. The solid line represents the saddlepoint mixture approximation to the density of the average. Right panel: bootstrap, saddlepoint and saddlepoint mixture approximation to the CDF.*
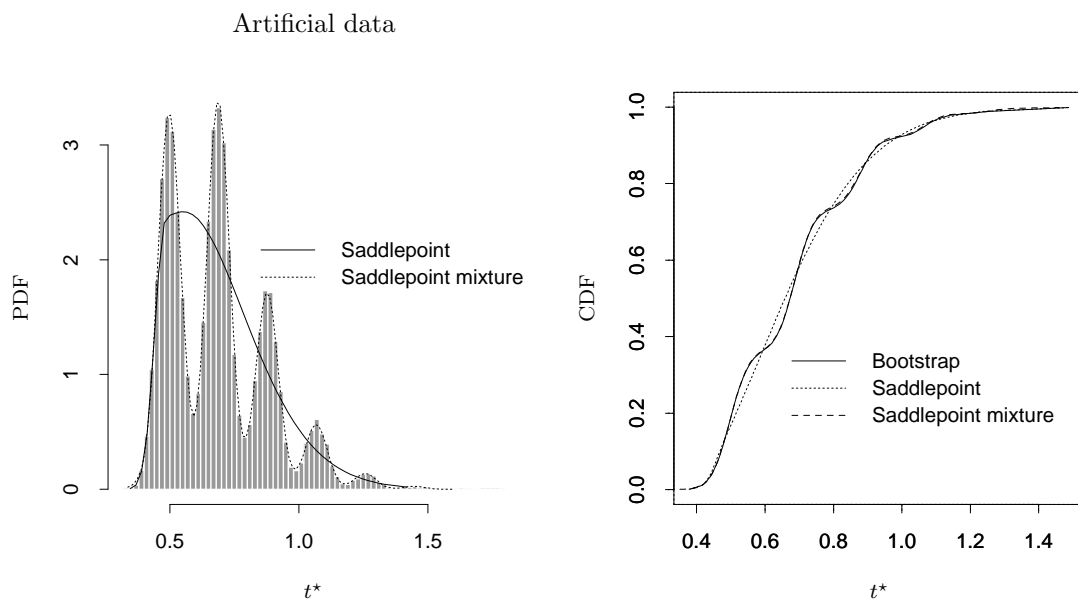
density is very accurate, whereas the usual saddlepoint approximation (solid) fails. For the approximation to the distribution given in the right panel of Figure 5.5 the same

conclusions can be drawn. □

5.2.2. Miscellany

In Example 5.2 the usual boxplot with a span of 3 times the interquartile range was used as an outlier detection rule. In order to develop a more general identification rule for outliers in the context of saddlepoint approximations one may find interesting ideas in the extensive statistical literature on the detection of outliers in a univariate sample; see, for instance, Rousseeuw and Leroy (1987), Davies and Gather (1993) or Barnett and Lewis (1994). One should also take into account the fact that the bootstrap statistic does not inherit the breakdown point of the original estimator. Moreover, Young and Daniels (1990) showed by means of two simple situations, involving the mean, that the bootstrap can be noticeably biased for small sample sizes.

The technique presented in this section is only valid if the data set contains a single outlier. But, it should be feasible to extend it to data sets containing two or more outliers. In this case the technique would consist in calculating a saddlepoint mixture approximation given $f_j = r$ and $f_k = s$, where the jth and the kth observations are outliers. This involves intensive calculations for the cumulant generating function. Moreover to get the mixture approximations more individual saddlepoint approximations are needed. The identification of the modes on the histogram of the bootstrap replicates is another unsolved problem. In considering a data set containing two extreme outliers, I noticed that an individual mode on the histogram of the replicates does not correspond to the case where both outliers have been resampled the same number of times. Finally, in considering data sets containing outliers, there may also be a *masking* effect, *e.g.* in considering an extreme outlier on the left and one on the right, it may be possible for the histogram not to show multi-modality. This is discussed in Barnett and Lewis (1994, Section 4.1.4), together with the related phenomenon of *swamping*.

One needs to question whether the saddlepoint mixture approximation is worth the trouble. First, who uses the average when data contains an extreme outlier? Having detected an extreme outlier, one would immediately use a robust estimate of location, or check if it is due to a measurement error and therefore omit it. Second, as the main advantage of saddlepoint approximations for bootstrap statistics is to deliver very accurate approximations far into the tails of the distribution, it seems to be interesting to know whether the saddlepoint mixture approximations outperforms the classical saddlepoint approximations in the tails. This is very important to know prior to any further generalization of the proposed method.

Let us consider Daniels and Young's data containing a single outlier, and the average as statistic of interest. As illustrated in Example 5.2 the mixture approximation works extremely well in the centre of the distribution, where the classical one fails (see Figures 5.1 and 5.3). In comparing the approximations to the PDF of the bootstrap distribution to the density of the average, these figures clearly illustrate the accuracy of the saddlepoint mixture approximation. But in considering the CDF approximation, one does not remark any big difference between them. Similar conclusions are suggested by other examples (Figures 5.4 and 5.5). As one is mostly interested in the distribution
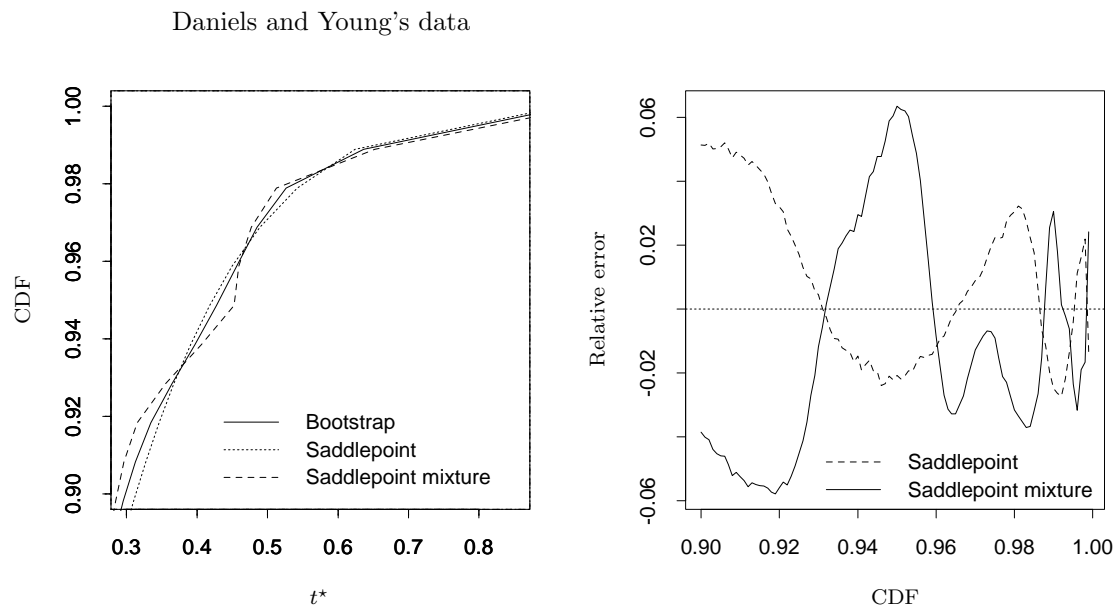
PSfrag replacements

Daniels and Young's data

Figure 5.6. *CDF approximations for Daniels and Young's data. Left panel: bootstrap ($R = 49,999$), saddlepoint and saddlepoint mixture approxima- tion to the CDF for values between 0.9 and 1. Right panel: relative errors* (Approximation − Bootstrap)/Bootstrap.

in the tails (*e.g.* for computing confidence intervals, or upper and lower quantiles), I try to answer this question in what follows.

The left panel of Figure 5.6 compares the bootstrap, saddlepoint and saddlepoint mixture approximation to the CDF for values between between 0.9 and 1. The sad- dlepoint approximation (dotted) is closer to the bootstrap distribution (solid) than the mixture approximation (dashed) is. The variability of the saddlepoint mixture approxi- mation seems to be larger. This fact was not reflected by the right panel of Figure 5.3, which suggested that the saddlepoint mixture approximation works better. The reason for this may be that the saddlepoint mixture approximation does not go as far into the tails as the saddlepoint approximation does. This may lie in the fact that for each r — the number of times the outlier has been resampled — a sort of confidence interval is calculated. As mentioned in Remark 5.4 it should be sufficient to consider a maximal value of $r = 4$. But the figure suggests that with $r = 4$ the value of the correspond- ing upper bound of the interval is not sufficiently far away in the tail. This was also reflected by the fact that in the present context the range of the saddlepoint mixture approximation was always included in that of the saddlepoint approximations — even for larger r. Moreover, if r is chosen too big we may get computational problems.

As another advantage of saddlepoint approximations is that they deliver relative and not absolute errors, it also may be interesting to compare the relative errors of the approximations. In the right panel of Figure 5.6 the CDF is plotted versus the relative errors of the approximations (computed by means of 100 evenly spaced points between 0.9 and 0.999). For the saddlepoint approximation (dashed) the variation of the relative error around zero is smaller than with the saddlepoint mixture approximation (solid).
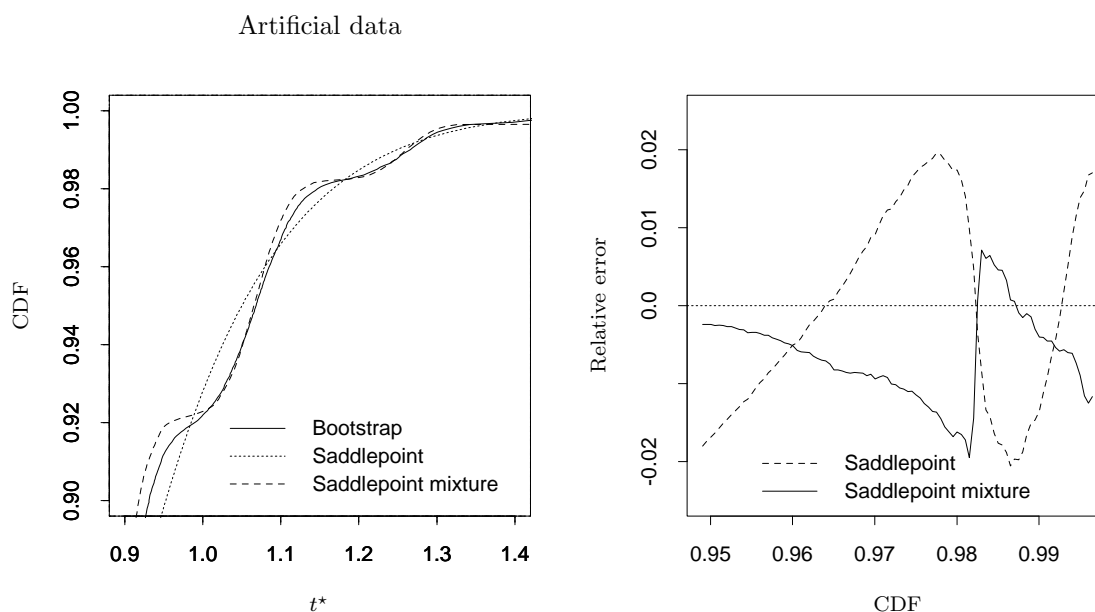
Artificial data

Figure 5.7. *CDF approximations for the artificial data of Example 5.5. Left panel: bootstrap (R = 49,999), saddlepoint and saddlepoint mixture approximation to the CDF for values between 0.9 and 1. Right panel: relative errors* (Approximation − Bootstrap)/Bootstrap.

Let us now perform the same analysis in the upper tail of the artificial data of Example 5.5; see Figure 5.7. In this case the saddlepoint mixture approximation seems to be more accurate in the extreme upper tail of the CDF. A possible explanation for this may be found in Young and Daniels (1990). They showed that for small sample sizes the bootstrap is biased. Note that Daniels and Young's data consist of $n = 10$ observations, whereas we have $n = 50$ for the artificial data set. Hence the sample size may be too small to compare both saddlepoint approximations appropriately, as already the underlying bootstrap distribution may be biased. As a result of this, a sample of size $n = 50$ should be sufficiently large to judge over the performance of the two approximations. Moreover, Jun Shao and Dongsheng Tu (1995, page 206) noted that

> 'The accuracy of various approximations to the bootstrap estimators is important. However, since the bootstrap estimators have their own errors in estimating the sampling distribution of a given statistic, it may not be worthwhile to reduce the error of the approximation to the bootstrap estimator much below the error of the bootstrap estimator.'

But, is the saddlepoint mixture really worth the trouble? Remember that having extreme outliers in the data would clearly lead us to use robust estimates of location and not the average. Robust statistics have been in use for hundreds of years (Stigler, 1973). More recently, a review of robust inference is given by Hampel (2000). Nowadays, robust statistics, described by Hampel (2001, page 1) as *'the stability theory of statistical procedures'*, have been proven to be very useful, and are widely used. For a discussion of the place of robustness (and statistics in general) in the field of tension between pure

mathematics and scientific applications, see Hampel (1997, 1998).

In the next section we look at the application of saddlepoint approximations for bootstrap statistics defined by scalar estimating equations. As applications we consider robust M-estimates of location. The generalization of the presented technique to studentized bootstrap statistics will be given in Chapter 6.

5.3. Estimating equations

This section illustrates how the saddlepoint approximations can be applied to the case in which the statistic is determined implicitly as the solution to estimating equations. As a nonlinear transformation is involved difficulties arise. Daniels (1983) compared two distinct ways of doing this which lead to different approximations of similar accuracy. He remarked that the one proposed in the next section appears to be the most convenient for approximating tail probabilities. Davison and Hinkley (1988, Section 4) applied this approximation method to the nonparametric bootstrap. They showed that saddlepoint approximations to bootstrap distributions of estimates based on monotone estimating equations can often replace simulation with excellent results; see also Davison and Hinkley (1997, Section 9.5).

The saddlepoint approximation to the density of an estimator defined by a scalar estimating equation (see Section 5.3.1 and Daniels (1983, Section 4) in more detail) is directly related to Field and Hampel's (1982) work. Indeed, Daniels (1983, Appendix) proved that his density approximation is the integrated form of Field and Hampel's approximation, *i.e.* by numerically integrating his saddlepoint density approximation we get directly their approximation; see also Field and Hampel (1982, Section 8). They extended the applicability of Hampel's (1974a) asymptotic expansion from the arithmetic mean to M-estimators of location with monotone estimating equations. Already Hampel (1974a) noted that integration of his approximation for the average gave precisely Daniel's (1954, page 633) saddlepoint approximation. That this is possible is indicated by the close relationship between the saddlepoint approximation and conjugate distributions (Daniels, 1954, page 639). In fact both approximations may yield identical results. A similar technique for densities of multivariate M-estimates was derived by Field (1982) as an adaptation of the saddlepoint technique of Daniels (1954) and the small sample method of Hampel (1974a). This can be seen also as a generalization of Field and Hampel (1982). Further details are given in Field and Ronchetti (1990, Chapters 4 and 6).

5.3.1. Saddlepoint approximations

The usual saddlepoint approximations can be extended to the case where the scalar statistic T is defined by an estimating equation of the form

$$\sum_{i=1}^{n} \psi(X_i, T) = 0, \tag{5.3.1}$$

where $\psi(x,\theta)$ is assumed to be a monotonic decreasing function in θ for all x, and $\mathrm{E}\{\psi(X,\theta)\} = 0$ for all θ.

Let us write $U(t) = \sum \psi(x_i, t)$. The corresponding bootstrap statistic $T^\star$ based on sampling from the data $x_1, \ldots, x_n$ is the solution of

$$U^\star(t) = \sum_{i=1}^{n} \psi(x_i, t) f_i^\star = 0,$$

where $f_i^\star$ denotes the bootstrap frequency of x_i. Since $U^\star(t)$ is monotone in t we have

$$\Pr(T^\star \le t) = \Pr\{U^\star(t) \le 0\}.$$

Indeed, if each $\psi(x_i, t)$ is monotonically decreasing in t, then

$$\Pr(T^\star \le t) = \Pr\left\{\sum_{i=1}^{n} \psi(x_i, T^\star) f_i^\star \ge \sum_{i=1}^{n} \psi(x_i, t) f_i^\star\right\} = \Pr\left\{\sum_{i=1}^{n} \psi(x_i, t) f_i^\star \le 0\right\}.$$

Hence, the saddlepoint approximation to the CDF of $T^\star$ can be approximated by that of $U^\star(t)$ by means of (2.4.2). Setting $d = 1$, $T = U^\star$, $t = 0$, $a_i = \psi(x_i, t)$ and $X_i = f_i^\star$ the approximation simply becomes $F_s(0)$. The cumulant generating function of $U^\star(t)$ is given by (5.1.6),

$$K(\zeta, t) = n \log\left[\sum_{i=1}^{n} p_i \exp\{\zeta\psi(x_i, t)\}\right]. \tag{5.3.2}$$

To calculate the saddlepoint approximation to the density of $T^\star$ we need the Jacobian of the transformation from $U^\star$ to $T^\star$. This Jacobian is denoted by $J(\zeta, t)$ and is $|\dot{K}(\zeta, t)/\zeta|$, where $\dot{K}(\zeta, t)$ is the first derivative of $K(\zeta, t)$ with respect to t. In our context we have

$$J(\zeta, t) = n\left|\frac{\displaystyle\sum_{i=1}^{n} p_i \exp\{\zeta\psi(x_i, t)\}\dot{\psi}(x_i, t)}{\displaystyle\sum_{i=1}^{n} p_i \exp\{\zeta\psi(x_i, t)\}}\right|, \tag{5.3.3}$$

where $\dot{\psi}(x_i, t)$ is the first derivative of $\psi(x_i, t)$ with respect to t. The saddlepoint $\hat{\zeta}$ is the solution to $K'(\zeta, t) = 0$, where prime denotes differentiation with respect to ζ. Thus the saddlepoint approximation to the density of $T^\star$ at t is (Daniels, 1983, equation (4.2))

$$f_s(t) = J(\hat{\zeta}, t)\left\{2\pi|K''(\hat{\zeta}, t)|\right\}^{-1/2} \exp\left\{K(\hat{\zeta}, t)\right\}. \tag{5.3.4}$$

This approximation can also be found by multiplying the saddlepoint approximation to the PDF of $U^\star(t)$ evaluated at 0, and given in (2.4.1), by the Jacobian evaluated at

$\hat{\zeta}$, $J(\hat{\zeta}, t)$. To compute the approximation we need the first and second derivatives of $K(\zeta, t)$ with respect to ζ, which are

$$K'(\zeta, t) = n \frac{\sum_{i=1}^{n} p_i \psi(x_i, t) \exp\{\zeta \psi(x_i, t)\}}{\sum_{i=1}^{n} p_i \exp\{\zeta \psi(x_i, t)\}}$$

and

$$K''(\zeta, t) = n \frac{\sum_{i=1}^{n} p_i \psi^2(x_i, t) \exp\{\zeta \psi(x_i, t)\}}{\sum_{i=1}^{n} p_i \exp\{\zeta \psi(x_i, t)\}} - \frac{1}{n} \{K'(\zeta, t)\}^2. \qquad (5.3.5)$$

The saddlepoint $\hat{\zeta}$ is the unique solution of $K'(\zeta, t) = 0$, hence at $\hat{\zeta}$ (5.3.5) becomes

$$K''(\hat{\zeta}, t) = n \frac{\sum_{i=1}^{n} p_i \psi^2(x_i, t) \exp\{\hat{\zeta} \psi(x_i, t)\}}{\sum_{i=1}^{n} p_i \exp\{\hat{\zeta} \psi(x_i, t)\}}.$$

Apart from the derivatives of the cumulant generating function, the only other quantity required is the derivative of $\psi(\cdot, t)$ with respect to t, which can generally be obtained easily from the definition of $\psi(\cdot, \cdot)$.

Remark 5.5. As the advantage of saddlepoint approximations is that they are very accurate far into the tails of the distribution, it is very important to use an appropriate range of values of t to estimate the entire distribution of $T^\star$. Recall that the quantiles of the distribution of $T^\star$ are estimated by obtaining the CDF approximation at m evenly spaced values of t, and then interpolating the CDF using a spline smoother. The range of values of t used is determined using a *binary search* (BS). Our modified version starts at the original value of the statistic and considers a value of t far away. Let $\mathcal{A}$ denote a sorted set of user defined requirements for the upper bound, for example that the saddlepoint approximation to the CDF at the upper bound is very close to 1. Now, we can check the midpoint of this interval against $\mathcal{A}$ and eliminate half of the interval from further consideration. The BS repeats this procedure, halving the size of the remaining portion of the interval each time. The algorithm stops if there is no significant difference between two consecutive values. Similarly, a BS is applied to find the lower bound with for example as requirement that the CDF approximation is very close to 0. Thus we find the effective range of t values needed to calculate the entire saddlepoint approximation. In practice, the effective range computed using the BS does not differ much from the range of the bootstrap replicates for data without outliers. More insight on the range of values of t will be given in Section 6.5.2.3 in the context of studentized bootstrap statistics. $\qquad \square$

5.3.2. M-estimates

An estimator defined by (5.3.1) is called an *M-estimator*. The name 'M-estimator' (Huber, 1964) comes from 'generalized *m*aximum likelihood'. A frequently used function ψ is *Huber's function*

$$\psi_k(x) = \begin{cases} x, & \text{if } |x| < k, \\ k\,\text{sign}(x), & \text{otherwise,} \end{cases} \tag{5.3.6}$$

where the cutoff point k is called the *tuning parameter*. The Huber estimators were motivated by studies in robust statistics concerning the influence of extreme data points on the estimate. Note that Huber's ψ is a monotone function, and that the resulting M-estimators have excellent properties (Hampel *et al.*, 1986, Section 2.4b).

For further details on M-estimates see, for example, Huber (1981, Chapter 3). There is a vast range of possible M-estimators. Many theoretical, numerical and simulation studies have been made. Some key references are the large-scale study by Andrews *et al.* (1972) resulting from the so-called *Princeton robustness year* and completed by Hampel (1995), see also Huber (1964), Hampel (1974b) and Hogg (1974) for additional studies. An important general treatment is given in Hampel *et al.* (1986).

Remark 5.6. The most common value of the tuning parameter in practice is $k = 1.345$. It is justified by the fact that the resulting M-estimate is about 95% as precise as the maximum likelihood estimate for Gaussian data. □

5.3.2.1. Huber's M-estimate of location

When estimating location it seems natural to use ψ-functions of the form $\psi(x, \theta) = \psi_k(x - \theta)$. We assume that (5.3.1) holds in order that T is Fisher consistent. To calculate the saddlepoint approximations to the CDF and PDF of Huber's M-estimate for location we only need

$$\frac{\partial}{\partial \theta} \psi_k(x - \theta) = \dot{\psi}_k(x - \theta) = \begin{cases} -1, & \text{if } |x - \theta| < k, \\ 0, & \text{otherwise.} \end{cases}$$

Using $\psi(x_i, t) = \psi_k(x - \theta)$ the saddlepoint approximation to the density of Huber's M-estimate of location can easily be obtained by means of (5.3.4), and the one for the CDF by calculating $F_s(0)$ in (2.4.2); see Section 5.3.1 for details.

Example 5.6 (Tuna data). Table 5.1 gives data from an aerial survey of schools of Southern Bluefin Tuna in the Great Australian Bight (Davison and Hinkley, 1997, Table 4.5). One interesting thing with this data set ($n = 64$) is that that there are four extreme values: 13.21, 13.27, 14.39 and 16.26. As our interest lies in location a robust estimate of location seems to be appropriate. The left panel of Figure 5.8 shows the histogram of the 49, 999 bootstrap replicates. The figure also shows the saddlepoint approximation to the bootstrap density of Huber's M-estimate with $k = 1.345$. In considering the data set we may have expected a multimodal structure, but as the figure clearly shows, the saddlepoint approximation works well when we use a robust estimate of location.

Table 5.1. *The tuna data come from an aerial line transect survey of Southern Bluefin Tuna in the Great Australian Bight. An aircraft with two spotters on board flies randomly allocated line transects. Each school of tuna sighted is counted and its perpendicular distance from the transect measured. The survey was conducted in summer when tuna tend to stay on the surface.*

0.19	0.28	0.29	0.45	0.64	0.65	0.78	0.85
1.00	1.16	1.17	1.29	1.31	1.34	1.55	1.60
1.83	1.91	1.97	2.05	2.10	2.17	2.28	2.41
2.46	2.51	2.89	2.89	2.90	2.92	3.03	3.19
3.48	3.79	3.83	3.94	3.95	4.11	4.14	4.19
4.36	4.53	4.97	5.02	5.13	5.75	6.03	6.19
6.19	6.45	7.13	7.35	7.77	7.80	8.81	9.22
9.29	9.78	10.15	11.32	13.21	13.27	14.39	16.26

PSfrag replacements

Tuna data

Figure 5.8. *Saddlepoint approximations for the tuna data. Left panel: histogram of $R = 49,999$ replicates of Huber's M-estimate of location. The solid line represents the saddlepoint approximation to the bootstrap density of Huber's M-estimate of location. Right panel: bootstrap and saddlepoint approximation estimates of the CDF.*

The right panel of Figure 5.8 illustrates that the saddlepoint approximation is very accurate far into the tails of the distribution. It took 21,031 seconds (about 5.8 hours) in CPU time to get the bootstrap distribution and 5 seconds to get the saddlepoint approximation. □

Table 5.2. *Newcomb's measurements of the passage time of light to traverse a known distance. The given values multiplied by 10^{-3} plus 24.8 give the time in millionth of a second for light to traverse a known distance.*

-44	-2	16	16	19	20	21	21	22	22	23
23	23	24	24	24	24	24	25	25	25	25
25	26	26	26	26	26	27	27	27	27	27
27	28	28	28	28	28	28	28	29	29	29
29	29	30	30	30	31	31	32	32	32	32
32	33	33	34	36	36	36	36	37	39	40

5.3.2.2.   Huber's M-estimate of location with initial MAD scaling

In many statistical models the parameter splits naturally into a main part and a nuisance part. It is intuitively clear that the maximal bias of the main part in a contamination model (as is usually the case with real data) depends on the maximal bias of the nuisance part. Therefore it would be interesting to treat both parameters. This problem can be solved by defining T through $\psi(x, \theta) = \psi_k\{(x - \theta)/S\}$ where S is a robust estimator of scale. Hampel *et al.* (1986, page 105) recommend the use of MAD scaling for M-estimators — MAD stands for 'median absolute deviation about the median'. The estimator is

$$S \;\; = \;\; \beta\mathrm{MAD}(x_1, \ldots, x_n) = \beta\mathrm{med}\{|x_1 - \mathrm{med}(x_1, \ldots, x_n)|, \ldots, |x_n - \mathrm{med}(x_1, \ldots, x_n)|\}, \tag{5.3.7}$$

where $\mathrm{med}(\cdot)$ denotes the median and β equals $1/\Phi^{-1}(3/4)$ to obtain Fisher consistency for a Gaussian sample. Throughout the rest of this thesis the estimator (5.3.7) will be simply referred to as MAD. The corresponding bootstrap statistic $T^\star$ solves

$$\sum_{i=1}^{n} \psi_k\left(\frac{x_i^\star - T^\star}{S}\right) = 0,$$

where S is the MAD of the original data set $x_1, \ldots, x_n$. Hence the saddlepoint approximations for this standardized M-estimator of location with MAD estimator of scale can be obtained with $\psi(x_i, t) = \psi_k\{(x_i - t)/s\}$ and $\dot{\psi}(x_i, t) = -I\{|(x_i - t)/s| < k|\}/s$, where dot denotes the derivative of ψ with respect to t.

Example 5.7 (Newcomb's data). Newcomb's measurements of the passage time of light are given in Table 5.2 (Stigler, 1977, Table 5, Data Set 23). This data set has $n = 66$ and contains two outliers: -44 and -2. In the left panel of Figure 5.9 the saddlepoint approximation to the PDF of Huber's M-estimate of location with initial MAD scaling is shown. The saddlepoint approximation works very well. And the CDF approximation given in the right panel of Figure 5.9 underlines the accuracy of the approximation.
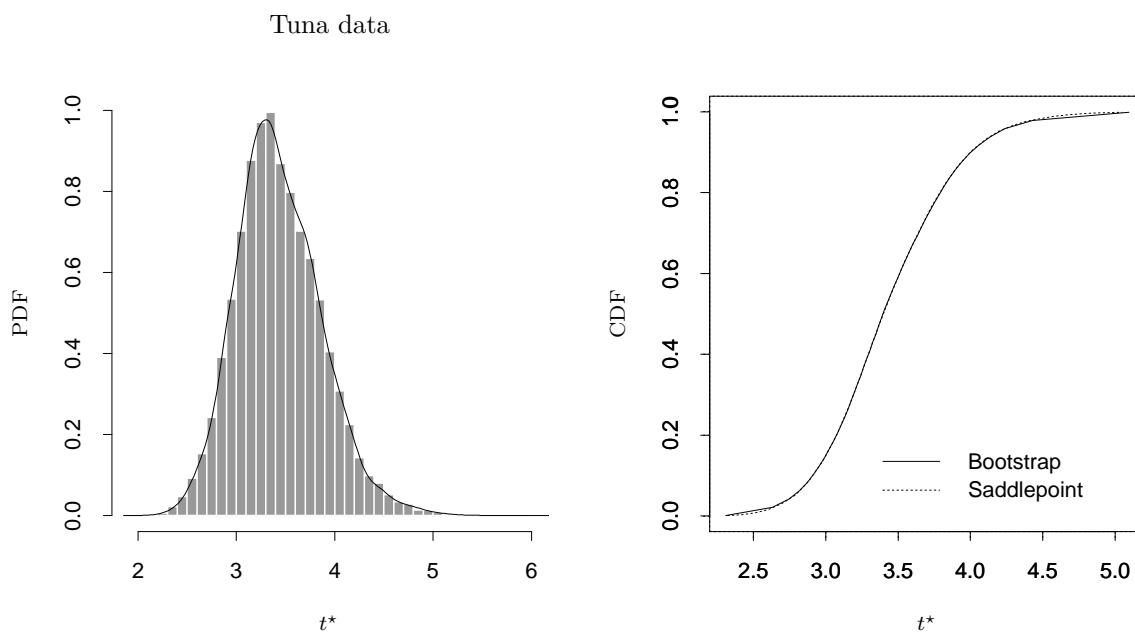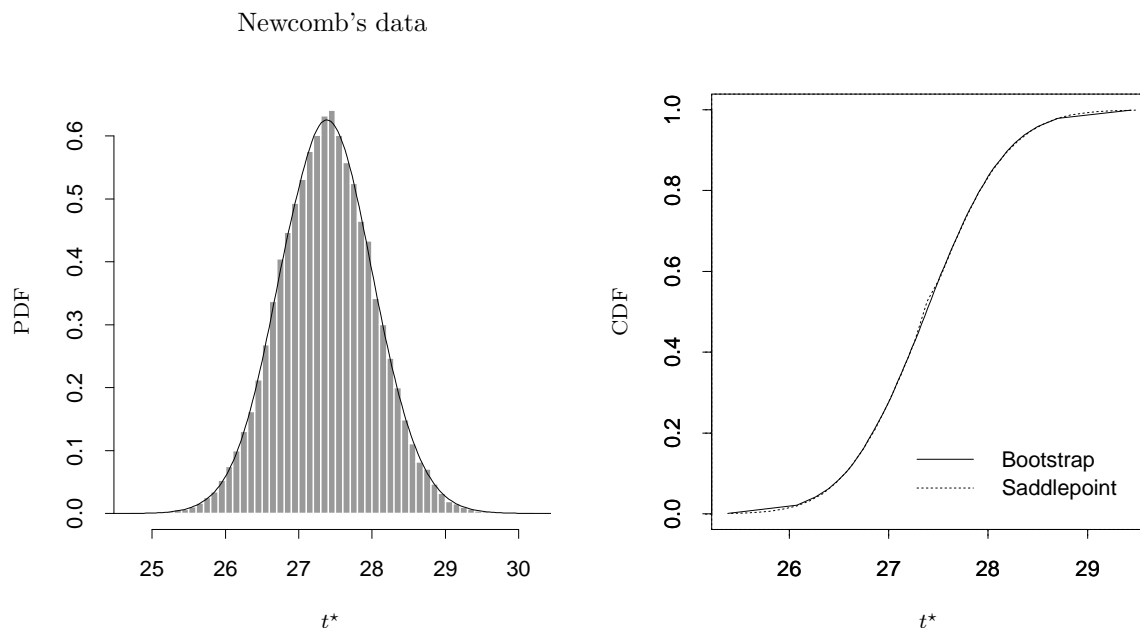
Newcomb's data

Figure 5.9. *Saddlepoint approximations for Newcomb's data. Left panel: histogram of $R = 49,999$ replicates of Huber's M-estimate of location with initial MAD scaling. The solid line represents the saddlepoint approximation to the bootstrap density of Huber's M-estimate of location with initial MAD scaling. Right panel: bootstrap and saddlepoint approximation estimates of the CDF.*

Moreover, the saddlepoint approximations are much faster in CPU time (about 3 seconds compared to the 3,786 seconds (about 63 minutes) to get the $R = 49,999$ bootstrap replicates). $\square$

Additional examples of the approach presented in this section can be found in Kuonen (1998a, Examples 3.1 and 3.3) and Kuonen (1998b, pages 15 and 16) for Huber's M-estimate of location, and in Kuonen (1998a, Examples 3.4 and 3.5) and Kuonen (1998b, pages 18 and 19) for Huber's M-estimate of location with initial MAD scaling. The conclusions remain the same, illustrating that the advantage of saddlepoint approximations is twofold: they deliver very accurate approximation, and they outperform the 'brute-force' bootstrap by a significant factor.

5.4.   Conclusion

xample 5.2 illustrated a limitation of the saddlepoint approximation presented in Section 5.1.3 in the case where the data set contains outliers. Therefore in Section 5.2 I generalized the classical approach by introducing a saddlepoint mixture approximation in conditioning on the number of times the outlier has been resampled. I showed in examples that this approach delivers very accurate approximations. Nevertheless, one may question the use of a statistic like the average when data contain outliers. To enable the use of robust estimates of location a technique

was presented in Section 5.3. As applications Huber's M-estimate of location and its initial MAD scaled version were considered. In the next chapter the approach will be generalized to give approximations for studentized bootstrap statistics, which are very important in the bootstrap context.

6. Studentized bootstrap distributions

'At present, some of the most interesting and at the same time most promising new methods and open problems have to do with small sample situations. ... In small sample situations simulation is indispensable. ... On the other hand, it was noted by Hampel that a variant of the saddle point method can give fantastically accurate approximations down to very small sample sizes. ... I hope that in the near future these approaches will bring to a close several open problems in the area of small samples: studentization, confidence intervals, and testing.'

Peter J. Huber (1996, page 62)

Any estimate of any parameter should be accompanied by an estimate of its own variability. Together they form the basis for the construction of confidence intervals. There are several ways of using bootstrap results in this context. The simplest consists in using a normal approximation to the distribution of a pivot and hence to derive normal confidence intervals. Another, more accurate and general device is the *studentized bootstrap* or *bootstrap-t method*; see Efron (1982) or Hall (1988). This method will be described briefly in Section 6.1 and in more detail in Chapter 7, whereas the corresponding saddlepoint approximations in order to replace the computer-intensive bootstrap computations will be presented in Section 6.2. These can be seen as a generalisation of the ones seen in Section 5.3.1, as in the studentized context the statistic of interest is the solution of a system of equations. The first such example (Daniels and Young, 1991) generalised Davison and Hinkley's (1988) saddlepoint approximations for the unstudentized bootstrap mean to the studentized bootstrap mean. Daniels and Young (1991) first obtained a bivariate saddlepoint approximation, then, after a non-linear transformation, integrated out an unwanted variable either numerically or by Laplace approximation to obtain the marginal distribution of the statistic of interest. This approach was extended by Davison *et al.* (1995, Section 3) and is summarised in Davison and Hinkley (1997, Section 9.5.3). This method is also known as the *integration approach* and will be discussed in more detail in Section 6.2. Concerning the accuracy and coverage of bootstrap confidence intervals Jing *et al.* (1994), based on results from Jing and Robinson (1994), used saddlepoint approximations in order to study relative errors in much the way Hall (1988) used Edgeworth methods. Based on these works

Feuerverger *et al.* (1999) showed the second-order relative accuracy, on bounded sets, of the studentized bootstrap by exploiting connections between Edgeworth expansions obtained by DiCiccio and Romano (1990) and saddlepoint approximations.

In the context of robust inference, related work started with Field's (1982) saddlepoint approximation to the joint density of multivariate M-estimates. Their relation was already discussed briefly in Section 5.3. To overcome the substantial computational requirement of numerical integration in order to obtain the marginal density Fan and Field (1995) derived saddlepoint approximations to marginal densities for M-estimators, which are based on Field (1982) and Tingley and Field (1990). A more general form and justifications of the regularity assumptions required in the derivation of their approximations can be found in Field and Ronchetti (1990, Chapter 6). In addition, Fan and Field (1995) computed the marginal distribution using formulae given in DiCiccio *et al.* (1991) and DiCiccio and Martin (1991). A review of saddlepoint approximations for marginal distributions in this context is given by Ronchetti (1996). More recently, an additional extension is provided by Almudevar *et al.* (2000).

After stating the integration approach in Section 6.2 I will consider in Section 6.3 the classical studentized statistic, and in Section 6.4 the studentized versions of Huber's M-estimate of location, of Huber's M-estimate of location with initial MAD scaling and of Huber's proposal 2. Finally, additional remarks on implementation and related problems are given in Section 6.5. This chapter forms the basis of Chapter 7, where the resulting studentized confidence intervals will be discussed in detail.

6.1. Introduction

ivotal quantities play a crucial role in the theory of confidence intervals. Much of the bootstrap literature concerns the construction of approximate confidence intervals in nonparametric settings. This raises the question: what is a good pivot in nonparametric estimation problems? The answer to this question is nicely illustrated in Efron (1992, Section 5); see also Davison and Hinkley (1997, Section 2.4).

Let T be an estimator of the parameter θ. The quantity $T - \theta$ is a poor choice in most situations as its distribution usually depends on unknowns. A second guess is a t-like statistic, say

$$Z = \frac{T - \theta}{V^{1/2}}, \tag{6.1.1}$$

where T is the estimator of location and V is a consistent estimator of its variance. The quantity Z is called an *approximate pivot*, whose distribution is approximately the same for each value of θ. This property enables the use of Z to construct accurate confidence intervals. Throughout the rest of this thesis I use Z to denote a studentized statistic. One simple approximation is to take Z to be $N(0, 1)$. This is often valid as $n \to \infty$ and is only an approximation for finite samples. A slightly better approximation is given by Student's t distribution on $n - 1$ degrees of freedom. One can expect that Z will be behave like a t-statistic, but there is no guarantee of having $n - 1$ degrees of freedom.

Huber (1981, Section 6.8) noted that (6.1.1) works remarkably well for M-estimates; see also the extensive simulation study by Shorack (1976), inspired by the Princeton study (Andrews *et al.*, 1972). But the use of Student's t distribution as an approximation does not adjust the resulting confidence interval to account for skewness in the underlying population or other errors; see Efron and Tibshirani (1993, Section 12.4). Recently, Boos and Hughes-Oliver (2000) tried to derive rules of thumb for the value of n needed for normal and t intervals in order to give appropriate coverage probabilities. By means of examples they illustrated that it depends mostly on the underlying skewness than on the kurtosis. For the t intervals this was already discussed in Johnson (1978).

A procedure which is known to be accurate is to estimate the distribution of Z from replicates of the studentized bootstrap statistic,

$$Z^\star = \frac{T^\star - t}{V^{\star 1/2}}, \tag{6.1.2}$$

where t denotes the observed value of the statistic T, and $T^\star$ and $V^\star$ are based on a simulated sample, $X_1^\star, \ldots, X_n^\star$. Hence we can obtain accurate intervals without having to make normal or Student theory assumptions. This procedure estimates the distribution of Z directly from the data. Following Remark 5.2 a $(1-2\alpha)$ basic bootstrap confidence interval has limits

$$z_{(\alpha(R+1))}^\star, \quad z_{((1-\alpha)(R+1))}^\star.$$

Using (6.1.1) the $(1-2\alpha)$ studentized bootstrap confidence interval limits for θ have the form

$$t - v^{1/2} z_{((1-\alpha)(R+1))}^\star, \quad t - v^{1/2} z_{(\alpha(R+1))}^\star, \tag{6.1.3}$$

where $z_{(\alpha(R+1))}^\star$ is the $\alpha(R+1)$th order statistic of the simulated values $z_1^\star, \ldots, z_R^\star$. Note that this method, also known as bootstrap-t method, gives second-order accurate and correct intervals in a wide variety of situations; see Hall (1988, 1992), Abramovitch and Singh (1985), and DiCiccio and Efron (1992, 1996). Bradley Efron (1992, page 120) concluded his investigations with:

> '*The Z statistic (6.1.1) is a reasonable answer to the question*
> '*what is a good nonparametric pivotal quantity?*"

In summary, studentization yields a stable pivotal that circumvents many problems.

Remark 6.1. Hu and Kalbfleisch (2000) proposed a bootstrap procedure which estimates the distribution of an estimating function by resampling its terms using bootstrap techniques. Their method concentrates attention on the estimating function and equation from which the estimator is obtained. Studentized versions of their so-called *estimating function bootstrap* yield methods which are invariant under reparametrizations. But, especially using M-estimates this may be difficult to compute in the presence of nuisance parameters (Léger, 2000). Additional drawbacks of this procedure were also illustrated by Canty and Davison (2000).   □

6.2. Saddlepoint approximations

n this section the saddlepoint approximations of Section 5.3.1 are extended to enable the computation of studentized statistics. Suppose that the statistic of interest, T, and some nuisance statistics, $S = (S_1, \ldots, S_{d-1})^T$, are the solution to the d estimating equations

$$U(t, s) = \sum_{i=1}^{n} \psi(x_i, t, s) = 0, \tag{6.2.1}$$

where $\psi(x_i, t, s)$ is a $d \times 1$ vector. The bootstrap quantities $T^\star$ and $S^\star$ are the solutions of the equation

$$U^\star(t, s) = \sum_{i=1}^{n} \psi(x_i, t, s) f_i^\star = 0, \tag{6.2.2}$$

where the frequencies $(f_1^\star, \ldots, f_n^\star)$ have a multinomial distribution with denominator n and probability vector $(p_1, \ldots, p_n)$; typically $p_i \equiv n^{-1}$. We assume that there exists an unique solution to (6.2.2). Following (5.3.2) the cumulant generating function of $U^\star(t, s)$ for fixed t and s is

$$K(\zeta, t, s) = n \log \left[\sum_{i=1}^{n} p_i \exp\{\zeta^T \psi(x_i, t, s)\} \right] \tag{6.2.3}$$

and the saddlepoint approximation to the PDF of $U^\star(t, s)$ is given by (2.4.1) using $T = U^\star$, $a_i = \psi(x_i, t, s)$ and $X_i = f_i^\star$. To get the saddlepoint approximations to the marginal PDF and CDF of $T^\star$ we need the Jacobian for the transformation from $U^\star$ to $T^\star$ and $S^\star$. This is hard to obtain exactly, but following the arguments by Spady (1991, Section 2) a good approximation is given by

$$J(\zeta, t, s) = \left| n \sum_{i=1}^{n} w(x_i, \zeta, t, s) \left\{ \frac{\partial \psi(x_i, t, s)}{\partial t}, \frac{\partial \psi(x_i, t, s)}{\partial s^T} \right\} \right|, \tag{6.2.4}$$

where

$$w(x_i, \zeta, t, s) = \frac{p_i \exp\{\zeta^T \psi(x_i, t, s)\}}{\displaystyle\sum_{k=1}^{n} p_k \exp\{\zeta^T \psi(x_k, t, s)\}}. \tag{6.2.5}$$

The Jacobian (6.2.4) reduces to the Jacobian (5.3.3) when s is not present. Thus, similar to the arguments in Section 5.3.1 one can use the fact that

$$\Pr(T^\star = t, \, S^\star = s) = J(\zeta, t, s) \, \Pr\{U^\star(t, s) = 0\}.$$

Hence using (2.4.1) the saddlepoint approximation to the joint density of $T^\star$ and $S^\star$ is

$$f_{T^\star, S^\star}(t, s) = J(\hat{\zeta}, t, s)(2\pi)^{-d/2} |K''(\hat{\zeta}, t, s)|^{-1/2} \exp\left\{ K(\hat{\zeta}, t, s) \right\}, \tag{6.2.6}$$

where $\hat{\zeta} = \hat{\zeta}(t,s)$ is the solution of the d equations $\partial K(\zeta,t,s)/\partial \zeta = 0$, and $K''(\hat{\zeta},t,s)$ is the $d \times d$ matrix with elements $\partial^2 K(\hat{\zeta},t,s)/\partial \zeta \partial \zeta^T$. We now apply Laplace's method to the integral of (6.2.6). Laplace's method is described in most books on asymptotic techniques; see for example Barndorff-Nielsen and Cox (1989, Sections 3.3 and 6.2). It is usually preferable to use the Laplace method instead of numerical integration; see also Section 6.5.2. Provided that the matrix $\partial^2 K(\hat{\zeta},t,s)/\partial s \partial s^T$ is negative definite, the Laplace approximation to the integral of (6.2.6) with respect to $s = (s_1, \ldots, s_{d-1})$ at t can be obtained in the following way. Let us write

$$f_{T^\star}(t) = \int f_{T^\star,S^\star}(t,s)ds = (2\pi)^{-d/2} \int r(s)\exp\{-h(s)\}ds, \qquad (6.2.7)$$

where $r(s) = J(\hat{\zeta},t,s)|K''(\hat{\zeta},t,s)|^{-1/2}$ and $h(s) = -K(\hat{\zeta},t,s)$. Then the approximation to $f_{T^\star}(t)$ is

$$f_{T^\star}(t) \doteq (2\pi)^{-d/2}(2\pi)^{(d-1)/2}r(\tilde{s})|h_2(\tilde{s})|^{-1/2}\exp\left\{-h(\tilde{s})\right\},$$

where $\tilde{s} = \tilde{s}(t)$ solves the $(d-1)\times 1$ system of equations $\partial h(s)/\partial s = 0$ and $|h_2(\tilde{s})|$ is the determinant of the $(d-1)\times(d-1)$ matrix of second derivatives of $h(s)$ with respect to s evaluated at $\tilde{s}$, which is assumed to be positive definite. Hence, the approximate marginal density of $T^\star$ at t is

$$f_s(t) = J(\tilde{\zeta},t,\tilde{s})\left\{2\pi|K''(\tilde{\zeta},t,\tilde{s})|\left|\frac{\partial^2 K(\hat{\zeta},t,\tilde{s})}{\partial s \partial s^T}\right|\right\}^{-1/2}\exp\left\{K(\tilde{\zeta},t,\tilde{s})\right\}, \qquad (6.2.8)$$

where $\tilde{\zeta} = \tilde{\zeta}(t)$ and $\tilde{s}$ are functions of t that simultaneously satisfy the $d\times 1$ and $(d-1)\times 1$ systems of equations

$$\frac{\partial K(\zeta,t,s)}{\partial \zeta} = 0, \qquad \frac{\partial K(\zeta,t,s)}{\partial s} = 0. \qquad (6.2.9)$$

These $2d-1$ equations can be solved using packaged routines. In practice, I used the S-PLUS function `nlmin`, which finds a local minimum using a general Newton optimiser. The R function `nlm` could be used as well. To do so one can minimise the sum of squares

$$\frac{\partial K(\zeta,t,s)}{\partial \zeta}^T\frac{\partial K(\zeta,t,s)}{\partial \zeta} + \frac{\partial K(\zeta,t,s)}{\partial s}^T\frac{\partial K(\zeta,t,s)}{\partial s}$$

and ensure that the minimum value is zero. For starting values, note that when t equals its sample value, say t_0, we have approximately that $\tilde{\zeta} = 0$ and that $\tilde{s}$ takes its values from the original data. Davison *et al.* (1995) noticed that this choice of initial values is about equally fast as one-step approximations to $\tilde{\zeta}$ and $\tilde{s}$ (DiCiccio *et al.*, 1992a). Moreover, the same initial values for the particular case of the studentized mean were already used by Daniels and Young (1991).

Furthermore, $\partial^2 K(\hat{\zeta},t,s)/\partial s \partial s^T$ in (6.2.8) denotes the $(d-1)\times(d-1)$ matrix whose elements are

$$\frac{\partial^2 K(\zeta,t,s)}{\partial s \partial s^T} - \frac{\partial^2 K(\zeta,t,s)}{\partial s \partial \zeta^T}\left\{\frac{\partial^2 K(\zeta,t,s)}{\partial \zeta \partial \zeta^T}\right\}^{-1}\frac{\partial^2 K(\zeta,t,s)}{\partial \zeta \partial s^T}. \qquad (6.2.10)$$

As $K(\zeta, t, s)$ is given by (6.2.3), we have

$$\frac{\partial K(\zeta, t, s)}{\partial \zeta} = n \sum_{i=1}^{n} w(x_i, \zeta, t, s)\psi(x_i, t, s), \qquad (6.2.11)$$

$$\frac{\partial K(\zeta, t, s)}{\partial s} = n \sum_{i=1}^{n} w(x_i, \zeta, t, s)\frac{\partial \psi(x_i, t, s)^T}{\partial s}\zeta.$$

In applying this to (6.2.9) we get $\tilde{\zeta}$ and $\tilde{s}$. Moreover,

$$\frac{\partial^2 K(\zeta, t, s)}{\partial \zeta \partial \zeta^T} = n \sum_{i=1}^{n} w(x_i, \zeta, t, s)\psi(x_i, t, s)\psi(x_i, t, s)^T - \frac{1}{n}\left\{\frac{\partial K(\zeta, t, s)}{\partial \zeta}\right\}^2.$$

In the same way, we obtain

$$\frac{\partial^2 K(\zeta, t, s)}{\partial \zeta \partial s^T} = n \sum_{i=1}^{n} w(x_i, \zeta, t, s)\left\{\frac{\partial \psi(x_i, t, s)}{\partial s^T} + \psi(x_i, t, s)\zeta^T\frac{\partial \psi(x_i, t, s)}{\partial s^T}\right\}$$
$$- \frac{1}{n}\left\{\frac{\partial K(\zeta, t, s)}{\partial \zeta}\frac{\partial K(\zeta, t, s)}{\partial s}\right\}.$$

Hence at $\tilde{\zeta}$ and $\tilde{s}$ the matrices in (6.2.10) become

$$\frac{\partial^2 K(\tilde{\zeta}, t, \tilde{s})}{\partial \zeta \partial \zeta^T} = n \sum_{i=1}^{n} w(x_i, \zeta, t, s)\psi(x_i, t, s)\psi(x_i, t, s)^T,$$

$$\frac{\partial^2 K(\tilde{\zeta}, t, \tilde{s})}{\partial \zeta \partial s^T} = n \sum_{i=1}^{n} w(x_i, \tilde{\zeta}, t, \tilde{s})\left\{\frac{\partial \psi(x_i, t, \tilde{s})}{\partial s^T} + \psi(x_i, t, \tilde{s})\zeta^T\frac{\partial \psi(x_i, t, \tilde{s})}{\partial s^T}\right\}$$

and

$$\frac{\partial^2 K(\tilde{\zeta}, t, \tilde{s})}{\partial s_l \partial s_k} = n \sum_{i=1}^{n} w(x_i, \tilde{\zeta}, t, \tilde{s})\left\{\zeta^T\frac{\partial^2 \psi(x_i, t, \tilde{s})}{\partial s_l \partial s_k} + \zeta^T\frac{\partial \psi(x_i, t, \tilde{s})}{\partial s_l}\zeta^T\frac{\partial \psi(x_i, t, \tilde{s})}{\partial s_k}\right\},$$

where s_l and s_k denote the lth and kth elements of s. The saddlepoint approximation to the CDF of $T^\star$ at t is $F_s(t)$ given in (2.4.2) with

$$w = \text{sign}(t - t_0)\left\{2K(\tilde{\zeta}, t, \tilde{s})\right\}^{1/2}, \qquad (6.2.12)$$

$$v = -\frac{\partial K(\tilde{\zeta}, t, \tilde{s})}{\partial t}|K''(\tilde{\zeta}, t, \tilde{s})|^{1/2}\left|\frac{\partial^2 K(\hat{\zeta}, t, \tilde{s})}{\partial s \partial s^T}\right|^{1/2}\{J(\tilde{\zeta}, t, \tilde{s})\}^{-1}. \qquad (6.2.13)$$

The only additional quantity needed is

$$\frac{\partial K(\tilde{\zeta}, t, \tilde{s})}{\partial t} = n \sum_{i=1}^{n} w(x_i, \tilde{\zeta}, t, \tilde{s})\zeta^T\frac{\partial \psi(x_i, t, \tilde{s})}{\partial t}.$$

One remarks that the problems in this approach may mainly be computational. The only analytical work that is required consists in the knowledge of the estimating equation function $\psi(x_i, t, s)$ and its derivatives, which define the statistic of interest $T^\star$ and the nuisance statistics $S^\star$.

6.3.   Classical studentized statistic

o illustrate the ideas in the previous section I first consider the classical studentized statistic, which consists in using (6.1.1) with $T = \bar{X}$ and $V = S = n^{-1}\sum(X_i - \bar{X})^2$. The resulting studentized bootstrap statistic is given in (6.1.2), where the variance of the bootstrap sample is $V^\star = S^\star = n^{-1}\sum(X_i^\star - \bar{X}^\star)^2 = n^{-1}\sum X_i^{\star 2} - \bar{X}^{\star 2}$, the average of the bootstrap sample is $T^\star = \bar{X}^\star$ and t denotes the observed value of the statistic T given by $t = \bar{x} = n^{-1}\sum x_i$. As this studentized statistic is location-invariant we can replace x_i with $x_i - \bar{x}$, and therefore we have $t = 0$. Thus (6.1.2) reduces to

$$Z^\star = \frac{\bar{X}^\star}{S^{\star 1/2}}, \tag{6.3.1}$$

which is the studentized mean for the bootstrap sample. This application was already treated by Daniels and Young (1991) and in Davison and Hinkley (1997, Example 9.19). It can be shown that Z and S are the solutions of the $d = 2$ estimating equations given in (6.2.1) with

$$\psi(x_i, z, s) = \left(\begin{array}{c} x_i - zs^{1/2} \\ x_i^2 - s(1 + z^2) \end{array} \right).$$

Indeed, $U(z, s) = 0$ splits up in $\sum x_i = zs^{1/2}n$ and $\sum x_i^2 = s(1 + z^2)n$, which yield $\bar{x}/s^{1/2} = z$ and $n^{-1}\sum x_i^2 - \bar{x}^2 = s$.

Using $T = Z$ and $S = S$ the marginal saddlepoint approximations of $Z^\star$ are given in (6.2.8) for the PDF and in (2.4.2) for the CDF using (6.2.12) and (6.2.13). The required 2×1 matrices of derivatives are

$$\frac{\partial \psi(y_i, z, s)}{\partial z} = \left(\begin{array}{c} -s^{1/2} \\ -2sz \end{array} \right), \quad \frac{\partial \psi(y_i, z, s)}{\partial s} = \left(\begin{array}{c} -zs^{-1/2}/2 \\ -(1 + z^2) \end{array} \right)$$

and

$$\frac{\partial^2 \psi(y_i, z, s)}{\partial s^2} = \left(\begin{array}{c} zs^{-3/2}/4 \\ 0 \end{array} \right).$$

Example 6.1 (Cushny and Peebles' data). Let us look at the data by Cushny and Peebles (1905) on the prolongation of sleep by means of two drugs. The original data set was bivariate: for ten subjects, two different values were recorded (one for each drug). These data are cited in numerous books as an example of a normally distributed sample; see Hampel *et al.* (1986, Section 2.0). The $n = 10$ pairwise differences, *i.e.* the differences between drug effects per subject, are the following:

$$0.0, 0.8, 1.0, 1.2, 1.3, 1.3, 1.4, 1.8, 2.4, 4.6.$$

However, a glance at these numbers reveals that the normality assumption is questionable, due to the occurrence of 4.6 which appears to be an outlier. As seen in Section 5.2
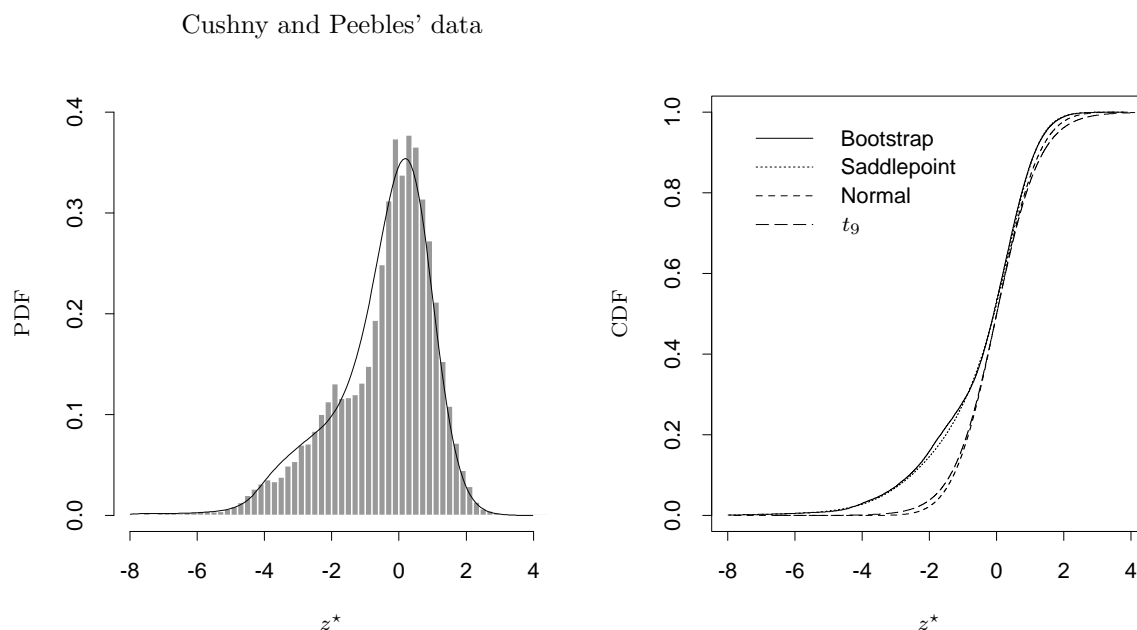
Cushny and Peebles' data

Figure 6.1. *Saddlepoint approximations for Cushny and Peebles' data. Left panel: histogram of $R = 49,999$ replicates of the classical studentized statistic. The solid line represents the saddlepoint approximation to the bootstrap density of the classical studentized statistic. Right panel: bootstrap, saddlepoint, normal and Student-t approximation estimates of the CDF.*

Table 6.1. *Short's determinations of the parallax of the sun.*

7.35	7.68	7.75	7.77	7.92	8.07	8.23
8.23	8.35	8.42	8.43	8.50	8.50	8.50
8.60	8.65	8.90	9.09	9.61	10.15	10.33

the saddlepoint approximation to the unstudentized average would break down. One might hope that using the studentized average this phenomenon does not occur. Indeed, the left panel of Figure 6.1 illustrates that the saddlepoint approximation to the PDF of the classical studentized statistics works reasonably well. The saddlepoint approximation (dotted) to the CDF given in the right panel of Figure 6.1 is also very accurate, whereas standard normal (dashed) and Student's t_9 approximations (long-dashed) would fail in this context. This is underlined by the normal Q-Q plot of the $R = 49,999$ replicates of the classical studentized bootstrap statistic in the left panel of Figure 6.2, where the saddlepoint approximation (dotted) and the normal approximation (dashed) are superposed. □

Example 6.2 (Short's data). Table 6.1 contains Short's 1763 determinations ($n = 21$) of the parallax of the sun (in seconds of a degree), based on the 1761 transit of Venus
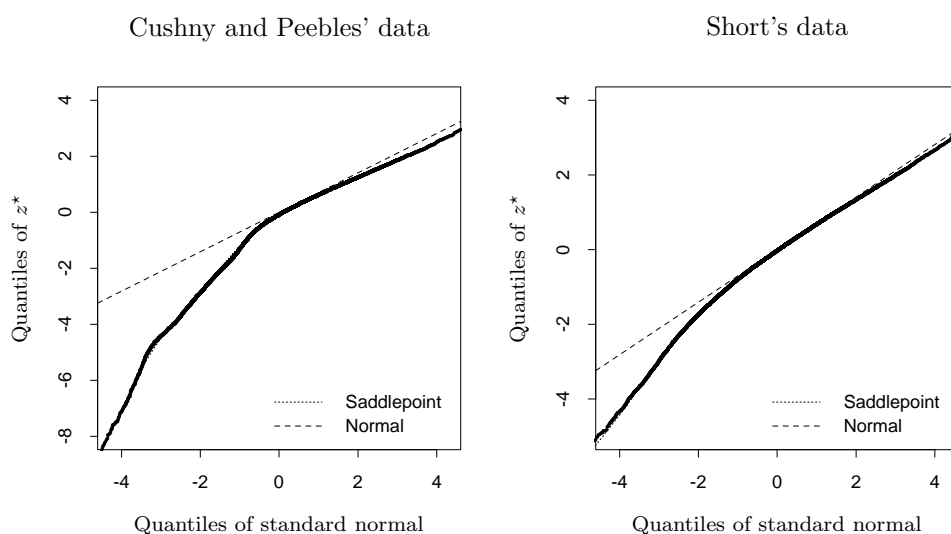
Figure 6.2. *Approximations for the classical studentized statistic for Cushny and Peebles' data (left panel) and for Short's data (right panel). Q-Q plot of the $R = 49,999$ replicates, with the saddlepoint approximation and normal approximation.*

(Stigler, 1977, Table 4, Data Set 4). These data for which we have $t = 8.525$ issue from a larger study. But already using this small subset one notes an amazing fact: although they were published in 1763 the precision of the measurements is incredible as '*recent radar determinations would lead to a value of 8.794*' (Stephen M. Stigler, 1977, page 1072). The values $9.61, 10.15$ and 10.33 can be seen as extreme and as the left panel of Figure 6.3 indicates studentization does a great job . The saddlepoint approximation to the PDF of the classical studentized bootstrap statistic is very accurate. A normal approximation to the CDF may not be as accurate as desired; see the dashed lines in the right panels of Figures 6.3 and 6.2. The same holds for the t_{20} approximation which is superposed in the right panel of Figure 6.3 (long-dashed). From both figures we see that the saddlepoint approximation (dotted) to the CDF works well. □

Additional examples of saddlepoint approximations to the bootstrap PDF and CDF of the classical studentized statistic can be found in Kuonen (1998a, Examples 4.1–4.4), Kuonen (2000b, page 9), Kuonen (2000c, pages 9 and 10) and Kuonen (2000d, page 9).

6.4. Studentized M-estimates

> '*Perhaps the most immediate problem now is the exploration (theoretical and empirical) of studentizing procedures. This should then satisfy the most important practical needs with regard to a location parameter.*'
>
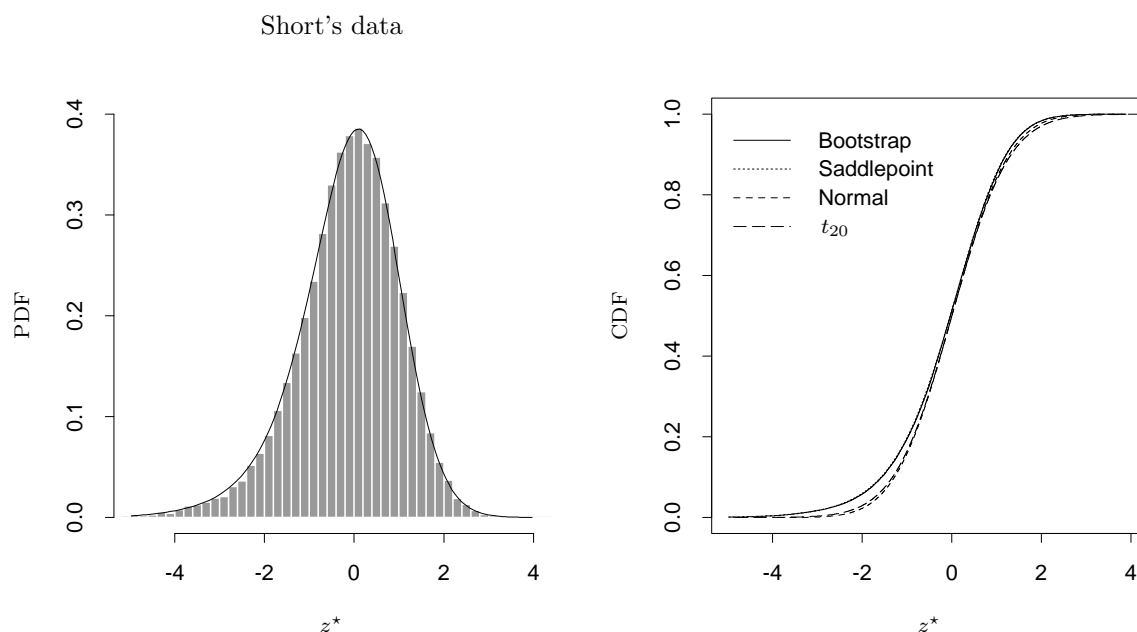> Frank R. Hampel (Andrews *et al.*, 1972, page 242)

Short's data

Figure 6.3. *Saddlepoint approximations to Short's data. Left panel: histogram of $R = 49,999$ replicates of the classical studentized statistic. The solid line represents the saddlepoint approximation to the bootstrap density of the classical studentized statistic. Right panel: bootstrap, saddlepoint, normal and Student-t approximation estimates of the CDF.*

obust M-estimators for a location parameter θ play an important role in robust inference. As a first example Huber's M-estimate of location was considered, but being not scale invariant I then considered the standardised M-estimator of location with initial MAD scaling. In this section we consider their studentized versions together with the studentized version of Huber's proposal 2, which is especially appropriate when both location and scale parameters are of interest.

For the classical studentized statistic the variance of the estimator of location was available, but in the context of M-estimators we need to use the nonparametric delta method estimate (5.1.5) as an approximation to the unknown variance. What follows holds for all three statistics considered in this section. Denote by $\hat{\sigma}$ a scale estimate whose value depends on the statistic in use. Following Remark 5.3 the nonparametric linear delta method estimate of the variability of the estimator $\hat{\theta}$ becomes

$$v_L = \hat{\sigma}^2 \sum_{i=1}^{n} \psi_k^2 \left(\frac{x_i - \hat{\theta}}{\hat{\sigma}} \right) \left\{ \sum_{i=1}^{n} \psi_k' \left(\frac{x_i - \hat{\theta}}{\hat{\sigma}} \right) \right\}^{-2}, \qquad (6.4.1)$$

where $\psi_k(\cdot)$ denotes Huber's function (5.3.6) and prime denotes differentiation with respect to x_i. Indeed, we have

$$\psi_k'(x) = \begin{cases} 1, & \text{if } |x| < k, \\ 0, & \text{otherwise.} \end{cases}$$

Note also that $\sum \psi_k'\{(x_i - \hat{\theta})/\hat{\sigma}\}$ is an integer, namely the number of observations

contained in the interval $(\hat{\theta} - k\hat{\sigma}, \hat{\theta} + k\hat{\sigma})$, and thus all derivatives of $\psi_k(\cdot)$ with respect to x_i beyond the first equal zero.

Using $T = \hat{\theta}$, $V = v_L$ and $\theta = \theta_0$ the studentized statistic (6.1.1) can be rewritten as

$$Z = \frac{\hat{\theta} - \theta_0}{v_L^{1/2}} = \frac{\hat{\theta} - \theta_0}{\hat{\sigma}} \left\{ \sum_{i=1}^{n} \psi_k^2 \left(\frac{x_i - \hat{\theta}}{\hat{\sigma}} \right) \right\}^{-1/2} \sum_{i=1}^{n} \psi_k' \left(\frac{x_i - \hat{\theta}}{\hat{\sigma}} \right). \tag{6.4.2}$$

I will use $k = 1.345$ throughout this thesis.

6.4.1. Huber's M-estimate of location

Let $\hat{\theta}$ be Huber's M-estimate of location as introduced in Section 5.3.2.1. Setting $\hat{\sigma} = 1$ in (6.4.2) yields the studentized version of Huber's M-estimate of location. Thus its bootstrap version is

$$Z^{\star} = (\hat{\theta}^{\star} - \theta_0) \left\{ \sum_{i=1}^{n} \psi_k^2(x_i^{\star} - \hat{\theta}^{\star}) \right\}^{-1/2} \sum_{i=1}^{n} \psi_k'(x_i^{\star} - \hat{\theta}^{\star}), \tag{6.4.3}$$

where $\hat{\theta}^{\star}$ is Huber's M-estimate issued from the bootstrap sample, and θ_0 is its value on the original data $x_1, \ldots, x_n$. For our purposes we now simply use (6.2.1) with the following $d = 2$ estimating equations

$$\psi(x_i, z, s) = \left(\begin{array}{c} \psi_k(x_i - s) \\ zn^{-1} \left\{ \sum_{i=1}^{n} \psi_k^2(x_i - s) \right\}^{1/2} - (s - \theta_0)\psi_k'(x_i - s) \end{array} \right), \tag{6.4.4}$$

where $s = \hat{\theta}$ is the nuisance statistic and z is the statistic of interest, *i.e.* the studentized version of Huber's M-estimate of location. Indeed the first estimating equation defines Huber's M-estimate of location and the second becomes

$$z \left\{ \sum_{i=1}^{n} \psi_k^2(x_i - s) \right\}^{1/2} = (s - \theta_0) \sum_{i=1}^{n} \psi_k'(x_i - s),$$

which at $s = \hat{\theta}$ gives (6.4.3). To calculate the marginal saddlepoint approximations of $Z^{\star}$, (6.2.8) and (2.4.2) using (6.2.12) and (6.2.13), we can get from (6.4.4) the needed matrices of derivatives. For ease of notation denote $\psi_k(x_i - s)$ by $\psi_{k,i}$, $i = 1, \ldots, n$. Hence

$$\frac{\partial}{\partial z} \psi(x_i, z, s) = \left(\begin{array}{c} 0 \\ n^{-1} \left\{ \sum \psi_{k,i}^2 \right\}^{1/2} \end{array} \right), \tag{6.4.5}$$

$$\frac{\partial}{\partial s} \psi(x_i, z, s) = \left(\begin{array}{c} -\psi_{k,i}' \\ -zn^{-1} \left\{ \sum \psi_{k,i}^2 \right\}^{-1/2} \sum \psi_{k,i} \psi_{k,i}' - \psi_{k,i}' \end{array} \right) \tag{6.4.6}$$

Table 6.2. *Pedigree of EMT6 cell lifetimes in hours.*

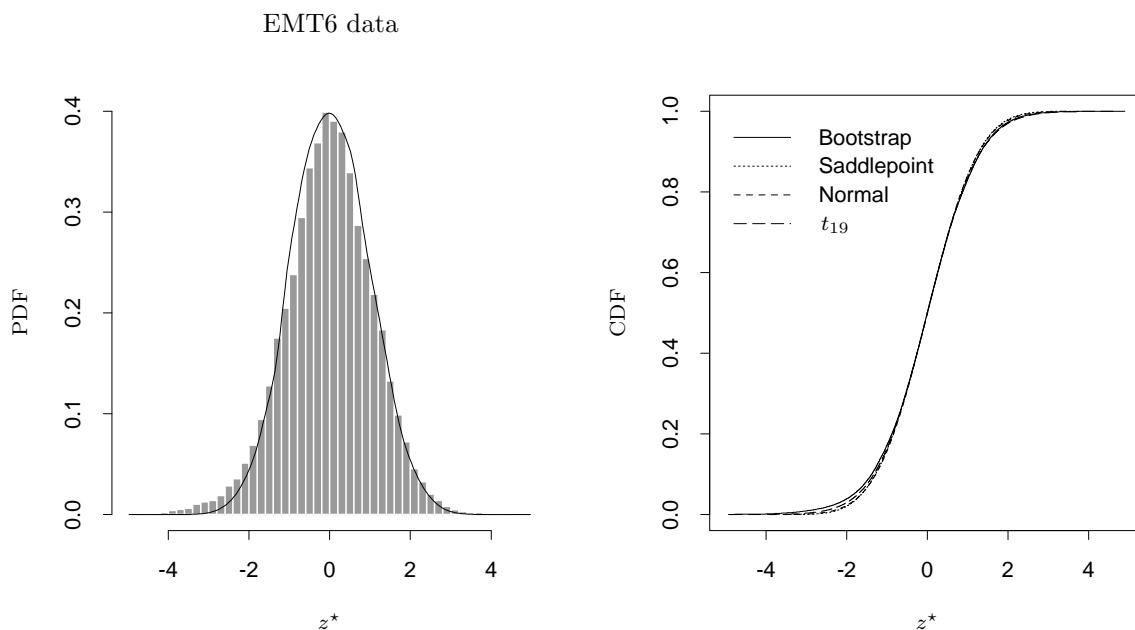| 7.	7.8	8.2	8.4	8.5	8.7	8.8	8.9	9.0	9.1
9.1	9.5	9.8	10.3	10.4	10.4	10.4	10.5	10.9	22.2

EMT6 data

Figure 6.4. *Saddlepoint approximations for the EMT6 data. Left panel: histogram of $R = 49,999$ replicates of the studentized version of Huber's M-estimate of location. The solid line represents the saddlepoint approximation to the bootstrap density of the studentized version of Huber's M-estimate of location. Right panel: bootstrap, saddlepoint, normal and Student-t approximation estimates of the CDF.*

and $\partial^2 \psi(x_i, z, s)/\partial s \partial s^T$ is

$$\begin{pmatrix} 0 \\ -zn^{-1} \left\{ \sum \psi_{k,i}^2 \right\}^{-1/2} \left[-\sum \psi_{k,i}'^2 + \left\{ \sum \psi_{k,i} \psi_{k,i}' \right\}^2 \left\{ \sum \psi_{k,i}^2 \right\}^{-1} \right] \end{pmatrix}. \tag{6.4.7}$$

Example 6.3 (EMT6 data). Table 6.2 lists a typical pedigree of $n = 20$ EMT6 cell lifetimes in hours (Staudte and Sheather, 1990, page 98). One notes the outlier at 22.2, hence a statistic like the studentized version of Huber's M-estimate of location seems appropriate. The accuracy of the saddlepoint approximation to the bootstrap PDF of the studentized version of Huber's M-estimate of location is illustrated in the left panel of Figure 6.4, whereas the approximations to the CDF are given in its right panel. All approximations to the CDF seem to work reasonably well far into the tails of the distribution. □

Example 6.4 (Shrimp data). Table 6.3 lists the results of $n = 18$ different collaborators

Table 6.3. *Percentage of shrimp in shrimp cocktail.*

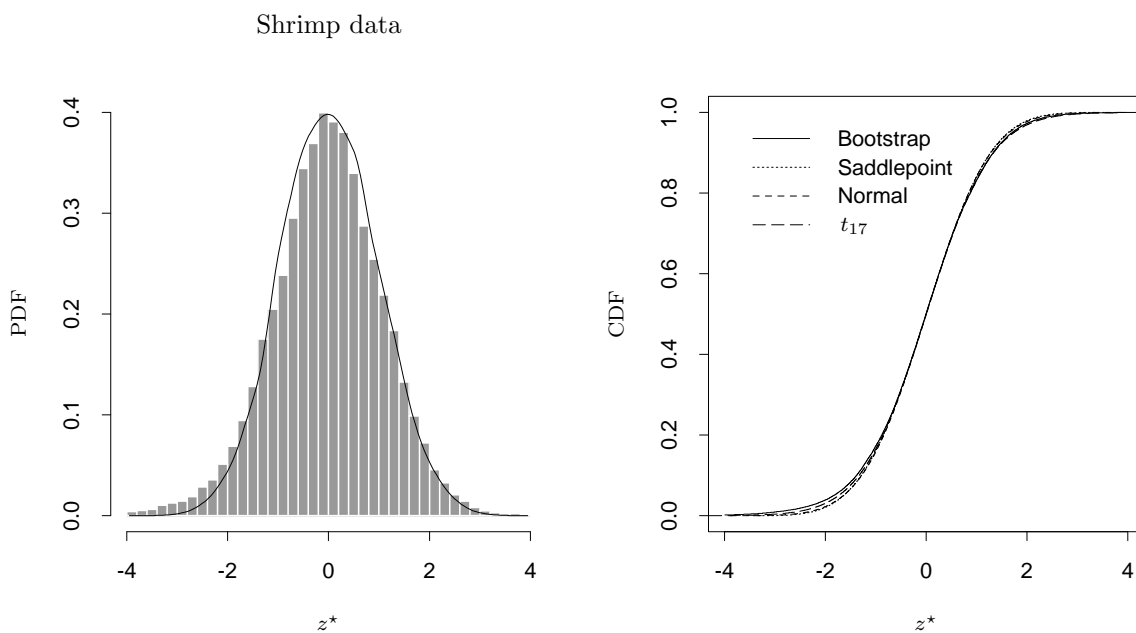| 26.6	30.3	30.7	30.7	30.8	31.2	31.2	31.7	31.7
32.2	32.2	32.3	32.4	32.5	33.0	33.3	33.8	35.7

Shrimp data

Figure 6.5. *Saddlepoint approximations for the shrimp data. Left panel: histogram of $R = 49,999$ replicates of the studentized version of Huber's M-estimate of location. The solid line represents the saddlepoint approximation to the bootstrap density of the studentized version of Huber's M-estimate of location. Right panel: bootstrap, saddlepoint, normal and Student-t approximation estimates of the CDF.*

applying one method of determining the amount of shrimp in the same size container of shrimp cocktail (Staudte and Sheather, 1990, page 134). There are two observations which are outlying, namely 26.6 and 35.7. Except a small underestimation of the bootstrap density in the lower tail the saddlepoint approximation to the PDF given in the right panel of Figure 6.5 is accurate. For the bootstrap distribution all approximations, including the saddlepoint approximation (dotted) to the CDF, work reasonably well (right panel of Figure 6.5).   □

Remark 6.2. Further applications to additional data sets suggest that bootstrapping the studentized version of Huber's M-estimate of location may be questionable. For data sets containing several outliers the saddlepoint approximation to the bootstrap density of the studentized version of Huber's M-estimate of location breaks down by not capturing the multi-modal structures on the histogram of the bootstrap replicates. For instance, this is the case with Newcomb's data ($n = 66$, Table 5.2). The saddlepoint
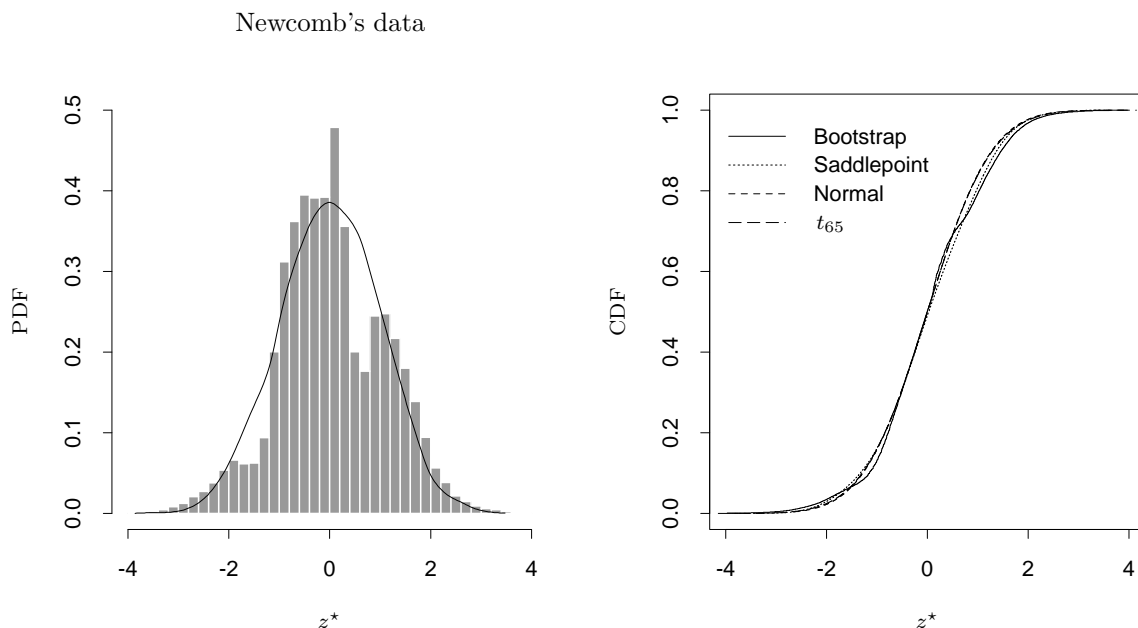
Newcomb's data

Figure 6.6. *Saddlepoint approximations for Newcomb's data. Left panel: histogram of $R = 49,999$ replicates of the studentized version of Huber's M-estimate of location. The solid line represents the saddlepoint approximation to the bootstrap density of the studentized version of Huber's M-estimate of location. Right panel: bootstrap, saddlepoint, normal and Student-t approximation estimates of the CDF.*

approximations are given in Figure 6.6. One notes in the left panel the multi-modality of the bootstrap density. Nevertheless the CDF approximations work well. Second, the bootstrap may break down by delivering extreme bootstrap values in the tails of the distribution. For illustration, the boxplot of the $R = 49,999$ replicates of the studentized version of Huber's M-estimate of location applied to Rosner's data (Example 5.4) is given in Figure 6.7. Moreover, using highly skewed data estimates of both location and scale are desired in order to overcome these problems. For example, this is the case with the studentized versions of Huber's M-estimate of location with initial MAD scaling, and with Huber's proposal 2. Both will be treated separately in the following sections. $\qquad\square$

6.4.2. Huber's M-estimate of location with initial MAD scaling

In the previous section the idea to use a scale estimate together with a location estimate came up. An example of an estimator which merges both is the standardised M-estimator of location with initial MAD scaling as seen in Section 5.3.2.2. Denoting this estimator by $\hat{\theta}$ and the MAD as defined in (5.3.7) by $\hat{\sigma}$, (6.4.2) delivers the studentized version of Huber's M-estimate of location. Its bootstrap version is

$$Z^\star = \frac{\hat{\theta}^\star - \theta_0}{\sigma_0} \left\{ \sum_{i=1}^n \psi_k^2 \left(\frac{x_i^\star - \hat{\theta}^\star}{\sigma_0} \right) \right\}^{-1/2} \sum_{i=1}^n \psi_k' \left(\frac{x_i^\star - \hat{\theta}^\star}{\sigma_0} \right), \qquad (6.4.8)$$
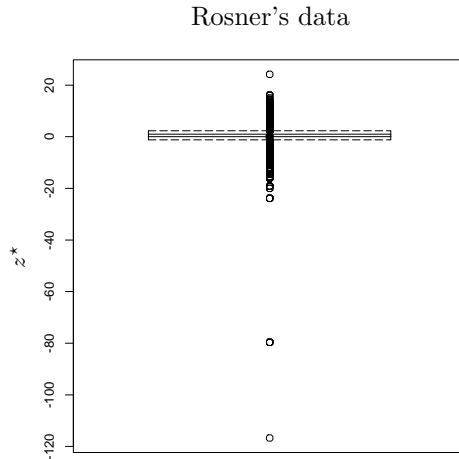
Rosner's data

Figure 6.7. *Boxplot of $R = 49,999$ replicates of the studentized version of Huber's M-estimate of location using Rosner's data. Outliers are represented by circles and the whiskers and staples by dashed lines.*

where $\hat{\theta}^\star$ is Huber's M-estimate of location with initial MAD scaling issued from the bootstrap sample, θ_0 is its observed value and σ_0 the observed MAD. Using $\sigma_0 = 1$ in (6.4.8) reduces to (6.4.3). This connection can be exploited in what follows. First, it can easily be verified that the $d = 2$ estimating equations are

$$\psi(x_i, z, s) = \left(\begin{array}{c} \psi_k\{(x_i - s)/\sigma_0\} \\ zn^{-1}\left\{ \sum_{i=1}^n \psi_k^2\left(\dfrac{x_i - s}{\sigma_0}\right) \right\}^{1/2} - \dfrac{s - \theta_0}{\sigma_0}\,\psi_k'\left(\dfrac{x_i - s}{\sigma_0}\right) \end{array} \right), \qquad (6.4.9)$$

where the nuisance statistic is $s = \hat{\theta}$ and z is the studentized version of Huber's M-estimate of location with initial MAD scaling. Second, the derivatives needed to get the saddlepoint approximations of $Z^\star$ can derived from those in the previous section. Indeed, denote (6.4.5) by Ξ_z, (6.4.6) by Ξ_s, (6.4.7) by Ξ_{ss} and, for fixed k, $\psi_k\{(x_i - s)/\sigma_0\}$ by $\psi_{k,i}$. The derivatives of (6.4.9) are then

$$\frac{\partial}{\partial z}\psi(x_i, z, s) = \Xi_z, \quad \frac{\partial}{\partial s}\psi(x_i, z, s) = \sigma_0^{-1}\Xi_s \quad \text{and} \quad \frac{\partial^2}{\partial s \partial s^T}\psi(x_i, z, s) = \sigma_0^{-2}\Xi_{ss}.$$

Hence the saddlepoint approximation for the PDF of $T^\star = Z^\star$ and $S^\star = S_1^\star = \hat{\theta}^\star$ can be obtained by means of (6.2.8), and for the CDF approximation one needs to apply (2.4.2) with (6.2.12) and (6.2.13).

Remark 6.3. In Remark 6.2 some limitations of the studentized version of Huber's M-estimate of location were illustrated. Concerning the multi-modal structure of the bootstrap density, shown for example in the left panel of Figure 6.6 using Newcomb's data, one remarks in the left panel of Figure 6.8 that this phenomenon does not occur when initial MAD scaling is applied. Moreover for Newcomb's data all CDF approximation shown in the right panel of Figure 6.8 work remarkably well; see also Kuonen (2000c, pages 18 and 19). Furthermore, in Figure 6.7 the failure of the bootstrap for the studen-
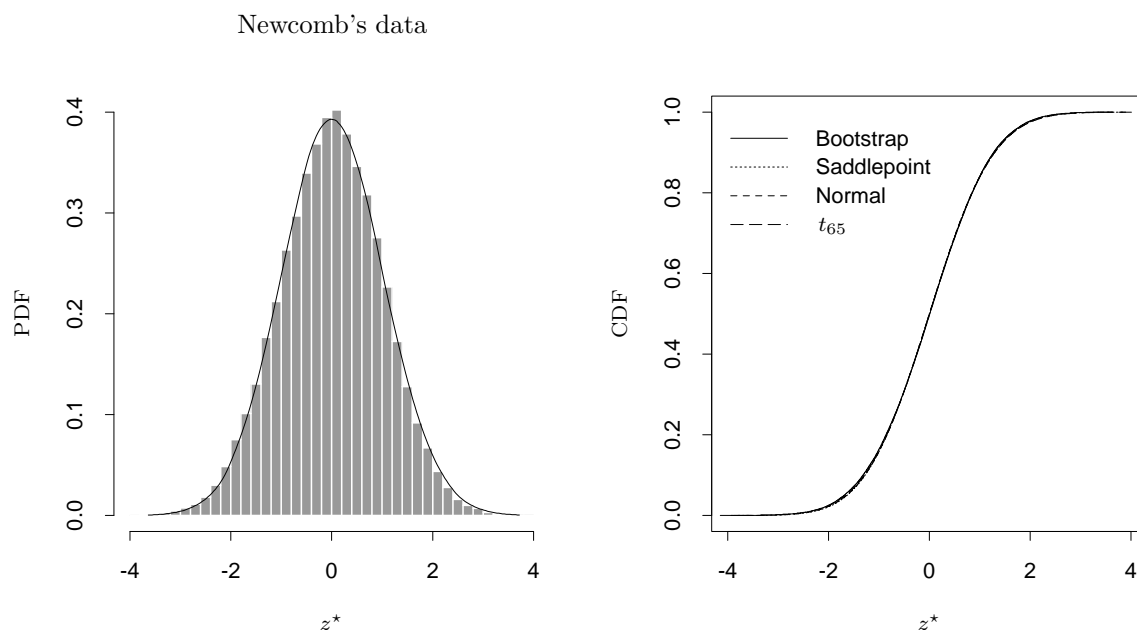
Newcomb's data

Figure 6.8. *Saddlepoint approximations for Newcomb's data. Left panel: histogram of $R = 49,999$ replicates of the studentized version of Huber's M-estimate of location with initial MAD scaling. The solid line represents the saddlepoint approximation to the bootstrap density of the studentized version of Huber's M-estimate of location with initial MAD scaling. Right panel: bootstrap, saddlepoint, normal and Student-t approximation estimates of the CDF.*

Table 6.4. *Michelson's determinations of the velocity of light in air, made in 1882. The given values $+299,000$ are Michelson's determinations in km/s.*

573	578	599	611	682	696	711	723
748	748	772	774	778	781	796	796
797	809	816	820	851	883	1051	

tized M-estimator of location was shown. By means of initial MAD scaling Figure 6.9 suggests an improvement of the range of the bootstrap replicates. Initial MAD scaling also provided a remedy for additional data sets where the non-scaled versions did not work properly. □

Example 6.5 (Michelson's data). Table 6.4 gives Michelson's $n = 23$ supplementary determinations of the velocity of light in air, made in 1882 (Stigler, 1977, Table 7). The value 1051 is an outlier. The first scientific study on the speed of light was carried out by Michelson in 1879. A careful and interesting examination of this study is given by MacKay and Oldford (2000, Sections 3 and 4) in the broader context of defining what a statistical method is. The left panel of Figure 6.10 shows the saddlepoint approxima-
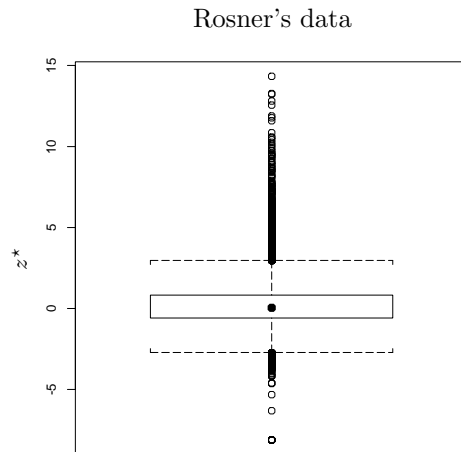
Rosner's data

Figure 6.9. *Boxplot of $R = 49,999$ replicates of the studentized version of Huber's M-estimate of location with initial MAD scaling using Rosner's data. Outliers are represented by circles, the whiskers and staples by dashed lines and the median by the filled octagon in the box.*

Michelson's data

Figure 6.10. *Saddlepoint approximations for Michelson's data. Left panel: histogram of $R = 49,999$ replicates of the studentized version of Huber's M-estimate of location with initial MAD scaling. The solid line represents the saddlepoint approximation to the bootstrap density of the studentized version of Huber's M-estimate of location with initial MAD scaling. Right panel: bootstrap, saddlepoint, normal and Student-t approximation estimates of the CDF.*

tion to the density of the studentized version of Huber's M-estimate of location with initial MAD scaling. The saddlepoint approximation overestimates its lower tail and underestimates its upper tail slightly. In the right panel of Figure 6.10 the estimated CDF of the bootstrap approach (solid) and the saddlepoint approximation (dotted) are

Michelson's data   Moriori data

Figure 6.11. *Approximations for Huber's M-estimate of location with initial MAD scaling for Michelson's data (left panel) and for the Moriori data (right panel). Q-Q plot of the $R = 49,999$ replicates, with the saddlepoint approximation and normal approximation.*

Table 6.5. *Capacities of a sample of seventeen male Moriori skulls in cc.*

1,230	1,260	1,318	1,348	1,360	1,364	1,378	1,380	1,380
1,410	1,410	1,420	1,445	1,470	1,540	1,545	1,630	

shown together with the normal (dashed) and the t_{22} (long-dashed) approximations. All three approximations deliver similar performance. To get a better impression of the approximation in the tails, it is interesting to look at them graphically by means of a normal Q-Q plot of the bootstrap replicates, which for Michelson's data is given in the left panel of Figure 6.11. One notes that the saddlepoint (dotted) and normal approximations (dashed) are not as accurate as desired in the lower and upper tails. □

Example 6.6 (Moriori data). The data in Table 6.5 give the capacities of $n = 17$ male Moriori skulls (Barnett and Lewis, 1994, page 40). Due to the value $1,630$ the data are skewed. The saddlepoint approximation to the PDF of the studentized version of Huber's M-estimate of location with initial MAD scaling overestimates the upper tail of the bootstrap density; see the left panel of Figure 6.12. The CDF approximations, given in the right panel of Figure 6.12, underestimate the upper tail of the bootstrap distribution. This is underlined by the normal Q-Q plot in the right panel of Figure 6.11. □

The previous examples illustrated that the saddlepoint approximation to the boot-
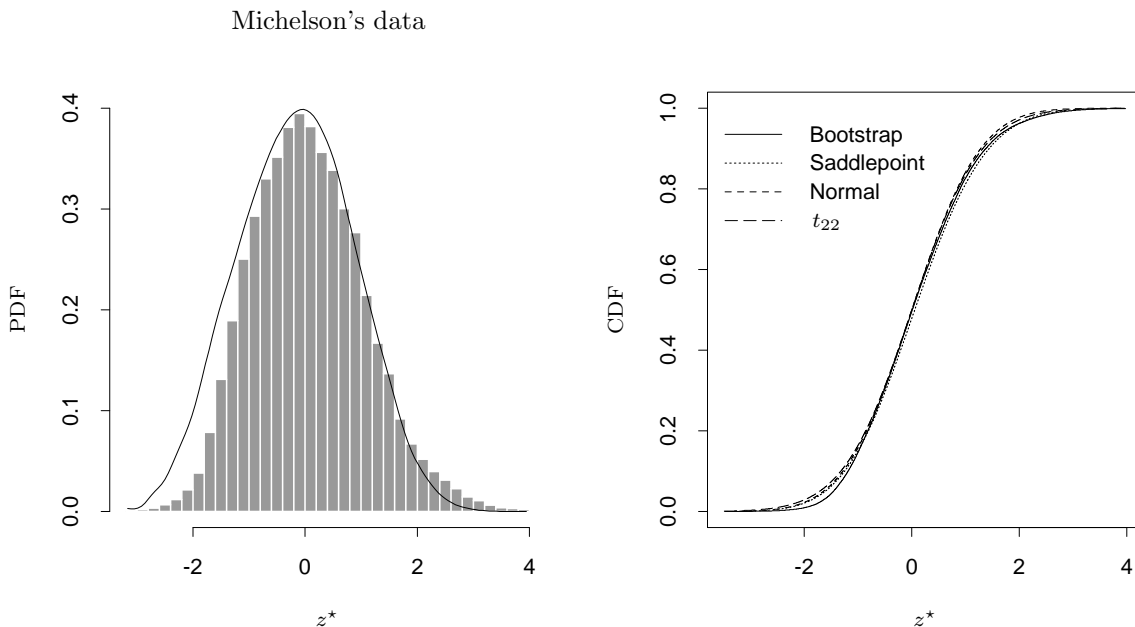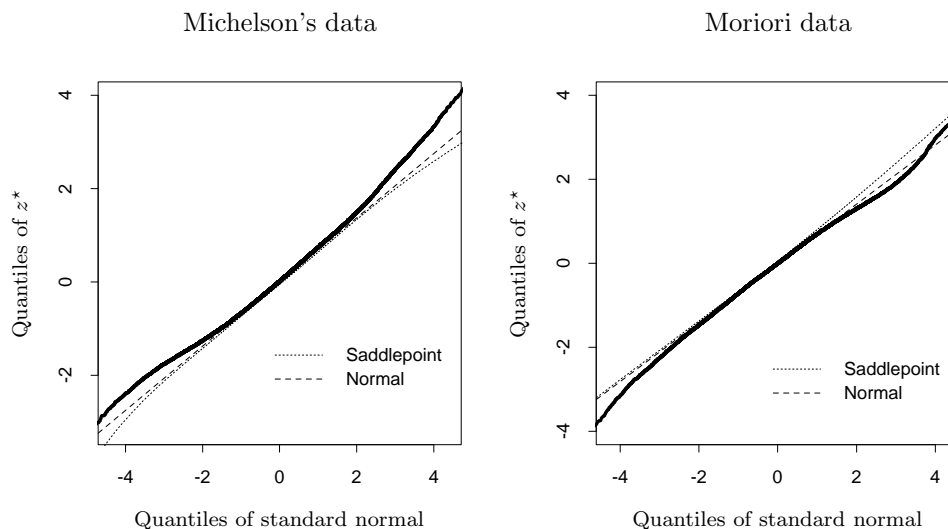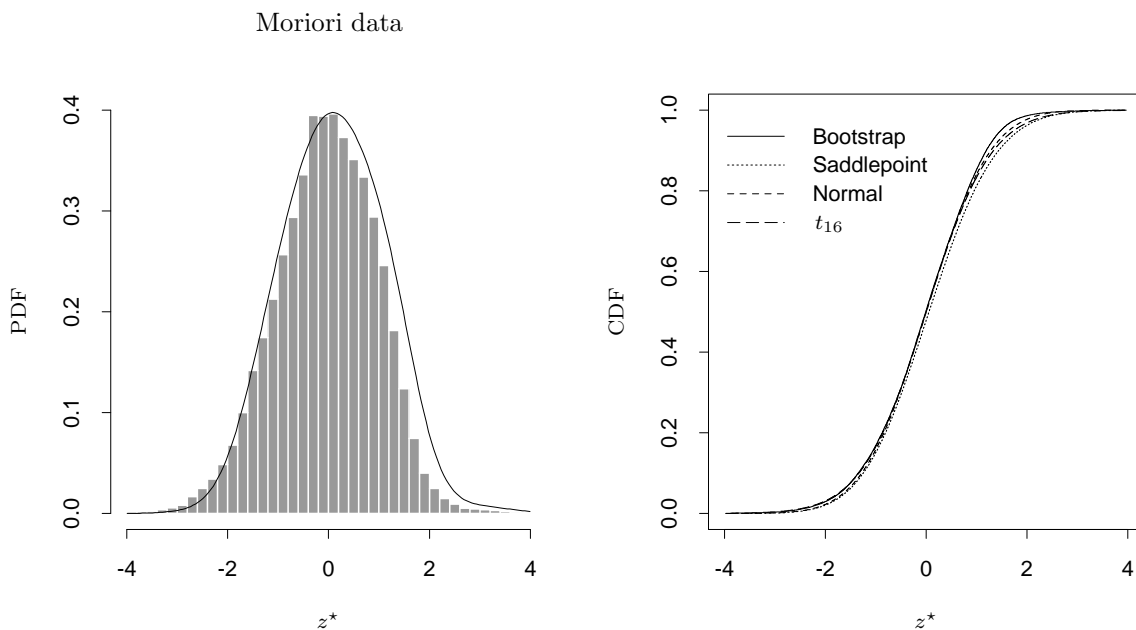
Moriori data

Figure 6.12. *Saddlepoint approximations for the Moriori data. Left panel: histogram of $R = 49,999$ replicates of the studentized version of Huber's M-estimate of location with initial MAD scaling. The solid line represents the saddlepoint approximation to the bootstrap density of the studentized version of Huber's M-estimate of location with initial MAD scaling. Right panel: bootstrap, saddlepoint, normal and Student-t approximation estimates of the CDF.*

strap PDF or CDF of the studentized version of Huber's M-estimate of location with initial MAD scaling delivered over- or underestimation in the tails. This was confirmed in considering additional data sets. Moreover, note that the scale estimate $\hat{\sigma}$ was fixed to the one from the original data set. This corresponds to assuming that the true variance is known. One may prefer to take a bootstrap version of the scale estimate, which also seems to be more natural as this would take into account a separate scale estimate for each bootstrap sample, and not only a single (overall) scale estimate. Finally, it is well known that location M-estimators are usually not scale-invariant. This problem can be solved by using an estimator like Huber's proposal 2.

Remark 6.4. One may also think that the problem could be resolved by using (6.4.8) with $\sigma_0 = \hat{\sigma}^\star$, where $\hat{\sigma}^\star$ is the MAD obtained from the bootstrap sample. But nonparametric resampling of the median, which forms the basis of the MAD defined in (5.3.7), is known to only work when n is quite large and the data do not contain too many outliers; see Davison and Hinkley (1997, Example 2.16). Hence the median does not have the property of being *bootstrappable* (Brown *et al.*, 2001). Bootstrappability is the ability to accurately estimate sampling characteristics. Remarks about the bootstrap performance of the median can also be found in Hall and Martin (1991) or Hall (1992, Appendix IV). Brown *et al.* (2001, Sections 3 and 6) suggested a smoothed median with a breakdown point of 0.341. Their results in favour of the smoothed median look promising but I

doubt that the use of the smoothed median will outperform Huber's proposal 2.   □

6.4.3.   Huber's proposal 2

In analogy to the Huber estimator for the location problem, a natural idea is to 'bring in' the observations which are too far from the estimated location, and to determine these estimates implicitly by the requirement that they should be the classical estimators taken from the transformed observations. More precisely, we suppose that $\hat\theta$ and $\hat\sigma$ are M-estimates found by simultaneous solution of the equations

$$\sum_{i=1}^{n} \psi_k\left(\frac{x_i - \theta}{\sigma}\right) = 0, \qquad \sum_{i=1}^{n} \psi_k^2\left(\frac{x_i - \theta}{\sigma}\right) = n\gamma, \qquad (6.4.10)$$

where $\psi_k(\cdot)$ denotes Huber's function. One noticed that in taking $k = \infty$ and $\gamma = 1$ we obtain the classical estimates $\hat\theta = \bar x$ and $\hat\sigma^2 = n^{-1}\sum(x_i - \bar x)^2$. Huber (1964, 1981) treats questions of existence and uniqueness of solutions to (6.4.10), as well as the derivation of their influence functions. Moreover, he noticed that these equations result in an underestimation of $\hat\sigma$ for standard normal observations and set therefore (Huber, 1981, page 137)

$$\gamma = 2\Phi(k) - 1 - 2k\phi(k) + 2k^2[1 - \Phi(k)]$$

in his *proposal 2* (Huber, 1964, page 96) in order to get Fisher consistency. For this choice we have $\gamma < k^2$ whenever $k > 0$. Huber (1964) remarked that his proposal 2 is not sensitive to a 'wrong' choice of k. In practice, I considered $k = 1.345$, thus $\gamma = 0.71$. Note that the resulting M-estimate of location, $\hat\theta$, is already scale invariant prior to studentization as it has been coupled with an estimate of scale, $\hat\sigma$.

Using $\hat\theta$ and $\hat\sigma$, the simultaneous solutions of (6.4.10), the studentized version of Huber's proposal 2 is given by (6.4.2). Its bootstrap version is

$$\begin{aligned}
Z^\star &= \frac{\hat\theta^\star - \theta_0}{\hat\sigma^\star} \left\{\sum_{i=1}^{n} \psi_k^2\left(\frac{x_i^\star - \hat\theta^\star}{\hat\sigma^\star}\right)\right\}^{-1/2} \sum_{i=1}^{n} \psi_k'\left(\frac{x_i^\star - \hat\theta^\star}{\hat\sigma^\star}\right) \\
&= \frac{\hat\theta^\star - \theta_0}{\hat\sigma^\star}\, (n\gamma)^{-1/2} \sum_{i=1}^{n} \psi_k'\left(\frac{x_i^\star - \hat\theta^\star}{\hat\sigma^\star}\right), \qquad (6.4.11)
\end{aligned}$$

where $\hat\theta^\star$ and $\hat\sigma^\star$ are based on the bootstrap sample $x_1^\star, \dots, x_n^\star$, and θ_0 is the location estimate of the original data set.

To approximate the bootstrap CDF and PDF of $Z^\star$ we need to apply the marginal saddlepoint approximations defined in Section 6.2. The most natural choice would be to use $T = Z$ and $S = (S_1, S_2) = (\hat\theta, \hat\sigma)$, being solutions to the $d = 3$ estimating equations

$$\psi(x_i, z, \theta, \sigma) = \begin{pmatrix} \psi_k\{(x_i - \theta)/\sigma\} \\ \psi_k^2\{(x_i - \theta)/\sigma\} - \gamma \\ \psi_k'\{(x_i - \theta)/\sigma\} - z(\gamma/n)^{1/2}\sigma(\theta - \theta_0)^{-1} \end{pmatrix}.$$

But Kuonen (1998a, Section 4.4.1) remarked that then the five equations in (6.2.9) can not be satisfied simultaneously. This is due to the violation of an assumption introduced by Huber (1964, page 97), which is needed to get unique solutions $\hat{\theta}$ and $\hat{\sigma}$ to (6.4.10).

An alternative approach is given in Davison and Hinkley (1997, Example 9.20). They used a different set of estimating equations and simplified them in order to get the marginal saddlepoint approximations to the studentized bootstrap version of Huber's proposal 2. Their approach will be illustrated in what follows. Note that equation (6.4.11) can also be written as

$$
Z^{\star} \;=\; \frac{\hat{\theta}^{\star} - \theta_0}{\hat{\sigma}^{\star}} \, (\gamma/n)^{-1/2} n^{-1} \sum_{i=1}^{n} \psi'_k \left(\frac{x_i^{\star} - \hat{\theta}^{\star}}{\hat{\sigma}^{\star}} \right).
$$

In denoting $n^{-1} \sum \psi'_k \{ (x_i^{\star} - \hat{\theta}^{\star})/\hat{\sigma}^{\star} \}$ by $s_2^{\star}$ we have at $Z^{\star} = z$ that

$$
\hat{\sigma}^{\star} z = (\hat{\theta}^{\star} - \theta_0)(\gamma/n)^{-1/2} s_2^{\star}.
$$

It yields that $\hat{\theta}^{\star}$ equals $\theta_0 + z(\gamma/n)^{1/2}\hat{\sigma}^{\star}/s_2^{\star}$ and therefore

$$
\tilde{e}_i = \frac{x_i^{\star} - \hat{\theta}^{\star}}{\hat{\sigma}^{\star}} = \frac{x_i^{\star} - \theta_0}{\hat{\sigma}^{\star}} - \frac{z(\gamma/n)^{1/2}\hat{\sigma}^{\star}/s_2^{\star}}{\hat{\sigma}^{\star}} = \sigma_0 e_i^{\star}/s_1^{\star} \,-\, z(\gamma/n)^{1/2}/s_2^{\star},
$$

where $s_1^{\star}$ denotes $\hat{\sigma}^{\star}$, $e_i^{\star} = (x_i^{\star} - \theta_0)/\sigma_0$ are the studentized bootstrap values and θ_0, σ_0 are the observed values of the location and of the scale estimate, respectively. Hence we can apply the marginal saddlepoint approximations (6.2.8) and (2.4.2), using (6.2.12) and (6.2.13), with $T = Z$, $S = (S_1, S_2) = (\hat{\sigma}, n^{-1} \sum \psi'_k \{ (x_i - \hat{\theta})/\hat{\sigma} \})$ and

$$
\psi(x_i, z, s_1, s_2) = \begin{pmatrix} \psi_k(\tilde{e}_i) \\ \psi_k^2(\tilde{e}_i) - \gamma \\ \psi'_k(\tilde{e}_i) - s_2 \end{pmatrix}. \tag{6.4.12}
$$

Note that $ns_2^{\star}$ is the number of observations contained in the interval $(\hat{\theta}^{\star} - k\hat{\sigma}^{\star}, \, \hat{\theta}^{\star} + k\hat{\sigma}^{\star})$, and thus it would be unwise to treat $s_2^{\star}$ as continuous. Hence the sampling version of s_2 is fixed to its observed value, *i.e.* $s_2^{\star} = s_2 = n^{-1} \sum \psi'_k \{ (x_i - \theta_0)/\sigma_0 \}$, and (6.4.12) can be modified by dropping off its third component. Therefore, $T = Z$ and $S = S_1 = \hat{\sigma}$ are the solutions to the remaining $d = 2$ estimating equations. By denoting $\psi_k(\tilde{e}_i)$ for fixed k by $\psi_{k,i}$, the derivatives needed to calculate the saddlepoint approximation are given by

$$
\frac{\partial \psi(x_i, z, s)}{\partial z} \;=\; \begin{pmatrix} -\psi'_{k,i}(\gamma/n)^{1/2}/s_2 \\ -2\psi_{k,i}\psi'_{k,i}(\gamma/n)^{1/2}/s_2 \end{pmatrix},
$$

$$
\frac{\partial \psi(x_i, z, s)}{\partial s} \;=\; \begin{pmatrix} -\sigma_0 e_i \psi'_{k,i}/s_1^2 \\ -2\sigma_0 e_i \psi_{k,i}\psi'_{k,i}/s_1^2 \end{pmatrix}
$$

and

$$
\frac{\partial^2 \psi(x_i, z, s)}{\partial s \partial s^T} \;=\; \begin{pmatrix} 2\sigma_0 e_i \psi'_{k,i}/s_1^3 \\ 4\sigma_0 e_i \psi_{k,i}\psi'_{k,i}/s_1^3 + 4\sigma_0^2 e_i^2 \psi'^2_{k,i}/s_1^4 \end{pmatrix}.
$$

Table 6.6. *Determinations of copper in whole meal flour.*

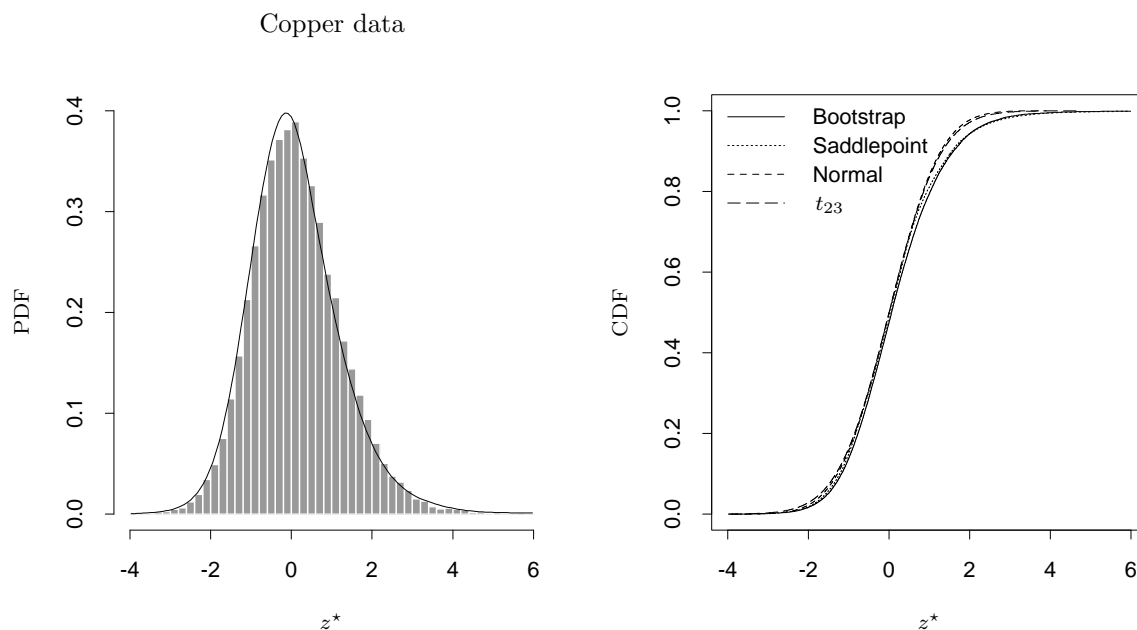|							
2.20	2.20	2.40	2.40	2.50	2.70	2.80	2.90
3.03	3.03	3.10	3.37	3.40	3.40	3.40	3.50
3.60	3.70	3.70	3.70	3.70	3.77	5.28	28.95

PSfrag replacements

Copper data

Figure 6.13. *Saddlepoint approximations for the copper data. Left panel: histogram of $R = 49,999$ replicates of the studentized version of Huber's proposal 2. The solid line represents the saddlepoint approximation to the bootstrap density of the studentized version of Huber's proposal 2. Right panel: bootstrap, saddlepoint, normal and Student-t approximation estimates of the CDF.*

Example 6.7 (Copper data). The copper data (Analytical Methods Committee, 1989) contain $n = 24$ determinations of copper in whole meal flour. The data set given in Table 6.6 is highly asymmetric with one value, 28.95, that appears to be out by a factor of ten. The saddlepoint approximation to the studentized bootstrap statistic of Huber's proposal 2 is given in Figure 6.13. As one can see in the left panel the approximation to the PDF is very accurate. The CDF approximations are shown in the right panel. The saddlepoint approximation (dotted) is very accurate, whereas the normal (dashed) and the t_{23} (long-dashed) approximations fail in the upper tail of the distribution. The left panel of Figure 6.14 shows the normal Q-Q plot of the bootstrap values for the copper data. Superposed on this plot are the quantiles of the saddlepoint (dotted) and normal approximations (dashed). Normal approximation is poor, whereas the saddlepoint approximation is good. □
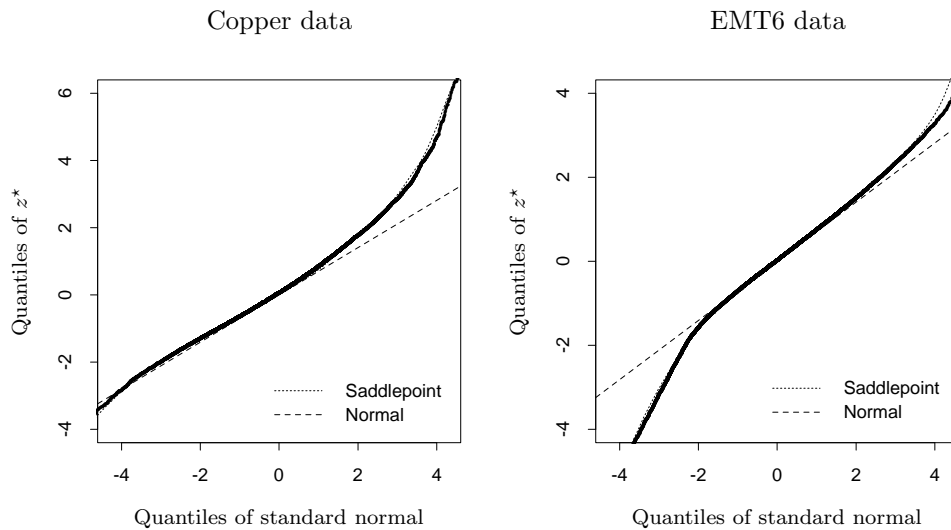
Copper data

EMT6 data

Figure 6.14. *Approximations for Huber's proposal 2 for the copper data (left panel) and for the EMT6 data (right panel). Q-Q plot of the $R = 49,999$ replicates, with the saddlepoint approximation and normal approximation.*

Example 6.8 (EMT6 data). The EMT6 data are given in Table 6.2. The left panel of Figure 6.15 shows the saddlepoint approximation to $Z^\star$ onto the histogram of the bootstrap replicates. One remarks that the saddlepoint approximation performs very well for the PDF as well as for the CDF (dotted line in the right panel of Figure 6.15). The normal (dashed) and the t_{19} (long-dashed) approximations to the CDF underestimate the bootstrap distribution in the lower tail. The approximations to the quantiles are given in the right panel of Figure 6.14. One clearly sees a large deviation from normality in the tails, suggesting that the normal approximation (dashed) would fail, whereas the saddlepoint approximation (dotted) performs very well. □

Additional examples of the application of saddlepoint approximations to the studentized bootstrap statistic of Huber's proposal 2 can be found in Kuonen (1998a, Examples 4.5–4.10), Kuonen (2000b, page 13), Kuonen (2000c, page 19) and Kuonen (2000d, page 13). All examples I considered convinced me that it is preferable to use the studentized version of Huber's proposal 2 instead of the studentized version of Huber's M-estimate of location with initial MAD scaling. This recommendation will be underlined in the next chapter in comparing the resulting studentized confidence intervals, hopefully.

6.5. Miscellany

n this section I will illustrate the problems that may occur by applying the integration approach presented in Section 6.2 and applied in Sections 6.3 and 6.4. The described problems may affect the accuracy of the saddlepoint approximations. As far as feasible I will present remedies to overcome these problems.
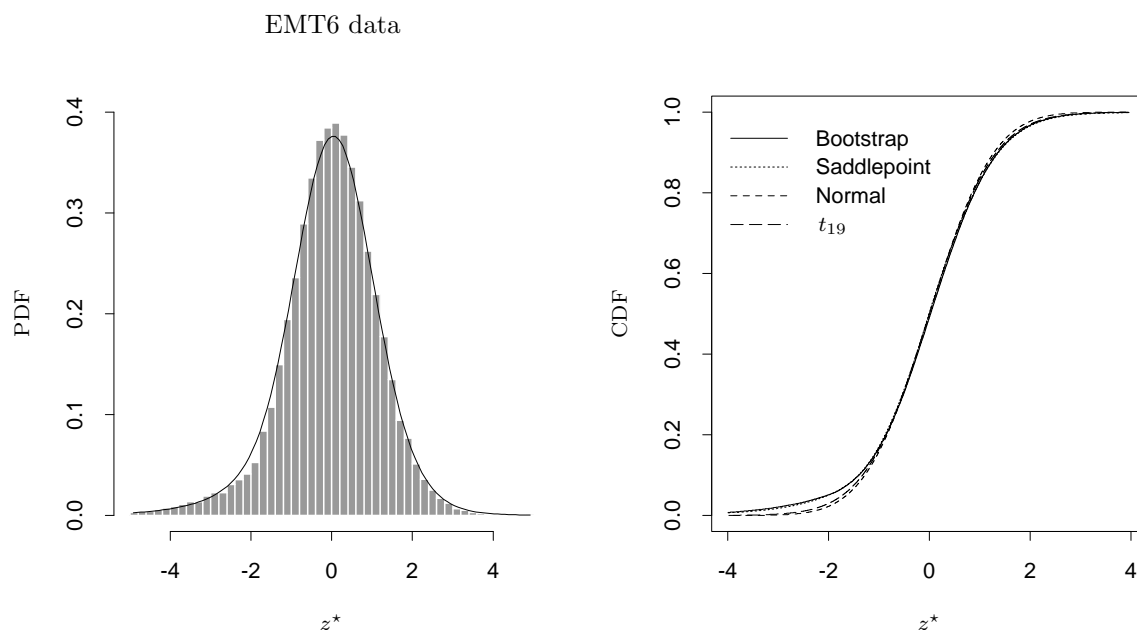
EMT6 data

Figure 6.15. *Saddlepoint approximations for the EMT6 data. Left panel: histogram of $R = 49,999$ replicates of the studentized version of Huber's proposal 2. The solid line represents the saddlepoint approximation to the bootstrap density of the studentized version of Huber's proposal 2. Right panel: bootstrap, saddlepoint, normal and Student-t approximation estimates of the CDF.*

6.5.1. Diagnostics

Daniels and Young (1991, Section 5) noticed that care is needed when using a Laplace approximation in (6.2.7) in order to obtain the marginal approximation to the density of $T^\star$ given in (6.2.8). This problem occurs because to apply a Laplace approximation we assume the matrix $\partial^2 K(\hat{\zeta}, t, s)/\partial s \partial s^T$ to be negative definite, *i.e.* the cumulant generating function $K(\hat{\zeta}, t, s)$ to be concave. But, there is no guarantee that $K(\hat{\zeta}, t, s)$ is a concave function of t and s. Daniels and Young (1991, page 171) and Davison and Hinkley (1997, page 482) noticed that fortunately this difficulty is much rarer in large samples. This is something I can confirm as well. The only times I encountered non-concavity of $K(\hat{\zeta}, t, s)$ was when I considered highly-skewed data sets with moderate values of n. This is for instance the case with Darwin's data using the studentized version of Huber's proposal 2. The data are given in Table 6.7 (Hand *et al.*, 1994, page 130). For fixed values of $t = z$, namely $z = -2$ and 1, Figure 6.16 shows $K(\hat{\zeta}, z, s)$ as a function of s. The left panel illustrates that there does not exist an unique maximum of $K(\hat{\zeta}, -2, s)$ or similarly an unique minimum of $h(s) = -K(\hat{\zeta}, -2, s)$ as defined in (6.2.7). The unique minimum of $h(s)$, $\tilde{s}$ say, should exist within the interior of the domain of integration for s. Uniqueness of $\tilde{s}$ is needed in order to apply the Laplace approximation and to ensure that the major contribution of the integral of (6.2.6) with respect to s is interior to the region of integration. Using the classical studentized statistic Davison and Hinkley (1997, Example 9.19) noted a bimodality in the contour plot of $K(\hat{\zeta}, z, s)$,

Table 6.7. *Observations issued from an experiment to examine the superiority of cross-fertilised plants over self-fertilised plants. The 15 pairs of plants were measured at a fixed time after planting and the difference in heights between the cross- and self-fertilised plants are recorded in eights of an inch.*

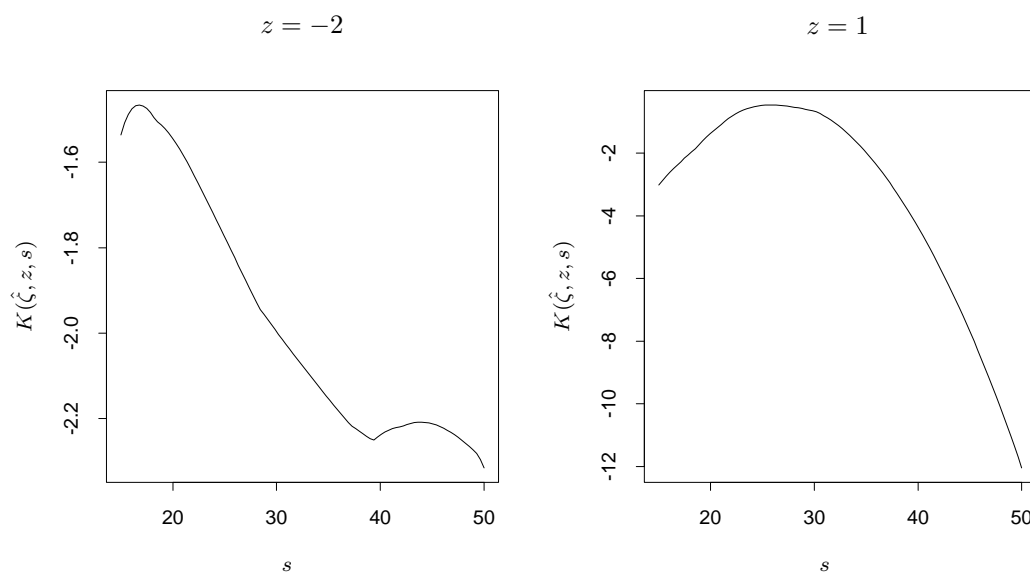| −67	−48	6	8	14	16	23	24
28	29	41	49	56	60	75	

PSfrag replacements

Figure 6.16. *The cumulant generating function (6.2.3) as a function of s, for $z = -2$ (left panel) and $z = 1$ (right panel), using the studentized version of Huber's proposal 2 applied to Darwin's data.*

but they also remarked that the resulting marginal saddlepoint approximation to the density of the studentized average $Z^\star$ remains accurate. This results from the fact that the bimodality in the cumulant generating function did not deliver a multi-modal joint density as its second peak adds little (Davison and Hinkley, 1997, Figure 9.11). This is nicely illustrated by the wireframe plot of the joint density (6.2.6) in Figure 6.17. One notes the unimodal structure of the joint density. The level plot of the joint density (6.2.6) is given in Figure 6.18 and shows that only values of z between about -2 and 4 need to be considered to get the entire marginal approximation to $Z^\star$. Davison and Hinkley (1997, page 481) remarked the bimodality at $z = -3.5$ which lies outside this range.

It is always wise to use diagnostic plots like the ones given in Figures 6.16–6.18. The drawback is the CPU time needed. For example to compute the joint density (6.2.6) using the classical studentized statistic on a 50×50 grid it took about 16 minutes by means of the method to be described in Section 6.5.2.1. This clearly slows down the integration approach and thus may not be suggested for routine use. Nevertheless, let
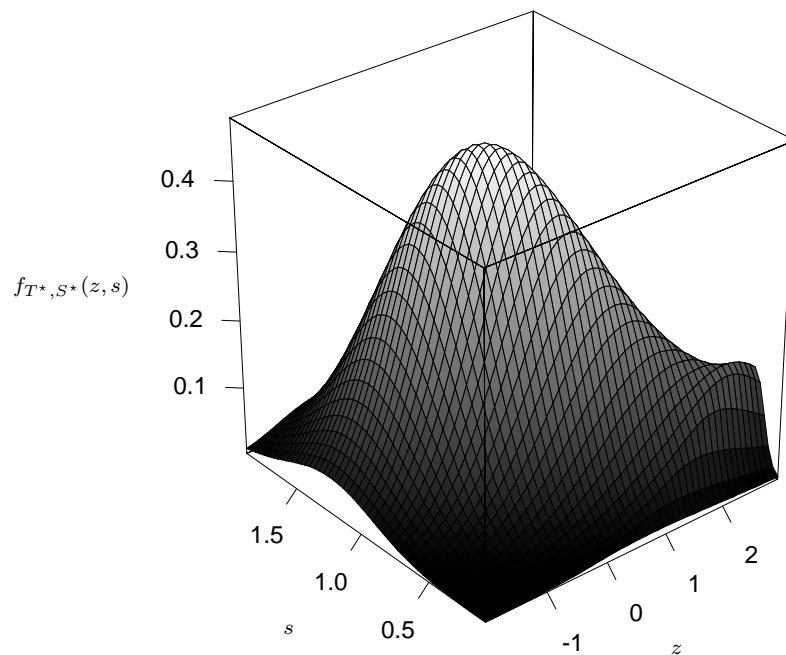
Figure 6.17. *Wireframe plot for the joint density (6.2.6) using the classical studentized statistic applied to Darwin's data.*
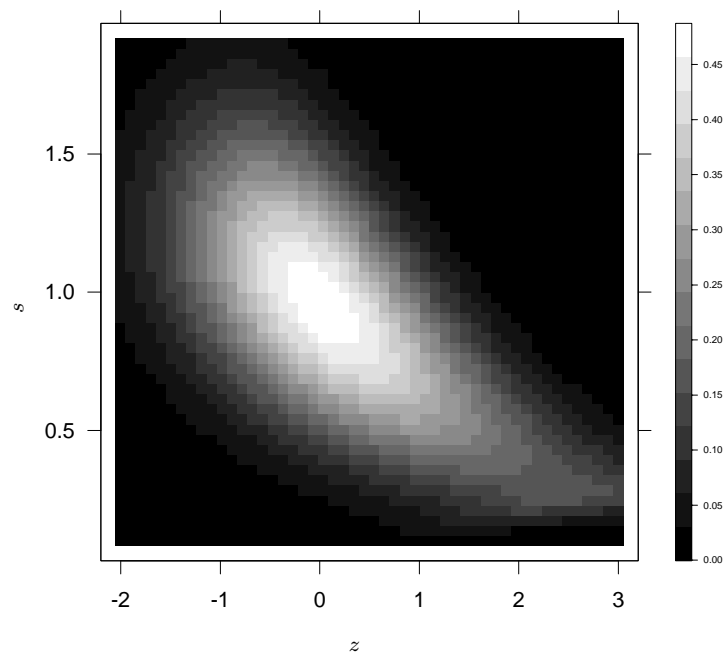
Figure 6.18. *Level plot for the joint density (6.2.6) using the classical studentized statistic applied to Darwin's data.*

us consider what will happen when we apply the studentized version of Huber's proposal 2 to Darwin's data. Wireframe plots under different viewing angles are given in Figure 6.19. The multi-modality of the joint density (6.2.6) is shown and underlined in the
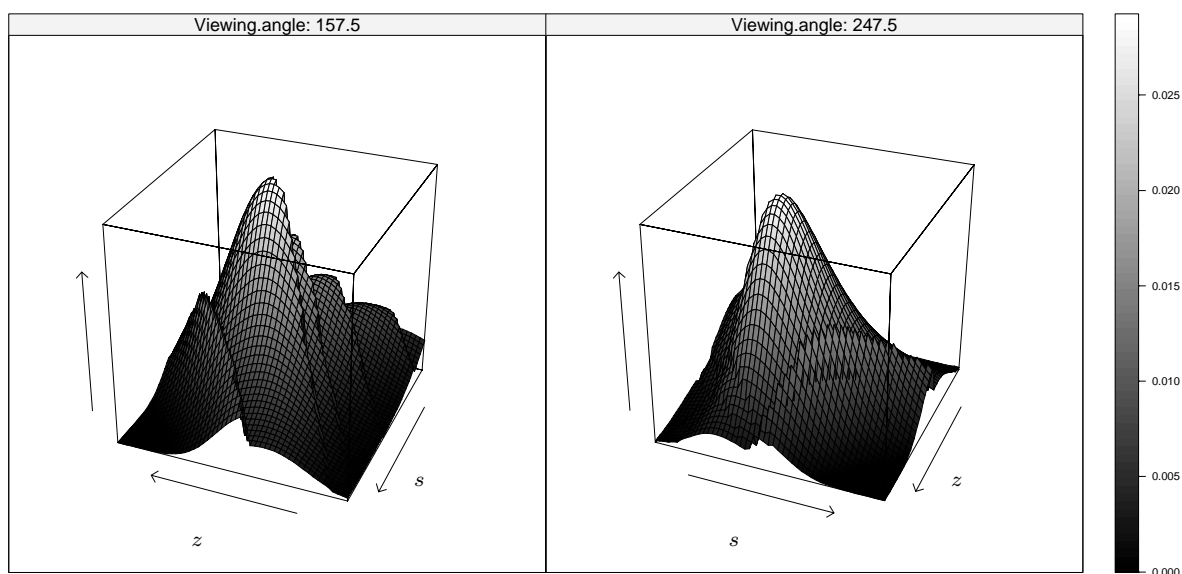
Figure 6.19. *Wireframe plots for the joint density (6.2.6) using the studentized version of Huber's proposal 2 applied to Darwin's data. Left panel: using a viewing angle of 157.5. Right panel: using a viewing angle of 247.5.*
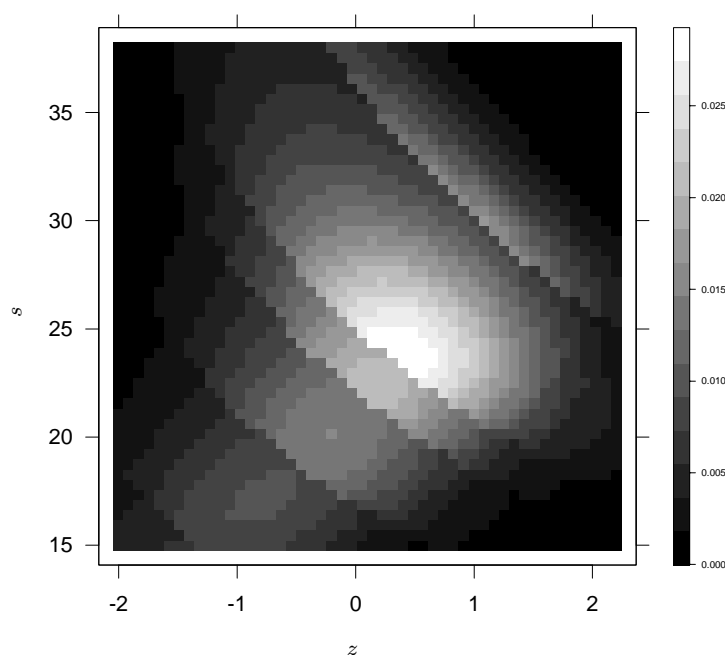
Figure 6.20. *Level plot for the joint density (6.2.6) using the studentized version of Huber's proposal 2 applied to Darwin's data.*

level plot given in Figure 6.20. Computation using a 50×50 grid took about 42 minutes in CPU time. The peaks shown in these figures would result in an inaccurate marginal approximation as the Laplace approximation only accounts for the dominant centre peak.

Another approach to diagnose a possible failure of Laplace's method is to check the

signs of the eigenvalues of the real symmetric $(d-1) \times (d-1)$ matrix $\partial^2 K(\hat{\zeta}, z, s)/\partial s \partial s^T$ at a range of values of s: if all eigenvalues are negative Laplace approximation can be used, as this fact ensures the convexity of $h(s)$, the matrix $-\partial^2 K(\hat{\zeta}, z, s)/\partial s \partial s^T$ being then positive definite, and hence the uniqueness of $\tilde{s}$.

Nevertheless, note that an approximation can always be found by a numerical integration of the joint density (6.2.6) with respect to s; especially if there exists at least a positive eigenvalue of $\partial^2 K(\hat{\zeta}, z, s)/\partial s \partial s^T$.

In the examples considered in Sections 6.3 and 6.4 this behaviour did not occur, so I always considered a Laplace approximation. Moreover, if the difficulty is thought to have arisen one could simply jitter the values of interest and recalculate the quantities needed for the integration approach. But, in order to stay as general as possible numerical integration of (6.2.6) should be used to find the marginal distribution of $T^\star = Z^\star$.

6.5.2. Numerical integration

The safest procedure to apply to the joint density (6.2.6) would be to integrate it numerically with respect to $s = (s_1, \ldots, s_{d-1})$. Computationally such numerical integration is most conveniently performed using direct function evaluation of (6.2.6) on a regular grid. Numerical integration methods for use in S-Plus and R are discussed in detail in Appendix A. Following the notations therein, to get the approximation to the PDF of $T^\star$ the evaluation of the following multi-dimensional integral is needed

$$f_{T^\star}(t) = \int \cdots \int_{R_{d-1}} f_{T^\star, S^\star}(t, s_1, \ldots, s_{d-1}) ds_1 \cdots ds_{d-1}, \tag{6.5.1}$$

where the $(d-1)$-dimensional integration region R_{d-1} is usually $[-\infty, \infty]^{d-1}$.

In order to get the approximation to the CDF of $T^\star$ at $\tilde{t}$, say, we need to integrate (6.5.1) with respect to t,

$$F_{T^\star}(\tilde{t}) = \Pr(T^\star \le \tilde{t}) = \int_\infty^{\tilde{t}} f_{T^\star}(t) dt, \tag{6.5.2}$$

where $f_{T^\star}(t)$ is as in (6.5.1). Hence this corresponds to a d-dimensional integration.

To apply the various quadrature methods given in Appendix A effectively we need to solve several problems prior to their use. First, I will describe in Section 6.5.2.1 how I implemented the computation of the joint density of $T^\star$ and $S^\star$, (6.2.6), which is needed to calculate the function evaluations in order to perform the numerical integration. Second, by default the integration region in (6.5.1) and the lower integration bound in (6.5.2) are infinite. Depending on the choice of the statistic of interest this domain may be finite with well-defined upper and lower limits of the intervals. If not, transformations need to be used. This will be discussed in Section 6.5.2.2. Especially for (6.5.2) we may need to have an idea of the effective range of $T^\star$; see Section 6.5.2.3. Finally, examples will be given in Section 6.5.2.4.

6.5.2.1. Computation of the joint density

From the definition of the joint density $f_{T^\star,S^\star}(t,s)$ in (6.2.6) one sees that it depends on $\hat\zeta = \hat\zeta(t,s)$, the saddlepoint, which is the solution of the d equations $\partial K(\zeta,t,s)/\partial\zeta = 0$ with $\partial K(\zeta,t,s)/\partial\zeta$ given in (6.2.11). Replacing in (6.2.11) the $w(x_i,\zeta,t,s)$ by their definition (6.2.5) the left-hand system of equations in (6.2.9) becomes

$$\frac{\partial K(\zeta,t,s)}{\partial\zeta} = \sum_{i=1}^{n} \psi(x_i,t,s)\exp\left\{\log p_i + \zeta^T\psi(x_i,t,s)\right\} = 0. \qquad (6.5.3)$$

In the notation introduced in Section 2.4 equation (6.5.3) can be written for fixed t and s as $A^T Y = 0$, where A^T is the $d \times n$ matrix whose ith column is $a_i = \psi(x_i,t,s)$, and Y is the $n \times 1$ vector with ith element $y_i = \exp\left\{\log p_i + \zeta^T a_i\right\}$. The problem of solving the d saddlepoint equations $A^T Y = 0$ can be solved by existing software for fitting a generalised linear model (GLM) as mentioned by Booth and Butler (1990) and used in Canty and Davison (1996, 1999). Indeed, this can be simply done by fitting a Poisson GLM with covariate matrix A and offset $\log p$, where $p = (p_1, \dots, p_n)$. The calculation can easily be performed using the S-PLUS or R function `glm` as follows

```
> fit.glm <- glm(rep(0,n) ~ A + offset(log(p)) - 1, family=poisson)
```

The saddlepoint $\hat\zeta$ is then returned as the parameter estimates,

```
> zeta.tilde <- coef(fit.glm)[1:d]
```

Practice showed that using `glm` instead of the minimiser functions `nlmin` in S-PLUS or `nlm` in R resulted in a significant speed-up in the computation of the saddlepoint $\hat\zeta$ in this context.

Remark 6.5. One could also use a GLM fit to solve the $2d-1$ equations (6.2.9) needed to calculate (6.2.8) at t. An algorithm which I tried starts by fixing s at $\check s$, say, and solving the left-hand equations in (6.2.8) using `glm` which gives $\hat\zeta(t,\check s)$. The latter is now fixed to $\check\zeta$ and used in the equations in the right-hand side of (6.2.8) to get $\check s$ by means of the Newton optimisers `nlmin` or `nlm`. If both sets of equations in (6.2.8) bracket zero the solutions to the $2d-1$ equations (6.2.9) are given by $\tilde\zeta = \check\zeta$ and $\tilde s = \check s$ and the resulting approximation to the PDF of $T^\star$ can then easily be computed. If not, iteration on $\check s$ is needed until convergence occurs. But the convergence of this algorithm can be very slow. It is outperformed by direct use of `nlmin` or `nlm` to solve (6.2.9). $\qquad\square$

6.5.2.2. Transformations

Integrals over infinite domains, like (6.5.1), should be transformed to a finite region in view of the accuracy and convergence of the quadrature method in use. This is especially true for adaptive integration algorithms (see Appendix A.3) which require repeated subdivision of the integration region. When the integration is infinite, the point along the axis where the current subregion is to be cut is not clearly defined. Hence it is convenient to consider some appropriate transformations from the infinite

integration region in (6.5.1) to a finite region. Then quadrature can be applied directly on the transformed integrand over the finite integration region.

There are a number of simple one-variable transformations that have been used for integration problems; see Davis and Rabinowitz (1984, Chapter 3) or Genz (1992, Section 3.2) for further discussion and examples. Care must be taken in the selection of the transformation. As a check on consistency and efficiency I used several transformations for different computations of the same integral, and I compared their results. The transformations listed below performed best and the other transformations which I tried will not be listed here.

The four studentized bootstrap statistics $Z^\star$ considered in Sections 6.3 and 6.4 all use as statistic of interest $T^\star = Z^\star$. Concerning the nuisance statistics $S^\star$ they can be split up into two groups. The first group (the *scale group*) consists of the classical studentized statistic (6.3.1) with the classical variance of the bootstrap sample, and of the studentized version of Huber's proposal 2 (6.4.11) with the M-estimate of scale as defined in (6.4.10). The second group (the *location group*) consists of the studentized version of Huber's M-estimate of location (6.4.3) and of its initially MAD scaled version (6.4.8), both using Huber's M-estimate of location as $S^\star$. For the scale group (6.5.1) reduces to

$$f_{T^\star}(t) = \int_0^\infty f_{T^\star,S^\star}(t,s)ds$$

as scale estimates are non-negative by definition. Transforming this half-infinite region of integration to the finite interval $[0,1]$ using the change of variables $s = r/(1-r)$ yields

$$f_{T^\star}(t) = \int_0^1 f_{T^\star,S^\star}(t,r)(1-r)^{-2}dr. \qquad (6.5.4)$$

For the location group a change of variables is not necessary as for both group members the nuisance statistics is a robust estimate of location, *i.e.* an estimate of the centre of the distribution and hence its range certainly is defined by the finite range of the underlying data. Therefore (6.5.1) becomes

$$f_{T^\star}(t) = \int_{\min(x_1,\dots,x_n)}^{\max(x_1,\dots,x_n)} f_{T^\star,S^\star}(t,s)ds. \qquad (6.5.5)$$

For the CDF approximation (6.5.2) all four statistics are treated in the same way. Using the change of variables $t = \{\tilde{t} - \check{t}(1+\tilde{t})\}/(1-\check{t})$ the integral in (6.5.2) can be written as

$$F_{T^\star}(\tilde{t}) = -\int_0^1 f_{T^\star}(\check{t})(1-\check{t})^{-2}d\check{t}, \qquad (6.5.6)$$

where $f_{T^\star}(\cdot)$ is given in (6.5.4) for the scale group and in (6.5.5) for the location group.

Remark 6.6. An important point to realize is that when using quadrature methods, like the Gauss–Legendre rule (see Appendix A.2), the only information such a method has

about the integrand is a sequence of numerical values for it. To get a definite result for the integral, such a procedure then effectively has to make certain assumptions about the smoothness and other properties of the integrand. If a sufficiently pathological integrand is given, these assumptions may not be valid, and as a result, we may simply obtain the wrong answer. This problem may occur, for example, if one tries to integrate numerically a function which has a very thin peak at a particular position. The numerical integration routine samples the function at a number of points, and then assumes that the function varies smoothly between these points. As a result, if none of the sample points come close to the peak, then the peak will go undetected, and its contribution will not be correctly included. Therefore it is very important to get an idea of the effective range of the studentized statistic $T^\star = Z^\star$ in (6.5.2). □

6.5.2.3. Range of the studentized statistic

Consider the studentized statistic (6.1.1) and represent the estimators as functionals, i.e. $Z = z(F), T = t(F)$ and $V = v(F)$. Then (6.1.1) becomes

$$z(F) = \frac{t(F) - \theta}{v(F)^{1/2}}.$$

Based on Section 5.1.2 and following Hinkley and Wei (1984, Section 2.1) and Davison and Hinkley (1997, Section 2.7.2 and Example 9.18) it can be shown that the influence function for the studentized statistic $Z = z(F)$ is $L_z(x; F) = v(F)^{-1/2}L_t(x; F)$, where $L_t(\cdot; F)$ is the influence function of $T = t(F)$. The corresponding empirical influence values for the studentized quantity $z(\hat{F})$ are $l_i = v^{-1/2}\tilde{l}_i$, where v is the observed value of V and $\tilde{l}_1, \ldots, \tilde{l}_n$ are the empirical influence values of the statistic T.

For the classical studentized statistic defined in Section 6.3 we have

$$l_i = v^{-1/2}(x_i - \bar{x}), \quad i = 1, \ldots, n, \tag{6.5.7}$$

with $v = s = n^{-1}\sum(x_i - \bar{x})^2$.

Using Remark 5.3 one sees that the $\tilde{l}_i$ for a M-estimate of location are given by $-n\sigma_0\psi_k\{(x_i - \theta_0)/\sigma_0\}/\sum\psi'_k\{(x_i - \theta_0)/\sigma_0\}$, where θ_0 and σ_0 are the values for the original data $x_1, \ldots, x_n$. Hence the empirical influence values of the studentized versions of the M-estimates given in Section 6.4 are

$$l_i = -n\sigma_0 v_L^{-1/2}\psi_k\left(\frac{x_i - \theta_0}{\sigma_0}\right)\left\{\sum_{i=1}^{n}\psi'_k\left(\frac{x_i - \theta_0}{\sigma_0}\right)\right\}^{-1}, \tag{6.5.8}$$

where v_L is the nonparametric linear delta method estimate (6.4.1) evaluated at $\hat{\theta} = \theta_0$ and $\hat{\sigma} = \sigma_0$. Substituting (6.4.1) in (6.5.8) yields

$$l_i = -n\psi_k\left(\frac{x_i - \theta_0}{\sigma_0}\right)\left\{\sum_{i=1}^{n}\psi_k^2\left(\frac{x_i - \theta_0}{\sigma_0}\right)\right\}^{-1}. \tag{6.5.9}$$

For the studentized version of Huber's M-estimate of location (6.4.3) equation (6.5.9) reduces to $l_i = -n\psi_k(x_i - \theta_0)/\sum \psi_k^2(x_i - \theta_0)$, whereas the ones for the studentized version of Huber's M-estimate of location with initial MAD scaling are (6.5.9) using as σ_0 the observed MAD, and the ones for the studentized version of Huber's proposal 2 are (6.5.9) using as θ_0 and σ_0 the solutions to (6.4.10).

A linear delta method estimate of the variability of the studentized statistic can now easily be obtained by means of equation (5.1.5) with l_i defined in (6.5.7) for the classical studentized statistic and in (6.5.9) for the studentized M-estimators. Moreover, these empirical influence values can be used to get an estimate of the effective range of the studentized statistics. Denote by l_{min} the smallest empirical influence value of the studentized statistic and by l_{max} its largest value. One may hope that all possible values of the studentized bootstrap statistic lie in the interval $[l_{min}, l_{max}]$. If so, no change of variables would be needed in (6.5.2) and one could approximate the CDF (6.5.2) by means of

$$F_{T^\star}(\tilde{t}) = \int_{l_{min}}^{\tilde{t}} f_{T^\star}(t)dt. \qquad (6.5.10)$$

Let us consider Newcomb's data (Table 5.2) and the shrimp data (Table 6.3). Besides $[l_{min}, l_{max}]$ I calculated for each of the four studentized statistics the range of the $R = 49,999$ bootstrap replicates and the one resulting from the saddlepoint approximation using the Laplace approximation (see Section 6.2). Note that the latter was determined as in Section 5.3.1 (Remark 5.5) using a binary search (BS). But, depending on the statistic of interest, this BS is very time-intensive and hence the investigations in this section are two-fold: we need an appropriate range of possible values not only for performing the numerical integration in (6.5.2) but also for the case when the Laplace approximation is used in order to replace the BS. Each of these ranges is represented by a shingle in Figure 6.21 for Newcomb's data and in Figure 6.22 for the shrimp data. From Figure 6.21 we see that for all four statistics the range of the empirical influence values covers the range of the bootstrap replicates and the one of the saddlepoint approximation. But, one also notices that the range of the empirical influence values is significantly larger than the other two. This is especially true for the classical studentized statistic due to the extreme outlier -44 in Newcomb's data (Table 5.2). Hence the use of (6.5.10) would not overcome the problems of the numerical quadrature methods mentioned in Remark 6.6.

A similar pattern can be seen in Figure 6.22 for the shrimp data. For the studentized versions of Huber's M-estimators of location all ranges are about the same. But, for Huber's proposal 2 the range of the bootstrap replicates is larger than the ones from the saddlepoint approximation and the one of the empirical influence values. This illustrates another limitation of this approach. Additional examples lead me to the conclusion that by means of the interval $[l_{min}, l_{max}]$ one gets a sort of idea where the effective range could be, and hence the $[l_{min}, l_{max}]$ are very useful for replacing the BS, but not for performing an optimal numerical integration. Nevertheless, I applied to (6.5.10) the adaptive integration schemes outlined in Appendix A.3 and noticed a slight improvement in speed,
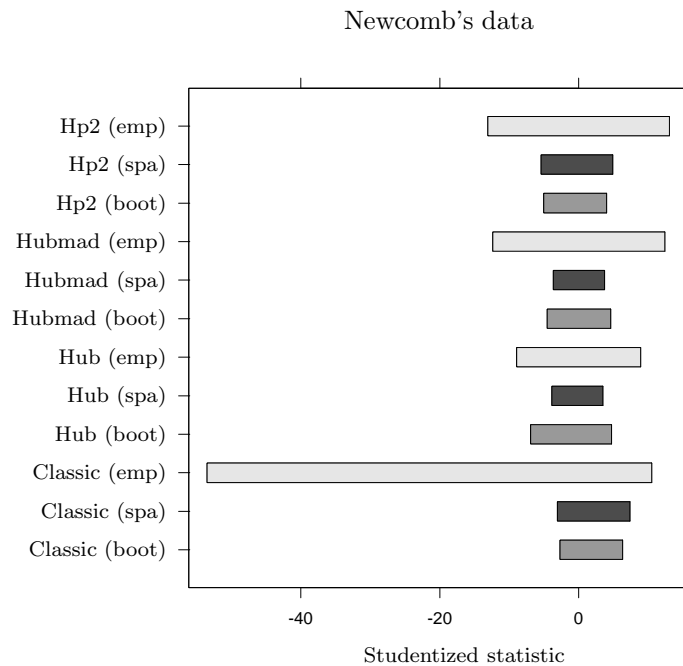
Newcomb's data

Figure 6.21. *Ranges using Newcomb's data of the classical studentized statis-*
tic (Classic), of the studentized versions of Huber's M-estimate of location
(Hub), of its initially MAD scaled version (Hubmad) and of the studen-
tized version of Huber's proposal 2 (Hp2). The type of range is given in the
brackets, where 'emp' denotes empirical influence values, 'spa' stands for
saddlepoint approximation and 'boot' denotes bootstrap.

but the lack of accuracy discussed in Remark 6.6 still remained in several examples. But, for such cases one could bypass these problems using the split-t transformations proposed in Genz and Kass (1997) prior to the use of adaptive numerical integration algorithms. Fortunately this does not occur often.

6.5.2.4.   Examples

To illustrate the comments made so far I considered Darwins's data (Table 6.7) and Short's data (Table 6.1). As studentized statistics I took the studentized version of Huber's proposal 2 for Darwins's data and the classical studentized statistic for Short's data.

PDF approximation

For both studentized bootstrap statistics I computed, besides the $R = 49,999$ bootstrap replicates, the following PDF approximations: the saddlepoint approximation based on Laplace's method (see Section 6.2) with the interval $[l_{min}, l_{max}]$ as the range of possible values, and approximations obtained through numerical integration of (6.5.4). Based on the comments in Appendix A.5 I used as quadrature methods the default S-Plus function `integrate`, which implements uni-dimensional adaptive 15-point Gauss–Kronrod quadrature, and a 128-point Gauss–Legendre (GL) rule, as implemented in the S-Plus function `GL.integrate.1D` (Table A.1). Note that to calculate an entire quadrature
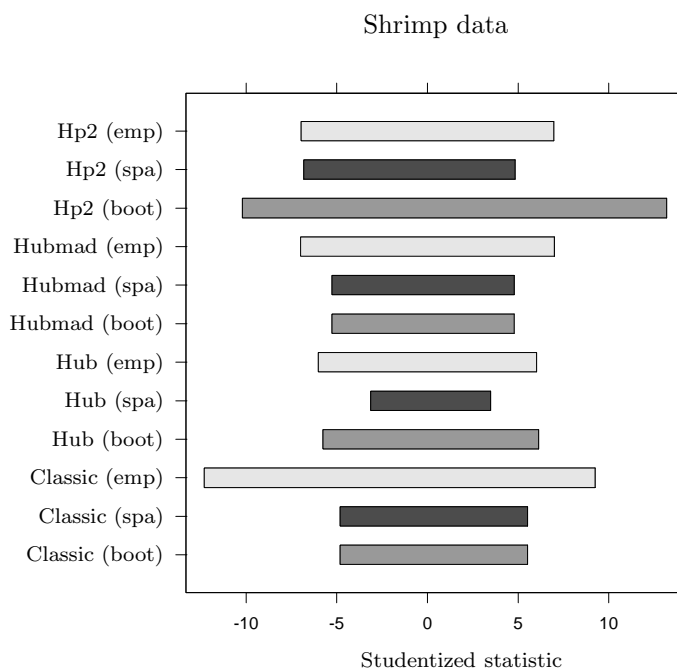
Shrimp data

Figure 6.22. *Ranges using the shrimp data of the classical studentized statistic (Classic), of the studentized versions of Huber's M-estimate of location (Hub), of its initially MAD scaled version (Hubmad) and of the studentized version of Huber's proposal 2 (Hp2). The type of range is given in the brackets, where 'emp' denotes empirical influence values, 'spa' stands for saddlepoint approximation and 'boot' denotes bootstrap.*

approximation, the values of (6.5.4) for 50 evenly spaced points between $[l_{min}, l_{max}]$ were calculated and a spline was fitted to these values. To improve the accuracy of the approximations the resulting approximation for the density was integrated by means of the trapezoidal rule and rescaled so that the density integrates to one. Remark also that for the 128-point GL rule 128 evaluations of the joint density (6.2.6) with each time the numerical minimisation of the $2d - 1$ equations (6.2.9) are needed.

For Darwin's data the results are shown in the left panel of Figure 6.23. The saddlepoint approximation (solid) underestimates the lower tail and overestimates the centre. The same happens with `integrate` (dotted), whereas the 128-point GL rule (dashed) seems to perform best, being the only approximation that captured the multi-modality in the centre and the behaviour in the lower tail. The approximate CPU time for these computations are given in the first row of Table 6.8. The default S-Plus function `integrate` is very slow compared to the other approximations, but it still is significantly faster than the brute-force bootstrap computation which took about 12 hours. The PDF approximations for Short's data are shown in the right panel of Figure 6.23. The saddlepoint approximation (solid) and the approximation obtained by means of a 128-point GL rule (dashed) seem to work very well, whereas the use of the S-Plus function `integrate` is very inaccurate. As shown in the second row of Table 6.8 the latter is about 10 times slower in CPU time than the direct simulation of the $R = 49,999$ replicates. Only the saddlepoint approximation outperforms the bootstrap simulation

Darwin's data   Short's data
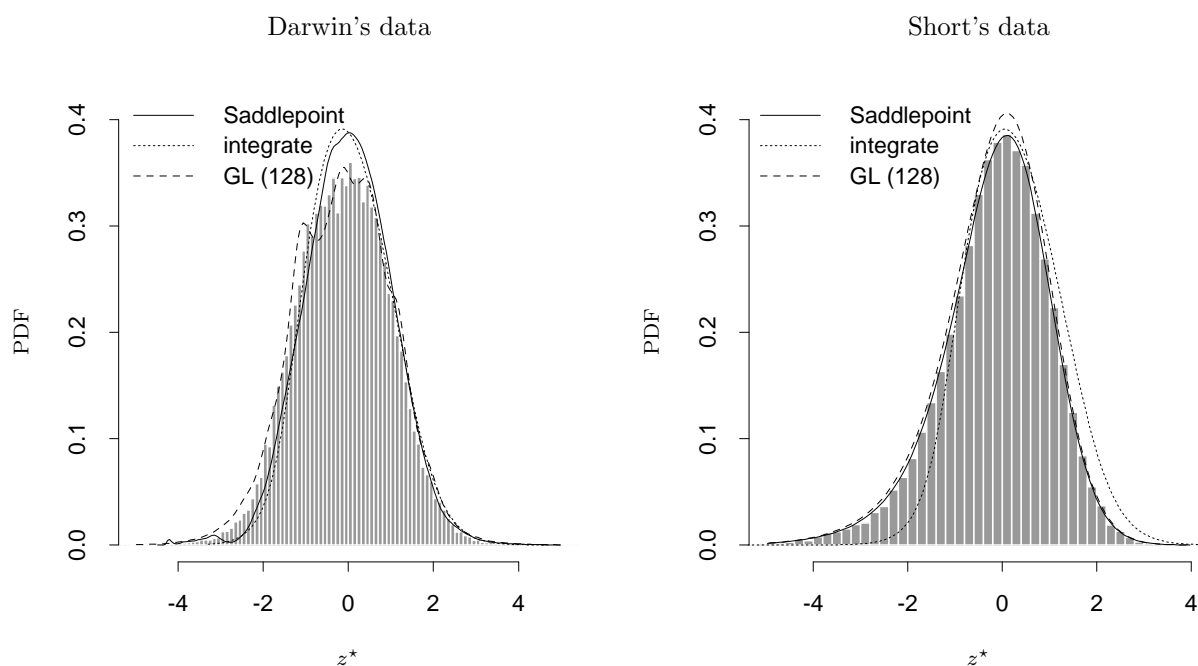
Figure 6.23. *PDF approximations. Left panel: histogram of $R = 49,999$ replicates of the studentized version of Huber's proposal 2 for Darwin's data with saddlepoint, `integrate` and 128-point Gauss–Legendre (GL) approximations of the PDF. Right panel: histogram of $R = 49,999$ replicates of the classical studentized statistic for Short's data with saddlepoint, `integrate` and 128-point Gauss–Legendre (GL) approximations of the PDF.*

Table 6.8. *Approximate CPU time in minutes of the PDF calculation using the bootstrap, the saddlepoint approximation, `integrate` and a 128-point Gauss–Legendre (GL) quadrature method.*

	Bootstrap ($R = 49,999$)	Saddlepoint	`integrate`	GL (128)
Darwins's data	712	39	184	37
Short's data	15	3	117	22

in CPU time in this example, being about 5 times faster.

Additional examples showed me that one should not use `integrate` as the computation may then become very time-intensive and may lead to inaccurate PDF approximations. But, a GL rule with 128 points seems to be a good choice in practice.

CDF approximation

The use of quadrature methods for the CDF approximation results in integrating numerically (6.5.10). In order to approximate the entire CDF of $Z^\star$ the values of $F_{T^\star}(t)$ are calculated for 50 values of t equally spaced between l_{min} and l_{max} and a spline is used to interpolate between these values. It is important to note that the use of a 128-point GL rule (S-Plus function `GL.integrate.2D`, Table A.3) implies 16,384 function evaluations solving each time the equations (6.2.9), which is very computer-intensive. For instance, at $t = z^\star = 1$, using Darwin's data and the studentized version of Huber's proposal 2, the approximation of $F_{T^\star}(1)$ given in (6.5.10) using a 128-point GL rule took about 1,238 minutes in CPU time. The use of a 64-point GL rule decreases this to about 36 minutes but increases the inaccuracy. Hence the time needed to get the approximation for the entire range of the 50 values between l_{min} and l_{max} is clearly outperformed by the bootstrap running time given in the first column of Table 6.8. Applying a 128-point GL rule to Short's data for the classical studentized statistics in order to get $F_{T^\star}(1)$ took about 87 minutes, whereas the computation in this context for the $R = 49,999$ bootstrap samples took about 15 minutes (see Table 6.8)! Besides a 128-point GL rule I suggested in Appendix A.5 also the use of adaptive integration schemes, like the S-Plus functions `dcuhre` and `adapt` described in Appendix A.3. But simulation showed that `dcuhre` is extremely slow. For example, already for the classical studentized statistic as applied to Short's data it took about 33 minutes to get a single estimate of (6.5.10). And for Darwin's data using the studentized version of Huber's proposal 2 about 1,649 minutes were needed! The function `adapt` was a bit faster: about 66 minutes were needed to get the entire approximation to the CDF of the classical studentized statistic for Short's data. The resulting `adapt` approximation is represented in Figure 6.24 as the dashed line, together with the saddlepoint approximation (dotted). The latter was already represented in the right panel of Figure 6.2. The saddlepoint approximation based on Laplace's method (dotted) works extremely well and the one obtained by means of `adapt` (dashed) is accurate in the lower tail but underestimates slightly the upper tail. For Darwin's data using the studentized version of Huber's proposal 2, `adapt` needed on average about 51 minutes for the computation of a single estimate of (6.5.10)!

The examples in this section clearly illustrate the drawbacks of numerical integration for the computation of the CDF of $T^\star$. They become useless in practice as their running time is outperformed by direct simulation of the $R = 49,999$ bootstrap replicates. One may think that this is due to the use of interpreted languages like S-Plus or R, but I do not think that this is the case as numerical integration in statistics and especially in multi-dimensional problems still raises many open questions.

6.6.   Conclusion

In this chapter I considered saddlepoint approximations to studentized bootstrap distributions. As examples I used in Section 6.3 the classical studentized statistic, and in Section 6.4 the studentized versions of Huber's M-estimate of location, of Huber's M-estimate of location with initial MAD scaling and of
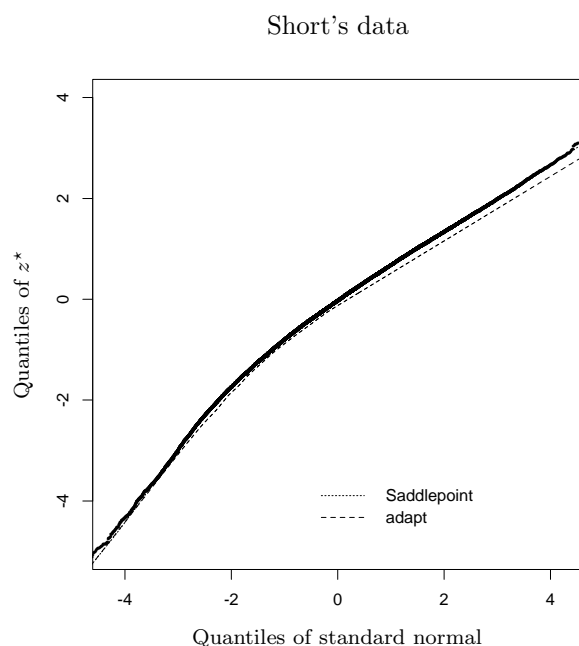
Short's data

Figure 6.24. *CDF approximations for the classical studentized statistic for Short's data. Q-Q plot of the $R = 49,999$ replicates, with the saddlepoint approximation and the* `adapt` *approximation.*

Huber's proposal 2. Their accuracy and their effectiveness for reducing the computational expense was illustrated by several examples. The saddlepoint approximations of Section 6.2 are based on a Laplace approximation and hence care is needed with their application. Fortunately, experience shows that the problems described in Section 6.5 do not occur often. Therefore the need to perform a very computer-intensive numerical integration as in Section 6.5.2 disappears. Especially concerning these numerical integration methods it was made clear that questions on convergence and efficient implementation remain. Some future approaches stated in Appendix A are promising.

Finally, note that Hampel *et al.* (1986, page 105) recommended the use of MAD scaling for M-estimators instead of Huber's proposal 2, and noted that the latter is less reliable. More recently, Thomas (2000) used the bootstrap to get an idea of the precision of various robust estimates of location. He noted that Huber's proposal 2 seems to be a preferable measure of location which is not unduly affected by outliers or asymmetry in the data. This is also what I observed in the bootstrap context when the estimators are in their studentized form. Even if the saddlepoint approximations to the studentized version of Huber's proposal 2 are more time-intensive, the resulting approximations are more accurate. Hence the hope is that the resulting studentized confidence intervals are very accurate as well and will deliver coverages that are close to nominal levels. This will be investigated in the next chapter.

7. Studentized bootstrap confidence intervals

> *'Confidence intervals have become familiar friends in the applied statistician's collection of data-analytical tools. They combine point estimation and hypothesis testing into a single inferential statement of great intuitive appeal. Recent advances in statistical methodology allow the construction of highly accurate approximate confidence intervals, even for very complicated probability models and elaborate data structures.'*
>
> Thomas J. DiCiccio and Bradley Efron (1996, page 189)

The major application for distributions and quantiles of an estimator, like the studentized statistic Z, is in the calculation of *confidence intervals*. Confidence intervals or, more generally, *confidence regions* form an important class of statistical methods. In these methods, the outcome of the statistical analysis is a subset of the set of possible values of unknown parameters. Confidence procedures are directly related to other kinds of standard statistical methods, in particular to point estimation and to hypothesis testing.

It was mentioned in Section 6.1 that through the use of the bootstrap we can obtain accurate confidence intervals without having to make normal or Student theory assumptions. One way to get such intervals is by means of the bootstrap-t method, which was described briefly in Section 6.1. Moreover, it was stated that the studentized statistic Z is a good approximate pivotal quantity. The remainder of Chapter 6 was then concerned with getting accurate saddlepoint approximations to the bootstrap PDF and CDF of the studentized bootstrap statistic $Z^\star$, which forms the basis of the studentized bootstrap confidence intervals. The aim of this chapter is to know whether the use of the saddlepoint approximations of Section 6.2 yields robust confidence intervals with coverages close to the nominal level and this with a short interval.

After an introduction in Section 7.1 the results of an extensive simulation study are given in Section 7.2. Finally, concluding remarks are given in Section 7.3.

7.1. Introduction

n many estimation situations it is of substantial interest to compute *reliable* confidence intervals for parameters of interest. A reliable confidence interval method is one giving intervals whose actual coverages are close to the nominal level. The aim is to construct a random interval $CI_{1-2\alpha}$ with nominal coverage $(1 - 2\alpha)$ such that

$$\Pr(\theta \in CI_{1-2\alpha}) = 1 - 2\alpha, \tag{7.1.1}$$

where θ is the true parameter value of the estimator of location T. The coverage probability $(1-2\alpha)$ is the relative frequency with which the confidence interval would include, or cover, the true parameter value in repetitions of the process that produced the data $x_1, \ldots, x_n$. Then the interval $CI_{1-2\alpha}$ has probability exactly $(1 - 2\alpha)$ of containing the true value of θ. More precisely, denoting the confidence interval $CI_{1-2\alpha}$ by $[\hat{\theta}_\alpha, \hat{\theta}_{1-\alpha}]$, the probability that θ lies below the lower limit of $CI_{1-2\alpha}$ is exactly α, as is the probability that θ exceeds the upper limit,

$$\Pr(\theta < \hat{\theta}_\alpha) = \alpha \quad \text{and} \quad \Pr(\theta > \hat{\theta}_{1-\alpha}) = \alpha. \tag{7.1.2}$$

A $(1 - 2\alpha)$ confidence interval $[\hat{\theta}_\alpha, \hat{\theta}_{1-\alpha}]$ with property (7.1.2) is called *equi-tailed*. This refers to the fact that the coverage error 2α is divided up evenly between the lower and upper ends of the interval. Confidence intervals are almost always constructed to be equi-tailed and hence I will restrict attention to equi-tailed intervals in this chapter. Note that property (7.1.2) implies (7.1.1), but not vice-versa. That is, (7.1.2) requires that the one-sided coverage error is α on each side, rather than just an overall coverage of $(1 - 2\alpha)$.

Unfortunately, it is not possible in general to find a confidence interval to satisfy (7.1.1) exactly. Hence approximate confidence intervals are needed which involve a discrepancy between the nominal coverage $(1-2\alpha)$ and the actual coverage probability. By far the most favourite is the standard confidence interval based on a normal approximation. Such standard intervals are useful tools, but they are based on an approximation that can be quite inaccurate in practice. Thomas J. DiCiccio and Bradley Efron (1996, page 223) underlined this well-known fact by stating that

> *'If the standard intervals were invented today, they might not be publishable.'*

Not only the choice of an approximate confidence interval is important but also the choice of an appropriate pivotal quantity. It has been illustrated in Section 6.1 that a good choice is provided by the *t*-like statistic (6.1.1),

$$Z = \frac{T - \theta}{V^{1/2}}, \tag{7.1.3}$$

where T is an estimator of location and V is an estimator of its variance. The resulting $(1 - 2\alpha)$ studentized confidence interval for θ has limits

$$\hat{\theta}_\alpha = t - v^{1/2} z_{1-\alpha} \quad \text{and} \quad \hat{\theta}_{1-\alpha} = t - v^{1/2} z_\alpha, \tag{7.1.4}$$

where v is the estimated variance, t denotes the observed value of T and z_α is the α-quantile of the distribution of Z. As the distribution of Z is usually unknown in practice, approximate methods based on approximations to the quantiles of Z are needed. The simplest method is a normal approximation by use of $z_{1-\alpha} = \Phi^{-1}(1-\alpha)$. Since then $z_\alpha = -z_{1-\alpha}$, the confidence limits (7.1.4) reduce to

$$t \mp v^{1/2}\Phi^{-1}(1-\alpha), \tag{7.1.5}$$

which are known as the limits of the *standard confidence interval*. Another approximation is given by the use of $z_\alpha = t_{n-1}(\alpha)$ in (7.1.4) with $t_{n-1}(\alpha)$ denoting the α-quantile of Student's t distribution on $n-1$ degrees of freedom. As nice as it is to be able to assume an underlying normal or a Student distribution, there are problems. Perhaps the studentized statistic does not follow a normal or a Student distribution. In that case, the resulting inference may well be wrong.

Bootstrap procedures are an alternative. One way to look at them is as procedures for handling data when one is not willing to make assumptions about the parameters of the populations from which one sampled. The most that one is willing to assume is that the data are a reasonable representation of the population from which they came. One then resamples from the data and draws inferences about the corresponding population and its parameters. The resulting confidence intervals have received the most theoretical study of any topic in the bootstrap analysis. A full discussion of this theory would go beyond the scope and intent of this chapter. Therefore I refer to the following overviews. The state of research into bootstrap confidence intervals is given and discussed by DiCiccio and Efron (1996). A brief introduction is also given in Davison and Hinkley (1997, Section 2.4) and a more thorough discussion of such methods in Davison and Hinkley (1997, Chapter 5); see also Efron and Tibshirani (1993, Chapters 12–14 and 22). More recently, an illustrative review is given by Carpenter and Bithell (2000).

Following Section 6.1 consider the studentized bootstrap statistic $Z^\star$ given in (6.1.2),

$$Z^\star = \frac{T^\star - t}{V^{\star 1/2}}, \tag{7.1.6}$$

where $T^\star$ and $V^\star$ are based on a simulated bootstrap sample, $X_1^\star, \ldots, X_n^\star$. The R simulated bootstrap values of $Z^\star$ are ordered and the α-quantile of Z, z_α, is estimated by the $\alpha(R+1)$th value of these. The approximate confidence limits for the $(1-2\alpha)$ *studentized bootstrap confidence interval* for θ are then given in (6.1.3) or (7.1.4),

$$t - v^{1/2}z_{((1-\alpha)(R+1))}^\star, \quad t - v^{1/2}z_{(\alpha(R+1))}^\star. \tag{7.1.7}$$

These are often also called bootstrap-t limits. The studentized bootstrap confidence intervals are approximate as well but they are theoretically much more accurate than the standard intervals and than the basic bootstrap intervals defined in Remark 5.2. Indeed, the standard or basic bootstrap methods are all first-order accurate in the sense that

$$\Pr(\theta > \hat{\theta}_{1-\alpha}) = \alpha + O(n^{-1/2}),$$

where $\hat{\theta}_{1-\alpha} = t + v^{1/2}\Phi^{-1}(1 - \alpha)$ or $\hat{\theta}_{1-\alpha} = 2t - t^{\star}_{(\alpha(R+1))}$ (Davison and Hinkley, 1997, equation (5.6)). The latter is based on the assumed pivotality of $T - \theta$. The studentized bootstrap intervals improve on this by making

$$\Pr(\theta > \hat{\theta}_{1-\alpha}) = \alpha + O(n^{-1})$$

with $\hat{\theta}_{1-\alpha} = t - v^{1/2}z^{\star}_{(\alpha(R+1))}$. The latter is referred to as second-order accuracy (Hall, 1988). Furthermore, DiCiccio and Efron (1992) showed that the bootstrap-t limits are also second-order correct in the sense that the asymptotic expansions of $\hat{\theta}_{1-\alpha}$ and the 'correct' limit $\tilde{\theta}_{1-\alpha} = t - v^{1/2}\tilde{z}_{\alpha}$, where $\tilde{z}_{\alpha}$ is the α-quantile of the true sampling distribution of Z, agree up to the second-order term, *i.e.* $\hat{\theta}_{1-\alpha} - \tilde{\theta}_{1-\alpha} = O(n^{-3/2})$; see also Efron and Tibshirani (1993, Section 22.3), DiCiccio and Efron (1996, Section 8), Davison and Hinkley (1997, Section 5.4) or van der Vaart (1998, Chapter 23). Moreover, Hall (1988), DiCiccio and Romano (1988), Beran (1987) and DiCiccio and Efron (1992) have shown that the various methods of construction confidence limits are equivalent to first-order or second-order to bootstrap-t methods. A comparison of their behaviour to second order is given by Bickel (1992).

Unfortunately, these are asymptotic results which are not always an useful guide for performance in practice. Hence simulation studies were conducted to compare the performance of the intervals especially on small samples. Empirical studies, like the ones in Shao and Tu (1995, Section 4.4.4), by Canty *et al.* (1996) and in Davison and Hinkley (1997, Example 5.7 and Section 5.7), have revealed that the bootstrap-t method performs well in terms of coverage error, even for small samples. However, the same studies typically show that the endpoints of these intervals can be highly unstable and wide, particularly in moderate samples. Moreover, Efron and Tibshirani (1993, footnote on page 160) remarked that the bootstrap-t method can be heavily influenced by outliers in the data. Bickel (1992) gave also some robustness remarks, which strongly mandate a robust estimate for scale as classical standard error estimates can be unstable. In order to guarantee reasonable solutions for small sample sizes one could also use a variance stabilizing transformation; see Tibshirani (1988), Efron and Tibshirani (1993, Section 12.6) and Polansky (2000). DiCiccio and Efron (1996) and Canty *et al.* (1996) argued that the superior coverage of the bootstrap-t method for small sample sizes is due to the fact that bootstrap-t confidence intervals tend to be long and highly variable. But Polansky (2000, Section 3) showed that this is not always the case.

What is most surprising is that in my knowledge none of the existing bootstrap-t simulation studies, expect the small one in Davison and Hinkley (1997, Table 9.10), considered robust estimators for $T^{\star}$ and $V^{\star}$ in (7.1.6), but most of them noted that the bootstrap-t method may be influenced by outliers. Hence an aim of the present chapter is to fill this gap by considering the studentized bootstrap versions of the M-estimates given in Section 6.4, together with the classical studentized statistic given in Section 6.3. To overcome the cruel computational burden of the bootstrap the saddlepoint approximations to these studentized bootstrap statistics as given in Section 6.2 will be considered.

Remark 7.1. As already mentioned in Section 5.3 and at the beginning of Chapter 6,

the sadddlepoint approximations based on Daniels (1954, 1983) are related to the ones of Field and Hampel (1982) or Field (1982). Using the latter Tingley and Field (1990) constructed confidence intervals, which are second-order correct in terms of coverage; see also Field and Ronchetti (1990, Section 6.3). Tingley and Field's (1990) use of M-estimates was motivated by the desire to have robust confidence intervals. Field and Ronchetti (1985) showed that the test procedures based on robust M-estimates will be robust as well; see also Hampel *et al.* (1986, Section 3.2) or Field and Ronchetti (1990, Section 7.3). □

7.2. Simulation study

The advantages of the use of simulation, particularly in robustness studies, are well-known: one can mimic a sample from any mathematically definable probability distribution. It results a great flexibility in the specification of distributions and one has the advantage of knowing exactly what mechanism produced the data and hence one can easily evaluate the performance of statistical procedures. But, it is clear that there is no guarantee that the pseudo-samples in simulation studies are representative of real data. This drawback is underlined by Stigler (1977, Section 1). In order to settle the question of whether robust estimators are necessary he then performed a study on real data sets and remarked that real data do exhibit somewhat different behaviour from that of the simulated data used in most robustness studies. Clearly this affects the consequent recommendations for the choice of an estimator and the assessments of the relative performance of estimators. Stigler's (1977) study is reviewed and extended by Rocke *et al.* (1982).

Following Hampel (2000, page 2) '*robust inference is the inference which is insensitive to (smaller or larger) deviations from the assumptions under which it is derived*'. Commonly used assumptions in statistics, like normality or symmetry, are only approximations to the reality and the questions arise what deviations tend to occur in practice, what effects they have on known statistical procedures, and how to develop better procedures. For answering these questions the bootstrap-t method using studentized M-estimators seems to be useful. But before giving any recommendations it is important to check if the procedure is robust when standard assumptions are not met. Suppose that one wishes to make inferences regarding a location parameter θ based on the studentized statistic Z as given in (7.1.3). Basically, there are two types of robustness we would hope to achieve. First, *robustness of performance* for T, and to then match T with an estimated standard error $V^{-1/2}$ yielding *robustness of validity* over a range of possible distributions for the sample.

In the non-bootstrap context studentized location estimators have been studied intensively in the statistical literature; see Barnett and Lewis (1994, pages 170–174) for a selective review. An extensive study is reported by Gross (1976) who studied a wide variety of confidence intervals procedures based on studentized statistics. Other studies of this nature have been proposed by Shorack (1976) and Gross (1977). Shorack (1976) remarked that studentized robust statistics outperform the classical studentized

statistic. Gross (1976) measured robustness of performance by the expected confidence interval length, and robustness of validity by the level of the conservative confidence interval based on the distributions examined. Already Leone *et al.* (1967) examined the sampling behaviour of studentized forms of robust location estimators for normal mixture models. Although not considering studentized statistics, the Princeton study (Andrews *et al.*, 1972) needs to be mentioned as well.

But the problem with all these studies is that asymmetry has been neglected. Given an asymmetric distribution any (robust) confidence interval, devised for symmetric distributions, will contain a fixed bias and the resulting inference will be wrong when no adjustment for this bias will be done. What is most surprising is that all studies noted the importance of asymmetry as symmetry is plainly unrealistic, but none of them considered it. Andrews *et al.* (1972, pages 109–111), for example, listed two tables containing the bias of two asymmetric distributions but they did not perform any analysis of these. This regrettable omission was explained by Hampel (1995, page 6): '*when it came to the analysis, the computer outputs for these cases arrived in Zurich after the deadline of the manuscript*'. Hampel (1995, Section 4) then completed the Princeton study for these two asymmetric cases.

An interesting generalization of (7.1.3) for the case of the classical studentized statistic (Section 6.3) is proposed by Johnson (1978), which corrects the resulting standard confidence interval for the asymmetry of the sampling distribution. Some additional tests and confidence intervals giving specific protection against outliers in normal samples are described in Barnett and Lewis (1994, Section 5.4).

Except Johnson (1978), Hampel (1995, Section 4) and Davison and Hinkley (1997, pages 484 and 485) I did not find additional studies were asymmetric distributions were used. All other studies only considered (heavy-tailed) symmetric distributions. This fact is very regrettable and strange as the aim of robust procedures is to systematically investigate the effects of deviations from modelling assumptions on known procedures, as this is the case with asymmetric distributions, and, if necessary, to develop new and hopefully better procedures.

Considering asymmetric distributions and the $(1 - 2\alpha)$ studentized bootstrap confidence interval for θ given in (7.1.7) one may think that one needs to add an empirical bias adjustment in the numerator of Z, for example based on the empirical second derivatives of $t(F)$. But Davison and Hinkley (1997, page 194) stated that '*an empirical bias adjustment could be incorporated, but this is usually not worthwhile, because the effect is implicitly adjusted for in the bootstrap distribution*' of $Z^\star$. This well-known fact was underlined by their good coverage results of (7.1.7) in considering an asymmetric distribution. Simulation under asymmetric distributions confirmed that it is not necessary to add such a bias adjustment in the numerator of Z, as I remarked that the studentized bootstrap accounts for this automatically.

7.2.1. Description of the study

Suppose that $X_1, \ldots, X_n$ are independent and identically distributed random variables whose PDF and CDF are denoted by f and F. Then $x_1, \ldots, x_n$ can be thought of

as their outcomes. In this simulation study samples will be generated from F and as studentized statistics I consider the classical studentized statistic (Section 6.3) and the three studentized versions of the M-estimates of location (Section 6.4).

For the classical studentized statistic (6.3.1) we have $T = \bar{X}$ and its variance v can be approximated by (5.1.5), $v_L = n^{-2} \sum l_i^2$ with $l_i = x_i - \bar{x}$ (see Example 5.1). Note that $v_L = n^{-1} s$, where s is the observed value of the classical variance, $s = n^{-1} \sum (x_i - \bar{x})^2$. Using (7.1.7) this yields the *classical* $(1 - 2\alpha)$ *studentized bootstrap confidence interval* limits for θ,

$$\bar{x} - n^{-1/2} s^{1/2} z^\star_{((1-\alpha)(R+1))} \quad \text{and} \quad \bar{x} - n^{-1/2} s^{1/2} z^\star_{(\alpha(R+1))}. \tag{7.2.1}$$

For the studentized versions of the M-estimates of location the variance of T, v, can again be estimated by the delta method variance estimate v_L given in (6.4.1). For the studentized version of Huber's M-estimate of location (6.4.3) the latter reduces to

$$v_L = \sum_{i=1}^{n} \psi_k^2(x_i - \hat{\theta}) \left\{ \sum_{i=1}^{n} \psi_k'(x_i - \hat{\theta}) \right\}^{-2},$$

where $\hat{\theta}$ is Huber's M-estimate of location issued from the simulated sample $x_1, \ldots, x_n$. For the studentized version of Huber's M-estimate of location with initial MAD scaling (6.4.8) the variance estimate of T is

$$v_L = \sigma_0^2 \sum_{i=1}^{n} \psi_k^2 \left(\frac{x_i - \hat{\theta}}{\sigma_0} \right) \left\{ \sum_{i=1}^{n} \psi_k' \left(\frac{x_i - \hat{\theta}}{\sigma_0} \right) \right\}^{-2},$$

where σ_0 is the observed MAD as defined in (5.3.7) and $\hat{\theta}$ is Huber's M-estimate of location with initial MAD scaling. Finally, using the studentized version of Huber's proposal 2 (6.4.11) gives

$$v_L = \hat{\sigma}^2 \sum_{i=1}^{n} \psi_k^2 \left(\frac{x_i - \hat{\theta}}{\hat{\sigma}} \right) \left\{ \sum_{i=1}^{n} \psi_k' \left(\frac{x_i - \hat{\theta}}{\hat{\sigma}} \right) \right\}^{-2},$$

where $\hat{\theta}$ and $\hat{\sigma}$ are the simultaneous solutions of (6.4.10). Hence using (7.1.7) the *robust* $(1 - 2\alpha)$ *studentized bootstrap confidence interval* limits for θ become

$$\hat{\theta} - v_L^{1/2} z^\star_{((1-\alpha)(R+1))}, \quad \hat{\theta} - v_L^{1/2} z^\star_{(\alpha(R+1))}. \tag{7.2.2}$$

The studentized bootstrap confidence intervals (7.2.1) and (7.2.2) are based on the $\alpha(R+1)$th and $(1-\alpha)(R+1)$th ordered values of the R simulated bootstrap replicates of $Z^\star$. To replace the computer-intensive bootstrap simulation in order to approximate the quantiles of Z I will use the saddlepoint approximations to the bootstrap CDF of $Z^\star$ as seen in Section 6.2. For example, $z^\star_{(\alpha(R+1))}$ will then be replaced with the approximate α-quantile of $Z^\star$ issued from the use of the saddlepoint approximation, $z_{s,\alpha}$ say. More detail will be given in Sections 7.2.3 and 7.2.4.

Table 7.1. *The distributions and their standardized skewness ϱ_3 and standardized kurtosis ϱ_4 (taken from Kokoska and Nevison (1989, Table 2)). The '—' indicates that the corresponding cumulant does not exist. The italic-faced values are the estimated standardized cumulants obtained by averaging over 100 sample cumulants from samples of size 10^6. The chosen values for n are given in the last column.*

	Distribution	ϱ_3	ϱ_4	n
F_1	$N(0,1)$	0	3	$5, 10, 15, 20, 40, 80$
F_2	t_3	—	—	$10, 15, 20, 40$
F_3	t_5	0	3.24	$10, 15, 20, 40$
F_4	Cauchy	—	—	$10, 15, 20, 40$
F_5	$0.9N(0,1)+0.1N(0,16)$	*0.0005*	*12.7109*	$10, 15, 20, 40$
F_6	$0.8N(0,1)+0.2N(10,1)$	*1.3696*	*3.2215*	$10, 15, 20, 40$
F_7	Slash	*-47.5133*	*521,136.5000*	$10, 15, 20, 40$
F_8	χ^2_5	0.04	0.054	$10, 15, 20, 40$
F_9	Folded t_7	*1.7962*	*9.4404*	$10, 15, 20, 40$
F_{10}	Folded $N(0,1)$	*0.9956*	*3.8698*	$10, 15, 20, 40$

Table 7.1 lists the distributions used in the simulation study. The corresponding standardized skewness $\varrho_3 = \kappa_3/\kappa_2^{3/2}$ and standardized kurtosis $\varrho_4 = \kappa_4/\kappa_2^2$ are listed as well as the different sample sizes n considered. Recall from Section 2.2 that κ_2 is the second cumulant of the random variable X, the skewness κ_3 its third cumulant and the kurtosis κ_4 its fourth cumulant. The skewness and the kurtosis are used as measures of how close a random variable is to normality. Skewness can be seen as a measure of asymmetry of a distribution and the kurtosis as a measure of its flatness or peakedness. Small values of their standardized versions suggest that X is close to normal; see also McCullagh (1987, Section 2.6). The symmetric distributions are F_1–F_7. Distributions F_1–F_4 are all standard. Distributions F_5 and F_6 are obtained by taking a $N(0,1)$ variate with probability 0.9 or 0.8 and taking a $N(0,16)$ or a $N(10,1)$ variate with probability 0.1 or 0.2. Distribution F_7 is the distribution of a $N(0,1)$ variate divided by an independent $U(0,1)$ variate. The asymmetric distributions are F_8–F_{10}. Distribution F_8 is standard and distribution F_{10} is the limiting form of distribution F_9 as the number of degrees of freedom tends to infinity. Other distributions were studied as well. The conclusions for these distributions were generally the same as the distributions I study here and hence have not been included.

Table 7.2. *The* S-Plus *and* R *function* `qqhnorm` *to produce half-normal plots.*

```
qqhnorm <- function(x, line=F, xlab="Half-normal quantiles",
                    ylab=deparse(substitute(x)), ...)
{
 x <- abs(x[!is.na(x)])
 y <- qnorm((1 + ppoints(length(x)))/2)[order(order(x))]
 plot(x=y, y=x, xlab=xlab, ylab=paste("|",ylab,"|"), ...)
 if (is.numeric(line)) abline(0, line)
 invisible()
}
```

7.2.2. Pivotal distributions

The first thing to check is to look whether or not a normal approximation method as used in (7.1.5) would work. This can be assessed by making a normal Q-Q plot of the sampled values of Z. By such a plot one gets an idea of how the pivotal distribution of Z appears. It is clear that if the Q-Q plot suggests that a normal approximation is poor then the latter should be replaced completely. This would also clearly justify the use of studentized bootstrap confidence interval methods. For distributions symmetrical about zero, this seems to be the case with the studentized statistics applied to symmetric distributions, I found it more appropriate to use a half-normal plot instead of a normal Q-Q plot. This can easily be done using the function `qqhnorm` given in Table 7.2. By means of a half-normal plot any information on symmetry will be lost and hence it is not suited for asymmetric distributions. The half-normal plot reveals the presence of outliers as outliers appear at the top right of the plot as distinct points, and departure from the straight line means that the normal model is not satisfactory.

I generated $20,000$ samples from the distributions listed in Table 7.1 and on each of the samples I calculated the value of Z for all four statistics, namely for the classical studentized statistic and for the three studentized statistics based on robust M-estimates. The half-normal plots of the $20,000$ simulated values of (7.1.3) using F_1 with $n = 5$ are shown in the top panels of Figure 7.1 and the ones for F_1 with $n = 80$ in the bottom panels of Figure 7.1. The patterns in using the other symmetric distributions (F_2–F_7) are about the same. One notes that independent of the statistic there seems to be a convergence to normality if n increases. When $n = 5$ (Figure 7.1, top panels) there are significant departures from normality in the pivotal distributions. Hence for moderate n a normal approximation clearly would fail. In using $n = 80$ (Figure 7.1, bottom panels) a normal distribution seems to provide a reasonable fit to the pivotal distribution of the studentized statistics. This illustrates in some sense the asymptotic normality of the studentized statistics when symmetry is assumed.

In considering the asymmetric distribution F_8 the Q-Q plots in Figure 7.2 illustrate
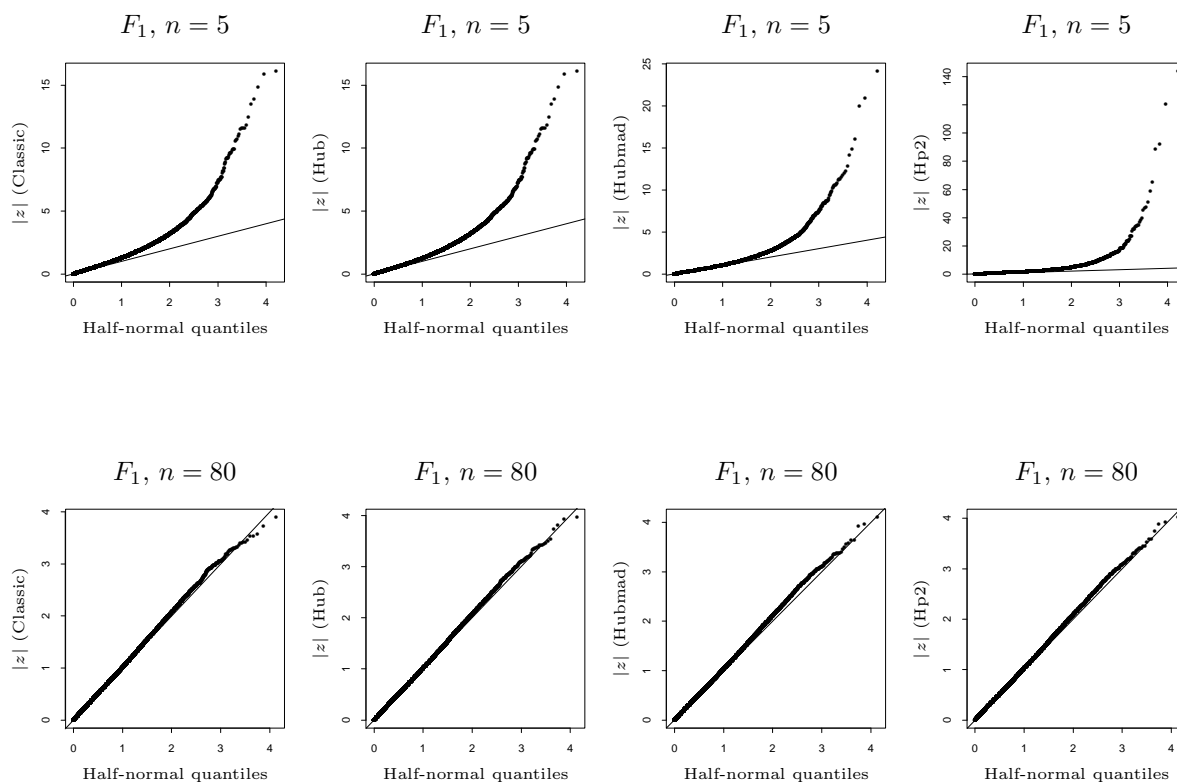
Figure 7.1. *Half-normal plots of* 20,000 *simulated values of the classical studentized statistic (Classic), of the studentized versions of Huber's M-estimate of location (Hub), of its initially MAD scaled version (Hubmad) and of the studentized version of Huber's proposal 2 (Hp2) (from left to right). Top panels: underlying $N(0,1)$ distribution (F_1) with $n = 5$. Bottom panels: underlying $N(0,1)$ distribution (F_1) with $n = 80$.*

that something similar is going on. In the top panels of Figure 7.2 ($n = 10$) normality seems to be not reasonable, whereas for $n = 40$ (Figure 7.2, bottom panels) the quantiles of Z are getting closer to the fitted normal approximation (solid line). Nevertheless one notes that the pivotal distributions are shifted to the right (on the vertical axes) compared to the ones of Figure 7.1 as a result of the underlying asymmetry of F_8. Similar patterns resulted from the use of F_9 and F_{10} and hence are not reproduced here.

In summary, one notes that the assumption of normality is unrealistic when sample sizes are small. One could try to improve the normal approximation in some way, or replace it completely. For the latter the studentized bootstrap can be used. From my point of view, studentized bootstrap methods as presented here have an important point in common with robust statistics: both deliver accurate results when standard methods work, but also when the standard methods fail. Furthermore, the use of the studentized bootstrap avoids restrictive parametric assumptions about the form of the underlying populations.
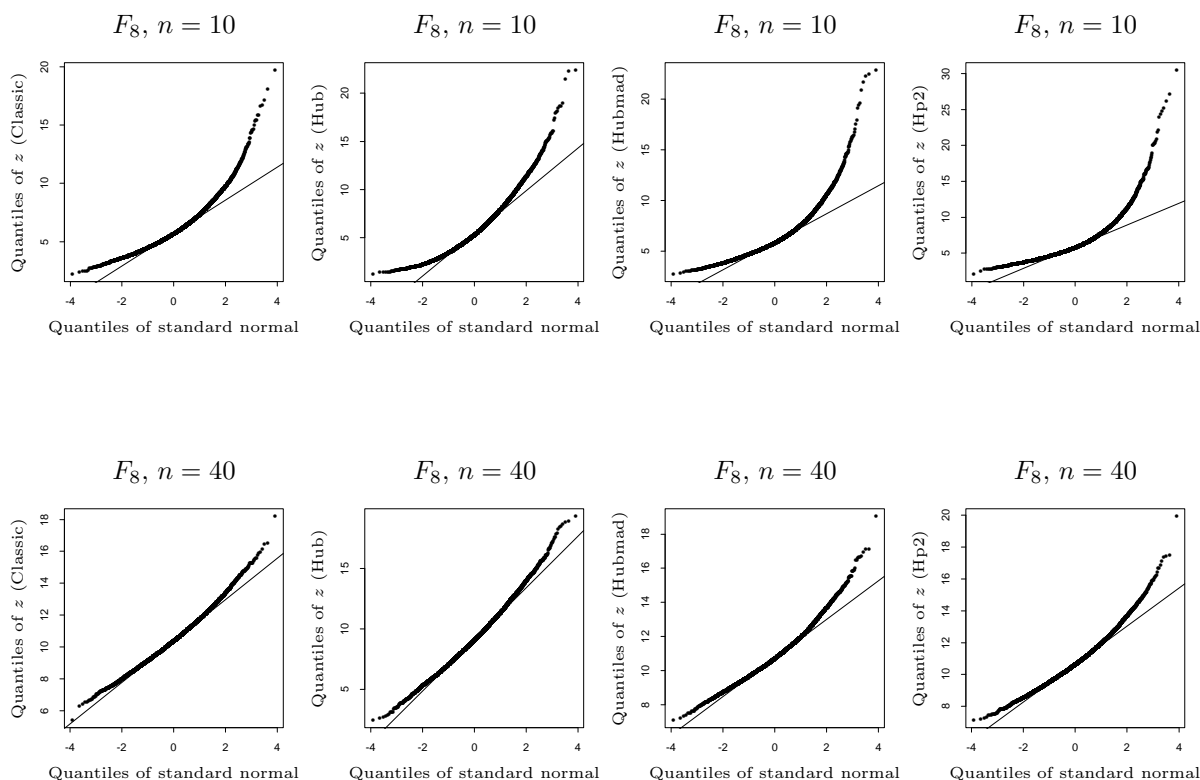
$F_8, n = 10$ $F_8, n = 10$ $F_8, n = 10$ $F_8, n = 10$

$F_8, n = 40$ $F_8, n = 40$ $F_8, n = 40$ $F_8, n = 40$

Figure 7.2. *Q-Q plots of* 20,000 *simulated values of the classical studentized statistic (Classic), of the studentized versions of Huber's M-estimate of location (Hub), of its initially MAD scaled version (Hubmad) and of the studentized version of Huber's proposal 2 (Hp2) (from left to right). Top panels: underlying χ_5^2 distribution (F_8) with $n = 10$. Bottom panels: underlying χ_5^2 distribution (F_8) with $n = 40$.*

7.2.3.   Robustness of validity: actual coverages

Coverage is the most important property of a confidence interval. In this section I will describe how I simulated the coverages of the nominal $(1 - 2\alpha)$ studentized bootstrap confidence intervals with limits (7.2.1) and (7.2.2). For the most interesting cases of Table 7.1 I will then list the results.

The actual coverages were obtained by using the relationship between confidence intervals and test of hypotheses. Namely, for a $(1-2\alpha)$ studentized bootstrap confidence interval one includes the values of θ at which one would not reject the null hypothesis at the two-sided significance level α. To be more precise, suppose one wants to test the null hypothesis $H_0 : \theta = \theta_0$ versus the alternative that θ differs from θ_0 at the two-sided significance level α. Using the fact that Z is a pivot, meaning that its distribution is the same for all relevant F, the p-value can be written as

$$\rho = \Pr(Z \geq z \mid H_0) = \Pr(Z \geq z_0 \mid F), \qquad (7.2.3)$$

where

$$z_0 = \frac{t - \theta_0}{v^{1/2}} \qquad (7.2.4)$$

127

is the observed studentized test statistic under H_0. The p-value (7.2.3) can now be easily approximated by means of the saddlepoint approximation to the CDF of the studentized statistic,

$$\rho \doteq 1 - F_s(z_0), \tag{7.2.5}$$

where $F_s(\cdot)$ is given in (2.4.2) with (6.2.12) and (6.2.13).

Recall that θ_0 is the theoretical true value of the parameter θ under H_0. When F is symmetric the bias of T as an estimator of θ is zero. In functional notation this translates to $t(F) = 0$ and therefore $\theta_0 = 0$ as θ_0 is an approximation for the expectation of the location estimator T. More precisely, we have that $\mathrm{E}(T) = \theta_0 + O(n^{-1})$. Care is needed when F is asymmetric. For the classical studentized statistic with $T = \bar{X}$ the fact that $\theta_0 = 0$ holds as we have $t(F) = 0$ independent of the type of symmetry of F. But when considering an asymmetric F and the robust studentized statistic θ_0 does not equal zero anymore. To illustrate what then needs to be done I consider the case where T is Huber's M-estimate of location given in Section 5.3.2.1. The applications for its initially MAD scaled version given in Section 5.3.2.2 and for Huber's proposal 2 M-estimate of location as defined in (6.4.10) are straightforward. The functional derived from (5.3.1) using $\psi_k(\cdot)$ for $\psi(\cdot)$ defines θ_0 uniquely by

$$\int \psi_k(x - \theta_0) f(x) dx = 0.$$

Following Huber (1981, Section 3.2) the value $\hat{\theta}_0$ defined implicitly by the equation

$$\sum_{i=1}^{N} \psi_k(x_i - \hat{\theta}_0) = 0 \tag{7.2.6}$$

is consistent at F, *i.e.* $\hat{\theta}_0 \to \theta_0$ in probability and almost surely (Huber, 1981, Corollary 2.2). Using $N = 10^6$ the values for $\hat{\theta}_0$ are listed in Table 7.3 for the asymmetric distributions F_8–F_{10} given in Table 7.1.

By replacing in (7.2.4) the θ_0 by $\hat{\theta}_0$ the saddlepoint approximation to the p-value (7.2.5) can easily be computed. To get a reasonable empirical estimate of coverage I repeated this whole procedure on $M = 1,000$ independent samples of size n. Denoting by $\rho_1, \ldots, \rho_M$ the resulting approximate p-values the *actual coverage* of the nominal $(1 - 2\alpha)$ studentized bootstrap confidence interval for θ can be calculated by means of

$$\varpi = \frac{\#\{(\rho_i \geq \alpha) \cap (\rho_i \leq 1 - \alpha)\}}{M}, \tag{7.2.7}$$

where $\#\{\cdot\}$ represents the number of times the event $\cdot$ occurs and $i = 1, \ldots, M$. The latter simply counts how many of the M hypotheses have been accepted, *i.e.* how many times the θ_0 was inside the $(1 - 2\alpha)$ studentized bootstrap confidence interval for θ.

As our aim is to estimate the probability $(1 - 2\alpha)$, the nominal coverage, we want to know how close the estimate ϖ, the actual coverage, is. To do so we can consider

Table 7.3. *The values of $\hat{\theta}_0$ as defined in (7.2.6) using $N = 10^6$ for the asymmetric distributions F_8–F_{10} of Table 7.1 in considering Huber's M-estimate of location (Hub), its initially MAD scaled version (Hubmad) and Huber's proposal 2 estimate of location (Hp2).*

Distribution	Hub	Hubmad	Hp2
F_8	4.40066	4.66068	4.66507
F_9	0.85309	0.80164	0.80028
F_{10}	0.78570	0.74423	0.74099

a simple two-sided 95% normal confidence interval for the simulation error, which has approximate limits

$$\varpi \pm 2\sqrt{\varpi(1 - \varpi)/M}. \tag{7.2.8}$$

What is nice with this approach is that one needs for each sample only to calculate one single saddlepoint approximation, namely $F_s(z_0)$. But as we will see in Section 7.2.4 one needs to compute the entire saddlepoint approximation to the CDF of Z over a likely range of values for z in order to get the limits of the studentized bootstrap confidence interval.

For the symmetric distributions F_1–F_3 and F_6 (given in Table 7.1) the actual coverages ϖ for nominal coverages of 90% and 95% and for different sample sizes n are presented in Table 7.4. The actual coverages of the $(1 - 2\alpha)$ studentized bootstrap confidence interval based on the studentized version of Huber's proposal 2 are not significantly different from the nominal coverages $(1 - 2\alpha)$, whereas the other three studentized statistics may deliver significantly different coverages. The latter can be easily seen by the italic-faced values which represent the actual coverages outside the 95% normal confidence interval for the simulation error (7.2.8). The studentized bootstrap confidence interval based on the classical studentized statistic encounters problems especially with distributions F_2 and F_6. For the intervals based on the studentized versions of Huber's M-estimate of location and on its initially MAD scaled version one notes that the use of F_1 with $n = 20$ results in significantly smaller coverages. Overall one remarks that the studentized bootstrap confidence interval based on the studentized version of Huber's proposal 2 delivers the best coverages independent of the underlying distribution.

The previous statement is underlined by considering additional examples shown in Figure 7.3. The top panels of Figure 7.3 show the actual coverages as a function of the sample size n in considering the symmetric distributions F_4 and F_5. In the top left panel of Figure 7.3 one notes that the studentized bootstrap confidence interval based on the classical studentized statistic completely fails to estimate the true nominal coverage of 80%, whereas the intervals based on the robust studentized statistics perform

Table 7.4. *Actual coverages ϖ (%) of the $(1-2\alpha)$ studentized bootstrap confidence intervals for chosen distributions and sample sizes n (Table 7.1) based on the classical studentized statistic (Classic), on the studentized versions of Huber's M-estimate of location (Hub), of its initially MAD scaled version (Hubmad) and on the studentized version of Huber's proposal 2 (Hp2). Coverages which are not included in the confidence interval for the simulation error (7.2.8) are italic-faced.*

Nominal coverage $(1 - 2\alpha)$

| Distribution | n | 90% | | | | 95% | | | |
		Classic	Hub	Hubmad	Hp2	Classic	Hub	Hubmad	Hp2
F_1	20	89.7	*86.8*	*83.2*	90.5	94.5	*91.9*	*88.6*	95.7
	40	91.2	90.6	*88.1*	89.6	95.1	94.5	*93.5*	95.3
F_2	10	*86.5*	*87.5*	*85.8*	90.9	*93.5*	*91.6*	*92.5*	94.7
	15	*85.6*	88.8	*86.5*	89.3	*92.9*	*93.0*	*91.9*	94.9
F_3	20	*87.8*	*85.9*	*87.7*	91.4	*93.5*	*91.5*	*91.1*	95.6
	40	90.4	*86.8*	88.7	89.9	94.8	*91.8*	94.7	94.4
F_6	20	*83.6*	*85.6*	*87.1*	89.9	*91.6*	*91.0*	*92.7*	96.2
	40	*86.0*	*87.9*	88.3	89.4	*92.2*	*92.5*	*93.6*	94.5

reasonable well. The actual coverages obtained using F_5 (Figure 7.3, top right panel) illustrate that the studentized bootstrap confidence interval based on the studentized version of Huber's proposal 2 is the only one giving accurate estimates of the nominal coverage of 99% throughout the range of n. The actual coverages in considering the asymmetric distributions F_8 and F_9 are represented by the bottom panels of Figure 7.3. As illustrated in both panels the studentized bootstrap confidence interval based on the studentized version of Huber's proposal 2 performs well. For the distribution F_9 (Figure 7.3, bottom right panel) only the one based on the studentized versions of Huber's M-estimate of location seems to compete with the latter. All other methods fail to achieve the claimed coverages of 95% (Figure 7.3, bottom left panel) and 90% (Figure 7.3, bottom right panel).

Another informative graphic can be obtained as follows. If T is continuous the p-value ρ defined in (7.2.3) has under H_0 a $U(0,1)$ distribution. Hence the corresponding random variable P has distribution

$$\Pr(P \leq \rho \mid H_0) = \rho.$$

This fact can now be used by plotting the quantiles of an uniform distribution on $[0, 1]$, which represent the nominal coverages, against the ordered values of the M actual
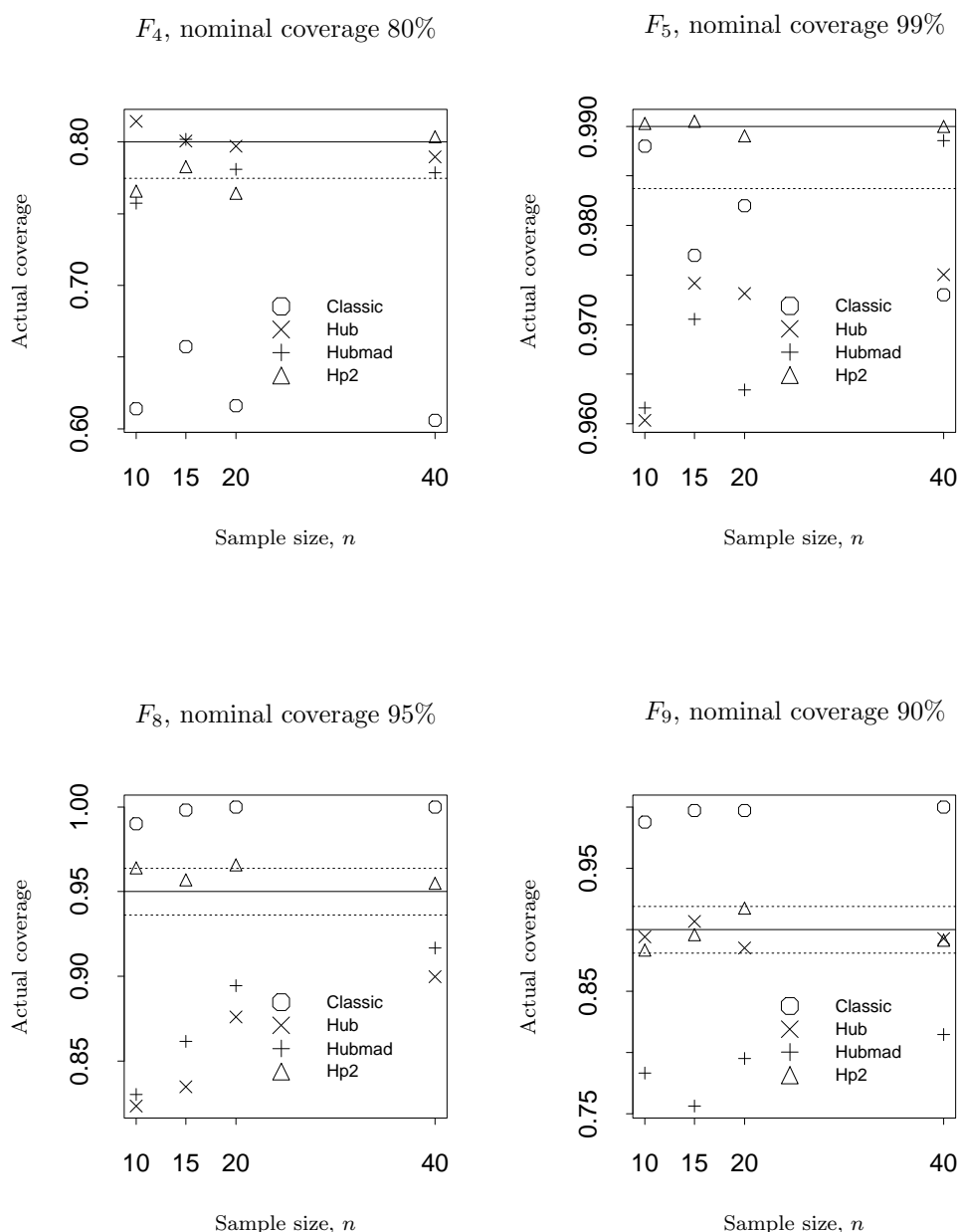
F_4, nominal coverage 80%

F_5, nominal coverage 99%

F_8, nominal coverage 95%

F_9, nominal coverage 90%

Figure 7.3. *Actual coverages for different sample sizes n of the $(1 - 2\alpha)$ studentized bootstrap confidence intervals based on the classical studentized statistic (Classic), on the studentized versions of Huber's M-estimate of location (Hub), of its initially MAD scaled version (Hubmad) and on the studentized version of Huber's proposal 2 (Hp2). The nominal coverages are represented by the solid lines and the limits of the confidence interval for the simulation error (7.2.8) by dotted lines. Top left panel: using the Cauchy distribution (F_4) and a nominal coverage of 80%. Top right panel: using a normal mixture distribution (F_5) and a nominal coverage of 99%. Bottom left panel: using the χ_5^2 distribution (F_8) and a nominal coverage of 95%. Bottom right panel: using the folded t_7 distribution (F_9) and a nominal coverage of 90%.*

coverages obtained through (7.2.7). This plot is very useful as one gets an overall idea of the coverage performance over the entire range of possible nominal levels. For the symmetric distributions F_1, F_4 and F_7 and for the asymmetric distribution F_{10} these plots are given in Figure 7.4. For F_1 and the moderate value $n = 15$ (Figure 7.4, top left panel) all four methods seem to perform reasonably well. For F_4 with $n = 40$ (Figure 7.4, top right panel) and F_7 with $n = 20$ (Figure 7.4, bottom left panel) the studentized bootstrap confidence intervals based on the classical studentized statistic (dotted) undercover in the bottom left of the plot and overcover in the top right of the plot. In both situations the studentized bootstrap confidence intervals based on the studentized versions of Huber's M-estimate of location, of its initially MAD scaled version and on the studentized version of Huber's proposal 2 deliver similar accuracy. Finally, in considering the asymmetric distribution F_{10} with $n = 40$ (Figure 7.4, bottom right panel) one remarks that the intervals based on the classical studentized statistic (dotted) completely fail. The ones based on the studentized version of Huber's M-estimate of location with initial MAD scaling (dashed) undercover significantly throughout the entire range of the nominal level. The studentized bootstrap confidence intervals based on the studentized version of Huber's M-estimate of location (large dashes) and on the studentized version of Huber's proposal 2 (small dashes) deliver similar accuracy.

Additional tables like Table 7.4 can be found in Kuonen (2000b, page 15) and Kuonen (2000c, page 22) and additional plots like Figure 7.4 in Kuonen (2000b, page 16), Kuonen (2000c, pages 23 and 24) and Kuonen (2000d, pages 20 and 22). The conclusions of the excellent coverage performance of the studentized bootstrap confidence intervals based on the studentized version of Huber's proposal 2 remain.

7.2.4.   Robustness of efficiency: confidence interval lengths

So far I only measured robustness of validity by assessing whether the actual coverage is close to the nominal coverage. As mentioned at the beginning of Section 7.2 it is very important to also measure the robustness of efficiency of a method once the robustness of validity has been achieved. The aim of a good method for robust inference is to achieve the claimed confidence with a short interval as one would expect, given good coverage properties, that a good approximate confidence interval method tracks the correct endpoints closely.

To do so we need to calculate the lower limit $\hat{\theta}_\alpha$ and the upper limit $\hat{\theta}_{1-\alpha}$ of the $(1 - 2\alpha)$ studentized bootstrap confidence intervals (7.2.1) and (7.2.2) given in Section 7.2.1. Following the notation therein the *confidence interval length* (CIL) becomes

$$\hat{\theta}_{1-\alpha} - \hat{\theta}_\alpha \doteq v_L^{1/2}(z_{s,1-\alpha} - z_{s,\alpha}), \tag{7.2.9}$$

where v_L is the nonparamametric delta method estimate of the variance of the location estimator T as defined in Section 7.2.1 and $z_{s,\alpha}$ denotes the α-quantile of $Z^\star$ issued from the use of the saddlepoint approximation to the CDF of $Z^\star$ as given in Section 6.2. To get $z_{s,\alpha}$ I approximated the entire bootstrap CDF of $Z^\star$ by calculating the values of $F_s(z)$ given in (2.4.2), using (6.2.12) and (6.2.13), for 50 values of z equally spaced between
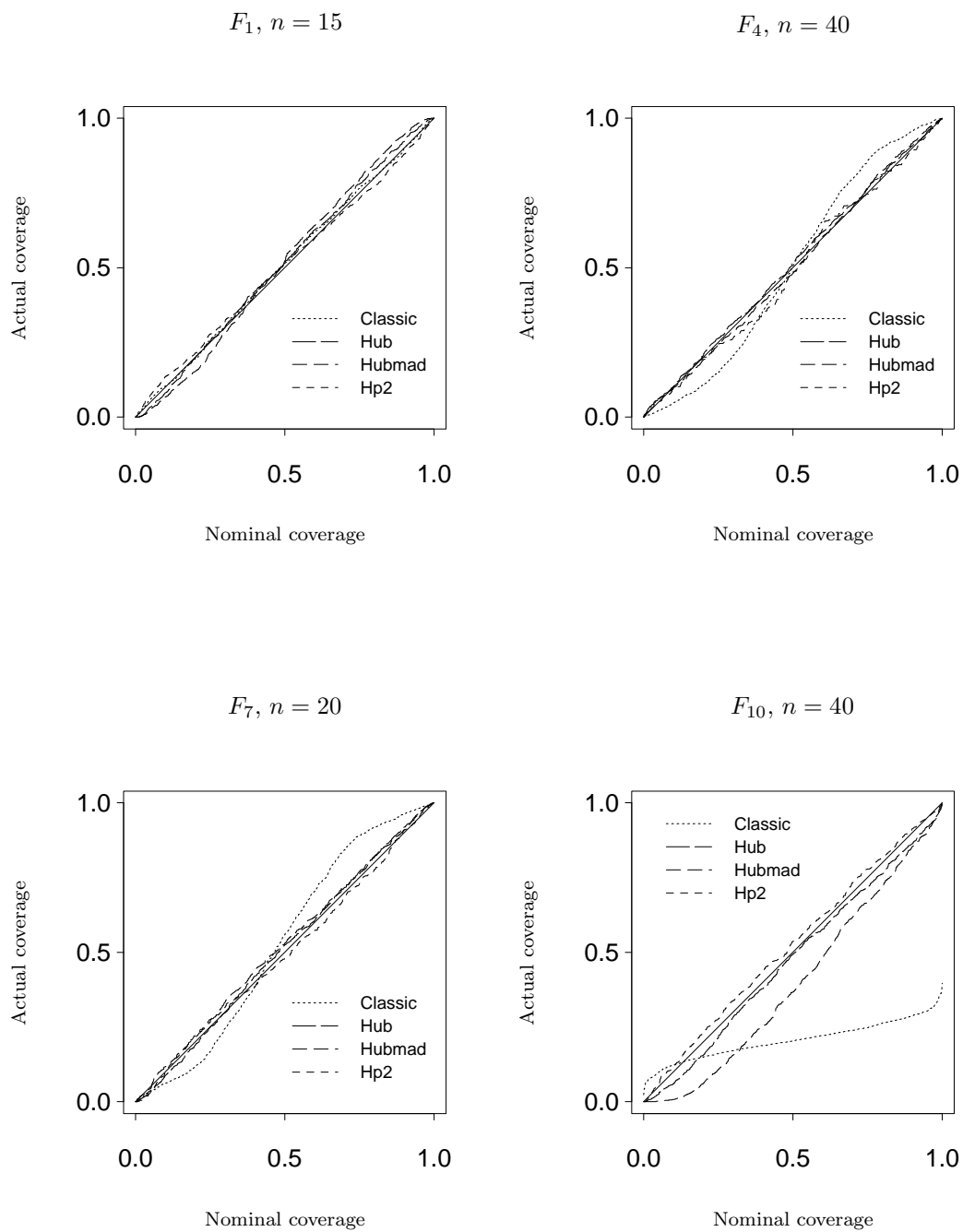
Figure 7.4. *Nominal against actual coverages of the studentized bootstrap confidence intervals based on the classical studentized statistic (Classic), on the studentized versions of Huber's M-estimate of location (Hub), of its initially MAD scaled version (Hubmad) and on the studentized version of Huber's proposal 2 (Hp2). The solid line is the line $x = y$. Top left panel: using the $N(0,1)$ distribution (F_1) with $n = 15$. Top right panel: using the Cauchy distribution (F_4) with $n = 40$. Bottom left panel: using the slash distribution (F_7) with $n = 20$. Bottom right panel: using the folded $N(0,1)$ distribution (F_{10}) with $n = 40$.*

l_{min} and l_{max}. The latter are the smallest and largest empirical influence values of the studentized statistic as defined in Section 6.5.2.3. Then I used a spline smoother to interpolate between the corresponding values of $\Phi^{-1}\{F_s(z)\}$ and the needed quantiles, $z_{s,\alpha}$ and $z_{s,1-\alpha}$, were then read off from the fitted curve.

Remark 7.2. To overcome the drawback of the calculation of the entire saddlepoint approximation to the CDF of $Z^\star$ one could also use the so-called *significance test method* as reviewed in Davison and Hinkley (1997, Section 5.5) and by Carpenter (1999). The idea is as follows. As in Section 7.2.3 assume that the associated test of $H_0 : \theta = \theta_0$ versus the alternative that θ differs from θ_0 is based on the statistic $Z(\theta_0)$. The corresponding p-value can be approximated by $\rho(\theta_0)$ by means of a saddlepoint approximation as in (7.2.5). The lower confidence limit $\hat{\theta}_\alpha$ can be seen as the smallest solution to $\rho(\theta_0) = \alpha$. Hence one could evaluate $\rho(\theta_0)$ for a grid of possible values of θ_0 and interpolate between these values to get $\hat{\theta}_\alpha$. The same interpolation method needs to be applied in order to get the upper confidence limit $\hat{\theta}_{1-\alpha}$. An efficient algorithm to do so is presented by Garthwaite and Buckland (1992). In the latter a separate sequential search is conducted for the lower and upper limits of the confidence interval. This approach seems promising and interesting but for each needed quantile a separate algorithm needs to be used. Moreover, in the context of this chapter we consider several values for the nominal level so this significance test method may become very time-intensive, whereas the entire saddlepoint approximation to the CDF of $Z^\star$ needs only to be computed once and one can then use the resulting spline fit to predict whatever quantile one wants. □

A similar approach to measure robustness of efficiency based on the *expected confidence interval lengths* (ECIL) has been proposed by Gross (1976). The use of ECIL can be criticized on various grounds, including the valid remark that ECIL is an average and it is well-known how poorly averages estimate location in long-tailed or asymmetric distributions. Nevertheless in his unpublished PhD thesis he found that ECIL can be used to compare the efficiency of the methods; see also the comments by Gross (1976, Section 7) and by Gross (1977, Section 7). Gross (1976) also introduced the concept of *deficiency* of a method A, say, having ECIL_A. Its deficiency can be calculated relative to the shortest confidence interval under study, *i.e.* the interval with the smallest ECIL (ECIL_{min}), as the ratio

$$\text{deficiency} = 1 - \text{ECIL}_{min}/\text{ECIL}_A.$$

These deficiencies should be close to zero, hopefully.

As mentioned before robustness of validity should be achieved prior to any investigations on the efficiency of a confidence interval method. Following the coverages results in Section 7.2.3 the number of possibilities for distributions and studentized bootstrap confidence intervals drops significantly. For instance the only situation where the studentized bootstrap confidence intervals based on all four studentized statistics delivered actual coverages close to nominal level is when the $N(0,1)$ distribution (F_1) has been considered. This is illustrated in the first two rows of Table 7.4 with $n = 20, 40$ and in the top left panel of Figure 7.4 with $n = 15$. For the case when $n = 40$ the boxplots of the confidence interval lengths of the resulting 95% studentized bootstrap confidence
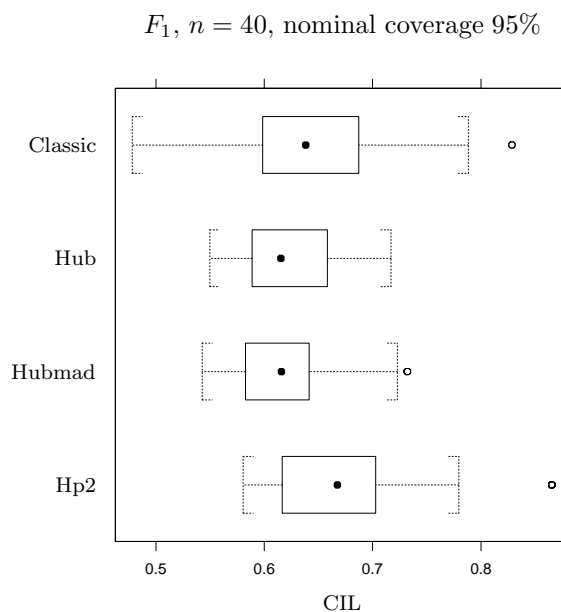
F_1, $n = 40$, nominal coverage 95%

Figure 7.5. *Boxplots of confidence interval lengths (CIL) of the 95% studentized bootstrap confidence intervals based on the classical studentized statistic (Classic), on the studentized versions of Huber's M-estimate of location (Hub), of its initially MAD scaled version (Hubmad) and on the studentized version of Huber's proposal 2 (Hp2). The $N(0,1)$ distribution (F_1) with $n = 40$ was used.*

intervals are shown in Figure 7.5. The simulations were based on $1,000$ samples from the underlying distribution and for each of the simulated datasets I computed the entire saddlepoint approximation to the CDF of all four studentized statistics. Figure 7.5 shows that the variance of the CIL seems to decrease when the robust studentized statistics are used, and that the CIL based on the studentized versions of Huber's M-estimate of location and on its initially MAD scaled version have a smaller CIL range compared to the others. The corresponding expected confidence interval lengths are given in the first row of Table 7.5. The 95% studentized bootstrap confidence intervals based on the studentized version of Huber's M-estimate of location with initial MAD scaling has the smallest ECIL and its deficiency compared to the other methods equals zero (Table 7.6, first row). The small deficiency of the intervals based on the studentized versions of Huber's M-estimates of location underline the small variability of the CIL seen in Figure 7.5. Recall also from the second row of Table 7.4 that the ones based on initial MAD scaling was the only of the four methods that had its actual coverage lying outside the confidence interval of the simulation error. Unfortunately using F_1 was the only situation in this simulation study where all four methods performed similarly. Moreover, for obvious reasons I will not consider studentized bootstrap confidence intervals based on the classical studentized statistic in the examples to follow.

A situation where all studentized bootstrap confidence intervals based on the robust studentized statistics delivered good coverage is given in the bottom left panel of Figure 7.4 when the slash distribution (F_7) with $n = 20$ was used. In considering F_7 with

Table 7.5. *Expected confidence interval lengths (ECIL) of the $(1-2\alpha)$ studentized bootstrap confidence intervals for chosen distributions, sample sizes n and nominal levels $(1-2\alpha)$ based on the classical studentized statistic (Classic), on the studentized versions of Huber's M-estimate of location (Hub), of its initially MAD scaled version (Hubmad) and on the studentized version of Huber's proposal 2 (Hp2).*

Distribution	n	Nominal level	Classic	Hub	Hubmad	Hp2
F_1	40	95%	0.64440	0.62635	0.61954	0.67317
F_7	40	99%	—	1.96347	2.13412	3.04439
F_9	20	90%	—	0.47255	—	0.53345

Table 7.6. *Deficiencies (%) of the $(1-2\alpha)$ studentized bootstrap confidence intervals for chosen distributions, sample sizes n and nominal levels $(1-2\alpha)$ based on the classical studentized statistic (Classic), on the studentized versions of Huber's M-estimate of location (Hub), of its initially MAD scaled version (Hubmad) and on the studentized version of Huber's proposal 2 (Hp2).*

Distribution	n	Nominal level	Classic	Hub	Hubmad	Hp2
F_1	40	95%	3.85	1.08	0.00	7.96
F_7	40	99%	—	0.00	7.99	35.50
F_9	20	90%	—	0.00	—	11.41

$n = 40$ the resulting CIL are represented in Figure 7.6. The CIL of the 99% studentized bootstrap confidence interval based on the studentized version of Huber's proposal 2 are larger than the others. This is underlined by its ECIL and its deficiency given in the second rows of Table 7.5 and 7.6. These large values of CIL seem to account for the good coverage performance of the studentized bootstrap confidence intervals based on the studentized version of Huber's proposal 2. One notes also that the 99% studentized bootstrap confidence interval based on the studentized versions of Huber's M-estimate of location seems to be most efficient in using F_7 with $n = 40$.

As last example consider the situation given in the bottom right panel of Figure 7.3 which is F_9 with $n = 20$ and a nominal level of 90%. For this asymmetric distribution only the studentized bootstrap confidence intervals based on the studentized versions of Huber's M-estimate of location and on the studentized version of Huber's proposal 2
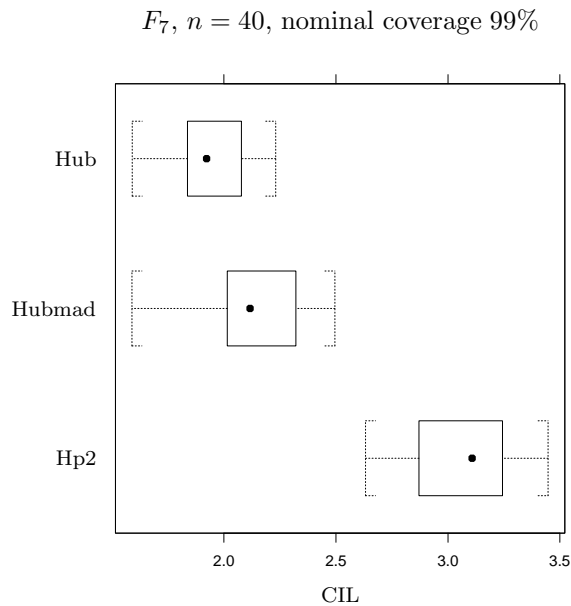
Figure 7.6. *Boxplots of confidence interval lengths (CIL) of the* 99% *studentized bootstrap confidence intervals based on the studentized versions of Huber's M-estimate of location (Hub), on its initially MAD scaled version (Hubmad) and on the studentized version of Huber's proposal 2 (Hp2). The slash distribution (F_7) with $n = 40$ was used.*

delivered accurate actual coverages. Figure 7.7 shows the CIL of both methods. Their ECIL (Table 7.5, last row) are similar. However, as illustrated in Figure 7.7 the variance of the interval based on the studentized version of Huber's proposal 2 is larger than the one using Huber's M-estimate of location. Their deficiencies are given in the last row of Table 7.6 but are less informative in this situation as only two methods have been compared.

Additional examples showed that the CIL of the $(1 - 2\alpha)$ studentized bootstrap confidence interval based on the studentized version of Huber's proposal 2 tend to be longer and more variable than the ones based on the other studentized statistics. This seems to account for being the method which worked best in Section 7.2.3, by delivering two-sided coverages only slightly less than nominal.

7.3. Conclusion

n this chapter I applied the saddlepoint approximations to studentized bootstrap distributions studied in Chapter 6 to the resulting studentized bootstrap confidence intervals. The latter are known to be very accurate without needing standard theory assumptions. The performance of the studentized bootstrap confidence intervals in considering various studentized statistics was then illustrated in Section 7.2 in the context of a simulation study.

In order to make robust inference about a location parameter θ there are three types of robustness one would like to achieve: robustness of performance for the estimator of
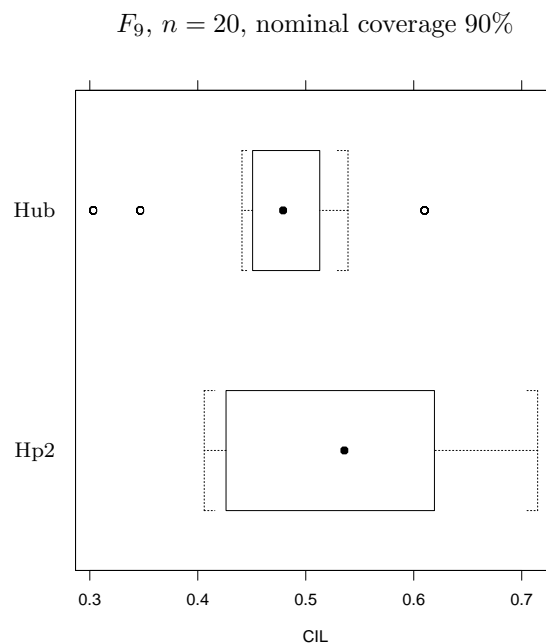
F_9, $n = 20$, nominal coverage 90%

Figure 7.7. *Boxplots of confidence interval lengths (CIL) of the* 90% *studentized bootstrap confidence intervals based on the studentized versions of Huber's M-estimate of location (Hub) and on the studentized version of Huber's proposal 2 (Hp2). The folded t_7 distribution (F_9) with $n = 20$ was used.*

θ, robustness of validity and robustness of efficiency for the resulting confidence interval method. These together form the necessary condition to enable robust inference. The robustness of performance for estimators of location was illustrated in Chapter 5 for the unstudentized versions and in Chapter 6 for the studentized versions. In both chapters it was shown by means of examples that the ones based on robust M-estimates performed best. This is especially true in Chapter 6 where in the bootstrap context Huber's proposal 2 seems to be the preferable measure of location (robustness of performance). In the present chapter I then considered the resulting studentized bootstrap confidence intervals. Their robustness of validity was measured by means of their actual coverages (Section 7.2.3) and their robustness of efficiency by their confidence interval lengths (Section 7.2.4). With similar actual coverage probabilities procedures that provide shorter intervals are better, being less deficient. In order to study the robustness of efficiency of a procedure it is preferable and necessary to compare intervals with nearly the same actual coverage. Hence for a method, like the bootstrap-t method, it is necessary that its robustness of validity is checked prior to any concerns about its robustness of efficiency. It was illustrated in Section 7.2.3 that the studentized bootstrap confidence intervals based on the studentized version of Huber's proposal 2 delivered actual coverages closest to the nominal ones (robustness of validity). Section 7.2.4 showed that these intervals were wider and more variable than the others, explaining the good coverage performance.

In summary, my recommendation is the following: if you want to make robust infer-

ence based on the very suitable candidate which is the studentized bootstrap then you should use the studentized version of Huber's proposal 2. It was the only statistic which delivered robustness of performance, robustness of validity and a sufficient robustness of efficiency — the three necessary ingredients for a robust inference.

8. Conclusion and outlook

'We shall not cease from exploration
and the end of all our exploring
will be to arrive where we started
and know the place for the first time.'

Thomas S. Eliot (1943, 'Little Gidding', Part V)

The present thesis can be split up into three parts. The first part consists of Chapter 3, the second of Chapter 4 and the third part of Chapters 5, 6 and 7. Their detailed dependencies were given in Figure 1.1. All three parts rely on Chapter 2. In the latter the ideas underlying the saddlepoint approximations were informally described for further use in the subsequent chapters.

Note that each of the chapters contained already a 'Conclusion' section, which were meant to summarize its contents and give an outlook to open questions or serve as bridge for the subsequent chapter. Nevertheless, I will here summarize the conclusions and provide a short list of possible further work.

In the first two parts of my thesis I focused on two of my developments in this area, namely on saddlepoint approximations for the distribution of quadratic forms in normal variables (Chapter 3) and for the distribution of the waiting time in the coupon collector's problem (Chapter 4).

Chapter 3 was concerned with quadratic forms, which enter into many statistics associated with normally distributed random variables. As in general quadratic forms are not positive definite, classical results such as Cochran's theorem implying a chi-squared distribution for the quadratic form do not apply, and another approach to the calculation of its distribution is needed. The various methods for computing the exact distribution of quadratic forms in normal variables require in general extensive numerical computations. Numerical integration methods, though sufficiently accurate in solving the general problem, require a considerable amount of computer time. Several approximation methods have been proposed to reduce those difficulties. However, it was illustrated in Section 3.4 that these approximate methods may be inaccurate. Hence I proposed a saddlepoint approximation for the distribution of quadratic forms in normal variates. It was shown to be comparable in speed with exact methods, almost as accurate and much easier to program. This was underlined by several examples, including applications to

nonparametric regression in Section 3.5. For example, in Section 3.5.2 I applied it to the assessment of significance when comparing the fit of a linear model with that of a local smoother. Other workers have noted that the corresponding pseudo-likelihood ratio test leads to a type of generalized F-statistic, and then have proposed chi-squared approximations to its distribution, based on the first few cumulants. The proposed saddlepoint approximation gives appreciably more accurate tail probabilities for the associated significance tests, which are then essentially exact. In Section 3.5.3 I listed a number of other cases to which the presented approach could be applied easily. Furthermore, one obvious direction in which this work could be extended is towards the application of saddlepoint approximations to the distributions of general quadratic forms, *i.e.* not necessary in normal variates as in Chapter 3. As the exact cumulant generating function is not known, one would need to derive some of its cumulants. The idea then would be to approximate the cumulant generating function by the leading four, say, terms and then to apply the saddlepoint approximation.

In Chapter 4 a standard combinatorial problem was considered, namely the estimation of the number of coupons, drawn at random with replacement, needed to complete a collection. This problem, known as the coupon collector's problem or the classical occupancy problem, has been investigated intensively. In Section 4.4 I used several examples to show that a saddlepoint approximation to the distribution of the waiting time in the coupon collector's problem is very accurate. Attention is given to the special case where all sampling probabilities are equal. An extension for the general case of unequal sampling probabilities is still missing.

The third part of this thesis was devoted to the use of saddlepoint approximations in order to replace the computer-intensive bootstrap. In Chapter 5 saddlepoint approximations for standard bootstrap distributions were presented. The basic notions of the saddlepoint approximation as applied to resampling were recalled in Section 5.1. Other workers have noted that the saddlepoint approximations to bootstrap distributions may fail if data contain outliers. Therefore, I proposed and studied in Section 5.2 a saddlepoint mixture approximation for the case of the average in conditioning on the number of times the outlier has been resampled. I showed in several examples that this approach delivers very accurate approximations. Nevertheless, one may question the use of a statistic like the average when data contain outliers. To enable the use of robust estimates of location defined by scalar estimating equations a technique was presented in Section 5.3. As applications I considered Huber's M-estimate for location and the standardized M-estimator of location with initial MAD scaling. I showed in several applications to real data that in both cases the saddlepoint approximations delivered very accurate approximations to their density and distribution functions. A possible future work resulting from Chapter 5 could be as follows. The saddlepoint mixture approximation presented in Section 5.2 is only valid if the data contain a single outlier. I think that it should be feasible to extend it to data containing two or more outliers; see also the comments given in Section 5.2.2. However, questions about its usefulness remain. Having detected an extreme outlier, one would immediately use a robust estimate of location, or check if it is due to a measurement error and if so omit it.

The natural continuation of Chapter 5 was then given in Chapter 6 by extending

the approach to studentized statistics. The method discussed in Section 6.2 is known as the integration saddlepoint approximation. The latter considers marginal saddlepoint approximations for studentized statistics, involving calculation of the joint cumulant generating function of the estimating equations that determine the bootstrapped estimator of interest and some nuisance statistics, followed by a saddlepoint approximation to their joint distribution. To get the marginal density of the studentized version of the estimator of interest, a Laplace approximation to the integral of the joint distribution is then used to avoid numerical integration. I considered in Section 6.3 the classical studentized statistic and in Section 6.4 the studentized versions of Huber's M-estimate of location, of Huber's M-estimate of location with initial MAD scaling and of Huber's proposal 2 estimate of location. The examples considered convinced me that it is preferable, in the bootstrap context, to use the studentized version of Huber's proposal 2 instead of the studentized version of Huber's M-estimate of location with initial MAD scaling. Actually the opposite is recommended by the main workers of robustness in the unstudentized and non-bootstrap context. Remarks on implementation and related problems resulting from the use of the integration saddlepoint approximation were given in Section 6.5. Care is needed with its application as it is based on a Laplace approximation, which fails when the joint cumulant generating function is concave. As illustrated in Section 6.5.1 it may be always wise to use diagnostic plots like the ones given therein. But they suffer from the drawback of being too time-intensive. Nevertheless, it was also noted that an approximation can always be found by numerical integration. Numerical integration methods for use in S-Plus and R were discussed in detail in Appendix A, and were applied in Section 6.5.2. Integrals over infinite domains should be transformed to a finite region in view of the accuracy and convergence of the quadrature method in use. Hence in Section 6.5.2.2 I listed the transformations which performed best, and in Section 6.5.2.3 I presented a method which can be used to get an idea of the range of the studentized statistics, namely by using the range of their empirical influence values. All these ideas where then used in Section 6.5.2.4 in order to perform the numerical integration. However, the examples considered clearly illustrated the drawbacks of numerical integration: it becomes useless in practice as their running time is outperformed by direct simulation of the bootstrap replicates. Fortunately, my experience shows that the problems described in Section 6.5 occur rarely. Therefore the need to perform a very computer-intensive numerical integration as in Section 6.5.2 disappears. Especially concerning these numerical integration methods it was made clear that questions on convergence and efficient implementation remain. Some future approaches stated in Appendix A are promising. Chapter 6 ended then with the recommendation that Huber's proposal 2 seems to be a preferable measure of location, being not unduly affected by outliers or asymmetry in the data.

Through the use of the bootstrap we can obtain accurate confidence intervals without having to make normal or Student theory assumptions. One way to get such intervals is by means of the bootstrap-t method. Chapter 6 underlined the fact that the studentized statistic is a good approximate pivotal quantity. The latter forms the basis of the studentized bootstrap confidence intervals. The aim of Chapter 7 was then to know whether the studentized statistics presented in Chapter 6 yield robust confidence

intervals with coverages close to nominal and this with short intervals. In robustness notations this can be translated into the following wish: knowing the robustness of performance of a location estimator one is interested to know whether the resulting confidence intervals are able to achieve robustness of validity and robustness of efficiency. All three concepts form the basis to enable the practitioner to make robust inference. To do so I listed in Section 7.2 the results of an extensive simulation study. More precisely, the robustness of validity of the bootstrap-t method was measured in Section 7.2.3 by means of the actual coverages, and its robustness of efficiency in Section 7.2.4 by the use of the confidence interval lengths. The best robustness of validity of the studentized bootstrap confidence intervals was achieved when they were based on the studentized version of Huber's proposal 2. I then stated in Section 7.2.4 that the resulting intervals were wider and more variable than the others in order to account for the good coverage performance. Finally, my recommendation in order to make robust inference based on the studentized bootstrap was then to use the studentized version of Huber's proposal 2, being the only estimator considered which delivered robustness of performance (Section 6.4.3), robustness of validity (Section 7.2.3) and robustness of efficiency (Section 7.2.4).

There are several directions in which the work presented in Chapters 5–7 may be extended. One possibility may be to extend the methods to inference in robust regression. Furthermore, I hope that the results, remarks and recommendations given in Chapters 5–7 may fill several gaps in the current literature and stimulate further discussion and research within the presented context.

A. Numerical integration

'Numerical analysis is an art, not a science.'

Herman Rubin (1999)

A.1. Introduction

umerical integration, which is also called *quadrature*, is the study of how the numerical value of an integral can be found. The purpose of this appendix is to discuss quadrature methods for approximate calculation of integrals within S-PLUS or R. All are based, in one way or another, on the obvious device of adding up the value of the integrand at a sequence of points within the range of integration. Hence, most of the approximations I consider have the form

$$
\int \cdots \int_{R_m} w(x_1, \ldots, x_m) f(x_1, \ldots, x_m) dx_1 \cdots dx_m \approx \sum_{i=1}^{M} W_i f(y_{i,1}, \ldots, y_{i,m}), \quad \text{(A.1.1)}
$$

where R_m is a given region in a m-dimensional Euclidean space E_m and $w(x_1, \ldots, x_m)$ is a given weight function. The $(y_{i,1}, \ldots, y_{i,m})$ lie in E_m and are called the *points* of the formula. The W_i are constants which do not depend on $f(x_1, \ldots, x_m)$ and are called the *coefficients* of the formula. We say that formula (A.1.1) has *degree r* (or *degree of exactness r*) if it is exact for all polynomials in $x_1, \ldots, x_m$ of degree $\leq r$ and there is at least one polynomial of degree $r + 1$ for which it is not exact. See also Stroud (1971) or Evans (1993, Chapter 6).

The theory of integration formulae for functions of one variable ($m = 1$) is well developed. A great deal of this theory can be found in the books by Engels (1980), Davis and Rabinowitz (1984), Evans (1993) or Press *et al.* (1993, Chapter 4). For $m = 1$ equation (A.1.1) can be written as

$$
\int_R w(x) f(x) dx = \int_a^b w(x) f(x) dx \approx \sum_{i=1}^{M} W_i f(y_i). \quad \text{(A.1.2)}
$$

In the classical formulae the integral of a function is approximated by the sum of its values at a set of equally spaced points, multiplied by certain aptly chosen coefficients of the formula. Examples include the trapezoidal and Simpson's rules. Hence only the W_i are free to be used to force the quadrature rule to have a certain degree of exactness. The freedom to fix the points y_i has been thrown away, presumably in the interests of getting nice linear equations for the W_i. If the y_i are also left free, the result is a set of non-linear equations which can be shown to have solutions based on the zeros of the associated sets of orthogonal polynomials for the given interval $[a, b]$ and weight function $w(x)$. This leads to the elegant theory of Gaussian quadrature, which will be discussed in Section A.2. Gaussian quadratures are formulae which are said to be *progressive* as the points for any point-number M are in general quite different from those for any other point-number. Another term used to describe quadrature rules is *adaptive*. A rule is adaptive if it compensates for a difficult subrange of an integrand by automatically increasing the number of quadrature points in the awkward region. As we will see in Section A.3 adaptive rules are usually based on a standard underlying quadrature rule, often a progressive one. For very high dimensionality Monte Carlo or random sampling methods (Section A.4) can begin to be competitive, though in this regime all methods tend to be very inaccurate for a reasonable computer effort. Finally, a comparison of the presented methods is given in Section A.5.

A.2.   Gaussian quadrature

The idea of Gaussian quadratures is to give ourselves the freedom to choose not only the coefficients W_i, but also the location of the points at which the function is to be evaluated. Moreover, the formula (A.1.2) is forced to have degree of exactness $2M - 1$. Because of the computational expense of generating a new Gaussian formula, only commonly used combinations of the interval and weight functions are normally tabulated, see Evans (1993, Section 2.3). Of these the most commonly used is the Gauss–Legendre rule with interval $[-1, 1]$ and weight function $w(x) = 1$.

Neither S-Plus or R offer the Gauss–Legendre rule, nor any other standard Gaussian quadrature rule, by default. Nevertheless, the `integrate2` library for S-Plus (`lib.stat.cmu.edu/S/integrate2`) contains the function `intgauss`, which performs the numerical integration of a function over a given region using a classical 10 point Gaussian formula. In order to enable more flexibility for the choice of M and the wish to extend it to higher dimensions, I wrote a S-Plus function `GL.integrate.1D` to compute (A.1.2) by means of the Gauss–Legendre rule and based on a modified version of the C function GAULEG given in Press *et al.* (1993, page 151). The function is entirely written in S-Plus; a C version is available. I used the fact that any finite range quadrature on the interval $[a, b]$, can be transformed using the linear transformation

$$x = \frac{b - a}{2}t + \frac{b + a}{2}$$

Table A.1. *The* S-PLUS *function* `GL.integrate.1D`.

```
GL.integrate.1D <- function(fct, low=-1, upp=1, order=10)
  {
  if(low==upp) stop("low = upp")
  name.YW<-paste("GL.YW", abs(low), abs(upp), order, sep=".")
  if (!exists(name.YW)) {
    assign(name.YW, GL.YW(order, xrange=c(low,upp)),
          where=1, immediate=T)
  }
  YW<- get(name.YW)
  approx<-numeric(order)
  for (i in 1:order)  approx[i] <-  YW[i,2] * fct(YW[i,1])
  sum(approx)
  }
```

to the standard interval $[-1, 1]$. The function `GL.integrate.1D` is given in Table A.1. It uses the function `GL.YW` which computes, if needed, the points y_i and the weights W_i for the interval $[\texttt{low}, \texttt{upp}]$, for $i = 1, \ldots, M$, where $M = \texttt{order}$. The function `GL.YW` is given in Table A.2.

For the multi-dimensional case we reduce the multiple integral in the left-hand side of (A.1.1) into the repeated integrals over $[-1, 1]$, namely

$$\int_{-1}^{1} dx_1 \int_{-1}^{1} dx_2 \cdots \int_{-1}^{1} f(x_1, \ldots, x_m) dx_m. \tag{A.2.1}$$

Then we can apply a classical quadrature formula to each integral in (A.2.1), which yields using the right-hand side of (A.1.1) a product rule of the form

$$\sum_{i_m=1}^{M} \cdots \sum_{i_1=1}^{M} W_{i_1} \cdots W_{i_n} f(y_{i_1}, \ldots, y_{i_m}), \tag{A.2.2}$$

where the weights W_{i_j} and the points y_{i_j}, $j = 1, \ldots, m$, are chosen to be appropriate for the specific dimension to which they are applied; see also Evans (1993, Chapter 6). The number of function evaluations using the M^m integration points may be quite large. For $m = 2$ this is illustrated by the S-PLUS function `GL.integrate.2D` given in Table A.3 using the fast version of the S-PLUS function `outer`, namely `outer2`, given in Table A.4 and also included in the latest R version ($\geq 1.3.0$) as a replacement of the initial R function `outer`. For example

```
> tmp.fct <- function(x,y) {1/(1-x*y)}
> GL.integrate.2D(tmp.fct, low=c(0,0), upp=c(1,1), order=128)
[1] 1.644886
```

Table A.2. *The* S-PLUS *function* `GL.YW`.

```
GL.YW <- function(M, xrange=NULL, epsilon=NULL)
{
 if (is.null(epsilon)) epsilon <- .Machine$double.eps
 if (M%%2 ==1) stop("M needs to be an even number")
 MM <- (M + 1) / 2
 Y <- W <- numeric(M)
 for (i in 1:floor(MM)) {
    ok<-F
    z  <- cos( pi * (i-0.25)/(M + 0.5) )
    while (ok == F) {
      p1 <- 1.0
      p2 <- 0.0
      for (j in 1:M) {
         p3 <- p2
         p2 <- p1
         p1 <- ((2.0*j - 1.0)*z*p2 - (j - 1.0)*p3)/j
         }
         pp <- M*(z*p1 - p2)/(z^2 - 1.0)
         z1 <- z
         z  <- z - p1/pp
         if (abs(z - z1) < epsilon) ok<-T
      }
    Y[i] <- -z
    Y[M+1-i] <- z
    W[i] <- 2.0 / (( 1 - z^2) * pp^2)
    W[M+1-i] <- W[i]
   }
 if(!is.null(xrange)) {
    xL<- (xrange[2]-xrange[1])/2.0
    W <- xL * W
    Y <- xL * Y + (xrange[1]+xrange[2])/2.0
   }
 cbind(Y, W)
}
```

took 5.48 seconds in CPU time to perform 16,384 function evaluations when the calculation of the array containing the points and the weights was needed, and 0.09 seconds in CPU time when previously tabulated values were taken.

Multi-dimensional Gaussian quadrature up to $m = 20$ over hyper-rectangles could also be computed with the subroutine D01FBF from the commercial NAG Fortran li-

Table A.3. *The* S-Plus *function* `GL.integrate.2D`.

```
GL.integrate.2D <- function(fct, low=c(-1,1), upp=c(1,1),
                            order=10)
  {
  YW.list<-as.list(1:2)
  for(i in 1:2) {
   name.YW<-paste("GL.YW", abs(low[i]), abs(upp[i]), order,
                   sep=".")
    if (!exists(name.YW)) {
        assign(name.YW, GL.YW(order,
               xrange=c(low[i],upp[i])),
               where=1, immediate=T)
     }
    YW.list[[i]]<- get(name.YW)
  }
  fcteval <- outer2(YW.list[[1]][,1], YW.list[[2]][,1], fct)
  sum(YW.list[[1]][,2] * apply(YW.list[[2]][,2] *
      fcteval, 2, sum))
}
```

brary (`www.nag.co.uk`), and loaded dynamically into S-Plus or R. But, for our purposes this will not be needed. It also may be interesting to notice that there is an ongoing development for a free software library of routines for numerical computing: the 'GNU Scientific Library' (GSL) available at `sources.redhat.com/gsl/`. Preliminary releases look promising.

There are many different ways in which the Gaussian quadrature has been extended. An example are the Gauss–Kronrod formulae; see, for instance, Davis and Rabinowitz (1984, Section 2.7.1.1). An optimal extension can be found for Gauss–Legendre quadrature, giving a degree of exactness of $3M + 1$. This is for instance the case with the S-Plus or R function `integrate`, which implements uni-dimensional adaptive 15-point Gauss–Kronrod quadrature based on the Fortran functions DQAGE and DQAGIE from QUADPACK (Piessens *et al.*, 1983; `www.netlib.org/quadpack/`). This function is the only numerical integration function implemented in the S-Plus or R standard packages. A similar function, `gkint` from the S-Plus library `integrate2` (`lib.stat.cmu.edu/S/integrate2`) uses a (7–15)-point Gauss–Kronrod pair by means of the routine DQAG from QUADPACK.

As mentioned, for M quadrature points in each dimension the sum in (A.2.2) is over M^m terms. Therefore the numerical effort of Gaussian quadrature techniques increases exponentially with the integral dimension. Hence when m is large this method is nearly useless. Furthermore, the trouble with Gaussian quadrature is that you have no real idea of how accurate the answer is. You can always increase the accuracy by using a

Table A.4. *The* S-Plus *function* outer2.

```
outer2 <- function (X, Y, FUN="*", ...)
{
  no.nx <- is.null(nx <- dimnames(X <- as.array(X)))
  dX <- dim(X)
  no.ny <- is.null(ny <- dimnames(Y <- as.array(Y)))
  dY <- dim(Y)
  if (is.character(FUN) && FUN=="*") {
    robj <- as.vector(X) %*% t(as.vector(Y))
    dim(robj) <- c(dX, dY)
  } else {
    match.fun <- function(FUN) return(FUN)
    FUN <- match.fun(FUN)
    Y <- rep(Y, rep(length(X), length(Y)))
    X <- rep(X, length.out = length(Y))
    robj <- array(FUN(X, Y, ...), c(dX, dY))
  }
  if (no.nx) nx <- vector("list", length(dX))
  else if (no.ny) ny <- vector("list", length(dY))
  if (!(no.nx && no.ny)) dimnames(robj) <- c(nx, ny)
  robj
}
```

higher order Gauss method or by applying it piecewise over smaller periods but you still do not know the accuracy in terms of correct decimal places. To get a prescribed accuracy one needs to look at adaptive integration, which keeps reducing the step size until a specified error has been achieved.

A.3.   Adaptive methods

daptive algorithms are now used widely for the numerical calculation of multiple integrals. These algorithms have been developed for a variety of integration regions, including hyper-rectangles, spheres and simplices. A globally adaptive algorithm for integration over hyper-rectangles was first described by van Dooren and de Ridder (1976) and programmed as a Fortran function HALF. It was improved by Genz and Malik (1980). Implementations of the Genz and Malik modified algorithm (programmed as a Fortran function ADAPT) have appeared in the NAG Fortran library (www.nag.co.uk, subroutine D01FCF). The routine operates by repeated subdivisions of the hyper-rectangular region into smaller hyper-rectangles. In each subregion, the integral is estimated using a rule of degree seven, and an error estimate is

obtained by comparison with a rule of degree five which uses a subset of the same points. These subdivisions are designed to dynamically concentrate the computational work in the subregions where the integrand is most irregular, and thus adapt to the behaviour of the integrand. Genz (1991) gives a detailed description in the context of adaptive numerical integration for simplices.

Berntsen *et al.* (1991a) improved the reliability of previous algorithms, and developed a new algorithm for adaptive multidimensional integration. Tests (Berntsen *et al.* 1988) of a Fortran implementation, DCUHRE (Berntsen *et al.*, 1991b), have shown that their goal has been achieved.

Both DCUHRE and ADAPT can be dynamically loaded into S-Plus or R. The Fortran routine DCUHRE is implemented in the S-Plus function `dcuhre`, which is contained in the `integrate2` library (`lib.stat.cmu.edu/S/integrate2`), and ADAPT comes with the S-Plus function `adapt` included in the S-Plus library `adapt` (`lib.stat.cmu.edu/S/adapt`). Note that for R the function `adapt` is in the package `integrate`, which is available on CRAN (`cran.r-project.org`).

Genz (1992) suggested that such subregion adaptive integration algorithms can be used effectively in some multiple integration problems arising in statistics. The key to good solutions for these problems is the choice of an appropriate transformation from the infinite integration region for the original problem to a suitable finite region for the subregion adaptive algorithm. Genz (1992, Section 3.2) also discussed different types of such transformations; see also Davis and Rabinowitz (1984) for further examples of possible transformations.

Traditional quadrature methods (even newer adaptive ones) have been almost forgotten in the recent rush to 'Markov Chain Monte Carlo' (MCMC) methods; Evans and Swartz (1995) provided a nice recent summary focusing on these methods. They indicate that significant progress has been made using five general techniques: asymptotic methods, importance sampling, adaptive importance sampling, multiple quadrature and Markov chain methods. More recently, Genz and Kass (1997) argued that the reason why existing quadrature methods have been largely overlooked in statistics, even though they are known to be more efficient than Monte Carlo methods for well-behaved problems of low dimensionality, may be that when applied they are poorly suited for peaked-integrand functions. Hence they proposed transformations based on split-t distributions to allow integrals to be efficiently computed using a subregion-adaptive numerical integration algorithm. Fortran routines are already available (BAYESPACK at `www.sci.wsu.edu/math/faculty/genz/genzhome/software.html`) and work on constructing a version for use with S-Plus is underway.

A.4.   Monte Carlo methods

umerical methods that are known as Monte Carlo (MC) methods can be loosely described as statistical simulation methods. For a complete introduction to MC integration I refer to Stroud (1971, Chapter 6), Kalos and Whitlock (1986) or Robert and Casella (1999, Chapter 3). The classical MC

method for approximating a multiple integral as given in the left-hand side of (A.1.1) with $w(x_1, \ldots, x_m) = 1$, denoted by $\mathcal{I}(f)$, is as follows. We choose M set of points $\{y_{1,1}, \ldots, y_{1,m}\}, \ldots, \{y_{M,1}, \ldots, y_{M,m}\}$ at random, uniformly distributed in R_m. The integral is then estimated using $W_i = V/M$ in the right-hand side of (A.1.1),

$$\mathcal{I}(f) \approx \hat{\mathcal{I}}(f) = \frac{V}{M} \sum_{i=1}^{M} f(y_{i,1}, \ldots, y_{i,m}), \tag{A.4.1}$$

where $V = \mathcal{I}(1)$ is the m-dimensional volume of R_m. One notes that the basic MC method iteratively approximates a definite integral by uniformly sampling from the domain of integration, and averaging the function values at the samples. The integrand is treated as a random variable, and the sampling scheme yields a parameter estimate of the mean, or expected value of the random variable. Since $\hat{\mathcal{I}}(f)$ in the right-hand side of (A.4.1) estimates $\mathcal{I}(f)$ the absolute error ϵ in this mean can be evaluated by considering the corresponding standard error of the mean,

$$\epsilon = \left| \hat{\mathcal{I}}(f) - \mathcal{I}(f) \right| \approx \frac{\sigma}{M^{1/2}}, \tag{A.4.2}$$

where σ^2 is $V\mathcal{I}(f^2) - \mathcal{I}^2(f)$. If the $\{y_{1,1}, \ldots, y_{1,m}\}, \ldots, \{y_{M,1}, \ldots, y_{M,m}\}$ are regarded as independent random variables then $\hat{\mathcal{I}}(f)$ is a random variable with mean $\mathcal{I}(f)$ and variance σ^2/M, which can also be estimated from the random sample through

$$\frac{V}{M^2} \sum_{i=1}^{M} \{f(y_{i,1}, \ldots, y_{i,m}) - \hat{\mathcal{I}}(f)\}^2.$$

Furthermore, the error estimate (A.4.2) may be inverted to show the number of samples needed to yield a desired error, $M = \sigma^2/\epsilon^2$. For $m = 1$ this is illustrated by the S-PLUS function `MC.integrate.1D` given in Table A.5, and by `MC.integrate.2D` in Table A.6 for the two-dimensional case. An example of their use is

```
> MC.integrate.1D(function(z) sqrt(z), 0, 1, 1000, 2/3)
To achieve an error of 0.0001 you need at least 314516 points.
[1] 0.6670135
```

This clearly reflects the slow convergence of the MC methods; the absolute error (A.4.2) has an average magnitude of $O(M^{-1/2})$. Hence to reduce the error, for example, by a factor of 10 requires a 100-fold increase in the number of sample points. In the previous example one would need $M = 314,516$ points to get an accuracy of 0.0001. Therefore, other methods have been studied for decreasing the error. Such approximations are called 'Quasi Monte Carlo' (QMC) methods. The QMC method uses a formula which is formally identical to that of the MC method, except that the points used for evaluating the function are generated deterministically. Unlike the MC method, the QMC method has a deterministic error bound, and the accuracy of the integral is generally significantly better than in the MC method. Many different QMC methods are known. One method

Table A.5. *The* S-PLUS *function* `MC.integrate.1D`*.*

```
MC.integrate.1D <- function(fct, low=-1, upp=1, npoints=100,
                            exact.value=NULL)
  {
  points <- runif(n=npoints, min=low, max=upp)
  approx.tmp <- apply(as.matrix(points), 2, fct)
  V.tmp <- diff(c(low, upp))
  approx <- mean(approx.tmp) * V.tmp
  varapprox <- var(approx.tmp) * V.tmp
  if(!is.null(exact.value)) {
    cat("To achieve an error of 0.0001 you need at least",
        floor(varapprox^2/(0.0001^2)), "points.\n")
    }
  approx
  }
```

Table A.6. *The* S-PLUS *function* `MC.integrate.2D`*.*

```
MC.integrate.2D <- function(fct, low=c(-1,1), upp=c(1,1),
                            npoints=100, exact.value=NULL)
  {
  points.x<-runif(n=npoints, min=low[1], max=upp[1])
  points.y<-runif(n=npoints, min=low[2], max=upp[2])
  approx.tmp <- fct(points.x, points.y)
  V.tmp <- diff(c(low[1], upp[1]))* diff(c(low[2], upp[2]))
  approx <- mean(approx.tmp) * V.tmp
  varapprox <- var(approx.tmp) * V.tmp
  if(!is.null(exact.value)) {
    cat("To achieve an error of 0.0001 you need at least",
        floor(varapprox^2/(0.0001^2)), "points.\n")
    }
  approx
  }
```

makes use of results from the the theory of numbers and is called the *number-theoretic* method; see Stroud (1971, Section 6.3), Fang and Wang (1994) or Fang *et al.* (1994).

Additional methods have been employed to reduce the error of the MC method, such as importance sampling, stratified sampling, antithetic variates and non-random sequences (Press *et al.*, 1993, Sections 7.6–7.8, Evans and Swartz, 1995). These methods are mostly concerned with finding point sets that yield smaller integration errors.

Importance sampling concentrates samples in the area where they are more effective by using *a priori* knowledge of the function. Stratified sampling tries to distribute samples evenly by subdividing the domain into subregions such as grids. It is possible to combine some of these techniques, or to apply them adaptively (Press *et al.*, 1993, Section 7.8). For example, uniformly distributed samples generated by stratification can be employed for importance sampling.

As described above, MC integration draws samples from the required distribution, and then forms sample averages to approximate expectations. MCMC methods draw these samples by running a constructed Markov chain for a long time. An example of a way to construct such a chain is the Gibbs sampler. An introduction to MCMC methods and their applications is given in Gilks *et al.* (1996) or Robert and Casella (1999). But questions on convergence of the chains and efficient implementation are still to be solved (Cappé and Robert, 2000).

It is well known that for high-dimensional integrals MC techniques should be preferred to the standard quadrature methods given in Sections A.2 and A.3 since the sum in (A.4.1) is only over M terms instead of the M^m terms in (A.1.1). Nevertheless, I do not feel so comfortable using the MC methods mentioned in this section for mainly two reasons: first, one needs too many function evaluations to get a certain accuracy and second, as in Chapters 5–7 of this thesis saddlepoint approximations are used to replace bootstrap simulation I do find it a little strange to use MC methods to get the saddlepoint approximations. So I will consider their simplest versions (Tables A.5 and A.6) in the next section only for purposes of illustration.

A.5. Comparison

The testing of numerical quadrature methods involves the practical realization of the theoretical claims, and is well illustrated by applying a method to a set of well-designed examples. For $m = 1$ two extensive sets of test integrals which appear in the numerical analysis literature have been used to make a comparative computation. The first set is due to Casaletto *et al.* (1969) and contains 50 functions ranging from polynomials up to degree 20 through functions with discontinuities; see also Evans (1993, Table 2.3). The second set of 21 examples is due to Kahaner (1971) which includes in addition some harder examples (Evans, 1993, Table 2.4). These 71 test examples have been integrated using the various S-Plus functions described in the previous sections. Namely, from Section A.2 the default S-Plus function `integrate`, the function `intgauss` using a 10-point Gaussian formula, the (7–15)-point Gauss–Kronrod method implemented in `gkint` and `GL.integrate.1D` given in Table A.1 which uses the Gauss–Legendre (GL) rule with M points. And from Section A.4 the function `MC.integrate.1D` (Table A.5).

In an extensive comparative study I applied them to the 71 test examples. I used GL with $4, 8, 16, 32, 64$ and 128 points, and MC with $10^3, 10^4, 10^5$ and 10^6 points. The polynomials were easily integrated with lower order (> 4) GL rules. For the other examples it appeared that a 64-point GL procedure is necessary to get reliable results.

Table A.7. *A selection of test integrals used in the comparison study.*

Integrals	Analytic values
$C_3 = \int_0^1 (x^2 - 2x + 3)\, dx$	2.333333
$C_{26} = \int_0^1 2/\{2 + \sin(10\pi x)\}\, dx$	1.154700
$C_{29} = \int_0^{2\pi} x \sin(30x) \cos x\, dx$	-0.209672
$C_{30} = \int_0^{2\pi} x \sin(30x) \cos(50x)\, dx$	0.117809
$C_{34} = \int_0^{100\pi} \{(100\pi)^2 - x^2\}^{1/2} \sin x\, dx$	298.435716
$C_{48} = \int_0^1 c_{48}(x)\, dx,\ c_{48}(x) = \begin{cases} 1/(x+2), & 0 \le x \le e-2 \\ 0, & e-2 < x \le 1 \end{cases}$	0.306852
$K_2 = \int_0^1 k_2(x)\, dx,\ k_2(x) = \begin{cases} 0, & 0 \le x < 0.3 \\ 1, & 0.3 \le\, < x \le 1 \end{cases}$	0.7
$K_7 = \int_0^1 x^{-1/2}\, dx$	2
$K_{15} = \int_0^{10} 25 \exp(-25x)\, dx$	1
$K_{16} = \int_0^{10} 50/\{\pi(1 + 2500x^2)\}\, dx$	0.499363
$K_{21} = \int_0^1 k_{21}(x)\, dx,\ k_{21}(x) = [1/\cosh\{10(x-0.2)\}]^2$ $+[1/\cosh\{100(x-0.4)\}]^4$ $+[1/\cosh\{1000(x-0.6)\}]^6$	0.210802
$E_1 = \int_0^1 \int_0^1 1/(1 - x_1 x_2)\, dx_1 dx_2$	$\pi^2/6$
$E_3 = \int_{-1}^1 \int_{-1}^1 (2 - x_1 - x_2)^{-1/2}\, dx_1 dx_2$	$16(2 - \sqrt{2})/3$
$E_4 = \int_{-1}^1 \int_{-1}^1 (3 - x_1 - 2x_2)^{-1/2}\, dx_1 dx_2$	$4\sqrt{2}(3\sqrt{3} - 2\sqrt{2} - 1)/3$
$E_6 = \int_{-1}^1 \int_{-1}^1 |x_1^2 + x_2^2 - 0.25|\, dx_1 dx_2$	$5/3 + \pi/16$
$E_7 = \int_{-1}^1 \int_{-1}^1 |x_1 - x_2|^{1/2}\, dx_1 dx_2$	$8/15$

For the MC methods choices of the number of points below $10,000$ were unsatisfactory.

The most interesting of the 71 test functions are given in the upper and middle blocks of Table A.7, where C_i denote the selected test integrals from Casaletto *et al.* (1969) and K_j the ones of Kahaner (1971). Note that $i = 3, 26, 29, 30, 34, 48$ or $j = 7, 15, 16, 21$ correspond to n in Evans' Tables 2.3 or 2.4 respectively. The resulting absolute errors of the procedures compared to their analytical values (right column of Table A.7) are given in Figure A.1 for $C_3, C_{26}, C_{29}, C_{48}, K_2$ and K_7, and in Table A.8 for $C_{30}, C_{34}, K_{15}, K_{16}$ and K_{21}.   As illustrated in Figure A.1 the polynomial C_3 was easily found, as well as C_{26} which contains an oscillation in the denominator (Figure A.2, top middle panel). Integrals C_{29} and C_{30}, C_{34} (Table A.8), represented in the top right, bottom left and bottom middle panels of Figure A.2, all exhibit oscillatory behaviour which gives only very high order methods any chance of success. This is especially true with C_{34} when
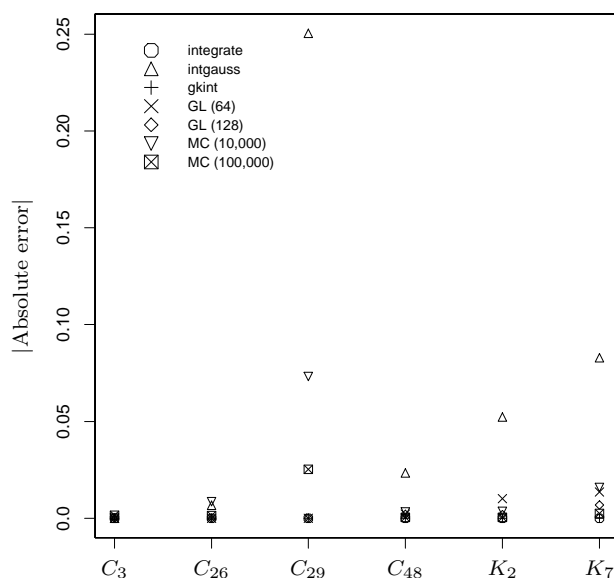
Figure A.1. *Plot of absolute integration errors for $C_3, C_{26}, C_{29}, C_{48}, K_2, K_7$ (given in Table A.7) using the* S-Plus *functions* integrate*,* intgauss*,* gkint*,* GL.integrate.1D *(GL) and* MC.integrate.1D *(MC) with the number of points in brackets.*

Table A.8. *Absolute integration errors for $C_{30}, C_{34}, K_{15}, K_{16}, K_{21}$ (given in Table A.7) using the* S-Plus *functions* integrate*,* intgauss*,* gkint*,* GL.integrate.1D *(GL) and* MC.integrate.1D *(MC) with the number of points in brackets. The values are rounded to three decimal places.*

	integrate	intgauss	gkint	GL (64)	GL (128)	MC (10,000)	MC (100,000)
C_{30}	0.000	3.186	0.000	0.239	0.176	0.228	0.064
C_{34}	0.000	5,476.455	0.000	3,425.142	0.000	886.795	131.527
K_{15}	0.000	0.681	0.000	0.000	0.000	0.022	0.024
K_{16}	0.000	0.362	0.000	0.001	0.000	0.061	0.047
K_{21}	0.211	0.211	0.210	0.211	0.211	0.210	0.210

MC integration and orders inferior to 128 are used. Similarly the discontinuities in C_{48} and K_2 did not cause surprising results in Figure A.1. Remark that an oddity occurred with K_{21} (Table A.8) which appears to defeat all the methods due to its nature shown in the bottom right panel of Figure A.2.

In summary, I noticed that the default S-Plus function integrate, which implements uni-dimensional adaptive 15-point Gauss–Kronrod quadrature, and a 128-point GL rule, as implemented in GL.integrate.1D, performed best, and this within the entire comparative computation of the 71 test examples. The S-Plus function intgauss
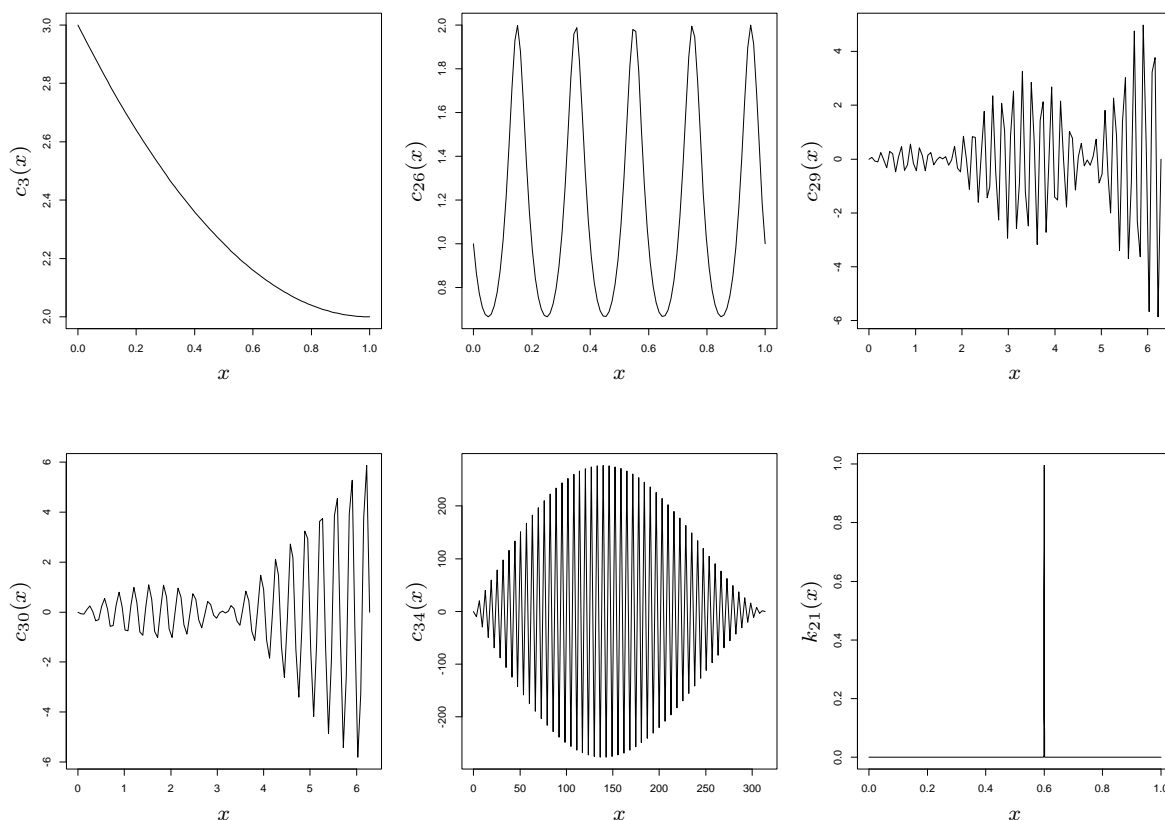
Figure A.2. *Plot of the integrands of $C_3, C_{26}, C_{29}, C_{30}, C_{34}$ and K_{21} (clockwise from top left) given in Table A.7.*

and the MC rules were less accurate.

In order to enable the testing of the other adaptive methods described in Section A.3 I considered the two-dimensional test integrals listed in Evans (1993, Table 6.2). A selection is given in the lower block of Table A.7, where E_i denote the I_i in Evans' Table 6.2 for $i = 1, 3, 4, 6, 7$. I considered the following adaptive rules: the ADAPT routine (Genz and Malik, 1980) using the S-Plus function `adapt` and DCUHRE (Berntsen *et al.*, 1991b) by means of the S-Plus function `dcuhre`. I compared both with `GL.integrate.2D` (GL) and `MC.integrate.2D` (MC), given in Tables A.3 and A.6, using several choices for the number of points per dimension. For example the use of a 32-point GL would result in 1024 function evaluations. The performance of these methods is illustrated in the left panel of Figure A.3. The maximal absolute error achieved over all integrals was 0.0326 with E_1 using `adapt`. This illustrates that all methods work reasonably well when applied to E_i, $i = 1, 3, 4, 6, 7$. Nevertheless, `adapt` and the MC methods (even with one million points) perform slightly worser than the others, whereas `dcuhre` and a 128-point GL seem to be the most accurate methods under consideration. But the right panel of Figure A.3 illustrates that the use of `dcuhre` results in significantly larger CPU times. This fact is not surprising for the MC methods as the number of points and hence the number of function evaluations is impractically large.

Once again, the use of a 128-point GL rule (`GL.integrate.2D`) delivered very ac-
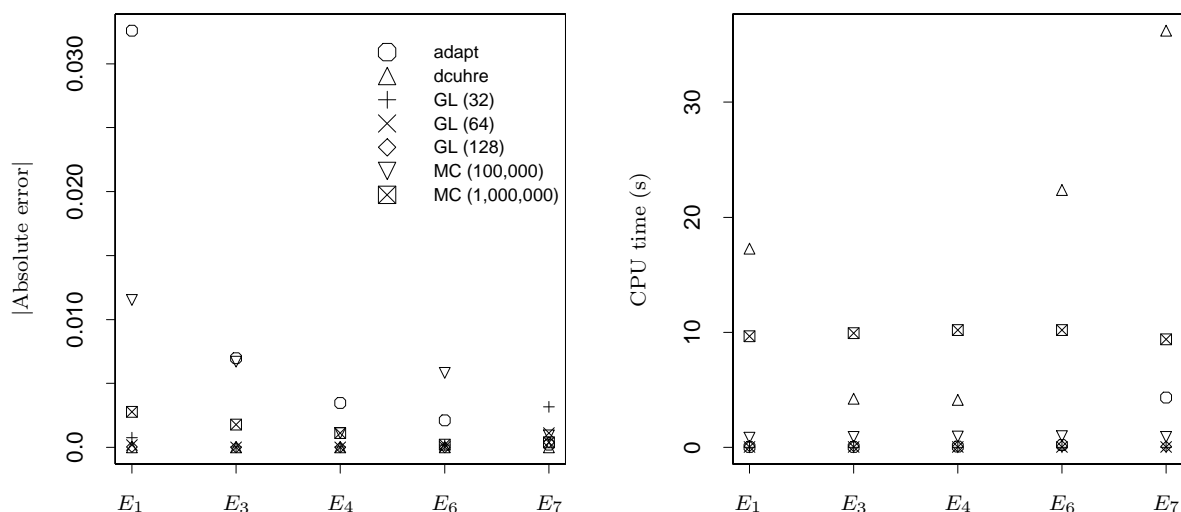
Figure A.3. *Integration of E_1, E_3, E_4, E_6, E_7 (given in Table A.7) using the* S-Plus *functions* adapt*,* dcuhre*,* GL.integrate.2D *(GL) and* MC.integrate.2D *(MC) with the number of points in brackets. Left panel: Absolute integration errors. Right panel: CPU times in seconds.*

curate results within small CPU times as well as did adapt. Moreover, I noticed that dcuhre outperformed adapt in accuracy, but used much more CPU time. This was underlined with additional examples which are not given here.

References

ABRAMOVITCH, L. & SINGH, K. (1985). Edgeworth corrected pivotal statistics and the bootstrap. *Annals of Statistics*, **13**, 116–132.

ALMUDEVAR, A., FIELD, C. & ROBINSON, J. (2000). The density of multivariate M-estimates. *Annals of Statistics*, **28**, 275–297.

ANALYTICAL METHODS COMMITTEE (1989). Robust statistics — how not to reject outliers. Part 1 and 2. *The Analyst*, **114**, 1693–1702.

ANDREWS, D. F, BICKEL, P. J., HAMPEL, F. R., HUBER, P. J., ROGERS, W. H. & TUKEY, J. W. (1972). *Robust Estimates of Location*. Princeton: Princeton University Press.

AZZALINI, A. (1998). Checking the validity of parametric models using nonparametric estimation: the PLRT approach. *Computing Science and Statistics; Proceedings of the 30th Symposium on the Interface*, 383–390.

AZZALINI, A. & BOWMAN, A. W. (1991). Nonparametric regression methods for repeated measurements. In *Nonparametric Functional Estimation and Related Topics, G. G. Roussas (Ed.)*. Dordrecht: Kluwer Academic Publishers, 377–387.

AZZALINI, A. & BOWMAN, A. W. (1993). On the use of nonparametric regression for checking linear relationships. *Journal of the Royal Statistical Society series B*, **55**, 549–557.

AZZALINI, A., BOWMAN, A. W. & HÄRDLE, W. (1989). On the use of nonparametric regression for model checking. *Biometrika*, **76**, 1–11.

BARBE, P. & BERTAIL, P. (1995). *The Weighted Bootstrap*. Volume 98 of *Lecture Notes in Statistics*. New York: Springer.

BARNDORFF-NIELSEN, O. E. (1986). Inference on full or partial parameters based on the standardized signed log likelihood ratio. *Biometrika*, **73**, 307–322.

BARNDORFF-NIELSEN, O. E. (1990). Approximate interval probabilities. *Journal of the Royal Statistical Society series B*, **52**, 485–496.

BARNDORFF-NIELSEN, O. E. (1991). Modified signed log likelihood ratio. *Biometrika*, **78**, 557–563.

BARNDORFF-NIELSEN, O. E. & COX, D. R. (1979). Edgeworth and saddle-point approximations with statistical applications (with Discussion). *Journal of the Royal Statistical Society series B*, **41**, 279–312.

BARNDORFF-NIELSEN, O. E. & COX, D. R. (1989). *Asymptotic Techniques for Use in Statistics*. London: Chapman & Hall.

BARNDORFF-NIELSEN, O. E. & COX, D. R. (1994). *Inference and Asymptotics*. London: Chapman & Hall.

References

BARNETT, V. & LEWIS, T. (1994). *Outliers in Statistical Data* (3rd ed.). New York: Wiley.

BAUM, L. E. & BILLINGSLEY, P. (1965). Asymptotic distributions for the coupon collector's problem. *Annals of Mathematical Statistics*, **36**, 1835–1839.

BECKER, R. A., CHAMBERS, J. M. & WILKS, A. R. (1988). *The New S Language*. London: Chapman & Hall.

BERAN, R. (1987). Prepivoting to reduce level error of confidence sets. *Biometrika*, **74**, 457–468.

BERAN, R. (1994). Seven stages of bootstrap. In *Computational Statistics — Papers Collected on the Occasion of the 25th Conference on Statistical Computing at Schloss Reisensburg, P. Dirschedl and R. Ostermann (Eds.)*. Physica: Heidelberg, 143–157.

BERNTSEN, J., ESPELID, T. O. & GENZ, A. (1988). *A Test of ADMINT*. Reports in Informatics 31, Department of Informatics, University of Bergen, Bergen, Norway.

BERNTSEN, J., ESPELID, T. O. & GENZ, A. (1991a). An adaptive algorithm for the approximate calculation of multiple integrals. *ACM Transactions on Mathematical Software*, **17**, 437–451.

BERNTSEN, J., ESPELID, T. O. & GENZ, A. (1991b). Algorithm 698: DCUHRE — An adaptive multidimensional integration routine for a vector of integrals. *ACM Transactions on Mathematical Software*, **17**, 452–456.

BICKEL, P. J. (1992). Theoretical comparison of different bootstrap-t confidence bounds. In *Exploring the Limits of Bootstrap, R. LePage and L. Billard (Eds.)*. New York: Wiley, 65–76.

BLEISTEIN, N. & HANDELSMAN, R. A. (1975). *Asymptotic Expansions of Integrals*. New York: Holt, Rinehart & Winston.

BOOS, D. D. & HUGHES-OLIVER, J. M. (2000). How large does n have to be for Z and t intervals? *The American Statistician*, **54**, 121–128.

BOOTH, J. G. & BUTLER, R. W. (1990). Randomization distributions and saddlepoint approximations in generalized linear models. *Biometrika*, **77**, 787–796.

BOOTH, J. G. & WOOD, A. T. A . (1995). An example in which the Lugannani–Rice saddlepoint formula fails. *Statistics & Probability Letters*, **23**, 53–61.

BOWMAN, A. W. & AZZALINI, A. (1997). *Applied Smoothing Techniques for Data Analysis: The Kernel Approach with* S-PLUS *Illustrations*. Oxford: Oxford University Press.

BRAZZALE, A. R. (2000). *Practical small-sample parametric inference*. PhD Thesis No. 2230, Department of Mathematics, Swiss Federal Institute of Technology, CH-1015 Lausanne.

BROWN, B. M., HALL, P. & YOUNG, G. A. (2001). The smoothed median and the bootstrap. *Biometrika*, **88**, 519–534.

BUCKLEY, M. J. & EAGLESON, G. K. (1998). An approximation to the distribution of quadratic forms in normal random variables. *Australian Journal of Statistics*, **30**, 150–159.

BUTLER, R. W. & SUTTON, R. K. (1998). Saddlepoint approximations for multivariate cumulative distribution functions and probability computations in sampling theory and outlier testing. *Journal of the American Statistical Association*, **93**, 596–604.

CANTY, A. J. & DAVISON, A. C. (1996). Implementation of saddlepoint approximations to bootstrap distributions. *Computing Science and Statistics; Proceedings of the 28th Symposium on the Interface*, 248–253.

CANTY, A. J. & DAVISON, A. C. (1999). Implementation of saddlepoint approximations in resampling problems. *Statistics and Computing*, **9**, 9–15.

CANTY, A. J. & DAVISON, A. C. (2000). Contribution to the discussion of 'The estimating function bootstrap' by F. Hu and J. D. Kalbfleisch. *The Canadian Journal of Statistics*, **28**, 488–493.

CANTY, A. J., DAVISON, A. C. & HINKLEY, D. V. (1996). Contribution to the discussion of 'Bootstrap confidence intervals' by T. J. DiCiccio and B. Efron. *Statistical Science*, **11**, 214–219.

CAPPÉ, O. & ROBERT, C. P. (2000). Ten years and still running! *Journal of the American Statistical Association*, **95**, 1282–1286.

CASALETTO, J., PICKETT, M. & RICE, J. R. (1969). A comparison of some numerical integration programs. *SIGNUM Newsletter*, **4**, 30–40.

CHAMBERS, J. M. (1998). *Programming with Data: A Guide to the S Language.* New York: Springer.

CARPENTER, J. (1999). Test inversion bootstrap confidence intervals. *Journal of the Royal Statistical Society series B*, **61**, 159–172.

CARPENTER, J. & BITHELL, J. (2000). Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians. *Statistics in Medicine*, **19**, 1141–1164.

CLEVELAND, W. S. & DEVLIN, S. J. (1988). Locally weighted regression: an approach to regression analysis by local fitting. *Journal of the American Statistical Association*, **83**, 596–610.

CLEVELAND, W. S. & LOADER, C. (1995). Smoothing by local regression: principles and methods. In *Statistical Theory and Computational Aspects of Smoothing, W. Härdle and M. G. Schimek (Eds.).* New York: Springer, 10–49.

COURANT, R. & HILBERT, D. (1950). *Methods of Mathematical Physics.* New York: Wiley.

COX, D. R. & REID, N. (1987). Approximations to noncentral distributions. *The Canadian Journal of Statistics*, **15**, 105–114.

CRAMÉR, H. (1938). Sur un nouveau théorème-limite de la théorie des probabilités. *Actualités Scientifiques et Industrielles*, Volume 736, Paris: Hermann.

CUSHNY, A. R. & PEEBLES, A. R. (1905). The action of optical isomers. II. Hyoscines. *Journal of Physiology*, **32**, 501–510.

DANIELS, H. E. (1954). Saddlepoint approximations in statistics. *Annals of Mathematical Statistics*, **25**, 631–650.

DANIELS, H. E. (1980). Exact saddlepoint approximations. *Biometrika*, **67**, 53–58.

DANIELS, H. E. (1983). Saddlepoint approximations for estimating equations. *Biometrika*, **70**, 89–96.

DANIELS, H. E. (1987). Tail probability approximations. *International Statistical Review*, **55**, 37–48.

References

DANIELS, H. E. (1997). Saddlepoint approximations in statistics. In *Breakthroughs in Statistics — Volume III, S. Kotz and N. L. Johnson (Eds.)*. New York: Springer, 177–200.

DANIELS, H. E. & YOUNG, G. A. (1991). Saddlepoint approximation for the studentized mean, with an application to the bootstrap. *Biometrika*, **78**, 169–179.

DAVIES, P. L. & GATHER, U. (1993). The identification of multiple outliers (with Discussion). *Journal of the American Statistical Association*, **88**, 782–801.

DAVIES, R. B. (1973). Numerical inversion of a characteristic function. *Biometrika*, **60**, 415–417.

DAVIES, R. B. (1980). Algorithm AS 155: the distribution of a linear combination of χ^2 random variables. *Applied Statistics*, **29**, 323–333.

DAVIS, P. J. & RABINOWITZ, P. (1984). *Methods of Numerical Integration*. New York: Academic Press.

DAVISON, A. C. (1988). Approximate conditional inference in generalized linear models. *Journal of the Royal Statistical Society series B*, **50**, 445–461.

DAVISON, A. C. (2001). Biometrika centenary: theory and general methodology. *Biometrika*, **88**, 13–52.

DAVISON, A. C. & HINKLEY, D. V. (1988). Saddlepoint approximations in resampling methods. *Biometrika*, **75**, 417–431.

DAVISON, A. C. & HINKLEY, D. V. (1997). *Bootstrap Methods and Their Application*. Cambridge: Cambridge University Press.

DAVISON, A. C., HINKLEY, D. V. & WORTON, B. J. (1995). Accurate and efficient construction of bootstrap likelihoods. *Statistics and Computing*, **5**, 257–264.

DAWKINS, B. (1991). Siobhan's problem: the coupon collector revisited. *The American Statistician*, **45**, 76–82.

DIACONIS, P. & HOLMES, S. (1994). Gray codes for randomization procedures. *Statistics & Computing*, **4**, 287–302.

DIBLASI, A. & BOWMAN, A. W. (1997). Testing for constant variance in a linear model. *Statistics & Probability Letters*, **33**, 95–103.

DICICCIO, T. J. & EFRON, B. (1992). More accurate confidence intervals in exponential families. *Biometrika*, **79**, 231–245.

DICICCIO, T. J. & EFRON, B. (1996). Bootstrap confidence intervals (with Discussion). *Statistical Science*, **11**, 189-228.

DICICCIO, T. J. & FIELD, C. A. (1991). An accurate method for approximate conditional and Bayesian inference about linear regression models from censored data. *Biometrika*, **78**, 903–910.

DICICCIO, T. J., FIELD, C. A. & FRASER, D. A. S. (1990). Approximations of marginal tail probabilities and inference for scalar parameters. *Biometrika*, **77**, 77–95.

DICICCIO, T. J. & MARTIN, M. A. (1991). Approximations of marginal tail probabilities for a class of smooth functions with applications to Bayesian and conditional inference. *Biometrika*, **78**, 891–902.

DiCiccio, T. J., Martin, M. A. & Young, G. A. (1991). An invariance property of marginal density and tail probability approximations for smooth functions. *Statistics & Probability Letters*, **12**, 249–255.

DiCiccio, T. J., Martin, M. A. & Young, G. A. (1992a). Analytical approximations for iterated boostrap confidence intervals. *Statistics and Computing*, **2**, 161–171.

DiCiccio, T. J., Martin, M. A. & Young, G. A. (1992b). Fast and accurate approximate double bootstrap confidence intervals. *Biometrika*, **79**, 285–295.

DiCiccio, T. J., Martin, M. A. & Young, G. A. (1994). Analytical approximations to bootstrap distribution functions using saddlepoint methods. *Statistica Sinica*, **4**, 281–296.

DiCiccio, T. J. & Romano, J. (1988). A review of bootstrap confidence intervals. *Journal of the Royal Statistical Society series B*, **50**, 338–355.

DiCiccio, T. J. & Romano, J. (1990). Nonparametric confidence limits by resampling methods and least favorable families. *International Statistical Review*, **58**, 59–76.

Easton, G. S. & Ronchetti, E. M. (1986). General saddlepoint approximations, with applications to L statistics. *Journal of the American Statistical Association*, **81**, 420–430.

Efron, B. (1979) Bootstrap methods: another look at the jackknife. *Annals of Statistics* **7**, 1–26.

Efron, B. (1981). Nonparametric standard errors and confidence intervals (with Discussion). *The Canadian Journal of Statistics*, **9**, 139–172.

Efron, B. (1982). *The Jackknife, The Bootstrap, and Other Resampling Plans.* CBMS-NSF Regional Conference Series in Applied Mathematics, 38, Philadelphia: SIAM.

Efron, B. (1992). Six questions raised by the bootstrap. In *Exploring the Limits of Bootstrap, R. LePage and L. Billard (Eds.).* New York: Wiley, 99–126.

Efron, B. & Tibshirani, R. J. (1993). *An Introduction to the Bootstrap.* New York: Chapman & Hall.

Eliot, T. S. (1943). *Four Quartets.* London: Faber & Faber.

Engels, H. (1980). *Numerical Quadrature and Cubature.* New York: Academic Press.

Esscher, F. (1932). On the probability function in collective risk theory. *Scandinavian Actuarial Journal*, **15**, 175–195.

Eubank, R. L. & Spiegelman, C. H. (1990). Testing the goodness of fit of a linear model via nonparametric regression techniques. *Journal of the American Statistical Association*, **85**, 387–392.

Evans, G. (1993). *Practical Numerical Integration.* New York: Wiley.

Evans, M. & Swartz, T. (1995). Methods for approximating integrals in statistics with special emphasis on Bayesian integration problems. *Statistical Science*, **10**, 254–272.

Fan, J. & Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications.* London: Chapman & Hall.

Fan, R. Y. K. & Field, C. A. (1995). Marginal densities of M-estimators. *The Canadian Journal of Statistics*, **23**, 185–197.

Fang, K. T. & Wang, Y. (1994). *Number-theoretic Methods in Statistics.* London: Chapman & Hall.

References

FANG, K. T., WANG, Y., BENTLER, P. M. (1994). Some applications of number-theoretic methods in statistics. *Statistical Science*, **9**, 416–428.

FAREBROTHER, R. W. (1990). Algorithm AS 256: the distribution of a quadratic form in normal variables. *Applied Statistics*, **39**, 294–309.

FELLER, W. (1970). *An Introduction to Probability Theory and Its Applications* (3rd ed., vol. I). New York: Wiley.

FEUERVERGER, A. (1989). On the empirical saddlepoint approximation. *Biometrika*, **76**, 457–464.

FEUERVERGER, A., ROBINSON, J. & WONG, A. (1999). On the relative accuracy of certain bootstrap procedures. *The Canadian Journal of Statistics*, **27**, 225–236.

FIELD, C. A. (1982). Small sample asymptotic expansions for multivariate *M*-estimates. *Annals of Statistics*, **10**, 672–689.

FIELD, C. A. (1997). Robust regression and small sample confidence intervals. *Journal of Statistical Planning and Inference*, **57**, 39–48.

FIELD, C. A. & HAMPEL, F. R. (1978). *Small sample asymptotic distributions of M-estimators of location.* Research Report No. 17, Fachgruppe für Statistik, ETH, Zürich, Switzerland.

FIELD, C. A. & HAMPEL, F. R. (1982). Small-sample asymptotic distributions of *M*-estimators of location. *Biometrika*, **69**, 29–46.

FIELD, C. A. & RONCHETTI, E. M. (1985). A tail area influence function and its application to testing. *Communications in Statistics series C*, **4**, 19–41.

FIELD, C. A. & RONCHETTI, E. M. (1990). *Small Sample Asymptotics.* Volume 13 of *Lecture Notes — Monograph Series.* Hayward: Institute of Mathematical Statistics.

FIELD, C. A. & TINGLEY, M. A. (1997). Small sample asymptotics: applications in robustness. In *Handbook of Statistics — Volume 15: Robust Inference, G. S. Maddala and C. R. Rao (Eds.).* Amsterdam: North-Holland, 513–536.

GARTHWAITE, P. H. & BUCKLAND, S. T. (1992). Generating Monte Carlo confidence intervals by the Robbins–Monro process. *Applied Statistics*, **41**, 159–171.

GATTO, R. (1994). *Saddlepoint methods and nonparametric approximations for econometric models.* PhD Thesis, Faculty of Economic and Social Sciences, University of Geneva, Switzerland.

GATTO, R. & RONCHETTI, E. M. (1996). General saddlepoint approximations of marginal densities and tail probabilities. *Journal of the American Statistical Association*, **91**, 666–673.

GENZ, A. (1991). An adaptive numerical integration algorithm for simplices. In *Computing in the 90s, Proceedings of the First Great Lake Computer Science Conference, N. A. Sherwani, E. de Doncker and J. A. Kapenga (Eds.).* Lecture Notes in Computer Science, 507, New York: Springer, 279–292.

GENZ, A. (1992). Statistics applications of subregion adaptive multiple numerical integration. In *Numerical Integration: Recent Developments, Software and Applications, T. O. Espelid and A. Genz (Eds.).* Dordrecht: Kluwer Academic Publishers, 267–280.

GENZ, A. & KASS, R. E. (1997). Subregion-adaptive integration of functions having a dominant peak. *Journal of Computational and Graphical Statistics*, **6**, 92–111.

GENZ, A. & MALIK, A. A. (1980). An adaptive algorithm for numerical integration over an N-dimensional rectangular region. *Journal of Computational and Applied Mathematics*, **6**, 295–302.

GILKS, W. R., RICHARDSON, S. & SPIEGELHALTER, D. J. (Eds.) (1996). *Markov Chain Monte Carlo in Practice*. London: Chapman & Hall.

GOOD, I. J. (1957). Saddlepoint methods for the multinomial distributions. *Annals of Mathematical Statistics*, **28**, 861–881.

GOOD, I. J. (1961). The multivariate saddlepoint method and chi-squared for the multinomial distribution. *Annals of Mathematical Statistics*, **32**, 535–548.

GOUTIS, C. & CASELLA, G. (1999). Explaining the saddlepoint approximation. *The American Statistician*, **53**, 216–224.

GROSS, A. M. (1976). Confidence interval robustness with long-tailed symmetric distributions. *Journal of the American Statistical Association*, **71**, 409–416.

GROSS, A. M. (1977). Confidence intervals for bisquare regression estimates. *Journal of the American Statistical Association*, **72**, 341–354.

HALL, P. (1988). Theoretical comparison of bootstrap confidence intervals (with Discussion). *Annals of Statistics*, **16**, 927–985.

HALL, P. (1992). *The Bootstrap and Edgeworth Expansion*. New York: Springer.

HALL, P. & MARTIN, M. A. (1991). On the error incurred using the bootstrap variance estimate when constructing confidence intervals for quantiles. *Journal of Multivariate Analysis*, **38**, 70–81.

HAMPEL, F. R. (1974a). Some small-sample asymptotics. In *Proceedings of the Prague Symposium on Asymptotic Statistics, J. Hajek (Ed.)*. 1973, Prague: Charles University, 109–126.

HAMPEL, F. R. (1974b). The influence curve and its role in robust estimation. *Journal of the American Statistical Association*, **69**, 383–393.

HAMPEL, F. R. (1995). *Some Additional Notes on the 'Princeton Robustness Year'*. Research Report No. 76, Seminar für Statistik, ETH, Zürich, Switzerland.

HAMPEL, F. R. (1997). *Is Statistics Too Difficult?* Research Report No. 81, Seminar für Statistik, ETH, Zürich, Switzerland.

HAMPEL, F. R. (1998). Is statistics too difficult? *The Canadian Journal of Statistics*, **26**, 497–513.

HAMPEL, F. R. (2000). *Robust Inference*. Research Report No. 93, Seminar für Statistik, ETH, Zürich, Switzerland.

HAMPEL, F. R. (2001). *Robust Statistics: A Brief Introduction and Overview*. Research Report No. 94, Seminar für Statistik, ETH, Zürich, Switzerland.

HAMPEL, F. R., RONCHETTI, E. M., ROUSSEEUW, P. J. & STAHEL, W. J. (1986). *Robust Statistics, the Approach Based on Influence Functions*. New York: Wiley.

HAND, D. J., DALY, F., LUNN, A. D., MCCONWAY, K. J. & OSTROWSKI, E. (1994). *A Handbook of Small Data Sets*. New York: Chapman & Hall.

HASTIE, T. J. & LOADER, C. (1993). Local regression: automatic kernel carpentry (with Discussion). *Statistical Science*, **8**, 120–143.

References

Hastie, T. J. & Tibshirani, R. J. (1990). *Generalized Additive Models*. London: Chapman & Hall.

Hesterberg, T. C. (1994). Saddlepoint quantiles and distribution curves, with bootstrap applications. *Computational Statistics*, **9**, 207–212.

Hinkley, D. V. & Wei, B-C. (1984). Improvements of jackknife confidence limit methods. *Biometrika*, **71**, 331–339.

Hogg, R. V. (1974). Adaptive robust procedures: a partial review and some suggestions for the future. *Journal of the American Statistical Association*, **69**, 909–923.

Hu, F. & Hu, J. (2000). A note on breakdown theory for bootstrap methods. *Statistics & Probability Letters*, **50**, 49–53.

Hu, F. & Kalbfleisch, J. D. (2000). The estimating function bootstrap (with Discussion). *The Canadian Journal of Statistics*, **28**, 449–499.

Huber, P. J. (1964). Robust estimation of a location parameter. *Annals of Mathematical Statistics*, **35**, 73–101.

Huber, P. J. (1981). *Robust Statistics*. New York: Wiley.

Huber, P. J. (1996). *Robust Statistical Procedures* (2nd ed.). CBMS-NSF Regional Conference Series in Applied Mathematics, 68, Philadelphia: SIAM.

Hughes, B. D. (1995). *Random Walks and Random Environments — Volume 1: Random Walks*. Oxford: Oxford University Press.

Huzurbazar, S. (1999). Practical saddlepoint approximations. *The American Statistician*, **53**, 225–232.

Ihaka, R. & Gentleman, R. (1996). R: a language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, **5**, 299–314.

Imhof, J. P. (1961). Computing the distribution of quadratic forms in normal variables. *Biometrika*, **48**, 419–426.

Jensen, D. R. & Solomon, H. (1972). A Gaussian approximation to the distribution of a definite quadratic form. *Journal of the American Statistical Association*, **67**, 898–902.

Jensen, D. R. & Solomon, H. (1994). Approximations to joint distributions of definite quadratic forms. *Journal of the American Statistical Association*, **89**, 480–486.

Jensen, J. L. (1988). Uniform saddlepoint approximations. *Advances in Applied Probability*, **20**, 622–634.

Jensen, J. L. (1995). *Saddlepoint Approximations*. Oxford: Oxford University Press.

Jing, B.-Y., Feuerverger, A. & Robinson, J. (1994). On the bootstrap saddlepoint approximations. *Biometrika*, **81**, 211–215.

Jing, B.-Y. & Robinson, J. (1994). Saddlepoint approximations for marginal and conditional probabilities of transformed variables. *Annals of Statistics*, **22**, 1115–1132.

Johnson, N. L. (1959). On an extension of the connexion between Poisson and χ^2 distributions. *Biometrika*, **46**, 352–63.

Johnson, N. L. (1978). Modified t tests and confidence intervals for asymmetrical populations. *Journal of the American Statistical Association*, **73**, 536–544.

Johnson, N. L. & Kotz, S. (1970). *Distributions in Statistics: Continuous Univariate Distributions — 2*. New York: Wiley.

Johnson, N. L., Kotz, S. & Kemp, A. W. (1994). *Univariate Discrete Distributions* (2nd ed.). New York: Wiley.

Kalos, M. H. & Whitlock, P. A. (1986). *Monte Carlo Methods Volume 1: Basics*. New York: Wiley.

Kahaner, D. K. (1971). Comparison of numerical quadrature formulae. In *Mathematical Software, J. R. Rice (Ed.)*. New York: Academic Press, 229–259.

Koerts, J. & Abrahamse, A. P. J. (1969). *On the Theory and Application of the General Linear Model*. Rotterdam: Rotterdam University Press.

Kokoska, S. & Nevison, C. (1989). *Statistical Tables and Formulae*. New York: Springer.

Kolassa, J. E. (1994). *Series Approximation Methods in Statistics*. Volume 88 of *Lecture Notes in Statistics*. New York: Springer.

Kuonen, D. (1998a). *Saddlepoint approximations for bootstrap distributions*. Undergraduate Thesis and Technical Report 1998.1, Department of Mathematics, Swiss Federal Institute of Technology, CH-1015 Lausanne.

Kuonen, D. (1998b). Saddlepoint approximations for bootstrap distributions. *Presentation of the Undergraduate Thesis*. March 16, 1998, Department of Mathematics, Swiss Federal Institute of Technology, Lausanne, Switzerland. (`stat.kuonen.com/talks/`)

Kuonen, D. (1998c). Saddlepoint approximation for testing a linear relationship. *Presentation at the Joint Session of 6th Prague Symposium on Asymptotic Statistics and 13th Prague Conference on Information Theory, Statistical Decision Functions and Random Processes*. August 27, 1999, Charles University, Prague, Czech Republic. (`stat.kuonen.com/talks/`)

Kuonen, D. (1998d). *Notes on saddlepoint mixture approximations*. Technical Report 1998.8, Department of Mathematics, Swiss Federal Institute of Technology, CH-1015 Lausanne.

Kuonen, D. (1999). Saddlepoint approximations for distributions of quadratic forms in normal variables. *Biometrika*, **86**, 929–935.

Kuonen, D. (2000a). A saddlepoint approximation for the collector's problem. *The American Statistician*, **54**, 165–169.

Kuonen, D. (2000b). Intervalles de confiance bootstrap studentisés basés sur des M-estimateurs robustes utilisant des approximations de point de selle. *Presentation at the XXIIe Journées de Statistiques*. May 16, 2000, Fès, Maroc. (`stat.kuonen.com/talks/`)

Kuonen, D. (2000c). Studentized bootstrap confidence intervals using saddlepoint approximations. *Presentation at the Workshop on Inference and Asymptotics*. July 14, 2000, Ascona, Switzerland. (`stat.kuonen.com/talks/`)

Kuonen, D. (2000d). Saddlepoint approximations of studentized bootstrap confidence intervals based on M-estimates. *Presentation at the International Conference of the Royal Statistical Society*. September 13, 2000, University of Reading, United Kingdom. (`stat.kuonen.com/talks/`)

Kuonen, D. (2000e). Fun with saddlepoint approximations. *Presentation at the Swiss Research Students in Statistics Meeting 2000*. November 3, 2000, University of Bern, Switzerland. (`stat.kuonen.com/talks/`)

LEE, S. M. S. & YOUNG, G. A. (1996). Asymptotics and resampling methods. *Computing Science and Statistics; Proceedings of the 28th Symposium on the Interface*, 221–227.

LÉGER, C. (2000). Contribution to the discussion of 'The estimating function bootstrap' by F. Hu and J. D. Kalbfleisch. *The Canadian Journal of Statistics*, **28**, 486–488.

LEONE, F. C., JAYACHANDRAN, T. & EISENSTAT, S. (1967). A study of robust estimators. *Technometrics*, **9**, 652–660.

LEVIN, B. (1981). A representation for multinomial cumulative distribution functions. *Annals of Statistics*, **9**, 1123–1126.

LEVIN, B. (1983). On calculations involving the maximum cell frequency. *Communications in Statistics — Theory and Methods*, **12**, 1299–1327.

LEVIN, B. (1992). Regarding 'Siobhan's problem: the coupon collector revisited'. *The American Statistician*, **46**, 76.

LIEBERMAN, O. (1994). Saddlepoint approximation for the distribution of the ratio of quadratic forms in normal variables. *Journal of the American Statistical Association*, **89**, 924–928.

LUGANNANI, R. & RICE, S. O. (1980). Saddlepoint approximation for the distribution of the sum of independent random variables. *Advances in Applied Probability*, **12**, 475–490.

LUGANNANI, R. & RICE, S. O. (1984). Distribution of the ratio of quadratic forms in normal variables — numerical methods. *SIAM Journal of Scientific and Statistical Computing*, **5**, 476–488.

MacKAY, R. J. & OLDFORD, R. W. (2000). Scientific method, statistical method and the speed of light. *Statistical Science*, **15**, 254–278.

MARTIN, M. A. (1996). Postcards from Asymptotica: a discussion of papers by Lee and Young, Ronchetti, and Kolossa and Tanner. *Computing Science and Statistics; Proceedings of the 28th Symposium on the Interface*, 239–241.

MASON, D. M. & NEWTON, M. A. (1992). A rank statistic approach to the consistency of a general bootstrap. *Annals of Statistics*, **20**, 1611–1624.

MATHAI, A. M. & PROVOST, S. B. (1992). *Quadratic Forms in Random Variables: Theory and Applications.* New York: Marcel Dekker.

McCULLAGH, P. (1987). *Tensor Methods in Statistics.* London: Chapman & Hall.

MONTI, A. C. & RONCHETTI, E. M. (1993). On the relationship between empirical likelihood and empirical saddlepoint approximations for multivariate M-estimators. *Biometrika*, **80**, 329–338.

PACE, L. & SALVAN, A. (1997). *Principles of Statistical Inference From a Neo-Fisherian Perspective.* Singapore: World Scientific.

PATNAIK, P. B. (1949). The non-central χ^2 and F-distributions and their applications. *Biometrika*, **36**, 202–232.

PEARSON, E. S. (1959). Note on an approximation to the distribution of non-central χ^2. *Biometrika*, **46**, 364.

PIESSENS, R., DE DONCKER-KAPENGA, E., UBERHUBER, C. & KAHANER, D. (1983). *QUAD-PACK: A Subroutine Package for Automatic Integration.* Berlin: Springer.

POLANSKY, A. M. (2000). Stabilizing bootstrap-t confidence intervals for small samples. *The Canadian Journal of Statistics*, **28**, 501–516.

PRESS, W. H., TEUKOLSKY, S. A., VETTERLING, W. T. & FLANNERY, B. P. (1993). *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge: Cambridge University Press.

PORT, S. C. (1994). *Theoretical Probability for Applications*. New York: Wiley.

READ, K. L. Q. (1998). A lognormal approximation for the collector's problem. *The American Statistician*, **52**, 175–180.

REID, N. (1988). Saddlepoint methods and statistical inference (with Discussion). *Statistical Science*, **3**, 213–238.

REID, N. (1991). Approximations and asymptotics. In *Statistical Theory and Modelling: In Honour of Sir David Cox, D. V. Hinkley, N. Reid and E. J. Snell (Eds.)*. London: Chapman & Hall, 287–334.

REID, N. (1995). The roles of conditioning in inference (with Discussion). *Statistical Science*, **10**, 138–199.

REID, N. (1996). Likelihood and higher order approximations to tail areas: a review and annotated bibliography. *The Canadian Journal of Statistics*, **24**, 141–166.

RICE, S. O. (1980). Distribution of quadratic forms in variables — Evaluation by numerical integration. *SIAM Journal of Scientific and Statistical Computing*, **1**, 438–448.

ROBERT, C. P. & CASELLA, G. (1999). *Monte Carlo Statistical Methods*. New York: Springer.

ROBINSON, J. (1982). Saddlepoint approximations for permutation tests and confidence intervals. *Journal of the Royal Statistical Society series B*, **44**, 91–101.

ROBINSON, J., HÖGLUND, T., HOLST, L. & QUINE, M. P. (1990). On approximating probabilities for small and large deviations in $\mathbb{R}^d$. *Annals of Probability*, **18**, 727–753.

ROBINSON, J. & SKOVGAARD, I. M. (1998). Bounds for probabilities of small relative errors for empirical saddlepoint and bootstrap tail approximations. *Annals of Statistics*, **26**, 2369–2394.

ROCKE, D. M., DOWNS, G. W. & ROCKE, A. J. (1982). Are robust estimators really necessary? *Technometrics*, **24**, 95–102.

RONCHETTI, E. M. (1996). A comparison of saddlepoint approximations for marginal distributions. *Computing Science and Statistics; Proceedings of the 28th Symposium on the Interface*, 228–233.

RONCHETTI, E. M. (1997). Introduction to Daniels (1954). Saddlepoint approximations in statistics. In *Breakthroughs in Statistics — Volume III, S. Kotz and N. L. Johnson (Eds.)*. New York: Springer, 171–176.

RONCHETTI, E. M. & WELSH, A. H. (1993). Empirical saddlepoint approximations for multivariate M-estimators. *Journal of the Royal Statistical Society series B*, **56**, 313–326.

ROSNER, B. (1977). Percentage points for the RST many outlier procedure. *Technometrics*, **19**, 307–313.

ROUSSEEUW, P. J. & LEROY, A. M. (1987). *Robust Regression and Outlier Detection*. New York: Wiley.

References

ROUTLEDGE, R. & TSAO, M. (1997a). On the relationship between two asymptotic expansions for the distribution of sample mean and its applications. *Annals of Statistics*, **25**, 2200–2209.

ROUTLEDGE, R. & TSAO, M. (1997b). Quadrature methods for computing distributions. *Journal of Computational Statistics & Data Analysis*, **26**, 149–161.

RUBEN, H. (1962). Probability content of regions under spherical normal distributions, IV: the distribution of homogeneous and non-homogeneous quadratic functions of normal variables. *Annals of Mathematical Statistics*, **33**, 542–570.

RUBEN, H. (1963). A new result on the distribution of quadratic forms. *Annals of Mathematical Statistics*, **34**, 1582–1584.

RUBIN, H. (1999). Posting in the newsgroup `sci.math.num-analysis`. December 9, 1999. (`www.math.niu.edu/~rusin/known-math/99/num_int`)

RUPPERT, D. & WAND, M. P. (1994). Multivariate locally weighted least squares regression. *Annals of Statistics*, **22**, 1346–1370.

SCHEFFÉ, H. (1959). *The Analysis of Variance.* New York: Wiley.

SHAO, J. & TU, D. (1995). *The Jackknife and Bootstrap.* New York: Springer.

SHORACK, G. R. (1976). Robust studentization of location estimates. *Statistica Neerlandica*, **30**, 119–141.

SINGH, K. (1998). Breakdown theory for bootstrap quantiles. *Annals of Statistics*, **26**, 1719–1732.

SKOVGAARD, I. M. (1987). Saddlepoint expansions for conditional distributions. *Journal of Applied Probability*, **24**, 875–887.

SKOVGAARD, I. M. (1996). An explicit large-deviation approximation to one-parameter tests. *Bernoulli*, **2**, 145–166.

SOLOMON, H. & STEPHENS, M. A. (1977). Distribution of a weighted sum of chi-squared variables. *Journal of the American Statistical Association*, **72**, 881–885.

SPADY, R. H. (1991). Saddlepoint approximations for regression models. *Biometrika*, **78**, 879–889.

STAUDTE, R. G. & SHEATHER, S. J. (1990). *Robust Estimation and Testing.* New York: Wiley.

STIGLER, S. M. (1973). Simon Newcombe, Percy Daniell, and the history of robust estimation 1885–1920. *Journal of the American Statistical Association*, **68**, 872–879.

STIGLER, S. M. (1977). Do robust estimators work with real data (with Discussion)? *Annals of Statistics*, **5**, 1055–1098.

STRAWDERMAN, R. L., CASELLA, G. & WELLS, M. T. (1996). Practical small-sample asymptotics for regression problems. *Journal of the American Statistical Association*, **91**, 643–654.

STROMBERG, A. J. (1997). Robust covariance estimates based on resampling. *Journal of Statistical Planning and Inference*, **57**, 321–334.

STROUD, A. H. (1971). *Approximation Caculation of Multiple Integrals.* New Jersey: Prentice–Hall.

THOMAS, G. E. (2000). Use of the bootstrap in robust estimation of location. *The Statistician*, **49**, 63–77.

TIBSHIRANI, R. J. (1988). Variance stabilization and the bootstrap. *Biometrika*, **75**, 433–444.

TINGLEY, M. A. & FIELD, C. A. (1990). Small-sample confidence intervals. *Journal of the American Statistical Association*, **85**, 427–434.

VAN DER VAART, A. W. (1998). *Asymptotic Statistics*. Cambridge: Cambridge University Press.

VAN DOOREN, P. & DE RIDDER, L. (1976). An adaptive algorithm for numerical integration over an N-dimensional cube. *Journal of Computational and Applied Mathematics*, **2**, 207–217.

VENABLES, W. N. & RIPLEY, B. D. (1999). *Modern Applied Statistics with S-Plus* (3rd ed.). New York: Springer.

VENABLES, W. N. & RIPLEY, B. D. (2000). *S Programming*. New York: Springer.

WANG, S. (1990a). Saddlepoint approximations in resampling analysis. *Annals of the Institute of Statistical Mathematics*, **42**, 115–131.

WANG, S. (1990b). Saddlepoint approximation for nonlinear statistics. *The Canadian Journal of Statistics*, **18**, 255–264.

WANG, S. (1992). General saddlepoint approximations in the bootstrap. *Statistics & Probability Letters*, **13**, 61–66.

WHITTLE, P. (1993). A conversation with Henry Daniels. *Statistical Science*, **8**, 342–353.

WOOD, A. T. A., BOOTH, J. G. & BUTLER, R. W. (1993). Saddlepoint approximations to the CDF for some statistics with non-normal limit distributions. *Journal of the American Statistical Association*, **88**, 680–686.

YOUNG, G. A. (1994). Bootstrap: more than a stab in the dark (with Discussion)? *Statistical Science*, **9**, 382–415.

YOUNG, G. A. & DANIELS, H. E. (1991). Bootstrap bias. *Biometrika*, **77**, 179–185.

Curriculum vitae of Diego Kuonen

Born on the 5th of August, 1973
Citizenship: Swiss
Single

Contact

Work address:	Chair of Statistics, Department of Mathematics (DMA), Swiss Federal Institute of Technology (EPFL), 1015 Lausanne, Switzerland.
Home address:	Haus FIS, 3920 Zermatt, Switzerland.
E-mail address:	`Diego@Kuonen.com`
Professional web page:	`stat.kuonen.com`
Personal web page:	`diego.kuonen.com`
WAP page:	`wap.kuonen.com`

Educational background

Sept. 2001	PhD in Statistics ('*Docteur ès sciences*'). PhD thesis: 'Computer-intensive statistical methods: saddlepoint approximations with applications in bootstrap and robust inference'. Supervisor: Professor A. C. Davison. Referees: Professor R. Dalang (DMA, EPFL), Professor T. J. DiCiccio (Cornell University, USA), Professor S. Morgenthaler (DMA, EPFL) and Dr G. A. Young (Cambridge University, UK).
1998 – 2001	PhD student in Statistics at the Department of Mathematics of the Swiss Federal Institute of Technology, Lausanne. Supervisor: Professor A. C. Davison.
March 1998	Diploma in Mathematics ('*Ingénieur mathématicien diplômé de l'EPFL*'). Undergraduate thesis: 'Saddlepoint approximations for bootstrap distributions'. Supervisor: Professor A. C. Davison. Referees: Professor E. M. Ronchetti (University of Geneva) and Dr G. A. Young (Cambridge University, UK).
1992 – 1997	Undergraduate student at the Department of Mathematics of the Swiss Federal Institute of Technology, Lausanne, Switzerland.
July 1992	Degree in Mathematics ('*Eidgenössische Maturität Typus* C'). Grammar school 'Spiritus Sanctus', Brig, Switzerland.

Teaching experience

Nov. 2001	*Lecturer* for 'Statistical Methods for Engineers' given in the context of the 'Postgraduate Course in Electrical Power Engineering', LRE, EPFL.
Oct. 2001 – March 2002	*Lecturer* for 'Probabilités et Statistique', DE and DMT, EPFL.
Oct. 2001	*Lecturer* for a one-day course 'Introduction to S-Plus for UNIX' given in the context of the 'Postgraduate Course in Mathematical Engineering', DMA, EPFL.
March 2001	*Instructor* (joint with Professor S. Morgenthaler and R. Furrer) for 'Probabilités & Statistique' given in the context of the '2001 Postgraduate Cycle in Biomedical Engineering', LGM, EPFL.
Oct. 2000 – Febr. 2001	*Teaching Assistant* for 'Probabilités et Statistique I' given by Professor A. C. Davison, DMA, EPFL.
Oct. 2000	*Instructor* for a one-day course 'Introduction to S-Plus for UNIX' given in the context of the 'Postgraduate Course in Mathematical Engineering', DMA, EPFL.
Oct. 1999 – Febr. 2000	*Supervision* of the semester projects 'Revisiting statistical applications in soccer' (B. Emonet, DMA, EPFL) and 'Statistical analysis of the serverlog of a website' (T. N. T. Ho, DMA, EPFL).
Oct. 1999	*Instructor* for a one-day course 'Introduction to S-Plus for UNIX' given in the context of the 'Postgraduate Course in Mathematical Engineering', DMA, EPFL.
April 1999 – Febr. 2001	*Organizer* and *Instructor* for a two-day course 'Introduction to S-Plus for UNIX', DMA, EPFL. Dates: April 8–9, 1999; Oct. 18–19, 1999; Oct. 10–11, 2000 and Febr. 13–14, 2001.
March 1999 – June 1999	*Teaching Assistant* for 'Monte Carlo Inference' given by Professor A. C. Davison, DMA, EPFL.
Oct. 1998 – March 2000	*Teaching Assistant* for 'Algèbre Linéaire' given by Professor A. C. Davison, DMA, EPFL.
March 1997 – May 1997	*Teaching Assistant* for 'Applied and Computational Statistics' given by Professor A. Marazzi, Faculty of Medicine, University of Lausanne.
Sept. 1996 – Febr. 1997	*Student Assistant* for 'Probabilités et Statistique' given by Dr J. M. Helbling, DMA, EPFL.
Sept. 1995 – Febr. 1998	*Student Assistant* for 'Probabilités et Statistique I et II' given by Professor S. Morgenthaler, DMA, EPFL.
Sept. 1995 – Febr. 1997	*Student Assistant* for 'Analysis I und II' (in German) given by Professor A. Wohlhauser, DMA, EPFL.

Research experience

1999 – present	*Statistics, gene transfer and DNA*. Joint work with Dr M. Jordan, Laboratory of Cellular Biotechnology, IGC, DC, EPFL.
2000 – 2001	*Analysis of T-cell immune activation in children with vertically transmitted hepatitis C virus infection.* Joint work with Dr A. Giovannetti, Universita degli Studi di Roma 'La Sapienza', Dipartimento di Medicina Clinica, Roma, Italy.

1999 – 2000	*Predicting wolf (Canis lupus) presence and reproduction sites, and simulation of spreading movements from different arrival points in the southern part of Switzerland (Canton of Valais).* Joint work with C. Glenz, GECOS, DGR, EFPL.
1999 – 2000	*Analysis of the effects exerted by highly active antiretroviral therapy on the immune system of HIV-1 infected individuals.* Joint work with Dr A. Giovannetti, Universita degli Studi di Roma 'La Sapienza', Dipartimento di Medicina Clinica, Roma, Italy.

Practical experience

1999 – present	*Expert in Computers and Programming* for the TESS (Testing Engineers Skills in Statistics) project, which was developed by the SEFI (European Society for Engineering Education) in partnership with the CRE (Association of European Universities).
1999 – present	*Webmaster* of the common web server of the Chairs of Statistics (`statwww.epfl.ch`), DMA, EPFL.
1999 – present	*Webmaster* of the Chair of Statistics of Professor A. C. Davison, DMA, EPFL.
1999 – 2001	*IT Responsible* of the Chair of Statistics of Professor A. C. Davison, DMA, EPFL.
1996 – 1997	*Creation of web pages* for the EPFL: 'The EPFL in 1996, annual report', 'The EPFL in 1995, annual report' and 'Postgraduate programmes and further education'.

Referred publications

Glenz, C., Massolo, A., Kuonen, D. & Schlaepfer, R. (2001). A wolf habitat suitability prediction study in Valais (Switzerland). *Landscape and Urban Planning*, **55**, 55–65.

Giovannetti, A., Mazzetta, F., Coviello, R., Casadei, A. M., Mattia, S., Marziali, M., Pierdominici, M., Kuonen, D., Pesce, A. M., Fiorilli, M., Aiuti, F. & Quinti, I. (2001). T-cell immune activation in children with vertically transmitted hepatitis C virus infection. *Viral Immunology*, **14**, 169–179.

Giovannetti, A., Pierdominici, M., Mazzetta, F., Salemi, S., Marziali, M., Kuonen, D., Iebba, F., Lusi, E. A., Cossarizza, A. & Aiuti, F. (2001). T cell responses to highly active antiretroviral therapy defined by chemokine receptors expression, cytokine production, T cell receptor repertoire and anti-HIV T-lymphocyte activity. *Clinical and Experimental Immunology*, **124**, 21–31.

Kuonen, D. (2000). A saddlepoint approximation for the collector's problem. *The American Statistician*, **54**, 165–169.

Giovannetti, A., Pierdominici, M., Mazzetta, F., Salemi, S., Marziali, M., Kuonen, D., Iebba, F., Lusi, E. A. & Aiuti, F. (2000). Contemporary evaluation of T cell subsets, chemokine receptors expression, cytokine production, anti-HIV cytotoxic activity and TCR V-beta repertoire in HIV-1-infected individuals during HAART. *Journal on Biotechnology and Molecular Biology*, **12**, 196.

Kuonen, D. & Roehrl, A. S. A (2000). Was France's World Cup win pure chance? *Student*, **3**, 153–166.

Kuonen, D. (1999). Saddlepoint approximations for distributions of quadratic forms in normal variables. *Biometrika*, **86**, 929–935.

Kuonen, D. & Roehrl, A. S. A (1998). Identification of key steps to evaluate statistical software. *InterStat*, [Online], October, **2**. (`interstat.stat.vt.edu/InterStat/`)

Kuonen, D. (1998). *Saddlepoint approximations for bootstrap distributions.* Undergraduate Thesis, Department of Mathematics, Swiss Federal Institute of Technology, CH-1015 Lausanne.

Other publications

Furrer, R. & Kuonen, D. (2001). GRASS GIS et R: main dans la main dans un monde libre. *Flash Informatique Spécial Été*, 51–56.

Kuonen, D. & Furrer, R. (2001). Data mining avec R dans un monde libre. *Flash Informatique Spécial Été*, 45–50.

Kuonen, D. (2001). *Introduction to* S-PLUS *for UNIX (with Exercises)* (4th ed.). Department of Mathematics, Swiss Federal Institute of Technology, CH–1015 Lausanne.

Kuonen, D. & Chavez–Demoulin, V. (2001). R — un exemple du succès des modèles libres. *Flash Informatique*, **2**, 3–7.

Kuonen, D. (2000). *Introduction to* S-PLUS *for UNIX (with Exercises)* (3rd ed.). Department of Mathematics, Swiss Federal Institute of Technology, CH-1015 Lausanne.

Kuonen, D., Roehrl, A. S. A., Schmiedl, S. & Verasani, T. (2000). WAPux — der moderne Pinguin: Auch Pinguine wollen mit WAP-Handys spielen! *Coverstory of 'Linux Magazin'*, **5**, 36–43.

Kuonen, D. (1999). *Introduction to* S-PLUS *for UNIX (with Exercises)* (2nd ed.). Department of Mathematics, Swiss Federal Institute of Technology, CH-1015 Lausanne.

Furrer, R., Kuonen, D. & Picasso, M. (1999). La rédaction de documents scientifiques avec LaTeX. *Coverstory of 'Flash Informatique'*, **2**, 1, 16–17.

Kuonen, D. (1999). *Introduction to* S-PLUS *for UNIX (with Exercises)* (1st ed.). Department of Mathematics, Swiss Federal Institute of Technology, CH-1015 Lausanne.

Proceedings

Kuonen, D. (2000). Saddlepoint approximations of studentized bootstrap confidence intervals based on M-estimates. *Abstracts of the International Conference of the Royal Statistical Society*, September 11–15, 2000, University of Reading, United Kingdom, 66.

Giovannetti, A., Kuonen, D., Mazzetta, F., Iebba, F., Lusi, E. A., Rosso, P., Pierdominici, M. & Aiuti, F. (2000). Improvement of CD4, but not CD8, TCR BV repertoire after HAART. *Proceedings of the XIIIe International AIDS Conference*, July 9–14, 2000, Durban, South Africa.

Giovannetti, A., Pierdominici, M., Mazzetta, F., Salemi, S., Marziali, M., Kuonen, D., Iebba, F., Lusi, E.A. & Aiuti, F. (2000). T cell responses to highly active antiretroviral therapy defined by chemokine receptors expression, cytokine production, T cell receptor repertoire and anti-HIV T-lymphocyte activity. *Proceedings of the Conference Highlights in Immunology, Oncology and Virology. AIDS, Cancer and Hepatitis: Scientific and Ethical Challenges*, June 14–17, 2000, Republic of San Marino.

Giovannetti, A., Pierdominici, M., Mazzetta, F., Salemi, S., Marziali, M., Kuonen, D., Iebba, F., Lusi, E. A. & Aiuti, F. (2000). Contemporary evaluation of T cell subsets, chemokine receptors expression, cytokine production, anti-HIV cytotoxic activity and TCR V-beta repertoire in HIV-1-infected individuals during HAART. *Proceedings of the Joint Symposium SI-SIIIC, BCG and EAACI*, June 7–10, 2000, Ferrara, Italy.

Kuonen, D. (2000). Intervalles de confiance bootstrap studentisés basés sur des M-estimateurs robustes utilisant des approximations de point de selle. *Proceedings of the XXXIIe Journées de Statistique*, Mai 15–19, 2000, Fez, Marocco, 501–502.

Kuonen, D. (1998). Saddlepoint approximation for testing a linear relationship. *Abstracts of Prague Stochastics '98: Joint Session of 6th Prague Symposium on Asymptotic Statistics and 13th Prague Conference on Information Theory, Statistical Decision Functions and Random Processes*, August 23–28, 1998, Charles University, Prague, Czech Republic, 52.

Technical reports

Emonet, B. & Kuonen, D. (2000). *Revisiting statistical applications in soccer*. Technical Report 2000.2, Department of Mathematics, Swiss Federal Institute of Technology, CH-1015 Lausanne.

Kuonen, D. & Roehrl, A. S. A (2000). *Lessons learnt from current statistics packages and a proposal for a new system for the new millennium.* Technical Report 2000.1, Department of Mathematics, Swiss Federal Institute of Technology, CH-1015 Lausanne.

Kuonen, D. (1999). *Analysis of the effects exerted by highly active antiretroviral therapy on the immune system of HIV-1 infected individuals.* Technical Report 1999.9, Department of Mathematics, Swiss Federal Institute of Technology, CH-1015 Lausanne.

Kuonen, D., Chavez–Demoulin, V., Roehrl, A. S. A. & Chavez, E. (1999). *La France, championne du monde de football: qui l'eût cru?* Technical Report 1999.1, Department of Mathematics, Swiss Federal Institute of Technology, CH-1015 Lausanne.

Kuonen, D. (1998). *Notes on saddlepoint mixture approximations.* Technical Report 1998.8, Department of Mathematics, Swiss Federal Institute of Technology, CH-1015 Lausanne.

Kuonen, D. (1998). *Saddlepoint approximations for bootstrap distributions.* Technical Report 1998.1, Department of Mathematics, Swiss Federal Institute of Technology, CH-1015 Lausanne.

Kuonen, D. (1997). *Statistical models for knock-out soccer tournaments.* Technical Report 1997.3, Department of Mathematics, Swiss Federal Institute of Technology, CH-1015 Lausanne.

Kuonen, D. & Marazzi, A. (1997). *Robust testing between two asymmetric distributions: a simulation study.* Technical Report 1997.2, Department of Mathematics, Swiss Federal Institute of Technology, CH-1015 Lausanne.

Kuonen, D. (1996). *Modelling the success of soccer teams in European championships.* Technical Report 1996.1, Department of Mathematics, Swiss Federal Institute of Technology, CH-1015 Lausanne.

Kuonen, D. (1995). *Pluie acide et racines des arbres.* Technical Report 1995.1, Department of Mathematics, Swiss Federal Institute of Technology, CH-1015 Lausanne.

Book reviews

Kuonen, D. (2000). Review of 'Medical Statistics at a Glance' by A. Petrie and C. Sabin, Blackwell Science. To appear in *Journal of Applied Statistics.*

Kuonen, D. (2000). Review of 'Stochastic Processes for Insurance and Finance' by T. Rolski *et al.*, New York: Wiley. To appear in *Journal of Applied Statistics.*

Kuonen, D. & Roehrl, A. S. A. (2000). Review of 'Introductory Statistics with Applications in General Insurance (2nd ed.)' by I. B. Hossack *et al.*, Cambridge: Cambridge University Press. *Journal of Applied Statistics*, **27**, 133.

Kuonen, D. & Roehrl, A. S. A. (1999). Review of 'Frontiers of Research in Economic Theory: The Nancy L. Schwartz Memorial Lectures, 1983–1997' by N. L. Schwartz *et al.* (Eds.), Econometric Society Monographs, No. 29, Cambridge: Cambridge University Press. To appear in *Journal of Applied Statistics.*

Kuonen, D. & Roehrl, A. S. A. (1999). Review of 'Computer Assisted Survey Information Collection' by Mick P. Couper *et al.* (Eds.), Wiley Series in Probability and Statistics, New York: Wiley. To appear in *Journal of Applied Statistics.*

Kuonen, D. & Roehrl, A. S. A. (1999). Review of 'Applied Regression Analysis' (3rd ed.) by N. R. Draper and H. Smith, Wiley Series in Probability and Statistics, New York: Wiley. To appear in *Journal of Applied Statistics.*

Presentations

November 3, 2000	*Elements of data visualization.* Invited speaker, Swiss Research Students in Statistics Meeting 2000, University of Bern, Switzerland.
November 3, 2000	*Fun with saddlepoint approximations.* Invited speaker, Swiss Research Students in Statistics Meeting 2000, University of Bern, Switzerland.

September 13, 2000	*Saddlepoint approximations of studentized bootstrap confidence intervals based on M-estimates.* International Conference of the Royal Statistical Society, University of Reading, United Kingdom.
August 21, 2000	*The best and worst of statistical graphics and what we learn from them.* Invited speaker, Serono Pharmaceutical Research Institute (SPRI), Plan-les-Ouates, Geneva, Switzerland.
July 14, 2000	*Studentized bootstrap confidence intervals using saddlepoint approximations.* Workshop on Inference and Asymptotics, Ascona, Switzerland.
May 16, 2000	*Intervalles de confiance bootstrap studentisés basés sur des M-estimateurs robustes utilisant des approximations de point de selle.* XXXIIe Journées de Statistique, Fez, Marocco.
August 27, 1998	*Saddlepoint approximation for testing a linear relationship.* Prague Stochastics '98 — Joint Session of 6th Prague Symposium on Asymptotic Statistics and 13th Prague Conference on Information Theory, Statistical Decision Functions and Random Processes, Charles University, Prague, Czech Republic.
May 11, 1998	*Estimation d'une densité: une introduction.* Department of Mathematics, Swiss Federal Institute of Technology, CH-1015 Lausanne
March 16, 1998	*Saddlepoint approximations for bootstrap distributions.* Presentation of the Diploma Report, Department of Mathematics, Swiss Federal Institute of Technology, CH-1015 Lausanne.

Posters

May 29 – June 3, 1999	*Predicting wolf (Canis lupus) presence and reproduction sites, and simulation of spreading movements from different arrival points.* Glenz, C., Massolo, A., Kuonen, D. & Schlaepfer, R., 3rd European Congress of Mammalogy, Jvskyla, Finland.

Conferences and workshops

November 20–21, 2000	*International Symposium on Data Mining and Statistics.* University of Augsburg, Germany.
November 8–10, 2000	*Statistics Culture and the Statistics Market in Switzerland, Swiss Days of Statistics.* St. Gallen, Switzerland.
September 17–19, 2000	*Ecole d'Été du 3e Cycle Romand de Statistique et Probabilités Appliquées.* Crans-Montana, Switzerland.
September 11–15, 2000	*International Conference of the Royal Statistical Society.* University of Reading, United Kingdom.
August 16, 2000	*Computational Methods in Decision-Making and Finance.* One-day meeting held within the 'International Workshop on Parallel Matrix Algorithms and Applications', Neuchâtel, Switzerland.
July 9–14, 2000	*Workshop on Inference and Asymptotics.* Ascona, Switzerland.
May 15–19, 2000	*XXXIIe Journées de Statistique.* Fez, Marocco.
September 1–3, 1999	*Engineering Education: Rediscovering the Centre — SEFI Annual Conference 1999.* Zürcher Hochschule Winterthur (ZHW) and Swiss Federal Institute of Technology Zürich (ETHZ), Switzerland.
May 7, 1999	*European Plan for Research in Official Statistics — Information-Workshop.* Swiss Federal Statistical Office, Neuchtel, Switzerland.

March 23–26, 1999	*State of the Art in Probability and Statistics — Symposium on the Occasion of the 65th Birthday of Willem van Zwet.* University of Leiden, Leiden, Netherlands.
August 23–28, 1998	*Prague Stochastics '98 — Joint Session of 6th Prague Symposium on Asymptotic Statistics and 13th Prague Conference on Information Theory, Statistical Decision Functions and Random Processes.* Charles University, Prague, Czech Republic.
July 26–30, 1998	*Warwick Randomised Algorithms and Stochastic Simulation — Tutorial and Workshop (WRASS).* University of Warwick, United Kingdom.

Committees

2001 – present	Member of the 'Free Software' working group of EPFL's IT 2001 project.
1999 – present	Member of the 'Conseil du DMA'.
1996 – present	Member of the 'Commission d'informatique du DMA'.

Professional memberships

2001 – present	Association for Computing Machinery Special Interest Group on Knowledge Discovery in Data and Data Mining (ACM SIGKDD).
2000 – present	'Association des Mathématiciens de l'EPFL' (AME).
1999 – present	Institute of Mathematical Statistics (IMS).
1998 – present	Swiss Statistical Society (SSS).
1998 – present	'Association amicale des anciens élèves de l'EPFL' (A3E2PFL).
1996 – present	American Statistical Association (ASA).
1998 – 2001	European Mathematical Society (EMS).
1998 – 2001	Swiss Mathematical Society (SMS).
1998 – 1999	'Association du personnel de la Confédération' (APC).

Other memberships

2001 – present	Founder and President of the organization 'Alumni der D'OBRU'.
2000 – present	Founder and Vice President of the 'Linux User Group Oberwallis' (LUGO).
1997 – present	Honorary President of the student organization 'Oberwalliser Studenten in Lausanne' (D'OBRU).
1995 – 1997	Founder and President of the student organization 'Oberwalliser Studenten in Lausanne' (D'OBRU).

Certificates

December 2000	*Entrepreneurship Course.* A 14-week evening course for pre-selected candidates given by CREATE ('The Branco Weiss Chair of Entrepreneurship and Innovation'), EPFL.
September 2000	*Statistics in Finance.* A two-day short course as part of the 'International Conference of the Royal Statistical Society' held at the University of Reading, United Kingdom.
September 1998	'*Programmation en Langage* C'. A five-day course (No. 3216) at the 'Service Informatique Centrale' of the EPFL.

Miscellany

2001 – present	Chief of the care service of the civil protection in Zermatt, Switzerland.
June 12, 2001	Invited participant in the round table 'Free Software at the EPFL' (`elle.epfl.ch`), salle Polyvalente, EPFL.
2000	Referee for *Student.*

Language skills

German	Mother tongue.
English	Fluently spoken, read and written.
French	Fluently spoken, read and written.
Italian	Spoken and read.

Computing skills

Softwares	S-Plus, R, LaTeX , XGobi, Xlisp-Stat, XploRe, BUGS, Matlab, Mathematica, Maple.
Programming languages	S, C, C++, Pascal, BASIC, shell scripts, Perl, CGI, HTML, XML, WML, JavaScript, SQL.
Operating systems	Linux, UNIX (Solaris, IRIX), DOS, Microsoft Windows 3.11, 95, 98, NT, 2000, ME and XP.
Strong points	Excellent control of the tools of the Linux and UNIX systems. Very good knowledge of the Internet applications. Skills in administrating Linux and UNIX systems.

Hobbies

Music, Soccer, Salsa, Golf, Linux operating system.

References

Professor A. C. Davison	Chair of Statistics. Department of Mathematics, Swiss Federal Institute of Technology, 1015 Lausanne, Switzerland. E-mail address: `anthony.davison@epfl.ch`.
Professor S. Morgenthaler	Chair of Applied Statistics. Department of Mathematics, Swiss Federal Institute of Technology, 1015 Lausanne, Switzerland. E-mail address: `stephan.morgenthaler@epfl.ch`.
Professor P. Nuesch	Chair of Statistics. Department of Mathematics, Swiss Federal Institute of Technology, 1015 Lausanne, Switzerland. E-mail address: `peter.nuesch@epfl.ch`.
Professor A. Marazzi	Unité de statistique, Institut de statistique universitaire de médecine sociale et préventive, Rue du Bugnon 17, 1005 Lausanne, Switzerland. E-mail address: `alfio.marazzi@unil.ch`.
Dr E. Chavez	Laboratoires Serono S.A. (LSA), Zone Industrielle de l'Ouriettaz, 1170 Aubonne, Switzerland. E-mail address: `enrico.chavez@serono.com`.
Mrs. Francoise Come	Secretary General SEFI (and Vice President EASK), European Society for Engineering Education (SEFI), rue de la Concorde 60, 1050 Brussels, Belgium. E-mail address: `francoise.come@sefi.be`.