

On the Bootstrap and Confidence Intervals

by

Michael A. Martin

**A thesis submitted for the degree of Doctor of Philosophy of the
Australian National University**

February, 1989

Preface

Chapter One of this thesis reviews work published by others in order to provide a background and motivation for the work that follows. Unless otherwise stated in the text, the remaining chapters describe my own work, supervised by Professor P. G. Hall and published jointly with him.

Michael A. Martin

Michael A. Martin

Acknowledgements

I could not have hoped to complete this work if it were not for the support, assistance, and friendship of a great many people.

Firstly, I thank Professor Peter Hall for his thoughtful and excellent supervision. I have benefited greatly from Peter's insight, and his friendship, and could not have wished for a better supervisor. I also wish to thank Professor William Schucany, who visited the Australian National University from Southern Methodist University, Dallas, Texas, during the course of this work, for his friendly advice and several helpful discussions. I am grateful as well to the staff in both Departments of Statistics at A.N.U. for their support and assistance, particularly Professors Chris Heyde and Chip Heathcote and Doctors Des Nicholls and Alan Welsh. I have been very fortunate to have worked in an extremely friendly and stimulating environment, which has made my time at A.N.U. most enjoyable.

A special word of thanks goes to Mrs Ann Milligan, who expertly and superbly tackled the difficult task of rendering my thesis into word-processed form. I am also very grateful to Barbara La Scala and Russell Standish for their help with the editing and printing of the thesis.

I am especially indebted to my family and friends, who have supported and encouraged me over the last few years; to Mum and Dad, and my dear grandmother, Nell, for their endless supply of love and encouragement; and to the many special people, too numerous to mention separately, who have talked with me, listened to me, and above all, accepted me as a friend.

Finally, I dedicate this thesis to my brother, Peter, who died of cancer in 1984, aged 16.

Related Publications

The following papers have been submitted for publication from the work in this thesis:

Hall, P. and Martin, M.A. (1988a) On the bootstrap and two-sample problems. *Austral. J. Statist.* **30A**, 179–192.

Hall, P. and Martin, M.A. (1988b) On bootstrap resampling and iteration. *Biometrika*, in press.

Hall, P. and Martin, M.A. (1988c) Exact convergence rate of bootstrap quantile variance estimator. *Probab. Theor. and Rel. Fields*, in press.

Hall, P. and Martin, M.A. (1988d) A note on the accuracy of bootstrap percentile method confidence intervals for a quantile. *Statist. Probab. Letters*, in press.

Hall, P. and Martin, M.A. (1988e) On the error incurred using the bootstrap variance estimate when constructing confidence intervals for quantiles. *J. Roy. Statist. Soc. Ser. B*, submitted, September, 1988.

Hall, P., Martin, M.A., and Schucany, W.R. (1988) Better nonparametric bootstrap confidence intervals for the correlation coefficient. *J. Statist. Computa. Simul.*, submitted, September, 1988.

Abstract

In this thesis, we discuss the use of bootstrap methods for constructing confidence intervals in a wide variety of circumstances. In particular, we provide asymptotic theory for bootstrap confidence intervals based on Edgeworth expansions in several problems, and develop a general bootstrap resampling scheme applicable to a large range of statistical problems. The thesis is divided into five chapters. The first gives a brief introduction to bootstrap methods and Edgeworth expansion theory to provide a background and motivation for our later work.

The second chapter presents asymptotic theory for the bootstrap in two-sample problems. If the samples are of sizes m and n , then our results show that one- and two-sided percentile- t confidence intervals for a difference between two means have coverage error $O(m^{-1} + n^{-1})$, and that symmetric two-sided intervals have coverage error $O(m^{-2} + n^{-2})$. Furthermore, when the populations are Normal, coverage error of all percentile- t intervals drops to $O(m^{-2} + n^{-2})$. Our asymptotic analysis, and the results of a simulation study, indicate that percentile- t is a respectable solution to the Behrens-Fisher problem: that of testing for the difference between means of two Normal populations, without making assumptions about variances.

A central contribution of our thesis is the introduction in Chapter 3 of a general resampling principle which encompasses bootstrap resampling and bootstrap iteration. Our principle is applicable to a wide range of statistical problems, such as the construction of confidence intervals and hypothesis tests, and bias reduction for point estimation. The principle is illustrated through several examples, including bias correction and L^1 shrinkage.

We extend our study of bootstrap iteration in Chapter 4, where we focus on iteration to correct for coverage error in confidence intervals. It is shown that, in general, bootstrap iteration reduces the order of coverage error by the factor $n^{-\frac{1}{2}}$ in the case of one-sided intervals, and by the factor n^{-1} in two-

sided intervals. We also investigate the effect of bootstrap coverage correction on position of critical point and mean interval length, and illustrate the application of bootstrap coverage correction to Bartlett-corrected likelihood-ratio tests and confidence intervals for correlation coefficients.

We treat in Chapter 5 the problem of using bootstrap methods to construct confidence intervals for quantiles. It is shown that standard bootstrap methods perform poorly in this problem, suffering from poor coverage levels. We then study the performance of Normal-theory confidence intervals for quantiles constructed using the bootstrap estimate of sample quantile variance to Studentize. It turns out that these intervals have coverage error $O(n^{-\frac{1}{2}})$ in both one- and two-sided cases. This result is in marked contrast with their behaviour in more classical problems, where coverage error is $O(n^{-\frac{1}{2}})$ in the one-sided case, but drops to $O(n^{-1})$ in the two-sided case, and arises from a curious feature of the Edgeworth expansion of the distribution of the bootstrap-Studentized quantile, in that its leading term, of order $n^{-\frac{1}{2}}$, is proportional to a polynomial containing both odd and even powers of the argument. We use the results of our asymptotic analysis, and a simulation study, to compare the performances of bootstrap-Studentized intervals and Normal-theory intervals constructed using the Siddiqui-Bloch-Gastwirth variance estimate.

Contents

Title page	
Preface	
Acknowledgements	
Related publications	
Abstract	
Contents	

Chapter One

The Bootstrap, Edgeworth Expansions and Confidence Intervals 1

1. Introduction	1
2. The bootstrap and confidence intervals	3
3. Principles of Edgeworth expansion	6
3.1. <i>Basic results</i>	6
3.2. <i>Sums of independent random variables</i>	8
3.3. <i>More general statistics</i>	12
3.4. <i>A model for valid Edgeworth expansions</i>	16
3.5. <i>Cornish-Fisher expansions</i>	20
3.6. <i>Bibliographical notes</i>	21
4. Summary of thesis	23

Chapter Two

On Bootstrap Confidence Intervals and Two-sample Problems 25

1. Introduction	25
2. Construction of bootstrap confidence intervals	26
3. Cornish-Fisher expansions of bootstrap critical points	28
4. Second-order correctness	31
5. Accelerated bias correction	32
6. Lengths of two-sided, equal-tailed intervals	33
7. Coverage error	34
8. Symmetric confidence intervals	36
9. Short confidence intervals	37
10. Simulation study	40

11. Proofs	41
Appendix A2.1: <i>Proof of Lemma 2.3</i>	49
Appendix A2.2: <i>Derivation of constants u and v in the percentile-t case</i>	53
Appendix A2.3: <i>Derivation of the polynomial q_{31} in the Cornish-Fisher expansion of percentile-t critical point</i>	55

Chapter Three

Bootstrap Resampling and Iteration 56

1. Introduction	56
2. The principle	57
2.1. <i>General formulation of statistical problems</i>	57
2.2. <i>Resampling solutions to statistical problems</i>	59
3. Iterating the principle	60
4. Bias correction	65
5. Lehmann-type confidence intervals	66
6. L^1 -shrinkage	71
7. Proofs	71
7.1. <i>Proof of Theorem 3.1</i>	71
7.2. <i>Proof of Lemma 3.2</i>	76
7.3. <i>Asymptotics for Lehmann-type intervals: proof of Theorem 3.3</i>	77
7.4. <i>Asymptotics for L^1-shrinkage</i>	80
Appendix A3.1: <i>Proof of Lemma A3.1</i>	82

Chapter Four

On Bootstrap Iteration for Coverage Correction in Confidence Intervals 83

1. Introduction, definitions and notation	83
1.1. <i>Introduction</i>	83
1.2. <i>Parametric and nonparametric bootstraps</i>	84
1.3. <i>Definitions and notation</i>	85
2. Effect of bootstrap iteration on coverage error	86

2.1. <i>The algorithm</i>	87
2.2. <i>One-sided intervals</i>	87
2.3. <i>Two-sided intervals</i>	88
2.4. <i>Coverage correction and the parametric bootstrap</i>	90
3. Effect of bootstrap coverage correction on critical point	91
3.1. <i>One-sided intervals</i>	91
3.2. <i>Two-sided intervals</i>	98
4. Effect of bootstrap coverage correction on interval length	101
5. Simulation study	104
6. Applications of coverage correction	104
6.1. <i>Bootstrap coverage correction for Bartlett-corrected likelihood-ratio tests</i>	105
6.2. <i>Better nonparametric bootstrap confidence intervals for the correlation coefficient</i>	107
7. Proofs	116
7.1. <i>Proof of Theorem 4.1</i>	116
7.2. <i>Proof of Theorem 4.2</i>	118
7.3. <i>Proof of Theorem 4.3</i>	123
7.4. <i>Proof of Theorem 4.4</i>	125
7.5. <i>Proof of Theorem 4.5</i>	126
7.6. <i>Proof of Theorem 4.6</i>	127
7.7. <i>Asymptotics for Bartlett correction: proof that terms of order $n^{-5/2}$ vanish in (4.49)</i>	127
Appendix A4.1: Proof of Lemma A4.1	134

Chapter Five

Bootstrap Methods for Constructing Confidence Intervals for Quantiles

136

1. Introduction	136
2. On the accuracy of bootstrap percentile method confidence intervals for a quantile	138
3. The exact convergence rate of the bootstrap quantile variance estimator	141
4. The error incurred using the bootstrap variance estimator when constructing confidence intervals for quantiles	143

5. On using the jackknife to estimate quantile variance	148
6. Proofs	150
6.1. <i>Proof of Theorem 5.1</i>	150
6.2. <i>Proof of Theorem 5.2</i>	151
6.3. <i>Proof of Theorem 5.3</i>	160
6.4. <i>Proof of Theorem 5.4</i>	171
6.5. <i>Proof of Theorem 5.5</i>	171
 Appendix A5.1: <i>Proof of (5.3)</i>	 173
 Bibliography	 174

"Well, I sort of made it up," said Pooh, . . . "it comes to me sometimes."

"Ah!", said Rabbit, who never let things come to him, but always went and fetched them.

A. A. Milne, *The House at Pooh Corner*, Chapter 5.

CHAPTER 1

THE BOOTSTRAP, EDGEWORTH EXPANSIONS AND CONFIDENCE INTERVALS

1. INTRODUCTION

In recent years, vast improvements in the speed and cost of computation have resulted in the development of a multitude of new statistical theories and techniques which in the past would have been dismissed as unthinkably slow and expensive to use. Even so, many of today's most commonly-used statistical methods were developed in the first few decades of the 20'th century, when most calculations had to be done laboriously by hand, or with the help of a mechanical desk calculator. As a result, these older methods often relied upon several simplifying assumptions which made them mathematically tractable.

Central to the development of many of these techniques is the assumption that the data conform to a Normal or Gaussian distribution. Indeed, the Normal distribution plays a dominating role in much of classical statistical theory. The reliance of these older methods on assumptions such as that of Normality places severe limitations on their wide applicability, and raises the serious question of what to do when the assumptions are not satisfied. Computer-intensive methods such as Efron's (1979a) bootstrap offer us freedom from these limiting assumptions by replacing standard model-based assumptions with an enormous amount of calculation. The bootstrap generates a model for the parent population directly from the data, and so reflects all the characteristics of the sample.

As a simple illustration of the way in which the assumption of Normality is often used, consider the problem of constructing a 95% confidence interval for the mean μ of an unknown population, using a sample of size n . This example is relevant to much of our later work. If we assume that the data come from a Normal distribution with known variance $\sigma^2 < \infty$, then, with probability exactly

0.95, the true mean lies within $1.96\sigma/\sqrt{n}$ of the sample mean, say $\bar{X}$. Of course, our exact confidence statement relies heavily on our assumptions of Normality and known variance. However, we may relax these assumptions and still make approximate confidence statements like the above. Suppose now that the variance σ^2 is unknown, and we estimate it by the sample variance, say $\hat{\sigma}^2$. Then, with probability approximately 0.95, the true mean lies within $1.96\hat{\sigma}/\sqrt{n}$ of the sample mean, the error in approximation arising from the difference between distribution functions of Normal and “Student’s” t with $n - 1$ degrees of freedom. We can also rid ourselves of the assumption that the data are Normal by using the Central Limit Theorem. Merely assuming that the data are independent and identically distributed observations from a distribution having finite variance, we have, with probability approximately 0.95 in large samples, that the true mean lies in the range $\bar{X} - 1.96\hat{\sigma}/\sqrt{n}$ to $\bar{X} + 1.96\hat{\sigma}/\sqrt{n}$. This statement is based on the fact that, as $n \rightarrow \infty$, the ratio $n^{\frac{1}{2}}(\bar{X} - \mu)/\hat{\sigma}$ converges in distribution to a Standard Normal $N(0, 1)$ distribution.

It might seem from our discussion in the last paragraph that we have solved our simple problem without too much trouble, but a number of questions remain to be answered. Firstly, confidence statements like the last are not exact, relying on a Normal approximation to the distribution of $n^{\frac{1}{2}}(\bar{X} - \mu)/\hat{\sigma}$. Two questions which immediately spring to mind are: how good is that Normal approximation, particularly in small samples, and can we do better by using a different method? If it is possible to do better, how close can we get to an exact solution? Another important issue is: what if, instead of constructing a confidence interval for a simple mean, we are interested in finding an interval for a more complex statistic, such as a correlation coefficient, or a quantile of an unknown distribution? In such cases, it may be difficult to find a transformation of the statistic of interest to one having an approximate Normal distribution, and even if such a transformation is available, it may not be useful for all but very large sample sizes.

Computer-intensive methods such as the bootstrap take advantage of high-speed computers to address questions such as these. In using the bootstrap, one can easily perform tens of thousands of arithmetic operations in the analysis of data sets containing a mere handful of data points. Such an idea would have seemed unbelievably wasteful to a statistician constrained to perform calculations by hand, albeit with the aid of a desk calculator! In Efron’s (1979b) words, the advent of fast, cheap computation allows us to “think the unthinkable” in analyzing data. In return for such intensive computation, we get freedom from

many of the standard assumptions which dominate classical statistics.

In this thesis, we are largely concerned with the problem of using the bootstrap to construct confidence intervals in a wide variety of circumstances. In particular, our primary aim is to construct confidence intervals whose coverage probability (that is, the probability that the true parameter lies within the interval) is close to some stated level. We provide theory, based on Edgeworth expansion, which enables us to compare theoretically the asymptotic performance of several different intervals for a parameter, and we reinforce our theory with several simulation studies which allow us to determine whether the results of our asymptotic analyses are reflected in practice. We also develop a general resampling principle which guides the operation of bootstrap resampling in a wide range of problems, and leads naturally to iterated bootstrap schemes which can be used to improve various aspects of a solution to a statistical problem.

In Section 2 of this chapter we give a brief introduction to bootstrap theory and its use in constructing confidence intervals. Section 3 discusses the principles of Edgeworth expansion, a tool which is used throughout the thesis. Finally, in Section 4 we provide a summary of the topics studied in the remainder of this thesis.

2. THE BOOTSTRAP AND CONFIDENCE INTERVALS

The idea behind the bootstrap is a simple one: if you wish to estimate a functional of a population distribution function F , for instance a population mean

$$\mu \equiv \int x \, dF(x),$$

one way to proceed is possibly to use the same functional of the sample distribution function $\hat{F}$, in our example equal to the sample mean

$$\bar{X} \equiv \int x \, d\hat{F}(x).$$

Such an argument does not work in all cases; estimation of a probability density function is a case in point. Nevertheless, the bootstrap principle is applicable in a wide range of problems, and has received much attention in the statistical literature.

Efron (1979a) first realized the merit and wide applicability of the bootstrap approach, and gave it its name. Although the use of similar methods was known prior to Efron's remarkable paper — in some senses, the idea is an old one — it

largely existed in descriptive and disparate form, rather than presented as a unified, general theory. A major feature of Efron's treatment of the bootstrap was its dependence on the computer. Efron pointed out that in many complex situations, where bootstrap quantities are awkward (if not impossible) to compute algebraically, they may be computed by using Monte Carlo methods to numerically estimate functionals of the empiric distribution function. More specifically, same-size samples (called "resamples") may be drawn with replacement from the original sample a large number of times, the value of the statistic of interest computed for each of the resamples, and the bootstrap quantity computed by averaging these values over all resamples. This approach works because, in the majority of applications of the bootstrap, bootstrap quantities may be expressed as expectations conditional on the sample. Our development in Chapter 3 of a general resampling principle takes advantage of this property.

References to work along lines similar to Efron's around the time of his seminal paper include Barnard (1963), Hope (1968) and Marriot (1979) who advocated the use of Monte Carlo methods for hypothesis testing; Hartigan (1969, 1971, 1975), who used resampled subsamples to construct point and interval estimates, and who stressed the connections with the jackknife of Quenouille (1949, 1956) and Tukey (1958); and Maritz and Jarrett (1978), who derived bootstrap estimates of the variance of the sample median. For some highly readable expositions of bootstrap theory for point and interval estimation, we refer the reader to Efron (1979a), Diaconis and Efron (1983), Efron and Gong (1983), Efron and Tibshirani (1986), and Efron's (1982) monograph. General accounts of bootstrap theory include Bickel and Freedman (1981) and Beran (1984b).

Hall (1988a) gives a detailed summary and comparison of bootstrap methods for constructing confidence intervals for a univariate parameter θ . He identifies five methods which are in common use: two so-called "percentile" methods, which he terms "backwards" and "hybrid" percentile methods, respectively; the "percentile-t" method; and Efron's bias-corrected and accelerated bias-corrected methods. References to the two percentile methods are numerous and include Efron (1979a, 1981a, 1982, 1987), and Efron and Tibshirani (1986), among others. The "percentile-t" method is discussed by Bickel and Freedman (1981), Hinkley and Wei (1984), Beran (1987a, 1987b, 1988), Hall (1987a, 1987b, 1988a) and Singh (1987). Bias-corrected and accelerated bias-corrected methods are introduced and discussed by Efron (1981a, 1987).

The five methods mentioned in the previous paragraph represent two major

schools of thought which dominate present approaches to the theory of bootstrap confidence intervals. The first goes right back to Efron's early development of the bootstrap. Efron's account has usually adhered closely to other resampling methods, especially the jackknife, in its early stages, and the percentile methods are central to his approach. The "backwards" percentile method is probably the most popular of the techniques amongst non-statisticians, and perhaps also with statisticians, and we shall often refer to this method as simply the percentile method. Hall (1988a) terms this method the backwards method because it is analogous to looking up the wrong statistical tables (that is, Normal rather than Student's t tables) backwards (that is, confusing upper and lower quantiles). He refers to the other percentile method as the hybrid method because it can be interpreted as looking up the wrong tables the right way around. Both "backwards" and "hybrid" intervals are called percentile-method intervals because they are based on the percentiles of the bootstrap distributions of $\hat{\theta}$ and $\hat{\theta} - \theta$, respectively, where $\hat{\theta}$ denotes our estimate of the parameter θ . See Hall (1988a) for more details. Efron pointed out as early as 1981 that the ordinary percentile methods can perform very poorly. He proposed (1981a, 1987) analytical corrections to percentile-method intervals based on transformation theory, to improve their performance, and he termed these intervals bias-corrected and accelerated bias-corrected intervals respectively.

The second major school of thought concerning bootstrap confidence intervals is directed at finding a unified approach to the problem which leads formally to the issues of pivotalness and bootstrap iteration, which do not emerge naturally from the percentile-method approach. The percentile- t approach falls into this category. Note that if $\hat{\theta}$ is our estimate of θ , based on a sample of size n , with asymptotic variance $n^{-1}\sigma^2$, and if $\hat{\sigma}^2$ is an estimate of σ^2 , then the statistic $n^{\frac{1}{2}}(\hat{\theta} - \theta)/\hat{\sigma}$ is asymptotically pivotal if σ is unknown (see for example, Hartigan, 1986). The percentile- t method is so called because it is based on the percentiles of the bootstrap distribution of the asymptotically pivotal "Studentized" statistic $n^{\frac{1}{2}}(\hat{\theta} - \theta)/\hat{\sigma}$. In the sense of Hall (1988a) the percentile- t method is equivalent to looking up the right tables the right way around. The importance of using a pivotal statistic when constructing a bootstrap confidence interval, at least in simple situations where the variance is easily estimated, is implicit in Hinkley and Wei (1984), and is stressed explicitly by Hall (1986a, 1988a), Hartigan (1986) and Beran (1987a). Our development in Chapter 3 of a general resampling principle and bootstrap iteration adheres closely to this second school of

thought. Beran (1987a) points out the close connection between his form of bootstrap iteration (which he calls “prepivotting”) and Studentizing to obtain an approximately pivotal statistic. Other accounts of bootstrap iteration are given by Hall (1986a), Loh (1987) and Hall and Martin (1988b).

Both of the philosophical standpoints that we have outlined have merit, although we favour the second over the first throughout this thesis. Our choice is by no means unequivocal, however, and is based on a number of philosophical preferences. Firstly, it seems to us more natural to base confidence intervals and hypothesis tests on pivotal rather than nonpivotal quantities. In Hall’s (1988a) terminology, we prefer to look up the right tables to start with, rather than look up the wrong tables and then correct for our error. The second is based on a view that computer-intensive methods such as the bootstrap, which are designed to avoid the need for tedious analytic corrections, should not have to appeal to such corrections. On the other hand, Efron’s (1981a, 1987) bias corrections do possess useful invariance properties under monotone transformations, not shared by percentile- t . It is easy to find examples where each of the percentile- t and accelerated bias-corrected methods fails. For example, it is well-known that the percentile- t method performs poorly in many cases where the variance σ^2 is difficult to estimate – see for example, Efron (1981a), p.154, and Hall, Martin and Schucany (1988) – while equal-tailed accelerated bias-corrected intervals for small samples and high nominal coverage probabilities can produce abnormally short intervals; compare Hall (1988a).

Further notation and details of how the various bootstrap confidence intervals are constructed are given as we need them. For an excellent and lucid exposition of the construction of and comparison between the various bootstrap confidence intervals for a univariate parameter θ , the reader is referred to Hall (1988a). In many places throughout this thesis we retain Hall’s (1988a) notation, and use several of his results in that paper.

3. THE PRINCIPLES OF EDGEWORTH EXPANSION

3.1. Basic results. In this section we give a brief account of Edgeworth expansions as approximations to distributions of estimates $\hat{\theta}$ of unknown parameters θ . A similar, more detailed account is given by Hall (1989). The basic result is that if $\hat{\theta}$ is constructed from a sample of size n , and if $n^{\frac{1}{2}}(\hat{\theta} - \theta)$ has a limiting Normal distribution with zero mean and variance σ^2 , then in many cases of interest we

may expand the distribution function of $n^{\frac{1}{2}}(\hat{\theta} - \theta)$ as a power series in $n^{-\frac{1}{2}}$:

$$P\{n^{\frac{1}{2}}(\hat{\theta} - \theta) \leq \sigma x\} = \Phi(x) + n^{-\frac{1}{2}}p_1(x)\phi(x) + \cdots + n^{-j/2}p_j(x)\phi(x) + \cdots, \quad (1.1)$$

where $\phi(x) \equiv (2\pi)^{-\frac{1}{2}}e^{-\frac{1}{2}x^2}$ is the Standard Normal density function and $\Phi(x) \equiv \int_{-\infty}^x \phi(y) dy$ is the Standard Normal distribution function. The functions p_j are polynomials whose coefficients depend on the cumulants of $\hat{\theta} - \theta$. We call an expansion such as (1.1) an Edgeworth expansion.

Equation (1.1) may be inverted to show that the solution $x = v_\alpha$ of the equation $P\{n^{\frac{1}{2}}(\hat{\theta} - \theta) \leq \sigma x\} = \alpha$ admits an expansion

$$v_\alpha = z_\alpha + n^{-\frac{1}{2}}p_{11}(z_\alpha) + n^{-1}p_{21}(z_\alpha) + \cdots + n^{-j/2}p_{j1}(z_\alpha) + \cdots, \quad (1.2)$$

where $0 < \alpha < 1$, the p_{j1} 's are polynomials depending on the p_j 's, and z_α is the solution of the equation $\Phi(z_\alpha) = \alpha$. We term inverse expansions like (1.2) Cornish-Fisher expansions.

So far, our discussion has focussed on the distribution of the non-Studentized statistic $n^{\frac{1}{2}}(\hat{\theta} - \theta)/\sigma$. However, in most cases of interest the asymptotic variance σ^2 is unknown, and is estimated by a function $\hat{\sigma}^2$ of the data. In that circumstance, we are interested in the distribution function of the Studentized statistic $n^{\frac{1}{2}}(\hat{\theta} - \theta)/\hat{\sigma}$, which admits an expansion like (1.1) but with different polynomials, which we call q_j to distinguish them from the polynomials p_j in the non-Studentized case. We adopt this notation throughout the thesis.

An important example of the types of results which we have discussed is that where $\hat{\theta}$ is a sample mean, $\hat{\sigma}^2$ a sample variance, and θ and σ^2 the corresponding population mean and variance, respectively. This example displays all of the major features of more general Edgeworth expansions: p_j (q_j) is a polynomial of degree at most $3j - 1$, and is an odd or even function according as j is even or odd, respectively; and expansion (1.1) is valid under appropriate moment and smoothness conditions on the underlying distribution. This example forms the basis for a discussion of more general Edgeworth expansions, and we study it in more detail in Subsection 3.2. We expand our treatment to more general expansions in Subsection 3.3 and place our arguments on a more formal level. More detail is given in Subsection 3.4 in the context of a general model under which formal Edgeworth expansions may be established rigorously, and we discuss inverse (Cornish-Fisher) expansions in Subsection 3.5.

Throughout the thesis, we are interested in Edgeworth expansions primarily for what they can tell us about asymptotic theory for the bootstrap. The most

common use for Edgeworth expansions in classical statistics is to make analytical corrections rather similar to those which the bootstrap makes by numerical means. Such applications of Edgeworth expansions are outside the scope of this thesis, and little comment is made on them. Although the theory discussed in this section is quite general, it is not universal, and there do exist important cases where our key principles are violated. A case in point is that of the Studentized quantile, which is treated separately, in Chapter 5.

3.2. Sums of independent random variables. Let $X, X_1, X_2, \dots$ denote independent and identically distributed random variables with mean $\theta \equiv \mu$ and finite variance σ^2 . One estimate of θ is the sample mean,

$$\hat{\theta} \equiv n^{-1} \sum_{i=1}^n X_i,$$

which has variance $n^{-1}\sigma^2$. Now, $S_n \equiv n^{\frac{1}{2}}(\hat{\theta} - \theta)/\sigma$ is asymptotically Normally distributed with zero mean and unit variance. Classical procedures, such as that described in Section 1 for the construction of a Normal-theory confidence interval, often use this approximation directly in inference about the mean. We are interested in the accuracy of this Normal approximation, for it determines, in the example of constructing a confidence interval for a mean, the coverage error of the interval.

We may approach the problem of describing errors in Normal approximations by using characteristic functions. Since S_n is asymptotically Standard Normal $N(0, 1)$ then its characteristic function, ψ_n , converges to that of a Standard Normal random variable, as $n \rightarrow \infty$; that is

$$\psi_n(t) \equiv E\{\exp(itS_n)\} \rightarrow e^{-\frac{1}{2}t^2}, \quad -\infty < t < \infty. \quad (1.3)$$

Now,

$$\psi_n(t) = \{\psi(t/n^{\frac{1}{2}})\}^n, \quad (1.4)$$

where ψ denotes the characteristic function of $Y \equiv (X - \mu)/\sigma$. The j 'th cumulant κ_j , of Y equals the coefficient of $(j!)^{-1}(it)^j$ in a power series expansion of $\log \psi(t)$; that is,

$$\psi(t) = \exp\{\kappa_1(it) + \frac{1}{2}\kappa_2(it)^2 + \dots + \frac{1}{j!}\kappa_j(it)^j + \dots\}. \quad (1.5)$$

However, since

$$\psi(t) = 1 + E(Y)it + \frac{1}{2}E(Y^2)(it)^2 + \dots + \frac{1}{j!}E(Y^j)(it)^j + \dots,$$

the cumulants of Y may be defined in terms of the moments of Y :

$$\begin{aligned}\kappa_1 &= E(Y), & \kappa_2 &= E(Y^2) - (EY)^2 = \text{var}(Y), \\ \kappa_3 &= E(Y^3) - 3E(Y^2)E(Y) + 2(EY)^3 = E(Y - EY)^3, \\ \kappa_4 &= E(Y^4) - 4E(Y^3)E(Y) - 3(EY^2)^2 + 12E(Y^2)(EY)^2 - 6(EY)^4 \\ &= E(Y - EY)^4 - 3(\text{var } Y)^2,\end{aligned}$$

and so on; see Kendall and Stuart (1977), p.72, equation (3.43), for formulae up to κ_{10} . Noting that $E(Y) = \kappa_1 = 0$ and $\text{var}(Y) = \kappa_2 = 1$, we have, by (1.4) and (1.5), that

$$\begin{aligned}\psi_n(t) &= \exp\left\{-\frac{1}{2}t^2 + n^{-\frac{1}{2}}\frac{1}{3!}\kappa_3(it)^3 + \cdots + n^{-(j-2)/2}\frac{1}{j!}\kappa_j(it)^j + \cdots\right\} \\ &= e^{-\frac{1}{2}t^2} \left[1 + n^{-\frac{1}{2}}\frac{1}{6}\kappa_3(it)^3 + n^{-1}\left\{\frac{1}{24}\kappa_4(it)^4 + \frac{1}{72}\kappa_3^2(it)^6\right\} + \cdots\right] \quad (1.6) \\ &= e^{-\frac{1}{2}t^2} \{1 + n^{-\frac{1}{2}}r_1(it) + n^{-1}r_2(it) + \cdots + n^{-j/2}r_j(it) + \cdots\}, \quad (1.7)\end{aligned}$$

where r_j is a polynomial of degree no more than $3j$ with real coefficients depending on the cumulants of Y up to the $(j+2)$ 'th, but not on n . Observe that r_j is an even function when j is even, and an odd function when j is odd. This fact is apparent from inspection of the argument leading to (1.7), and will prove vital to our future work. It turns out that for our purposes we need at most the first three polynomials r_j in (1.7).

We now discuss inversion of the characteristic function ψ_n . If we rewrite (1.7) in the form

$$\psi_n(t) = e^{-\frac{1}{2}t^2} + n^{-\frac{1}{2}}r_1(it)e^{-\frac{1}{2}t^2} + n^{-1}r_2(it)e^{-\frac{1}{2}t^2} + \cdots + n^{-j/2}r_j(it)e^{-\frac{1}{2}t^2} + \cdots, \quad (1.8)$$

then since

$$\psi_n(t) \equiv \int_{-\infty}^{\infty} e^{itx} dP(S_n \leq x),$$

and $e^{-\frac{1}{2}t^2} = \int_{-\infty}^{\infty} e^{itx} d\Phi(x)$, expansion (1.8) suggests the inverse expansion

$$\begin{aligned}P(S_n \leq x) &= \Phi(x) + n^{-\frac{1}{2}}R_1(x) + n^{-1}R_2(x) + \cdots \\ &\quad + n^{-j/2}R_j(x) + \cdots, \quad (1.9)\end{aligned}$$

where $R_j(x)$ denotes a function whose Fourier-Stieltjes transform equals $r_j(it)e^{-\frac{1}{2}t^2}$; that is,

$$\int_{-\infty}^{\infty} e^{itx} dR_j(x) = r_j(it)e^{-\frac{1}{2}t^2}.$$

We next compute the R_j 's. Integrating by parts repeatedly in $\int_{-\infty}^{\infty} e^{itx} d\Phi(x)$, we have

$$\begin{aligned} e^{-\frac{1}{2}t^2} &\equiv \int_{-\infty}^{\infty} e^{itx} d\Phi(x) \\ &= (-it)^{-1} \int_{-\infty}^{\infty} e^{itx} d\Phi^{(1)}(x) = \dots = (-it)^{-j} \int_{-\infty}^{\infty} e^{itx} d\Phi^{(j)}(x), \end{aligned}$$

from which it follows that

$$\int_{-\infty}^{\infty} e^{itx} d\{(-D)^j \Phi(x)\} = (it)^j e^{-\frac{1}{2}t^2}, \quad (1.10)$$

where D is the differential operator d/dx . If we interpret $r_j(-D)$ in the obvious way as a polynomial in D , then $r_j(-D)$ is a differential operator and, by (1.10),

$$\int_{-\infty}^{\infty} e^{itx} d\{r_j(-D)\Phi(x)\} = r_j(it) e^{-\frac{1}{2}t^2}. \quad (1.11)$$

Consequently, from (1.11), the functions R_j are given by

$$R_j(x) = r_j(-d/dx)\Phi(x). \quad (1.12)$$

Now observe that for $j \geq 1$, $(-D)^j \Phi(x) = -He_{j-1}(x)\phi(x)$, where the functions He_j are the Hermite polynomials:

$$He_0(x) \equiv 1, \quad He_1(x) \equiv x, \quad He_2(x) \equiv x^2 - 1, \quad He_3(x) \equiv x(x^2 - 3),$$

$$He_4(x) \equiv x^4 - 6x^2 + 3, \quad He_5(x) \equiv x(x^4 - 10x^2 + 15),$$

and so on; see, for example, Petrov (1975), p.137. They are orthogonal with respect to the weight function ϕ , and are normalized so that the leading coefficient is unity. Note that He_j is of degree j and is an even or odd function according as j is even or odd. As a result, for general $j \geq 1$, $R_j(x) \equiv p_j(x)\phi(x)$, where p_j is of degree no more than $3j - 1$ and is odd for even j and even for odd j , since r_j is of degree no more than $3j$ and is even/odd for even/odd j . In particular, observe from (1.6) that

$$p_1(x) \equiv -\frac{1}{6}\kappa_3(x^2-1), \quad p_2(x) \equiv -x\left\{\frac{1}{24}\kappa_4(x^2-3) + \frac{1}{72}\kappa_3^2(x^4-10x^2+15)\right\}. \quad (1.13)$$

Formula (1.9) may now be written

$$P(S_n \leq x) = \Phi(x) + n^{-\frac{1}{2}}p_1(x)\phi(x) + \dots + n^{-j/2}p_j(x)\phi(x) + \dots, \quad (1.14)$$

and is called an Edgeworth expansion of the distribution function of S_n . Third and fourth cumulants, κ_3 and κ_4 , are known as skewness and kurtosis respectively. The term of order $n^{-\frac{1}{2}}$ in (1.14) is a correction of the basic Normal approximation for the first order effect of skewness, while the term of order n^{-1} corrects for the first order effect of kurtosis and the second order effect of skewness.

The expansion (1.14) only rarely converges as an infinite series. Cramér (1928) showed that if X is absolutely continuous, then a necessary and sufficient condition for convergence is that $E\{\exp(\frac{1}{4}Y^2)\} < \infty$, where $Y \equiv (X - \mu)/\sigma$. This condition places severe limitations on the tails of the distribution of Y , failing even if X is exponentially distributed. Expansion (1.14) is usually interpreted as an *asymptotic expansion*, meaning that if the series is terminated after a given number of terms then the remainder is of smaller order than the last included term; that is,

$$P(S_n \leq x) = \Phi(x) + n^{-\frac{1}{2}}p_1(x)\phi(x) + \cdots + n^{-j/2}p_j(x)\phi(x) + o(n^{-j/2}). \quad (1.15)$$

Bhattacharya and Rao (1976), Theorem 20.1, p.208 prove that sufficient regularity conditions for (1.15) to hold, with the remainder of the stated order uniformly in all x , are $E(|X|^{j+2}) < \infty$ and

$$\limsup_{|t| \rightarrow \infty} |\psi(t)| < 1.$$

The second restriction, often called Cramér's condition, holds if the distribution of X has a nondegenerate absolutely continuous component – in particular, if X has a proper density function. As a consequence, our formal inversion of the characteristic function expansion (1.8) is valid if X has sufficiently many finite moments, and a smooth distribution. It fails if the distribution of X is not smooth, for example if X has a lattice distribution. There, extra terms of all orders $n^{-j/2}$, $j \geq 1$, must be added to (1.14) to allow for the fact that we are approximating an unsmooth function, $P(S_n \leq x)$, by a smooth function, the right-hand side of (1.14). These terms are discontinuous functions of x , having jumps at integral values of x . This was first proved by Esseen (1945) in his remarkable thesis. Esseen (1945) discussed expansions for lattice distributions, although there is a minor error in the first order term he gave. Gnedenko and Kolmogorov (1945), Section 43, correctly give that term.

3.3. More general statistics. Let S_n denote a statistic having an asymptotic Standard Normal distribution, for example, $S_n \equiv (\hat{\theta} - \theta)/\hat{s}$, where $\hat{s}$ is a consistent estimate of the asymptotic variance of $\hat{\theta}$. If we denote the j 'th cumulant of S_n by $\kappa_{j,n}$, the notation indicating its dependence on n , and if ψ_n denotes the characteristic function of S_n , then we have

$$\begin{aligned}\psi_n(t) &\equiv E\{\exp(itS_n)\} \\ &= \exp\{\kappa_{1,n}it + \frac{1}{2}\kappa_{2,n}(it)^2 + \cdots + \frac{1}{j!}\kappa_{j,n}(it)^j + \cdots\}.\end{aligned}\quad (1.16)$$

In many cases of practical interest, $\kappa_{j,n}$ is of order $n^{-(j-2)/2}$ and may be expanded as a power series in n^{-1} :

$$\kappa_{j,n} = n^{-(j-2)/2}(k_{j,1} + n^{-1}k_{j,2} + \cdots), \quad j \geq 1, \quad (1.17)$$

where $k_{1,1} = 0$ and $k_{2,1} = 1$, these relations reflecting the fact that S_n has asymptotically zero mean and unit variance; see James (1955, 1958) and James and Mayne (1962). In the simpler case where S_n is a normalized sum of independent and identically distributed random variables, formula (1.17) is implicit in the first identity of (1.6), where $\kappa_{j,n} = n^{-(j-2)/2}\kappa_j$, for $j \geq 2$, where κ_j is the j 'th cumulant of Y . Combining (1.16) and (1.17) we see that

$$\begin{aligned}\psi_n(t) &= \exp\left[-\frac{1}{2}t^2 + n^{-\frac{1}{2}}\{k_{1,2}it + \frac{1}{3!}k_{3,1}(it)^3\} \right. \\ &\quad \left. + n^{-1}\{\frac{1}{2}k_{2,2}(it)^2 + \frac{1}{4!}k_{4,1}(it)^4\} + \cdots\right] \\ &= e^{-\frac{1}{2}t^2} \left(1 + n^{-\frac{1}{2}}\{k_{1,2}it + \frac{1}{6}k_{3,1}(it)^3\} \right. \\ &\quad \left. + n^{-1}\left[\frac{1}{2}k_{2,2}(it)^2 + \frac{1}{24}k_{4,1}(it)^4 + \frac{1}{2}\{k_{1,2}it + \frac{1}{6}k_{3,1}(it)^3\}^2\right] \right. \\ &\quad \left. + O(n^{-3/2})\right).\end{aligned}\quad (1.18)$$

More generally,

$$\psi_n(t) = e^{-\frac{1}{2}t^2} \left(1 + n^{-\frac{1}{2}}r_1(it) + n^{-1}r_2(it) + \cdots + n^{-j/2}r_j(it) + \cdots\right),$$

where r_j is a polynomial of degree no more than $3j$, and is an even or odd function according as j is even or odd. This expansion is formally identical to (1.8), which was derived in the special case where S_n is a sum of independent random variables, and may be inverted to obtain an analogue of (1.14):

$$P(S_n \leq x) = \Phi(x) + n^{-\frac{1}{2}}p_1(x)\phi(x) + \cdots + n^{-j/2}p_j(x)\phi(x) + \cdots, \quad (1.19)$$

where p_j is a polynomial of degree no more than $3j - 1$, and is odd/even for even/odd j .

Recall from Subsection 3.2 that

$$\int_{-\infty}^{\infty} e^{itx} d\{-He_{j-1}(x)\phi(x)\} = (it)^j e^{-\frac{1}{2}t^2}, \quad j \geq 1,$$

where He_j is the j 'th Hermite polynomial. From this fact, and inverting (1.18), we get

$$p_1(x) \equiv -\{k_{1,2} + \frac{1}{6}k_{3,1}He_2(x)\} = -\{k_{1,2} + \frac{1}{6}k_{3,1}(x^2 - 1)\}, \quad (1.20)$$

and

$$\begin{aligned} p_2(x) &\equiv -\{\frac{1}{2}(k_{2,2} + k_{1,2}^2)He_1(x) + \frac{1}{24}(k_{4,1} + 4k_{1,2}k_{3,1})He_3(x) \\ &\quad + \frac{1}{72}k_{3,1}^2He_5(x)\} \\ &= -x\{\frac{1}{2}(k_{2,2} + k_{1,2}^2) + \frac{1}{24}(k_{4,1} + 4k_{1,2}k_{3,1})(x^2 - 3) \\ &\quad + \frac{1}{72}k_{3,1}^2(x^4 - 10x^2 + 15)\}. \end{aligned} \quad (1.21)$$

We interpret expansion (1.19) as an asymptotic series; that is,

$$P(S_n \leq x) = \Phi(x) + n^{-\frac{1}{2}}p_1(x)\phi(x) + \cdots + n^{-j/2}p_j(x)\phi(x) + o(n^{-j/2}),$$

for $j \geq 1$, uniformly in x under appropriate regularity conditions on X . We discuss such regularity conditions in more detail in the next subsection.

Next we illustrate the application of Edgeworth expansions to describing coverage accuracy of confidence intervals for a parameter θ . Suppose that we define $S_n \equiv n^{\frac{1}{2}}(\hat{\theta} - \theta)/\sigma$, where $n^{-1}\sigma^2$ is the asymptotic variance of $\hat{\theta}$. Then one- and two-sided nominal $(1 - \alpha)$ -level Normal-theory confidence intervals for θ are respectively

$$I_1 \equiv (-\infty, \hat{\theta} + n^{-\frac{1}{2}}\sigma z_{1-\alpha}) \quad \text{and} \quad I_2 \equiv (\hat{\theta} - n^{-\frac{1}{2}}\sigma z_{1-(\alpha/2)}, \hat{\theta} + n^{-\frac{1}{2}}\sigma z_{1-(\alpha/2)}),$$

where z_γ denotes the solution of the equation $\Phi(z_\gamma) = \gamma$, for $0 < \gamma < 1$. Bearing in mind that even/odd indexed p_j 's are odd/even functions respectively, we see from expansion (1.19) that the actual coverage probabilities of these intervals are

$$\begin{aligned} P(\theta \in I_1) &\equiv P(\theta < \hat{\theta} + n^{-\frac{1}{2}}\sigma z_{1-\alpha}) = P(S_n > -z_{1-\alpha}) \\ &= 1 - \{\Phi(-z_{1-\alpha}) + n^{-\frac{1}{2}}p_1(-z_{1-\alpha})\phi(-z_{1-\alpha}) + O(n^{-1})\} \\ &= 1 - \alpha - n^{-\frac{1}{2}}p_1(z_{1-\alpha})\phi(z_{1-\alpha}) + O(n^{-1}), \end{aligned} \quad (1.22)$$

and

$$\begin{aligned}
P(\theta \in I_2) &\equiv P(S_n > -z_{1-(\alpha/2)}) - P(S_n > z_{1-(\alpha/2)}) \\
&= P(S_n \leq z_{1-(\alpha/2)}) - P(S_n \leq -z_{1-(\alpha/2)}) \\
&= \Phi(z_{1-(\alpha/2)}) - \Phi(-z_{1-(\alpha/2)}) \\
&\quad + n^{-\frac{1}{2}} \{p_1(z_{1-(\alpha/2)})\phi(z_{1-(\alpha/2)}) - p_1(-z_{1-(\alpha/2)})\phi(-z_{1-(\alpha/2)})\} \\
&\quad + n^{-1} \{p_2(z_{1-(\alpha/2)})\phi(z_{1-(\alpha/2)}) - p_2(-z_{1-(\alpha/2)})\phi(-z_{1-(\alpha/2)})\} \\
&\quad + n^{-3/2} \{p_3(z_{1-(\alpha/2)})\phi(z_{1-(\alpha/2)}) - p_3(-z_{1-(\alpha/2)})\phi(-z_{1-(\alpha/2)})\} \\
&\quad + O(n^{-2}) \\
&= 1 - \alpha + 2n^{-1}p_2(z_{1-(\alpha/2)})\phi(z_{1-(\alpha/2)}) + O(n^{-2}). \tag{1.23}
\end{aligned}$$

In particular, one-sided Normal-theory confidence intervals have coverage errors of order $n^{-\frac{1}{2}}$, whereas two-sided intervals have coverage errors of order n^{-1} . Using (1.19) and the argument which we employed to obtain (1.22) and (1.23), it follows that, more generally,

$$P(\theta \in I_1) = 1 - \alpha + \sum_{j \geq 1} n^{-j/2} (-1)^j p_j(z_{1-\alpha}) \phi(z_{1-\alpha}), \tag{1.24}$$

and

$$P(\theta \in I_2) = 1 - \alpha + 2 \sum_{j \geq 1} n^{-j} p_{2j}(z_{1-(\alpha/2)}) \phi(z_{1-(\alpha/2)}), \tag{1.25}$$

where, as usual, we interpret these expansions as asymptotic series. Therefore, as a result of the crucial property that even/odd indexed p_j 's are odd/even functions, respectively, we see that the coverage probability of one-sided intervals may be expressed as a power series in $n^{-\frac{1}{2}}$, while the coverage probability of two-sided intervals may be written as a series in n^{-1} . This fact will prove critical in Chapters 3 and 4 when we discuss bootstrap iteration, for it means that each iteration to reduce coverage error reduces the order of error by a factor of $n^{-\frac{1}{2}}$ for one-sided intervals, but by a factor of n^{-1} for two-sided intervals.

Thus far, our discussion has centred on the non-Studentized statistic $n^{\frac{1}{2}}(\hat{\theta} - \theta)/\sigma$. If σ is unknown, and if we estimate it by $\hat{\sigma}$, say, then it makes sense to work with the Studentized statistic $S_n \equiv n^{\frac{1}{2}}(\hat{\theta} - \theta)/\hat{\sigma}$, instead of $n^{\frac{1}{2}}(\hat{\theta} - \theta)/\sigma$. The distribution function of the new S_n admits an expansion of the same form as (1.19), but with different polynomials which we denote by q_j . Hence,

$$P\{n^{\frac{1}{2}}(\hat{\theta} - \theta)/\hat{\sigma} \leq x\} = \Phi(x) + n^{-\frac{1}{2}}q_1(x)\phi(x) + \cdots + n^{-j/2}q_j(x)\phi(x) + \cdots,$$

where, as for the p_j 's, even/odd indexed q_j 's are odd/even functions respectively. Examples of one- and two-sided nominal $(1 - \alpha)$ -level Normal-theory confidence intervals for θ based on the asymptotic Normal distribution of S_n are

$$J_1 \equiv (-\infty, \hat{\theta} + n^{-\frac{1}{2}} \hat{\sigma} z_{1-\alpha}) \quad \text{and} \quad J_2 \equiv (\hat{\theta} - n^{-\frac{1}{2}} \hat{\sigma} z_{1-(\alpha/2)}, \hat{\theta} + n^{-\frac{1}{2}} \hat{\sigma} z_{1-(\alpha/2)}),$$

respectively. Their actual coverages are given by (1.22)–(1.25), except that p_j 's there are now replaced by q_j 's.

We may apply Edgeworth expansions to many other functionals of the distribution of $\hat{\theta}$. An example applicable to bias reduction, among other things, is the use of Edgeworth expansions to obtain expansions of moments. Note that if $r \geq 1$ is an integer and we define $S_n \equiv n^{\frac{1}{2}}(\hat{\theta} - \theta)/\sigma$, then

$$\begin{aligned} E\{(\hat{\theta} - \theta)^r\} &= (n^{-\frac{1}{2}}\sigma)^r E(S_n^r) \\ &= (n^{-\frac{1}{2}}\sigma)^r \left\{ \int_0^\infty x^r dP(S_n \leq x) - \int_0^\infty (-x)^r dP(S_n \leq -x) \right\} \\ &= (n^{-\frac{1}{2}}\sigma)^r r \int_0^\infty x^{r-1} \{P(S_n > x) + (-1)^r P(S_n \leq -x)\} dx. \end{aligned}$$

Now, when r is even,

$$P(S_n > x) + (-1)^r P(S_n \leq -x) = P(|N| > x) - 2 \sum_{j \geq 1} n^{-j} p_{2j}(x) \phi(x),$$

and when r is odd,

$$P(S_n > x) + (-1)^r P(S_n \leq -x) = -2n^{-\frac{1}{2}} \sum_{j \geq 1} n^{-j} p_{2j-1}(x) \phi(x),$$

where N denotes a Standard Normal $N(0, 1)$ random variable. Consequently,

$$E\{(\hat{\theta} - \theta)^r\} = \begin{cases} (n^{-\frac{1}{2}}\sigma)^r \{E(N^r) - 2r \sum_{j \geq 1} n^{-j} \int_0^\infty x^{r-1} p_{2j}(x) \phi(x) dx\}, & r \text{ even,} \\ -2(n^{-\frac{1}{2}}\sigma)^r r n^{\frac{1}{2}} \sum_{j \geq 1} n^{-j} \int_0^\infty x^{r-1} p_{2j-1}(x) \phi(x) dx, & r \text{ odd,} \end{cases} \quad (1.26)$$

where the infinite series are to be thought of as asymptotic expansions. We may write (1.26) more concisely as

$$E\{(\hat{\theta} - \theta)^r\} = n^{-[(r+1)/2]} (c_1 + n^{-1} c_2 + n^{-2} c_3 + \dots),$$

where $[\cdot]$ denotes the integer part function, and the constants $c_1, c_2, \dots$, depending on r , are derivable from polynomials in the Edgeworth expansion of the

distribution of S_n . The fact that these expansions of moments are series in n^{-1} explains why bootstrap iteration to correct for bias reduces the order of bias by a factor of n^{-1} rather than $n^{-\frac{1}{2}}$.

A rigorous proof of expansions such as (1.26) requires a nonuniform estimate of the form

$$\sup_{-\infty < x < \infty} (1 + |x|)^l \left| P(S_n \leq x) - \left\{ \Phi(x) + \sum_{k=1}^j n^{-k/2} p_k(x) \phi(x) \right\} \right| = O(n^{-(j+1)/2}),$$

for large l , of the difference between $P(S_n \leq x)$ and its Edgeworth expansion up to the j 'th term. Although such calculations may be carried out, expansions like (1.26) are often available more generally than are the Edgeworth expansions from which they are derived. For example, if θ and $\hat{\theta}$ denote population and sample means respectively, and if the r 'th order population moments are finite then (1.26) holds without further assumptions on the population, such as Cramér's condition. For this reason, and to make our exposition clearer, throughout this thesis we shall work only formally with Edgeworth expansions without giving rigorous proofs.

3.4. A model for valid Edgeworth expansions. In this subsection we provide a model under which the Edgeworth expansions discussed in the previous subsection are valid. Let $\mathbf{X}, \mathbf{X}_1, \mathbf{X}_2, \dots$ denote independent and identically distributed random d -vectors with mean μ and put $\bar{\mathbf{X}} = n^{-1} \sum_{1 \leq i \leq n} \mathbf{X}_i$. Let $A : R^d \rightarrow R$ be a smooth function satisfying $A(\mu) = 0$. Typically we shall be looking at functions such as $A(\mathbf{x}) \equiv \{g(\mathbf{x}) - g(\mu)\}/\tau(\mu)$ or $A(\mathbf{x}) \equiv \{g(\mathbf{x}) - g(\mu)\}/\tau(\mathbf{x})$, where $\theta \equiv g(\mu)$ is the parameter estimated by $\hat{\theta} \equiv g(\bar{\mathbf{X}})$, and where $\tau(\mu)^2$ is the asymptotic variance of $n^{\frac{1}{2}}\hat{\theta}$ and $\tau(\bar{\mathbf{X}})$ is an estimate of $\tau(\mu)$. This model is applicable to many problems of practical interest including those where θ is a mean, or a variance, or a ratio of means or variances, or a difference between means or variances, or a correlation coefficient, etc., in all of which θ has the form $g(\mu)$ for a smooth function g and a mean vector μ .

Denote the i 'th component of a d -vector $\mathbf{v}$ by $v^{(i)}$, and put $\mathbf{Z} \equiv n^{\frac{1}{2}}(\bar{\mathbf{X}} - \mu)$. We write

$$a_{i_1 \dots i_j} \equiv \left(\partial^j / \partial x^{(i_1)} \dots \partial x^{(i_j)} \right) A(\mathbf{x}) \Big|_{\mathbf{x}=\mu}.$$

Since $A(\mu) = 0$ and $\mathbf{Z} = O_p(1)$, we have, by Taylor expansion,

$$S_n \equiv n^{\frac{1}{2}} A(\bar{\mathbf{X}}) = S_{nr} + O_p(n^{-r/2}), \quad r \geq 1, \quad (1.27)$$

where

$$S_{nr} \equiv \sum_{i=1}^d a_i Z^{(i)} + n^{-\frac{1}{2}} \sum_{i_1=1}^d \sum_{i_2=1}^d a_{i_1 i_2} Z^{(i_1)} Z^{(i_2)} + \dots \\ + n^{-(r-1)/2} (r!)^{-1} \sum_{i_1=1}^d \dots \sum_{i_r=1}^d a_{i_1 \dots i_r} Z^{(i_1)} \dots Z^{(i_r)}. \quad (1.28)$$

The following theorem shows that the cumulants of S_{nr} admit expansion (1.17).

THEOREM 1.1. (Hall (1989), Theorem 2.1.) Let $\mathbf{Z} \equiv n^{\frac{1}{2}}(\bar{\mathbf{X}} - \mu)$, where $\bar{\mathbf{X}}$ denotes the mean of n independent random vectors distributed as $\mathbf{X}$ with finite $j \cdot r$ 'th moments and mean vector μ . Put

$$U_{nr} \equiv \sum_{i=1}^d b_i Z^{(i)} + n^{-\frac{1}{2}} \sum_{i_1=1}^d \sum_{i_2=1}^d b_{i_1 i_2} Z^{(i_1)} Z^{(i_2)} + \dots \\ + n^{-(r-1)/2} \sum_{i_1=1}^d \dots \sum_{i_r=1}^d b_{i_1 \dots i_r} Z^{(i_1)} \dots Z^{(i_r)},$$

for fixed arbitrary constants b . Then the j 'th cumulant of U_{nr} has the form

$$\kappa_{j,nr} = n^{-(j-2)/2} (k_{j,1} + n^{-1} k_{j,2} + n^{-2} k_{j,3} + \dots),$$

where the constants $k_{j,l}$ depend only on the b 's, the moments of $\mathbf{X}$ up to the $j \cdot r$ 'th and r and where the series contains only a finite number of non-zero terms.

A proof of Theorem 1.1 is given in Hall (1989).

Provided the first j moments of the $O_p(n^{-r/2})$ term in (1.27) are all of order $n^{-r/2}$, which is guaranteed under sufficiently stringent moment conditions on $\mathbf{X}$ and mild assumptions about A , the j 'th cumulant $\kappa_{j,n}$ equals the j 'th cumulant of S_{nr} up to order $n^{-r/2}$. This remainder may be made arbitrarily small by choosing r sufficiently large. Therefore, $\kappa_{j,n}$ admits asymptotic expansion (1.17).

The next step is to compute the variance of S_n so that, upon rescaling, S_n has unit asymptotic variance. The latter condition is implicit in (1.19). Put

$$\mu_{i_1 \dots i_j} \equiv E\{(X - \mu)^{(i_1)} \dots (X - \mu)^{(i_j)}\}, \quad j \geq 1.$$

Then $\mu_i = 0$ for each i , $E(Z^{(i)} Z^{(j)}) = \mu_{ij}$, $E(Z^{(i)} Z^{(j)} Z^{(k)}) = n^{-\frac{1}{2}} \mu_{ijk}$, and

$$E(Z^{(i)} Z^{(j)} Z^{(k)} Z^{(l)}) = \mu_{ij} \mu_{kl} + \mu_{ik} \mu_{jl} + \mu_{il} \mu_{jk} + O(n^{-1}).$$

Taking $r = 2$ in (1.27) so that

$$S_n = \sum_{i=1}^d a_i Z^{(i)} + n^{-\frac{1}{2}} \frac{1}{2} \sum_{i=1}^d \sum_{j=1}^d a_{ij} Z^{(i)} Z^{(j)} + O_p(n^{-1}),$$

we have

$$E(S_n) = n^{-\frac{1}{2}} \frac{1}{2} \sum_i \sum_j a_{ij} \mu_{ij} + O(n^{-1}),$$

$$E(S_n^2) = \sum_i \sum_j a_i a_j \mu_{ij} + O(n^{-1}),$$

and

$$\begin{aligned} E(S_n^3) = n^{-\frac{1}{2}} & \left\{ \sum_i \sum_j \sum_k a_i a_j a_k \mu_{ijk} \right. \\ & + \frac{3}{2} \sum_i \sum_j \sum_k \sum_l a_i a_j a_k l (\mu_{ij} \mu_{kl} + \mu_{ik} \mu_{jl} + \mu_{il} \mu_{jk}) \left. \right\} \\ & + O(n^{-1}). \end{aligned}$$

Consequently, the first three cumulants of S_n are

$$\kappa_{1,n} = E(S_n) = n^{-\frac{1}{2}} A_1 + O(n^{-1}),$$

$$\kappa_{2,n} = \text{var}(S_n) = \sigma^2 + O(n^{-1}),$$

and

$$\kappa_{3,n} = E(S_n^3) - 3E(S_n^2)E(S_n) + 2(E S_n)^3 = n^{-\frac{1}{2}} A_2 + O(n^{-1}),$$

where

$$\sigma^2 \equiv \sum_i \sum_j a_i a_j \mu_{ij}, \quad (1.29)$$

$$A_1 \equiv \frac{1}{2} \sum_i \sum_j a_{ij} \mu_{ij}, \quad (1.30)$$

and

$$A_2 \equiv \sum_i \sum_j \sum_k a_i a_j a_k \mu_{ijk} + 3 \sum_i \sum_j \sum_k \sum_l a_i a_j a_k l \mu_{ik} \mu_{jl}. \quad (1.31)$$

If the asymptotic variance of S_n , σ^2 , is not unity, then we redefine $S_n \equiv n^{\frac{1}{2}} A(\bar{X})/\sigma$. It then follows from the above calculations that for this definition of S_n , the constants $k_{1,2}$ and $k_{3,1}$ appearing in formula (1.20) for $p_1(x)$ are

$$k_{1,2} \equiv A_1 \sigma^{-1} \quad \text{and} \quad k_{3,1} \equiv A_2 \sigma^{-3}.$$

Theorem 1.2 below gives explicit regularity conditions under which expansion (1.19) is valid as an asymptotic series, assuming the above model for S_n .

THEOREM 1.2. (Bhattacharya and Ghosh (1978), Theorem 2.1.) Assume that the function A has $j + 2$ continuous derivatives in a neighbourhood of $\mu = E(\mathbf{X})$, that $A(\mu) = 0$, that $E(\|\mathbf{X}\|^{j+2}) < \infty$, and that the characteristic function ψ of $\mathbf{X}$ satisfies

$$\limsup_{\|\mathbf{t}\| \rightarrow \infty} |\psi(\mathbf{t})| < 1. \quad (1.32)$$

Define $a_{i_1 \dots i_r}$, $\mu_{i_1 \dots i_r}$, σ , A_1 , and A_2 as above. Suppose $\sigma > 0$. Then for $j \geq 1$,

$$P\{n^{\frac{1}{2}} A(\bar{\mathbf{X}})/\sigma \leq x\} = \Phi(x) + n^{-\frac{1}{2}} p_1(x) \phi(x) + \dots + n^{-j/2} p_j(x) \phi(x) + o(n^{-j/2}), \quad (1.33)$$

uniformly in x , where p_j is a polynomial of degree at most $3j - 1$, odd for even j and even for odd j . In particular,

$$p_1(x) \equiv -\{A_1 \sigma^{-1} + \frac{1}{6} A_2 \sigma^{-3} (x^2 - 1)\}. \quad (1.34)$$

Bhattacharya and Ghosh give a proof of this result. Condition (1.32) is a multivariate form of Cramér's condition, discussed earlier. Hall (1989) proves that it is satisfied if the distribution of $\mathbf{X}$ has a nondegenerate absolutely continuous component.

It remains to discuss models in which σ^2 is unknown, and we must estimate it by $\hat{\sigma}^2 \equiv \tau^2(\bar{\mathbf{X}})$. We must first verify that the asymptotic variance of an estimator $\hat{\theta} \equiv g(\bar{\mathbf{X}})$ of $\theta \equiv g(\mu)$ might be taken equal to a function τ^2 of the mean μ and might be estimated by the same function of $\bar{\mathbf{X}}$. If $A(\mathbf{x}) \equiv g(\mathbf{x}) - g(\mu)$, then by (1.29),

$$\sigma^2 = \sum_{i=1}^d \sum_{j=1}^d g_{(i)}(\mu) g_{(j)}(\mu) \{E(X^{(i)} X^{(j)}) - \mu^{(i)} \mu^{(j)}\},$$

where $g_{(i)}(\mathbf{x}) = (\partial/\partial x^{(i)})g(\mathbf{x})$. Suppose we adjoin to the vector $\mathbf{X}$ all those products $X^{(i)} X^{(j)}$ for which $g_{(i)}(\mu) g_{(j)}(\mu) \neq 0$, and which do not already appear in $\mathbf{X}$, and adjoin analogous terms to the vectors $\mathbf{X}_i$ and $\bar{\mathbf{X}}$. Let μ denote the mean vector of the new, lengthened $\mathbf{X}$, and put $\bar{\mathbf{X}} \equiv n^{-1} \sum_i \mathbf{X}_i$. Then σ^2 is a function of μ , say $\tau^2(\mu)$ and is estimated $\sqrt{n}$ -consistently by $\tau^2(\bar{\mathbf{X}})$. We have, by analogy with Theorem 1.2, the following asymptotic expansion for the distribution function of $S_n \equiv n^{\frac{1}{2}} A(\bar{\mathbf{X}})/\hat{\sigma}$:

$$P\{n^{\frac{1}{2}} A(\bar{\mathbf{X}})/\hat{\sigma} \leq x\} = \Phi(x) + n^{-\frac{1}{2}} q_1(x) \phi(x) + \dots + n^{-j/2} q_j(x) \phi(x) + o(n^{-j/2}), \quad (1.35)$$

uniformly in x , where q_j is a polynomial of degree at most $3j - 1$, odd for even j and even for odd j . If Theorem 1.2 is applied to the Studentized mean, expansion (1.35) holds assuming $(2j + 2)$ 'th moments on the underlying distribution. Hall (1987c) gives a more refined argument making extensive use of the special properties of the Studentized mean to show that $(j + 2)$ 'th moments suffice in that case. More generally, if we define $B \equiv A/\tau$, $b_{i_1 \dots i_r} \equiv B_{i_1 \dots i_r}(\mu)$, $B_1 \equiv \frac{1}{2} \sum_i \sum_j b_{ij} \mu_{ij}$ and

$$B_2 \equiv \sum_i \sum_j \sum_k b_i b_j b_k \mu_{ijk} + 3 \sum_i \sum_j \sum_k \sum_l b_i b_j b_k l \mu_{ik} \mu_{jl},$$

we have $q_1(x) \equiv -\{B_1 + \frac{1}{6} B_2(x^2 - 1)\}$; see Hall (1988a), Section 1.4. Hall (1988a) also shows that $p_1(x) - q_1(x) = -\frac{1}{2} \sigma^{-3} (\sum_i \sum_j a_i c_j \mu_{ij} x^2)$, where $c_i \equiv g_{(i)}(\mu)$.

Thus far in this section we have made little mention of the bootstrap. However, the present model for S_n fits well into the general framework for bootstrap resampling developed in Chapter 3. The reason for this is that μ , and hence $g(\mu)$, equals a functional $\theta(\cdot)$ of the population distribution function F . In all nonparametric cases, and in many parametric examples, $g(\bar{X})$ is the same functional of the empiric distribution function $\hat{F}$. This fact will prove vital to our development of Edgeworth expansions in the bootstrap context.

3.5. Cornish-Fisher expansions. Let S_n denote a statistic having a distribution admitting Edgeworth expansion (1.19), where we interpret this expansion as an asymptotic series. Write $w_{1-\alpha} = w_{1-\alpha}(n)$ for the $(1 - \alpha)$ -level quantile of S_n ,

$$P(S_n \leq w_{1-\alpha}) = 1 - \alpha,$$

and let $z_{1-\alpha}$ be the $(1 - \alpha)$ -level quantile of the Standard Normal distribution. We may invert expansion (1.19) to obtain an expansion of $w_{1-\alpha}$ in terms of $z_{1-\alpha}$, and vice-versa:

$$\begin{aligned} w_{1-\alpha} = z_{1-\alpha} + n^{-\frac{1}{2}} p_{11}(z_{1-\alpha}) + n^{-1} p_{21}(z_{1-\alpha}) + \dots \\ + n^{-j/2} p_{j1}(z_{1-\alpha}) + \dots, \end{aligned} \quad (1.36)$$

and

$$\begin{aligned} z_{1-\alpha} = w_{1-\alpha} + n^{-\frac{1}{2}} p_{12}(z_{1-\alpha}) + n^{-1} p_{22}(z_{1-\alpha}) + \dots \\ + n^{-j/2} p_{j2}(z_{1-\alpha}) + \dots. \end{aligned} \quad (1.37)$$

Such expansions, called Cornish-Fisher expansions and interpreted here as asymptotic series, are generally available uniformly in $\epsilon < \alpha < 1 - \epsilon$ for any

$0 < \epsilon < \frac{1}{2}$. The polynomials p_{j1} and p_{j2} are of degree at most $j + 1$ and are odd/even functions for even/odd j . They are determined completely by the p_j 's in (1.19) and satisfy the formal relations

$$\begin{aligned} 1 - \alpha \equiv & \Phi \left\{ z_{1-\alpha} + \sum_{j \geq 1} n^{-j/2} p_{j1}(z_{1-\alpha}) \right\} \\ & + \sum_{i \geq 1} n^{-i/2} p_i \left\{ z_{1-\alpha} + \sum_{j \geq 1} n^{-j/2} p_{j1}(z_{1-\alpha}) \right\} \\ & \times \phi \left\{ z_{1-\alpha} + \sum_{j \geq 1} n^{-j/2} p_{j1}(z_{1-\alpha}) \right\}, \end{aligned} \quad (1.38)$$

for $0 < \alpha < 1$, and

$$\sum_{j \geq 1} n^{-j/2} p_{j1}(x) + \sum_{i \geq 1} n^{-i/2} p_{i2} \left\{ x + \sum_{j \geq 1} n^{-j/2} p_{j1}(x) \right\} \equiv 0. \quad (1.39)$$

In particular, Taylor-expanding the left-hand side of (1.38) about $z_{1-\alpha}$, we get

$$\begin{aligned} 1 - \alpha = & 1 - \alpha + n^{-\frac{1}{2}} \{ p_{11}(z_{1-\alpha}) + p_1(z_{1-\alpha}) \} \phi(z_{1-\alpha}) \\ & + n^{-1} \left[p_{21}(z_{1-\alpha}) - \frac{1}{2} p_{11}^2(z_{1-\alpha}) z_{1-\alpha} \right. \\ & + p_{11}(z_{1-\alpha}) \{ p_1'(z_{1-\alpha}) - p_1(z_{1-\alpha}) z_{1-\alpha} \} \\ & \left. + p_2(z_{1-\alpha}) \right] \phi(z_{1-\alpha}) + O(n^{-3/2}), \end{aligned}$$

from which we conclude that

$$p_{11}(x) = -p_1(x) \quad \text{and} \quad p_{21}(x) = p_1(x)p_1'(x) - \frac{1}{2}xp_1(x)^2 - p_2(x).$$

If the distribution of S_n admits Edgeworth expansion (1.35) rather than (1.19) (for example, if σ is unknown and we estimate it by $\hat{\sigma}$), then expansions (1.36) and (1.37) hold, except that polynomials p_{j1} and p_{j2} there are replaced by different polynomials, which we denote by q_{j1} and q_{j2} respectively. Relations (1.38) and (1.39) then hold with q 's instead of p 's and we deduce that

$$q_{11}(x) = -q_1(x) \quad \text{and} \quad q_{21}(x) = q_1(x)q_1'(x) - \frac{1}{2}xq_1(x)^2 - q_2(x).$$

In most of our work we require only the first two polynomials in Cornish-Fisher inversions of (1.19) and (1.35). The polynomial q_{31} , which we need in Chapter 2, is derived in Appendix A2.3.

3.6. Bibliographical Notes. This subsection is intended to give the reader a brief summary of accounts in the literature of Edgeworth expansions and related topics.

The idea of expanding a distribution function in what we now call an Edgeworth expansion originated with the work of Chebyshev (1890) and Edgeworth (1904), who discussed formal construction of expansions for sums of independent random variables. Cramér (1928; 1946, Chapter VII) gave rigorous theory in that case. Later work in this area was carried out by Geary (1936, 1947), Gayen (1949, 1950) and Wallace (1958), before formal expansions were put on a rigorous footing, largely by the work of Hsu (1945), Bickel (1974), Chibishov (1972, 1973a, 1973b), and Bhattacharya and Ghosh (1978), among others. See also Chambers (1967), Pfanzagl (1973a, 1973b, 1979), Davis (1976), Sargan (1976) and Phillips (1977).

Our development of Edgeworth expansion theory has relied heavily on the basic cumulant expansion (1.17). Fisher (1930) recognized that formula without giving a comprehensive proof, and James (1955, 1958) and James and Mayne (1962) have given more thorough arguments. See also Kendall and Stuart (1977), Chapters 12 and 13.

Petrov (1975), Chapter VI has given a detailed account of Edgeworth expansions for distributions of sums of independent random variables, and his Section 1, Chapter VI is a useful introduction to the polynomials appearing in these expansions.

Cramér's condition has played a major role in the theory of Edgeworth expansions, even in the simplest cases. Esseen (1945) showed that one-term expansions with remainder $o(n^{-\frac{1}{2}})$ are valid under the weaker condition of non-latticeness, and he also treated expansions for lattice random variables. Bhattacharya and Rao (1976), Chapter 5 have provided a comprehensive treatment of Edgeworth expansions for sums of independent lattice-valued random variables.

Bhattacharya and Rao (1976) also gave an excellent treatment of expansions in the case of sums of vector-valued random variables. That work formed the basis for Bhattacharya and Ghosh's (1978) rigorous treatment of Edgeworth expansions. Withers (1983, 1984) presented general formulae for early terms in Edgeworth expansions under models including that treated in Subsection 3.4. Inverse (Cornish-Fisher) expansions were first studied by Cornish and Fisher (1937), and have been discussed by Kendall and Stuart (1977), Section 16.21, and Hall (1983).

Work on Edgeworth expansions in the bootstrap context has been carried out by Bickel and Freedman (1980, 1981), Singh (1981, 1987), Babu and Singh (1983), Abramovitch and Singh (1985), Hall (1986a, 1986b, 1987a, 1988a, 1988b,

1989), Liu and Singh (1987), and Hall and Martin (1988a, 1988b, 1988c), among others.

4. SUMMARY OF THESIS

The aim of this chapter has been to provide the reader with a fairly general background to the bootstrap method and the theory of Edgeworth expansions. We stress that throughout this thesis we use Edgeworth expansions primarily as a tool for describing and studying asymptotic theory for bootstrap confidence intervals in a wide variety of circumstances. In some respects our study of bootstrap confidence intervals follows on from the work of Hall (1988a), who developed a unified framework for bootstrap confidence intervals and critical points in estimation of a univariate parameter θ .

Chapter 2 provides asymptotic theory for the bootstrap in two-sample problems. In some ways our work there parallels Hall's (1988a) study of bootstrap confidence intervals in the simpler one-sample case. It is shown that if the samples are of sizes m and n , one- and two-sided percentile- t confidence intervals for a difference between two means has coverage error $O(m^{-1} + n^{-1})$, and that symmetric intervals have coverage error $O(m^{-2} + n^{-2})$. Furthermore, when the populations are Normal, coverage errors of all percentile- t intervals drop to $O(m^{-2} + n^{-2})$. Therefore, percentile- t provides a respectable solution to the Behrens-Fisher problem. We also introduce several notions such as second-order correctness, coverage error and mean interval length for use throughout the thesis, and we detail the construction of the various types of interval discussed earlier in Section 2.

A major contribution of this thesis is the introduction in Chapter 3 of a general resampling principle which encompasses bootstrap resampling and bootstrap iteration. The principle is widely applicable to many areas of statistics, such as the construction of confidence intervals and hypothesis tests, and to bias reduction for point estimation. Our principle is illustrated using several examples, including the application of the principle to bias-correction, L^1 shrinkage, and to solving complex interval construction problems such as a suggestion by Lehmann (1986), p.223, for "conditionally" short confidence intervals.

We extend our study of bootstrap iteration in Chapter 4, where we discuss in detail bootstrap iteration to correct for coverage error in confidence intervals. We give theory showing that bootstrap iteration reduces the order of coverage error by a factor of $n^{-\frac{1}{2}}$ in the one-sided case, and by a factor of n^{-1} for two-

sided intervals. We also discuss the effect of bootstrap coverage correction on other indicators of interval performance such as position of critical point and average interval length. Towards the end of that chapter we show how bootstrap coverage correction may be applied to Bartlett-corrected likelihood ratio tests, and to confidence intervals for the correlation coefficient.

The final chapter, Chapter 5, treats the problem of using the bootstrap to construct confidence intervals for quantiles. It is shown that many of the standard bootstrap methods, such as those outlined earlier in Section 2, perform relatively poorly in this problem compared with their performance in more standard problems. We then study the performance of Normal-theory confidence intervals for quantiles, constructed using the bootstrap estimate of sample quantile variance to "Studentize". It turns out that these intervals have coverage error $O(n^{-\frac{1}{2}})$ in both one- and two-sided cases, in marked distinction from more standard problems, where coverage error is $O(n^{-\frac{1}{2}})$ in the one-sided case but drops to $O(n^{-1})$ for two-sided intervals. Our result arises from a curious feature of the Edgeworth expansion of the distribution of the bootstrap-Studentized quantile, which distinguishes it from the expansions discussed earlier, in that the leading term, of order $n^{-\frac{1}{2}}$, is proportional to a polynomial containing both odd and even powers of the argument. The results of our asymptotic analysis are used to compare the performances in large samples of bootstrap-Studentized intervals and Normal-theory intervals constructed using the Siddiqui-Bloch-Gastwirth variance estimator.

CHAPTER 2

ON BOOTSTRAP CONFIDENCE INTERVALS AND TWO-SAMPLE PROBLEMS

1. INTRODUCTION

The Behrens-Fisher problem is that of testing whether the means of two Normal populations are equal, without making any assumptions about the variances. The solution proposed by Behrens (1929) and elucidated by Fisher (1935), is as follows. Let $\mathcal{X} \equiv \{X_1, \dots, X_m\}$ and $\mathcal{Y} \equiv \{Y_1, \dots, Y_n\}$ be independent random samples from x - and y -populations respectively, with sample means $\bar{X} \equiv m^{-1} \sum_i X_i$ and $\bar{Y} \equiv n^{-1} \sum_j Y_j$, and sample variances $\hat{\sigma}_X^2 \equiv m^{-1} \sum_i (X_i - \bar{X})^2$ and $\hat{\sigma}_Y^2 \equiv n^{-1} \sum_j (Y_j - \bar{Y})^2$. Tests and confidence intervals are based on the statistic

$$T' \equiv \{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)\} \{\hat{\sigma}_X^2(m-1)^{-1} + \hat{\sigma}_Y^2(n-1)^{-1}\}^{-\frac{1}{2}}. \quad (2.1)$$

If the populations are Normal (μ_X, σ_X^2) and (μ_Y, σ_Y^2) respectively, then the distribution of T' is that of a mixture of two Student's t -distributions. Indeed, we may write $T' = T_1 \sin \theta + T_2 \cos \theta$, where T_1, T_2 and θ are independent random variables, and T_1 and T_2 have Student's t -distributions with $m-1$ and $n-1$ degrees of freedom respectively. If the populations are Normal with equal variances, then T' has a Student's t -distribution with $m+n-2$ degrees of freedom. However, in general, a naïve usage of Student's t -tables results in tests and confidence intervals for $\mu_X - \mu_Y$ with inaccurate levels. Several devices which partially overcome this problem have been suggested, the best-known being the Welch-Aspin solution (Welch, 1947; Aspin, 1948), but none of these approaches addresses a major source of potential error: the possibility that one or other of the x - and y -populations is not Normal. In this chapter, we examine the

nonparametric bootstrap as a tool for solving the two-sample problem, with particular emphasis on its performance with general populations.

We show that one-sided percentile- t bootstrap confidence intervals for $\mu_X - \mu_Y$ have coverage error $O(m^{-1} + n^{-1})$, compared to $O(m^{-\frac{1}{2}} + n^{-\frac{1}{2}})$ when simplistic Normal or Student's t approximations are used; that two-sided symmetric percentile- t confidence intervals for $\mu_X - \mu_Y$ have coverage error $O(m^{-2} + n^{-2})$; and that when populations are Normal, all the common percentile- t intervals for $\mu_X - \mu_Y$ (both one- and two-sided) have coverage error $O(m^{-2} + n^{-2})$. The latter result suggests that even if populations are Normal, bootstrap procedures will be competitive with classical alternatives. A simulation study lends support to this conclusion. We also derive a formula for the acceleration constant required to implement Efron's (1987) accelerated bias correction. In some respects, our work on the two-sample problem parallels a study of bootstrap confidence intervals in the one-sample case by Hall (1988a). Beran (1987a) has also discussed the Behrens-Fisher problem in the context of pre pivoting the bootstrap.

We close this section with notation. It is convenient to work with

$$T \equiv \{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)\}(\hat{\sigma}_X^2 m^{-1} + \hat{\sigma}_Y^2 n^{-1})^{-\frac{1}{2}}, \quad (2.2)$$

instead of T' (defined at (2.1)), but all of our conclusions apply equally to T and T' . We denote (mean, variance) of x - and y -populations by (μ_X, σ_X^2) and (μ_Y, σ_Y^2) respectively, and define $\sigma^2 \equiv \sigma_X^2 m^{-1} + \sigma_Y^2 n^{-1}$ and $\hat{\sigma}^2 \equiv \hat{\sigma}_X^2 m^{-1} + \hat{\sigma}_Y^2 n^{-1}$. Standard Normal distribution and density functions are indicated by Φ and ϕ respectively. Let z_α denote the solution of the equation $\Phi(z_\alpha) = \alpha$, for $0 < \alpha < 1$. We denote generic random variables drawn from x - and y -populations by X and Y respectively.

2. CONSTRUCTION OF BOOTSTRAP CONFIDENCE INTERVALS

Let $\mathcal{X}^* \equiv \{X_1^*, \dots, X_m^*\}$ and $\mathcal{Y}^* \equiv \{Y_1^*, \dots, Y_n^*\}$ be samples (sometimes called "resamples") drawn independently and randomly, with replacement, from $\mathcal{X}$ and $\mathcal{Y}$ respectively. Put $\bar{X}^* \equiv m^{-1} \sum_i X_i^*$, $\bar{Y}^* \equiv n^{-1} \sum_j Y_j^*$, $\hat{\sigma}_X^{*2} \equiv m^{-1} \sum_i (X_i^* - \bar{X}^*)^2$, $\hat{\sigma}_Y^{*2} \equiv n^{-1} \sum_j (Y_j^* - \bar{Y}^*)^2$ and $\hat{\sigma}^{*2} \equiv \hat{\sigma}_X^{*2} m^{-1} + \hat{\sigma}_Y^{*2} n^{-1}$. The sample analogues $\bar{X}$, $\bar{Y}$, $\hat{\sigma}_X^2$, $\hat{\sigma}_Y^2$ and $\hat{\sigma}^2$ are defined by deleting the superscript $*$ throughout.

In the context of the two-sample problem, the percentile- t argument for constructing confidence intervals is based on the idea that the distribution of

$$T^* \equiv \{\bar{X}^* - \bar{Y}^* - (\bar{X} - \bar{Y})\} / \hat{\sigma}^*,$$

conditional on $\mathcal{X}$ and $\mathcal{Y}$, should approximate the unconditional distribution of $T \equiv \{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)\} / \hat{\sigma}$. Therefore, if we define

$$\hat{t}_\alpha \equiv \sup\{t : P(T^* \leq t \mid \mathcal{X}, \mathcal{Y}) \leq \alpha\}, \quad (2.3)$$

then each of $I_1 \equiv (-\infty, \bar{X} - \bar{Y} - \hat{t}_\alpha \hat{\sigma}]$, $I_2 \equiv [\bar{X} - \bar{Y} - \hat{t}_{1-\alpha} \hat{\sigma}, \infty)$ and $I_3 \equiv [\bar{X} - \bar{Y} - \hat{t}_{1-(\alpha/2)} \hat{\sigma}, \bar{X} - \bar{Y} - \hat{t}_{\alpha/2} \hat{\sigma}]$ is a confidence interval for $\mu_X - \mu_Y$ with nominal coverage $1 - \alpha$. These intervals are called “percentile- t ” confidence intervals because their endpoints are derived from percentiles of the Studentized statistic, T^* .

An alternative, non-Studentized approach is the so-called “percentile method”, which is as follows. Define

$$\hat{u}_\alpha \equiv \sup\{u : P(\bar{X}^* - \bar{Y}^* \leq u \mid \mathcal{X}, \mathcal{Y}) \leq \alpha\}.$$

Then $J_{11} \equiv (-\infty, \hat{u}_{1-\alpha}]$, $J_{12} \equiv [\hat{u}_\alpha, \infty)$ and $J_{13} \equiv [\hat{u}_{\alpha/2}, \hat{u}_{1-(\alpha/2)}]$ are percentile-method confidence intervals for $\mu_X - \mu_Y$ with nominal coverage $1 - \alpha$. They are based on the notion that confidence bounds placed on the bootstrap distribution of $\bar{X}^* - \bar{Y}^*$ should be close in some sense to confidence bounds for $\mu_X - \mu_Y$. Slightly different percentile-method intervals are obtained by arguing that the distribution of $\bar{X}^* - \bar{Y}^* - (\bar{X} - \bar{Y})$, conditional on $\mathcal{X}$ and $\mathcal{Y}$, should be close to the unconditional distribution of $\bar{X} - \bar{Y} - (\mu_X - \mu_Y)$. In particular, if we define

$$\hat{v}_\alpha \equiv \sup\{v : P[\bar{X}^* - \bar{Y}^* - (\bar{X} - \bar{Y}) \leq v \mid \mathcal{X}, \mathcal{Y}] \leq \alpha\},$$

we obtain confidence intervals $J_{21} \equiv (-\infty, \bar{X} - \bar{Y} - \hat{v}_\alpha]$, $J_{22} \equiv [\bar{X} - \bar{Y} - \hat{v}_{1-\alpha}, \infty)$ and $J_{23} \equiv [\bar{X} - \bar{Y} - \hat{v}_{1-(\alpha/2)}, \bar{X} - \bar{Y} - \hat{v}_{\alpha/2}]$, with nominal coverage $(1 - \alpha)$.

Efron (1987) noted that while the first type of percentile-method interval has some attractive features (for example, invariance under monotone transformations), it has poor coverage properties. In the two-sample case, Efron’s “accelerated bias-corrected” procedure takes the following form. Put

$$\hat{d} \equiv \Phi^{-1}\{P(\bar{X}^* - \bar{Y}^* \leq \bar{X} - \bar{Y} \mid \mathcal{X}, \mathcal{Y})\}, \quad (2.4)$$

and

$$\hat{\beta}_a = \hat{\beta}_a(\alpha) \equiv \Phi[\hat{d} + (\hat{d} + z_\alpha)\{1 - \hat{a}(\hat{d} + z_\alpha)\}^{-1}], \quad (2.5)$$

where $\hat{a}$ is the bootstrap estimate of Efron's "acceleration constant". Let K_j ($j = 1, 2, 3$) be the same as interval J_{1j} , except that percentiles $\hat{u}_\alpha$, $\hat{u}_{1-\alpha}$, etc. are replaced by $\hat{u}_{\hat{\beta}_a(\alpha)}$, $\hat{u}_{\hat{\beta}_a(1-\alpha)}$, and so on. Taking $\hat{a} \equiv 0$ yields (ordinary) bias correction, proposed and studied by Efron (1981a, 1982). A more judicious choice of $\hat{a}$, directed at providing more accurate coverage, depends on detailed properties of the distribution of T . It is studied in Section 5, where we give an explicit formula for the acceleration constant which yields second-order correct critical points. In Section 8 we discuss a symmetric version of I_3 . We do not discuss it here as it fails to have a one-sided analogue. We consider in Section 9 the construction of "short" percentile- t confidence intervals.

3. CORNISH-FISHER EXPANSIONS OF BOOTSTRAP CRITICAL POINTS

In this section we develop Cornish-Fisher expansions of bootstrap critical points, as m and n diverge. We assume throughout, without any loss of generality, that $m \geq n$. Then we let $n \rightarrow \infty$, assuming that $m \geq n$, and study the properties of bootstrap critical points for various different values of the ratio

$$r = r(m, n) \equiv m/n.$$

As a guide to the sort of behaviour we can expect, consider a one-sample problem, in which an unknown parameter θ , is estimated by $\hat{\theta}$ with asymptotic variance $n^{-1}\tau^2$. Suppose $\hat{\tau}^2$ is a $\sqrt{n}$ -consistent estimator of τ^2 , and put $S \equiv n^{\frac{1}{2}}(\hat{\theta} - \theta)/\tau$ and $T \equiv n^{\frac{1}{2}}(\hat{\theta} - \theta)/\hat{\tau}$. Under the kinds of regularity conditions discussed in Section 3 of Chapter 1, and used later in this section, distributions of both S and T admit Edgeworth expansions of the following type:

$$P(S \leq x) = \Phi(x) + n^{-\frac{1}{2}}p_1(x)\phi(x) + n^{-1}p_2(x)\phi(x) + O(n^{-3/2}), \quad (2.6)$$

and

$$P(T \leq x) = \Phi(x) + n^{-\frac{1}{2}}q_1(x)\phi(x) + n^{-1}q_2(x)\phi(x) + O(n^{-3/2}), \quad (2.7)$$

where p_j and q_j are polynomials of degree $3j - 1$, even for odd j and odd for even j (see, for example, Wallace, 1958; Bhattacharya and Ghosh, 1978; Withers, 1983; and Hall, 1988a). Inversions of these expansions give expressions for the quantiles x_α and y_α of the distributions of S and T , defined by

$$P(S \leq x_\alpha) = P(T \leq y_\alpha) = \alpha.$$

Indeed, it follows that

$$x_\alpha = z_\alpha + n^{-\frac{1}{2}}p_{11}(z_\alpha) + n^{-1}p_{21}(z_\alpha) + O(n^{-3/2}), \quad (2.8)$$

and

$$y_\alpha = z_\alpha + n^{-\frac{1}{2}}q_{11}(z_\alpha) + n^{-1}q_{21}(z_\alpha) + O(n^{-3/2}), \quad (2.9)$$

where

$$p_{11} \equiv -p_1, \quad q_{11} \equiv -q_1, \quad (2.10)$$

$$p_{21}(x) \equiv p_1(x)p'_1(x) - \frac{1}{2}x\{p_1(x)\}^2 - p_2(x), \quad (2.11)$$

$$q_{21}(x) \equiv q_1(x)q'_1(x) - \frac{1}{2}x\{q_1(x)\}^2 - q_2(x);$$

see Appendix A2.3 for a derivation of the polynomials q_{i1} , for $i = 1, 2, 3$.

In principle, the situation described above also applies to our estimate $\hat{\theta} \equiv \bar{X} - \bar{Y}$ of $\theta \equiv \mu_X - \mu_Y$. In that case, we should take $\tau^2 \equiv r^{-1}\sigma_X^2 + \sigma_Y^2$ and $\hat{\tau}^2 \equiv r^{-1}\hat{\sigma}_X^2 + \hat{\sigma}_Y^2$. Edgeworth and Cornish-Fisher expansions are series in $m^{-i/2}n^{-j/2}$ for $i, j \geq 0$, so it is clear that, by noting $m^{-i/2}n^{-j/2} = n^{-(i+j)/2}r^{-i/2}$, if we allow the polynomials p_j and q_j to depend on the variable “parameter” r , the expansions obtained for the two-sample problem may be put into the same form as those for the one-sample problem. This is done in Theorem 2.1 below.

Let $S \equiv \{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)\}\sigma^{-1}$ and $T \equiv \{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)\}\hat{\sigma}^{-1}$. Write $\gamma_X \equiv E\{(X - \mu_X)^3\}$ and $\kappa_X \equiv E\{(X - \mu_X)^4\} - 3\sigma_X^4$ for the skewness and kurtosis, respectively, of the X -sample. Define γ_Y and κ_Y similarly for the Y -sample.

Suppose that the distributions of (X, X^2) and (Y, Y^2) each satisfy bivariate versions of Cramér’s continuity condition — weaker than the assumption that X and Y be non-singular — and that $E(X^{10}) < \infty$ and $E(Y^{10}) < \infty$. These regularity conditions are sufficient for all the results in this chapter, although it is our aim to focus on the *form* of expansions, rather than to give detailed proofs under explicit regularity conditions. In fact, all the results given in this chapter hold under the assumption that X and Y have non-singular distributions with sufficiently many finite moments. Formal, rigorous justification of analogous results in the one-sample case may be found in the proof of Theorem 2.1 of Hall (1986a).

THEOREM 2.1. *In the two-sample problem, under the above regularity conditions, expansions (2.6)–(2.9) hold with*

$$\begin{aligned} p_1(x) &\equiv -\frac{1}{6}(\sigma_X^2 + r\sigma_Y^2)^{-3/2}(r^{-1/2}\gamma_X - r^{3/2}\gamma_Y)(x^2 - 1), \\ q_1(x) &\equiv \frac{1}{6}(\sigma_X^2 + r\sigma_Y^2)^{-3/2}(r^{-1/2}\gamma_X - r^{3/2}\gamma_Y)(2x^2 + 1), \\ p_2(x) &\equiv -x\left\{\frac{1}{24}(\sigma_X^2 + r\sigma_Y^2)^{-2}(r^{-1}\kappa_X + r^2\kappa_Y)(x^2 - 3) + \frac{1}{72}(\sigma_X^2 + r\sigma_Y^2)^{-3}\right. \\ &\quad \left.\times (r^{-1/2}\gamma_X - r^{3/2}\gamma_Y)^2(x^4 - 10x^2 + 15)\right\}, \end{aligned}$$

and

$$\begin{aligned} q_2(x) &\equiv x\left[\frac{1}{12}(\sigma_X^2 + r\sigma_Y^2)^{-2}(r^{-1}\kappa_X + r^2\kappa_Y)(x^2 - 3)\right. \\ &\quad - \frac{1}{18}(\sigma_X^2 + r\sigma_Y^2)^{-3}(r^{-1/2}\gamma_X - r^{3/2}\gamma_Y)^2(x^4 + 2x^2 - 3) \\ &\quad - \frac{1}{4}(\sigma_X^2 + r\sigma_Y^2)^{-2}\{x^2(r^{-1}\sigma_X^4 + r^2\sigma_Y^4) + 3r^{-1}\sigma_X^4 \\ &\quad \left.+ 2\sigma_X^2\sigma_Y^2(1 + r) + 3r^2\sigma_Y^4\}\right], \end{aligned}$$

and p_{11} , q_{11} , p_{21} and q_{21} given by (2.10) and (2.11).

Theorem 2.1 is proved in Section 11.

Define $\hat{\sigma}^{*2} \equiv \hat{\sigma}_X^{*2}m^{-1} + \hat{\sigma}_Y^{*2}n^{-1}$,

$$\begin{aligned} S^* &\equiv \{\bar{X}^* - \bar{Y}^* - (\bar{X} - \bar{Y})\}\hat{\sigma}^{-1}, \quad T^* \equiv \{\bar{X}^* - \bar{Y}^* - (\bar{X} - \bar{Y})\}\hat{\sigma}^{*-1}, \\ \hat{x}_\alpha &\equiv \sup\{x : P(S^* \leq x \mid \mathcal{X}, \mathcal{Y}) \leq \alpha\}, \quad \hat{y}_\alpha \equiv \sup\{y : P(T^* \leq y \mid \mathcal{X}, \mathcal{Y}) \leq \alpha\}. \end{aligned}$$

Then $\hat{x}_\alpha$ and $\hat{y}_\alpha$ are bootstrap estimates of x_α and y_α respectively. The bootstrap analogues of (2.6)–(2.9) are

$$\begin{aligned} P(S^* \leq x \mid \mathcal{X}, \mathcal{Y}) &= \Phi(x) + n^{-1/2}\hat{p}_1(x)\phi(x) + n^{-1}\hat{p}_2(x)\phi(x) + O_p(n^{-3/2}), \\ P(T^* \leq x \mid \mathcal{X}, \mathcal{Y}) &= \Phi(x) + n^{-1/2}\hat{q}_1(x)\phi(x) + n^{-1}\hat{q}_2(x)\phi(x) \\ &\quad + O_p(n^{-3/2}), \end{aligned} \tag{2.12}$$

$$\hat{x}_\alpha = z_\alpha + n^{-1/2}\hat{p}_{11}(z_\alpha) + n^{-1}\hat{p}_{21}(z_\alpha) + O_p(n^{-3/2}), \tag{2.13}$$

and

$$\hat{y}_\alpha = z_\alpha + n^{-1/2}\hat{q}_{11}(z_\alpha) + n^{-1}\hat{q}_{21}(z_\alpha) + O_p(n^{-3/2}), \tag{2.14}$$

respectively, where $\hat{p}_j$ and $\hat{q}_j$ are given in Theorem 2.1, except that population moments there should be replaced by the corresponding sample moments. In particular, γ_X and κ_X should be replaced by

$$\hat{\gamma}_X \equiv m^{-1} \sum_i (X_i - \bar{X})^3, \quad \hat{\kappa}_X \equiv m^{-1} \sum_i (X_i - \bar{X})^4 - 3\hat{\sigma}_X^4, \tag{2.15}$$

respectively, while $\hat{\gamma}_Y$ and $\hat{\kappa}_Y$, defined similarly, replace γ_Y and κ_Y , respectively. Expansions (2.13) and (2.14) provide us with expansions of endpoints of percentile and percentile- t confidence intervals. In fact, $\hat{y}_\alpha$ is just $\hat{t}_\alpha$, defined at (2.3), and thus

$$\bar{X} - \bar{Y} - \hat{t}_\alpha \hat{\sigma} = \bar{X} - \bar{Y} - \{z_\alpha + n^{-\frac{1}{2}} \hat{q}_{11}(z_\alpha) + n^{-1} \hat{q}_{21}(z_\alpha) + O_p(n^{-3/2})\} \hat{\sigma}. \quad (2.16)$$

Observe also that the points $\hat{u}_\alpha$ and $\hat{v}_\alpha$ used to construct percentile-method confidence intervals are given by

$$\begin{aligned} \hat{v}_\alpha &= \sup\{v : P[\bar{X}^* - \bar{Y}^* - (\bar{X} - \bar{Y}) \leq v \mid \mathcal{X}, \mathcal{Y}] \leq \alpha\} \\ &= \sup\{v : P(S^* \leq v \hat{\sigma}^{-1} \mid \mathcal{X}, \mathcal{Y}) \leq \alpha\} = \hat{x}_\alpha \hat{\sigma}, \end{aligned}$$

and

$$\begin{aligned} \hat{u}_\alpha &= \sup\{u : P(\bar{X}^* - \bar{Y}^* \leq u \mid \mathcal{X}, \mathcal{Y}) \leq \alpha\} \\ &= \bar{X} - \bar{Y} + \hat{v}_\alpha = \bar{X} - \bar{Y} + \hat{x}_\alpha \hat{\sigma}. \end{aligned} \quad (2.17)$$

Therefore, Cornish-Fisher expansions of the endpoints of percentile-method confidence intervals follow immediately from (2.13).

4. SECOND-ORDER CORRECTNESS

In “ideal” circumstances, a test or confidence interval for $\mu_X - \mu_Y$ would be based on the exact distribution of T , defined at (2.2). In particular, an exact, $(1 - \alpha)$ -level confidence interval for $\mu_X - \mu_Y$, would be

$$\mathcal{I} \equiv (-\infty, \bar{X} - \bar{Y} - y_\alpha \hat{\sigma}]. \quad (2.18)$$

The bootstrap confidence intervals listed in Section 2 are practical alternatives to “ideal” intervals. For example, the percentile- t alternative to $\mathcal{I}$ is $I_1 \equiv (-\infty, \bar{X} - \bar{Y} - \hat{y}_\alpha \hat{\sigma}]$. Noting that polynomials q_{11} and $\hat{q}_{11}$ differ only by $O_p(n^{-\frac{1}{2}})$, and that $\hat{\sigma} = O_p(n^{-\frac{1}{2}})$, it follows from (2.9) and (2.14) that the endpoints of $\mathcal{I}$ and I_1 differ by the amount

$$\begin{aligned} \hat{\sigma}(\hat{y}_\alpha - y_\alpha) &= \hat{\sigma} \left[z_\alpha + n^{-\frac{1}{2}} \hat{q}_{11}(z_\alpha) + n^{-1} \hat{q}_{21}(z_\alpha) + O_p(n^{-3/2}) \right. \\ &\quad \left. - \{z_\alpha + n^{-\frac{1}{2}} q_{11}(z_\alpha) + n^{-1} q_{21}(z_\alpha) + O(n^{-3/2})\} \right] \\ &= \hat{\sigma} n^{-\frac{1}{2}} \{\hat{q}_{11}(z_\alpha) - q_{11}(z_\alpha)\} + O_p(n^{-3/2}) = O_p(n^{-3/2}). \end{aligned}$$

Consequently, endpoints of $\mathcal{I}$ and I_1 agree in terms of second order, that is, in terms of order $(n^{-\frac{1}{2}})^2 = n^{-1}$. For this reason, we say that I_1 is *second-order*

correct. The other percentile- t intervals I_2 and I_3 are also second-order correct, relative to their “ideal” counterparts, based on the exact distribution of T . On the other hand, we see from (2.9) and (2.17), bearing in mind that $z_\alpha = -z_{1-\alpha}$ and that p_{11} is an even polynomial, that the endpoints of $\mathcal{I}$ and its percentile-method alternatives, $J_{11} \equiv (-\infty, \hat{u}_{1-\alpha}]$ and $J_{21} \equiv (-\infty, \bar{X} - \bar{Y} - \hat{v}_\alpha]$, differ by the amounts

$$\begin{aligned} -\hat{\sigma}(y_\alpha + \hat{x}_{1-\alpha}) &= -\hat{\sigma}[z_\alpha + n^{-\frac{1}{2}}q_{11}(z_\alpha) + n^{-1}q_{21}(z_\alpha) + O(n^{-3/2}) \\ &\quad + \{z_{1-\alpha} + n^{-\frac{1}{2}}\hat{p}_{11}(z_{1-\alpha}) + n^{-1}\hat{p}_{21}(z_{1-\alpha}) + O_p(n^{-3/2})\}] \\ &= -\hat{\sigma}n^{-\frac{1}{2}}\{q_{11}(z_\alpha) + p_{11}(z_\alpha)\} + O_p(n^{-3/2}), \end{aligned}$$

and

$$\begin{aligned} \hat{\sigma}(\hat{x}_\alpha - y_\alpha) &= \hat{\sigma}[z_\alpha + n^{-\frac{1}{2}}\hat{p}_{11}(z_\alpha) + n^{-1}\hat{p}_{21}(z_\alpha) + O_p(n^{-3/2}) \\ &\quad - \{z_\alpha + n^{-\frac{1}{2}}q_{11}(z_\alpha) + n^{-1}q_{21}(z_\alpha) + O(n^{-3/2})\}] \\ &= \hat{\sigma}n^{-\frac{1}{2}}\{p_{11}(z_\alpha) - q_{11}(z_\alpha)\} + O_p(n^{-3/2}), \end{aligned}$$

respectively. Therefore, only in special circumstances such as when $\gamma_X = \gamma_Y = 0$, will percentile-method confidence intervals be second-order correct.

5. ACCELERATED BIAS CORRECTION

In this section we show that choosing the acceleration constant $\hat{a}$ to be

$$\hat{a} \equiv \frac{1}{6}\hat{\sigma}^{-3/2}(m^{-2}\hat{\gamma}_X - n^{-2}\hat{\gamma}_Y), \quad (2.19)$$

renders the accelerated bias-corrected interval $K_1 \equiv (-\infty, \hat{u}_{\hat{\beta}_a(1-\alpha)}]$ second-order correct relative to the “ideal” interval, $\mathcal{I}$, defined at (2.18). Similarly, defining $\hat{a}$ as at (2.19) makes the accelerated bias-corrected intervals K_2 and K_3 second-order correct relative to their “ideal” counterparts based on the exact distribution of T .

Observe from (2.4) and (2.12) that

$$\begin{aligned} \hat{d} &\equiv \Phi^{-1}\{P(S^* \leq 0 \mid \mathcal{X}, \mathcal{Y})\} = \Phi^{-1}\{\frac{1}{2} + n^{-\frac{1}{2}}\hat{p}_1(0)\phi(0) + O_p(n^{-1})\} \\ &= n^{-\frac{1}{2}}\hat{p}_1(0) + O_p(n^{-1}). \end{aligned}$$

Thus, if we take $\hat{a} \equiv n^{-\frac{1}{2}}\hat{c}$, for some $\hat{c}$ yet to be determined, we see from the definition (2.5) of $\hat{\beta}_a(\alpha)$ that

$$z_{\hat{\beta}_a(\alpha)} \equiv \hat{d} + (\hat{d} + z_\alpha)\{1 - \hat{a}(\hat{d} + z_\alpha)\}^{-1} = z_\alpha + n^{-\frac{1}{2}}\{2\hat{p}_1(0) + z_\alpha^2\hat{c}\} + O_p(n^{-1}).$$

Therefore, by (2.13) and (2.17),

$$\begin{aligned}
 \hat{u}_{\hat{\beta}_a(\alpha)} &= \bar{X} - \bar{Y} + \hat{x}_{\hat{\beta}_a(\alpha)} \hat{\sigma} \\
 &= \bar{X} - \bar{Y} + \hat{\sigma} \{ z_{\hat{\beta}_a(\alpha)} + n^{-\frac{1}{2}} \hat{p}_{11}(z_{\hat{\beta}_a(\alpha)}) + n^{-1} \hat{p}_{21}(z_{\hat{\beta}_a(\alpha)}) + O_p(n^{-3/2}) \} \\
 &= \bar{X} - \bar{Y} + \hat{\sigma} [z_\alpha + n^{-\frac{1}{2}} \{ 2\hat{p}_1(0) + z_\alpha^2 \hat{c} \} + n^{-\frac{1}{2}} \hat{p}_{11}(z_\alpha) + O_p(n^{-1})] \\
 &= \bar{X} - \bar{Y} + \hat{\sigma} [z_\alpha - n^{-\frac{1}{2}} q_{11}(z_\alpha) + n^{-\frac{1}{2}} \{ 2\hat{p}_1(0) + z_\alpha^2 \hat{c} + \hat{p}_{11}(z_\alpha) \\
 &\quad + \hat{q}_{11}(z_\alpha) \} + O_p(n^{-1})].
 \end{aligned}$$

Noting that $z_\alpha = -z_{1-\alpha}$ and q_{11} is an even polynomial, the “ideal” interval, $\mathcal{I}$, has endpoint $\bar{X} - \bar{Y} + \hat{\sigma} \{ z_{1-\alpha} - n^{-\frac{1}{2}} q_{11}(z_{1-\alpha}) + O(n^{-1}) \}$. Comparing this with the formula above for $\hat{u}_{\hat{\beta}_a(1-\alpha)}$, with α replaced by $1 - \alpha$, we see that for second-order correctness we require

$$2\hat{p}_1(0) + z_\alpha^2 \hat{c} + \hat{p}_{11}(z_\alpha) + \hat{q}_{11}(z_\alpha) = 0 ;$$

that is,

$$\begin{aligned}
 \hat{c} &\equiv z_\alpha^{-2} \{ \hat{p}_1(z_\alpha) + \hat{q}_1(z_\alpha) - 2\hat{p}_1(0) \} \\
 &= n^{\frac{1}{2}} \cdot \frac{1}{6} \hat{\sigma}^{-3/2} (m^{-2} \hat{\gamma}_X - n^{-2} \hat{\gamma}_Y) ,
 \end{aligned}$$

on noting (2.10) and formulae for p_1 and q_1 in Theorem 2.1.

6. LENGTHS OF TWO-SIDED, EQUAL-TAILED INTERVALS

The Cornish-Fisher expansions developed in Section 3 allow us to deduce that the endpoints $\hat{b}_1$ and $\hat{b}_2$, of a two-sided, equal-tailed, $(1 - \alpha)$ -level bootstrap confidence interval $[\hat{b}_1, \hat{b}_2]$ have the form

$$\begin{aligned}
 \hat{b}_i &= \bar{X} - \bar{Y} + \hat{\sigma} \{ (-1)^i z_{1-(\alpha/2)} + n^{-\frac{1}{2}} \hat{s}_1(z_{1-(\alpha/2)}) + (-1)^i n^{-1} \hat{s}_2(z_{1-(\alpha/2)}) \\
 &\quad + O_p(n^{-3/2}) \} ,
 \end{aligned} \tag{2.20}$$

where $\hat{s}_1$ is an even polynomial and $\hat{s}_2$ is an odd polynomial; see for example equation (2.16). The polynomials $\hat{s}_1$ and $\hat{s}_2$ are “sample” versions of “theoretical” polynomials, s_1 and s_2 ; that is, $\hat{s}_1$ and $\hat{s}_2$ are obtained from s_1 and s_2 on replacing population moments by sample moments. Since sample moments converge to population moments at rate $n^{-\frac{1}{2}}$, then (2.20) may be written as

$$\begin{aligned}
 \hat{b}_i &= \bar{X} - \bar{Y} + \hat{\sigma} \{ (-1)^i z_{1-(\alpha/2)} + n^{-\frac{1}{2}} \hat{s}_1(z_{1-(\alpha/2)}) + (-1)^i n^{-1} s_2(z_{1-(\alpha/2)}) \\
 &\quad + O_p(n^{-3/2}) \} .
 \end{aligned} \tag{2.21}$$

From (2.21), it follows that the length of the interval $[\hat{b}_1, \hat{b}_2]$ is given by

$$(\hat{b}_2 - \hat{b}_1) = 2\hat{\sigma}\{z_{1-(\alpha/2)} + n^{-1}s_2(z_{1-(\alpha/2)})\} + O_p(n^{-2}). \quad (2.22)$$

Bearing in mind that the “ideal” interval, $\mathcal{I}_1 \equiv [\bar{X} - \bar{Y} - y_{1-(\alpha/2)}\hat{\sigma}, \bar{X} - \bar{Y} - y_{\alpha/2}\hat{\sigma}]$, has length $2\hat{\sigma}\{z_{1-(\alpha/2)} + n^{-1}q_{21}(z_{1-(\alpha/2)})\} + O(n^{-2})$, and that the percentile- t alternative I_3 (defined in Section 2) has $s_2 = q_{21}$, it follows from (2.22) that the lengths of $\mathcal{I}_1$ and I_3 differ only by a term of $O_p(n^{-2})$. In general, none of the other bootstrap intervals track the “ideal” interval, $\mathcal{I}_1$, as closely as this; the error in length is usually $O_p(n^{-3/2})$, since in those cases, $s_2 \neq q_{21}$.

7. COVERAGE ERROR

As in Section 6, using the Cornish-Fisher expansions developed in Section 3, we deduce that the endpoint, $\hat{b}$, of a one-sided, $(1 - \alpha)$ -level bootstrap confidence interval $(-\infty, \hat{b}]$ has the form

$$\hat{b} = \bar{X} - \bar{Y} + \hat{\sigma}\{z_{1-\alpha} + n^{-\frac{1}{2}}\hat{s}_1(z_{1-\alpha}) + n^{-1}\hat{s}_2(z_{1-\alpha}) + O_p(n^{-3/2})\}, \quad (2.23)$$

where $\hat{s}_1$ and $\hat{s}_2$ are polynomials described in Section 6. In this section we show that the chance that the interval $(-\infty, \hat{b}]$ covers the mean difference, $\mu_X - \mu_Y$, equals

$$\begin{aligned} P(\mu_X - \mu_Y \leq \hat{b}) &= 1 - \alpha + n^{-\frac{1}{2}}\{s_1(z_{1-\alpha}) - q_1(z_{1-\alpha})\}\phi(z_{1-\alpha}) \\ &\quad - n^{-1}\left[\frac{1}{2}\{s_1(z_{1-\alpha})\}^2 + s_1(z_{1-\alpha})\{q_1'(z_{1-\alpha}) - q_1(z_{1-\alpha})z_{1-\alpha}\} \right. \\ &\quad \left. - q_2(z_{1-\alpha}) - s_2(z_{1-\alpha}) + uz_{1-\alpha}\right]\phi(z_{1-\alpha}) \\ &\quad + O(n^{-3/2}), \end{aligned} \quad (2.24)$$

where $u = u(\alpha)$ is the constant such that

$$n^{\frac{1}{2}}E[T\{\hat{s}_1(z_{1-\alpha}) - s_1(z_{1-\alpha})\}] = u + O(n^{-1}). \quad (2.25)$$

Since sample moments converge at rate $n^{-\frac{1}{2}}$ to population moments, we may write (2.23) as

$$\hat{b} = \bar{X} - \bar{Y} + \hat{\sigma}\{z_{1-\alpha} + n^{-\frac{1}{2}}\hat{s}_1(z_{1-\alpha}) + n^{-1}s_2(z_{1-\alpha}) + O_p(n^{-3/2})\}.$$

Under the regularity conditions imposed prior to Theorem 2.1, using the argument employed in Step (iv) of the proof of Theorem 2.1 of Hall (1986a), it follows

that the $O_p(n^{-3/2})$ term makes a $O(n^{-3/2})$ contribution to coverage probability, so that the interval $(-\infty, \hat{b}]$ covers the mean difference with probability

$$P \left[T + n^{-\frac{1}{2}} \{ \hat{s}_1(z_{1-\alpha}) - s_1(z_{1-\alpha}) \} \geq -z_{1-\alpha} - n^{-\frac{1}{2}} s_1(z_{1-\alpha}) - n^{-1} s_2(z_{1-\alpha}) \right] + O(n^{-3/2}).$$

The following theorem, proved in Section 11, allows us to simplify the last-written probability.

THEOREM 2.2. Put $\tilde{T} \equiv T + n^{-\frac{1}{2}} \{ \hat{s}_1(z_{1-\alpha}) - s_1(z_{1-\alpha}) \}$, and define u by (2.25). Then, under the regularity conditions stated prior to Theorem 2.1,

$$P(\tilde{T} \leq x) = P(T \leq x) - n^{-1} u x \phi(x) + O(n^{-3/2}).$$

Taking $x = -z_{1-\alpha} - n^{-1/2} s_1(z_{1-\alpha}) - n^{-1} s_2(z_{1-\alpha})$ in Theorem 2.2, noting expansion (2.7) for $P(T \leq x)$, and Taylor-expanding, we obtain (2.23).

When the interval $(-\infty, \hat{b}]$ is second-order correct (as is the case for percentile- t and accelerated bias-corrected intervals), $s_1(z_{1-\alpha})$ must be identical to $-q_{11}(z_\alpha)$, which, in turn, equals $q_1(z_{1-\alpha})$ (since q_1 is even and $q_{11} \equiv -q_1$). In that circumstance, it follows from (2.23) that the interval $(-\infty, \hat{b}]$ has coverage error $O(n^{-1})$, whereas the error is $O(n^{-\frac{1}{2}})$ when the interval fails to be second-order correct.

The issue of second-order correctness is not as crucial when considering two-sided intervals $[\hat{b}_1, \hat{b}_2]$ where $\hat{b}_i$ is defined by (2.20). In that case, the probability that the interval $[\hat{b}_1, \hat{b}_2]$ covers the mean difference equals

$$\begin{aligned} P(\hat{b}_1 \leq \mu_X - \mu_Y \leq \hat{b}_2) &= 1 - \alpha - 2n^{-1} \left[\frac{1}{2} \{ s_1(z_{1-(\alpha/2)}) \}^2 z_{1-(\alpha/2)} \right. \\ &\quad + s_1(z_{1-(\alpha/2)}) \{ q_1'(z_{1-(\alpha/2)}) - q_1(z_{1-(\alpha/2)}) z_{1-(\alpha/2)} \} \\ &\quad \left. - q_2(z_{1-(\alpha/2)}) - s_2(z_{1-(\alpha/2)}) + u z_{1-(\alpha/2)} \right] \phi(z_{1-(\alpha/2)}) \\ &\quad + O(n^{-2}), \end{aligned} \tag{2.26}$$

by noticing that the polynomial $s_1 - q_1$ appearing in (2.24) is even, and that the $O(n^{-3/2})$ remainder in (2.24) may be written as $n^{-3/2} r(z_{1-(\alpha/2)}) \phi(z_{1-(\alpha/2)}) + O(n^{-2})$, where r is an even polynomial; see equation (1.24) of Chapter 1. We see from (2.26) that two-sided percentile-method, percentile- t and accelerated bias-corrected intervals all have coverage error $O(n^{-1})$. In the percentile- t case,

the constant u , defined by (2.25) using $z_{1-(\alpha/2)}$ rather than $z_{1-\alpha}$, is shown in Appendix A2.2 to be

$$\frac{1}{6}(2z_{1-(\alpha/2)}^2 + 1)\{(\sigma_X^2 + r\sigma_Y^2)^{-2}(r^{-1}\kappa_X + r^2\kappa_Y) - \frac{3}{2}(\sigma_X^2 + r\sigma_Y^2)^{-3}(r^{-\frac{1}{2}}\gamma_X - r^{3/2}\gamma_Y)^2\}, \quad (2.27)$$

which vanishes if $\gamma_X, \gamma_Y, \kappa_X$ and κ_Y are all zero. Likewise, in the percentile- t case, each of q_1 and q_2 is identically zero if first, third and fourth cumulants vanish. Consequently, s_1 and s_2 are also identically zero, and so from (2.24) we see that $P(\mu_X - \mu_Y \leq \hat{b}) = 1 - \alpha + O(n^{-3/2})$. Lengthy algebra, given in the proof of Lemma 2.4 in Section 11, shows that when first, third, fourth and fifth cumulants vanish (for example, when both populations are Normal), the $n^{-3/2}$ term is also zero, and

$$P(\mu_X - \mu_Y \leq \hat{b}) = 1 - \alpha + O(n^{-2}).$$

Indeed, this level of accuracy is exhibited by each of the percentile- t intervals I_1, I_2 and I_3 , when the underlying population is Normal. However, this level of accuracy is not available for accelerated bias-corrected intervals since there, the polynomial s_2 does not vanish when cumulants are set equal to zero. When used with Normal populations, accelerated bias-corrected one- and two-sided intervals have coverage accuracy no better than order n^{-1} .

8. SYMMETRIC CONFIDENCE INTERVALS

The two-sided percentile- t interval I_3 is equal-tailed since it places equal probability $\alpha/2$ in each tail. We show now how to construct a two-sided interval, L , which is symmetric about the sample mean difference. Put

$$\hat{w}_\alpha \equiv \sup\{w : P(|T^*| \leq w \mid \mathcal{X}, \mathcal{Y}) \leq \alpha\},$$

which estimates the percentile w_α defined by $P(|T| \leq w_\alpha) = \alpha$. Then L is given by

$$L \equiv [\bar{X} - \bar{Y} - \hat{w}_{1-(\alpha/2)}\hat{\sigma}, \bar{X} - \bar{Y} - \hat{w}_{\alpha/2}\hat{\sigma}].$$

Note that L is a percentile- t interval, since it is based on the Studentized statistic, T^* . The interval L has nominal coverage $1 - \alpha$, and as we show in the next paragraph, it has coverage error $O(n^{-2})$, even without the assumption of Normality. In this sense, it represents an improvement on equal-tailed percentile- t

intervals which were shown in Section 7 to have coverage accuracy only of order n^{-1} .

Since in an Edgeworth expansion of $P(T \leq x)$, the odd-indexed q_j 's are even functions, an Edgeworth expansion of $P(|T| \leq x)$ is a power series in n^{-j} , containing no terms in $n^{-j/2}$ for odd j . Inverting this expression, we see that a Cornish-Fisher expansion of $w_{1-(\alpha/2)}$ is also a power series in n^{-j} , the first two terms being $w_{1-(\alpha/2)} = z_{1-(\alpha/2)} + n^{-1}r(z_{1-(\alpha/2)}) + O(n^{-2})$, where r is an odd polynomial. Then $\hat{w}_{1-(\alpha/2)} = z_{1-(\alpha/2)} + n^{-1}\hat{r}(z_{1-(\alpha/2)}) + O_p(n^{-2})$, whence

$$\begin{aligned} P(\mu_X - \mu_Y \in L) &= P(|T| \leq \hat{w}_{1-(\alpha/2)}) \\ &= P\{|T| \leq z_{1-(\alpha/2)} + n^{-1}\hat{r}(z_{1-(\alpha/2)})\} + O(n^{-2}) \\ &= P(T_- \leq w_{1-(\alpha/2)}) - P(T_+ \leq -w_{1-(\alpha/2)}) \\ &\quad + O(n^{-2}), \end{aligned} \tag{2.28}$$

where $T_{\pm} = T \pm n^{-1}\{\hat{r}(z_{1-(\alpha/2)}) - r(z_{1-(\alpha/2)})\}$. An argument similar to that used to prove Theorem 2.2 shows that

$$P(T_{\pm} \leq x) = P(T \leq x) + n^{-3/2}a(x)\phi(x) + O(n^{-2}), \tag{2.29}$$

where a is an odd polynomial. The same technique is used in Section 3 of Hall (1988b) for the one-sample case. The desired result, $P(\mu_X - \mu_Y \in L) = 1 - \alpha + O(n^{-2})$, follows on taking $x = \mp w_{1-(\alpha/2)}$ in (2.29) and substituting into (2.28).

9. SHORT CONFIDENCE INTERVALS

We next show how to construct "short" confidence intervals for the mean difference. By "short" we mean in terms of actual interval length, although we could just as easily use *expected* interval length. Discussions of "short" confidence intervals are given by Pratt (1961, 1963), Harter (1964), Guenther (1969), Wilson and Tonascia (1971), and Kendall and Stuart (1977), vol.2, pp.125-129.

In "ideal" circumstances, for example, if we know the distribution of T , we may choose a, b to minimize $a + b$ subject to

$$P(-b \leq T \leq a) = 1 - \alpha, \tag{2.30}$$

and call $\mathcal{I}_0 \equiv [\bar{X} - \bar{Y} - \hat{\sigma}a, \bar{X} - \bar{Y} + \hat{\sigma}b]$ the "shortest" $(1 - \alpha)$ -level, two-sided confidence interval for $\mu_X - \mu_Y$. Since the distribution function of T admits an

Edgeworth expansion given by (2.7), the points a and b both converge to the solution $z_{1-(\alpha/2)}$ of the equation $\Phi(z_{1-(\alpha/2)}) = 1 - \alpha/2$, and we may write

$$a \sim z_{1-(\alpha/2)} + \sum_{j \geq 1} n^{-j/2} a_j, \quad b \sim z_{1-(\alpha/2)} + \sum_{j \geq 1} n^{-j/2} b_j, \quad (2.31)$$

where these expansions are to be interpreted as asymptotic series; see Chapter 1 for the definition of an asymptotic series. Observing that $[-(\bar{X} - \bar{Y}) - \hat{\sigma}b, -(\bar{X} - \bar{Y}) + \hat{\sigma}a]$ is the shortest confidence interval for $\mu_Y - \mu_X$, and that, since even-indexed q_j 's are odd functions and odd-indexed q_j 's are even functions,

$$P(-T \leq x) = \Phi(x) + (-n^{-\frac{1}{2}})q_1(x)\phi(x) + n^{-1}q_2(x)\phi(x) + O(n^{-3/2}),$$

we have $b \sim z_{1-(\alpha/2)} + \sum_{i \geq 1} (-n^{-\frac{1}{2}})^i a_i$ and $a \sim z_{1-(\alpha/2)} + \sum_{i \geq 1} (-n^{-\frac{1}{2}})^i b_i$. Therefore $b_i = (-1)^i a_i$, for $i \geq 1$. Put

$$\begin{aligned} \phi_0(x) &\equiv \Phi(x), \quad \phi_i(x) \equiv q_i(x)\phi(x) \quad \text{for } i \geq 1, \quad \phi_{ik}(x) \equiv (\partial/\partial x)^k \phi_i(x), \\ \text{and } \psi_{ik} &\equiv \phi_{ik}(z_{1-(\alpha/2)}). \end{aligned} \quad (2.32)$$

Noting that ϕ_i is an odd function for even i and an even function for odd i , it follows using (2.7) and (2.31) in (2.30) that

$$\begin{aligned} 1 - \alpha &= P(T \leq a) - P(T \leq -b) \\ &\sim \sum_{i \geq 0} n^{-i/2} \left[\phi_i \left(z_{1-(\alpha/2)} + \sum_{j \geq 1} n^{-j/2} a_j \right) \right. \\ &\quad \left. - \phi_i \left\{ -z_{1-(\alpha/2)} - \sum_{j \geq 1} (-n^{-\frac{1}{2}})^j a_j \right\} \right] \\ &\sim \sum_{i \geq 0} \sum_{k \geq 0} n^{-i/2} \frac{1}{k!} \left[\left(\sum_{j \geq 1} n^{-j/2} a_j \right)^k + (-1)^i \left\{ \sum_{j \geq 1} (-n^{-\frac{1}{2}})^j a_j \right\}^k \right] \psi_{ik}. \end{aligned}$$

On the right-hand side, all coefficients of $n^{-j/2}$ for odd j are identically zero. Equating to zero the coefficients of n^{-l} for $l \geq 1$, we conclude that

$$\psi_{2l,0} + \sum_{i=1}^{2l-1} \sum_{k=1}^{2l-i} \frac{1}{k!} \psi_{ik} \sum_{(k,2l-i)} a_{j_1} \cdots a_{j_k} = 0, \quad (2.33)$$

where $\sum_{(k,r)}$ denotes summation over $j_1, \dots, j_k \geq 1$ such that $j_1 + \cdots + j_k = r$. The requirement that $a + b$ be minimized is equivalent to the constraint that a_{2j} , $j \geq 1$ be minimized, and subject to this constraint, solutions $a_1, a_2, \dots$ of (2.33)

may be obtained recursively. For our purposes, it is sufficient to know only the first two a_j 's:

$$a_1 \equiv -\psi_{11}\psi_{02}^{-1}, \quad a_2 \equiv (\tfrac{1}{2}\psi_{11}^2\psi_{02}^{-1} - \psi_{20})\psi_{01}^{-1}. \quad (2.34)$$

The above derivation comes from Hall (1989). By substituting the formulae in Theorem 2.1 into (2.32) and the resulting formulae into (2.34), we get

$$a_1 \equiv -\tfrac{1}{6}(\sigma_X^2 + r\sigma_Y^2)^{-3/2}(r^{-\frac{1}{2}}\gamma_X - r^{3/2}\gamma_Y)(2z_{1-(\alpha/2)}^2 - 3),$$

and

$$\begin{aligned} a_2 \equiv & z_{1-(\alpha/2)} \left[-\tfrac{1}{12}(\sigma_X^2 + r\sigma_Y^2)^{-2}(r^{-1}\kappa_X + r^2\kappa_Y)(z_{1-(\alpha/2)}^2 - 3) \right. \\ & + \tfrac{1}{72}(\sigma_X^2 + r\sigma_Y^2)^{-3}(r^{-\frac{1}{2}}\gamma_X - r^{3/2}\gamma_Y)^2(20z_{1-(\alpha/2)}^2 - 21) \\ & + \tfrac{1}{4}(\sigma_X^2 + r\sigma_Y^2)^{-2}\{z_{1-(\alpha/2)}^2(r^{-1}\sigma_X^4 + r^2\sigma_Y^4) \\ & \left. + 3r^{-1}\sigma_X^4 + 2\sigma_X^2\sigma_Y^2(1+r) + 3r^2\sigma_Y^4\} \right]. \end{aligned}$$

The percentile- t alternative to $\mathcal{I}_0$ is $I_0 \equiv [\bar{X} - \bar{Y} - \hat{\sigma}\hat{a}, \bar{X} - \bar{Y} + \hat{\sigma}\hat{b}]$, where $\hat{a}$ and $\hat{b}$ are chosen to minimize $\hat{a} + \hat{b}$ subject to

$$P(-\hat{b} \leq T^* \leq \hat{a} \mid \mathcal{X}, \mathcal{Y}) = 1 - \alpha.$$

Such $\hat{a}$ and $\hat{b}$ admit expansions

$$\begin{aligned} \hat{a} &= z_{1-(\alpha/2)} + n^{-\frac{1}{2}}\hat{a}_1 + n^{-1}\hat{a}_2 + O_p(n^{-3/2}), \\ \hat{b} &= z_{1-(\alpha/2)} - n^{-\frac{1}{2}}\hat{a}_1 + n^{-1}\hat{a}_2 + O_p(n^{-3/2}), \end{aligned}$$

where $\hat{a}_1$ and $\hat{a}_2$ are obtained from a_1 and a_2 by replacing population moments by sample moments. Therefore,

$$\begin{aligned} P(\mu_X - \mu_Y \in I_0) &= P(-\hat{b} \leq T \leq \hat{a}) \\ &= P(R \leq a) - P(R \leq -b) + O(n^{-3/2}), \end{aligned} \quad (2.35)$$

where $R \equiv T + n^{-\frac{1}{2}}(a_1 - \hat{a}_1)$. Note that, under the regularity conditions imposed prior to Theorem 2.1, the $O_p(n^{-3/2})$ term makes a $O(n^{-3/2})$ contribution to coverage probability.

By the argument used to prove Theorem 2.2, we have

$$P(R \leq x) = P(T \leq x) - n^{-1}vx\phi(x) + O(n^{-3/2}), \quad (2.36)$$

where v is the constant such that

$$n^{\frac{1}{2}} E \{T(a_1 - \hat{a}_1)\} = v + O(n^{-1}). \quad (2.37)$$

We show in Appendix A2.2 that, in the percentile- t case, v is given by

$$v \equiv \frac{1}{6}(2z_{1-(\alpha/2)}^2 - 3)\{(\sigma_X^2 + r\sigma_Y^2)^{-2}(r^{-1}\kappa_X + r^2\kappa_Y) - \frac{3}{2}(\sigma_X^2 + r\sigma_Y^2)^{-3}(r^{-\frac{1}{2}}\gamma_X - r^{3/2}\gamma_Y)^2\}.$$

Combining (2.35) and (2.36), and using the fact that a and b are distant only $O(n^{-\frac{1}{2}})$ from $z_{1-(\alpha/2)}$, we conclude that the coverage probability of I_0 equals

$$\begin{aligned} P(\mu_X - \mu_Y \in I_0) &= P(-b \leq T \leq a) - 2n^{-1}vz_{1-(\alpha/2)}\phi(z_{1-(\alpha/2)}) + O(n^{-3/2}) \\ &= 1 - \alpha - 2n^{-1}vz_{1-(\alpha/2)}\phi(z_{1-(\alpha/2)}) + O(n^{-3/2}). \end{aligned} \quad (2.38)$$

In comparison, the coverage probability of the usual percentile- t interval I_3 , computed by substituting $s_1 \equiv -q_{11}$ and $s_2 \equiv -q_{21}$ into (2.26), is

$$P(\mu_X - \mu_Y \in I_3) = 1 - \alpha - 2n^{-1}uz_{1-(\alpha/2)}\phi(z_{1-(\alpha/2)}) + O(n^{-3/2}), \quad (2.39)$$

where u is defined by (2.25). Combining (2.38) and (2.39), and using formulae for u and v in the percentile- t case, we find that the ratio of coverage errors for shortest and equal-tailed percentile- t confidence intervals converges to

$$\lim_{n \rightarrow \infty} \frac{P(\mu_X - \mu_Y \in I_0) - (1 - \alpha)}{P(\mu_X - \mu_Y \in I_3) - (1 - \alpha)} = \frac{2z_{1-(\alpha/2)}^2 - 3}{2z_{1-(\alpha/2)}^2 + 1}.$$

This ratio is always less than one, meaning that not only is the shortest confidence interval shorter than its equal-tailed counterpart, but it also enjoys better coverage accuracy, at least in large samples.

10. SIMULATION STUDY

To ascertain whether our “small sample asymptotics” reflect properties of actual sample sizes, we conducted a simulation study for samples of sizes between 8 and 12, taking $m = n$ throughout. We treated two cases: firstly where both populations are Normal $N(0, 1)$, and secondly, where the x -population is negative folded Normal $-|N(0, 1)|$ and the y -population is folded Normal $|N(0, 1)|$. For each value of $m(=n) = 8, 10, 12$, we drew 4,000 independent paired samples $(\mathcal{X}, \mathcal{Y})$, and for each pair we constructed bootstrap critical points using $B = 399$

resampling operations. Our estimates of coverage error, mean interval length and standard deviation of interval length were averages over the 4,000 pairs. In the case of one-sided intervals, "length" was defined as distance of interval endpoint from sample mean difference, $\bar{X} - \bar{Y}$. All intervals had nominal coverage 95% and all one-sided intervals were of the form $(-\infty, \hat{b}]$. Computations were carried out on a VAX 11/785 computer, and Normal pseudo-random values were generated using NAG routine G05DDF.

Table 2.1 summarizes the results of the study, which are in general agreement with the conclusions of our asymptotic analysis. With Normal data, all percentile- t methods (including the symmetric method) work exceptionally well, notably better than ordinary percentile methods. With folded Normal data, percentile- t better accounts for skewness than do the ordinary percentile methods, and the symmetric interval provides slightly better coverage accuracy than does the equal-tailed interval. (The case of Normal populations is an exception; there equal-tailed and symmetric intervals both have coverage error $O(n^{-2})$.) The "shortest" intervals generally have high coverage accuracy, although not consistently higher than their equal-tailed counterparts, reflecting the point that larger sample sizes are required for the "short" intervals to enjoy better coverage accuracy.

11. PROOFS

Proof of Theorem 2.1. We shall use the so-called "delta method", described briefly by Bhattacharya and Ghosh (1978), p.435, to identify Edgeworth expansions of the distribution functions of S and T . Assuming that (X, X^2) and (Y, Y^2) each satisfy bivariate versions of Cramér's condition, and that $E(X^{10}) < \infty$ and $E(Y^{10}) < \infty$ we may apply Theorem 2(b) of Bhattacharya and Ghosh (1978), with $s = 5$, to verify that remainder terms in (2.6) and (2.7) are of the stated order. (In the case of (2.6), the conditions introduced prior to Theorem 2.1 may be relaxed. There, it is only necessary to assume that X and Y satisfy Cramér's condition and that $E(X^5) < \infty$ and $E(Y^5) < \infty$, and then use Bhattacharya and Rao (1976), Corollary 20.4, p.215 with $s = 5$.) To identify the polynomials p_1, p_2, q_1 and q_2 in (2.6)–(2.9), it is sufficient to compute the first four cumulants of S and T . This will be done next.

Put $U_1 \equiv m^{-\frac{1}{2}} \sum_i (X_i^2 - \sigma_X^2)$, $U_2 \equiv m^{-\frac{1}{2}} \sum_i X_i$, $V_1 \equiv n^{-\frac{1}{2}} \sum_j (Y_j^2 - \sigma_Y^2)$ and $V_2 \equiv n^{-\frac{1}{2}} \sum_j Y_j$. Assuming, without loss of generality, that $\mu_X - \mu_Y = 0$, U_1, U_2, V_1 and V_2 are sums of independent random variables with zero means.

TABLE 2.1. *Summary of simulation study.* Data are averages of 4,000 simulations of each instance, 399 resampling operations being used to compute each critical point. Abbreviations cov., lth., s.d. indicate coverage probability, mean length, standard deviation of length. Abbreviations N.app., p'ile(1), p'ile(2), p'ile- t , symm. and short indicate interval type, being naïve Normal approximation, type 1 percentile (intervals J_{1i}), type 2 percentile (intervals J_{2i}), percentile- t (intervals I_i), symmetric (interval L), and shortest confidence interval (interval I_0).

n	One-sided intervals				Two-sided intervals					
	N.app.	p'ile(1)	p'ile(2)	p'ile- t	N.app.	p'ile(1)	p'ile(2)	p'ile- t	symm.	short
Normal populations										
8	cov.	.921	.923	.923	.948	.911	.912	.905	.952	.958
	lth.	.757	.763	.761	.896	1.80	1.82	1.82	2.23	2.22
	s.d.	.146	.155	.157	.210	.348	.360	.360	.465	.465
10	cov.	.926	.926	.925	.947	.913	.912	.910	.944	.950
	lth.	.688	.692	.693	.783	1.64	1.65	1.65	1.92	1.91
	s.d.	.116	.126	.126	.161	.277	.290	.290	.345	.347
12	cov.	.932	.933	.932	.949	.921	.923	.927	.951	.955
	lth.	.636	.640	.639	.704	1.52	1.53	1.53	1.73	1.72
	s.d.	.097	.105	.105	.128	.231	.242	.242	.280	.278
Folded Normal populations										
8	cov.	.901	.904	.893	.949	.903	.904	.895	.954	.949
	lth.	.320	.331	.312	.455	.763	.765	.765	.993	1.00
	s.d.	.071	.080	.070	.165	.169	.173	.173	.280	.310
10	cov.	.903	.910	.896	.949	.914	.914	.906	.956	.949
	lth.	.292	.302	.285	.387	.695	.699	.699	.848	.851
	s.d.	.057	.065	.057	.111	.137	.142	.142	.200	.209
12	cov.	.914	.920	.909	.951	.917	.918	.916	.957	.952
	lth.	.270	.279	.263	.345	.643	.647	.647	.757	.758
	s.d.	.047	.054	.048	.087	.113	.119	.119	.157	.164

Put $R_1 \equiv \sigma^{-3}(m^{-3/2}U_1 + n^{-3/2}V_1)$, $R_2 \equiv \sigma^{-4}(m^{-2}U_2^2 + n^{-2}V_2^2)$ and $Z \equiv \sigma^{-1}(m^{-1/2}U_2 - n^{-1/2}V_2)$. Notice that $S = Z$. Then

$$\begin{aligned}\hat{\sigma}^{-1} &= (\hat{\sigma}_X^2 m^{-1} + \hat{\sigma}_Y^2 n^{-1})^{-\frac{1}{2}} = (\sigma^2 + \sigma^3 R_1 - \sigma^4 R_2)^{-\frac{1}{2}} \\ &= \sigma^{-1} \left\{ 1 - \frac{1}{2} \sigma R_1 + \sigma^2 \left(\frac{3}{8} R_1^2 + \frac{1}{2} R_2 \right) + O_p(n^{-3/2}) \right\},\end{aligned}$$

whence

$$T = Z \left\{ 1 - \frac{1}{2} \sigma R_1 + \sigma^2 \left(\frac{3}{8} R_1^2 + \frac{1}{2} R_2 \right) + O_p(n^{-3/2}) \right\}. \quad (2.40)$$

It is a tedious matter to obtain the first four moments of T and S , and we give these in Lemma 2.3, leaving a proof of this Lemma to Appendix A2.1.

LEMMA 2.3. *The first four moments of T , given by (2.40), are*

$$\begin{aligned}E(T) &= -n^{-\frac{1}{2}} \frac{1}{2} s^{-3} g + O(n^{-3/2}), \\ E(T^2) &= 1 + n^{-1} (2s^{-6} g^2 + s^{-4} h^2) + O(n^{-2}), \\ E(T^3) &= -n^{-\frac{1}{2}} \frac{7}{2} s^{-3} g + O(n^{-3/2}),\end{aligned} \quad (2.41)$$

and

$$E(T^4) = 3 + n^{-1} (28s^{-6} g^2 + 6s^{-4} k^2 - 2s^{-4} l^2) + O(n^{-2}),$$

where $s^2 \equiv \sigma_X^2 + r\sigma_Y^2$, $g \equiv r^{-\frac{1}{2}} \gamma_X - r^{3/2} \gamma_Y$, $h^2 \equiv 3r^{-1} \sigma_X^4 + (r+1) \sigma_X^2 \sigma_Y^2 + 3r^2 \sigma_Y^4$, $k^2 \equiv 5r^{-1} \sigma_X^4 + (r+1) \sigma_X^2 \sigma_Y^2 + 5r^2 \sigma_Y^4$ and $l^2 \equiv r^{-1} E(X^4) + r^2 E(Y^4)$. The first four moments of S are

$$E(S) = 0, E(S^2) = 1, E(S^3) = n^{-\frac{1}{2}} s^{-3} g,$$

and

$$E(S^4) = 3 + n^{-1} (s^{-4} l^2 - 3s^{-4} a) = 3 + n^{-1} s^{-4} b,$$

} (2.42)

where $a \equiv r^{-1} \sigma_X^4 + r^2 \sigma_Y^4$, and $b \equiv r^{-1} \kappa_X - r^2 \kappa_Y$.

Since the first four cumulants of T may be expressed in terms of the first four moments of T via the formulae $k_1(t) = E(T)$, $k_2(t) = E(T^2) - (ET)^2$, $k_3(t) = E(T^3) - 3E(T^2)E(T) + 2(ET)^3$, and $k_4(T) = E(T^4) - 4E(T^3)E(T) - 3(ET^2)^2 + 12E(T^2)(ET)^2 - 6(ET)^4$ (see Kendall and Stuart (1977), vol.1, p.72), then the first four cumulants of T are

$$\begin{aligned}k_1(T) &= -n^{-\frac{1}{2}} \frac{1}{2} s^{-3} g + O(n^{-3/2}), \quad k_3(T) = -n^{-\frac{1}{2}} 2s^{-3} g + O(n^{-3/2}), \\ k_2(T) &= 1 + n^{-1} \left\{ \frac{7}{4} s^{-6} g^2 + s^{-4} h^2 \right\} + O(n^{-2}), \quad \text{and} \\ k_4(T) &= n^{-1} (12s^{-6} g^2 + 6s^{-4} a^2 - 2s^{-4} b) + O(n^{-2}).\end{aligned}$$

Applying the same procedure to the moments of S gives

$$k_1(S) = 0, \quad k_2(S) = 1, \quad k_3(S) = n^{-\frac{1}{2}}s^{-3}g, \quad \text{and } k_4(S) = n^{-1}s^{-4}b.$$

The characteristic functions of T and S are

$$\begin{aligned} E(e^{itT}) = \exp\{itk_1(T) + \frac{1}{2}(it)^2k_2(T) + \frac{1}{6}(it)^3k_3(T) + \frac{1}{24}(it)^4k_4(T) \\ + O(n^{-3/2})\}, \end{aligned} \quad (2.43a)$$

$$\begin{aligned} E(e^{itS}) = \exp\{itk_1(S) + \frac{1}{2}(it)^2k_2(S) + \frac{1}{6}(it)^3k_3(S) + \frac{1}{24}(it)^4k_4(S) \\ + O(n^{-3/2})\}. \end{aligned} \quad (2.43b)$$

Substituting formulae for cumulants given above into (2.43a) and (2.43b), we may expand the right-hand sides of (2.43a) and (2.43b) as

$$\begin{aligned} E(e^{itT}) &= \exp\left[-t^2/2 - n^{-\frac{1}{2}}\left\{\frac{1}{2}it + \frac{1}{3}(it)^3\right\}s^{-3}g \right. \\ &\quad \left. + n^{-1}\left\{\frac{1}{2}(it)^2\left(\frac{7}{4}s^{-6}g^2 + s^{-4}h^2\right) \right. \right. \\ &\quad \left. \left. + \frac{1}{24}(it)^4(12s^{-6}g^2 + 6s^{-4}a^2 - 2s^{-4}b)\right\} + O(n^{-3/2})\right] \\ &= e^{-t^2/2} \left[1 - n^{-\frac{1}{2}}s^{-3}g\left\{\frac{1}{2}it + \frac{1}{3}(it)^3\right\} + n^{-1}\left\{(it)^2(s^{-6}g^2 + \frac{1}{2}s^{-4}h^2) \right. \right. \\ &\quad \left. \left. + \frac{1}{12}(it)^4(8s^{-6}g^2 + 3s^{-4}a^2 - s^{-4}b) + \frac{1}{18}(it)^6s^{-6}g^2\right\} \right. \\ &\quad \left. + O(n^{-3/2})\right], \end{aligned} \quad (2.44a)$$

and

$$\begin{aligned} E(e^{itS}) &= \exp\left\{-s^2/2 + n^{-\frac{1}{2}}\frac{1}{6}(it)^3s^{-3}g + n^{-1}\frac{1}{24}(it)^4s^{-4}b + O(n^{-3/2})\right\} \\ &= e^{-s^2/2} \left[1 + n^{-\frac{1}{2}}\frac{1}{6}(it)^3s^{-3}g + n^{-1}\left\{\frac{1}{24}(it)^4s^{-4}b + \frac{1}{72}(it)^6s^{-6}g^2\right\} \right. \\ &\quad \left. + O(n^{-3/2})\right], \end{aligned} \quad (2.44b)$$

respectively. Inverting these Fourier transforms, and recalling from Chapter 1 that the inverse Fourier transform of $(it)^je^{-\frac{1}{2}t^2}$ is $-He_{j-1}(x)\phi(x)$ for $j \geq 1$, where $He_i(x)$ is the i 'th Hermite polynomial (see, for example, Petrov (1975), p.137), we arrive at (2.6) and (2.7) with the stated formulae for q_1 , p_1 , q_2 and p_2 .

Proof of Theorem 2.2. We prove Theorem 2.2 by relating Edgeworth expansions of the distribution functions of T and $\tilde{T}$. It suffices to show that first, third and fourth cumulants of T and $\tilde{T}$ agree up to (but not necessarily including) terms of order $n^{-3/2}$ — that is, $k_j(\tilde{T}) = k_j(T) + O(n^{-3/2})$ for $j = 1, 3$, and 4 — and that $k_2(\tilde{T}) = k_2(T) + 2n^{-1}u + O(n^{-2})$. For then, if we express characteristic

functions of T and $\tilde{T}$ as in (2.43a), expand in polynomials as in (2.44a), invert to obtain Edgeworth expansions for the distributions of T and $\tilde{T}$, and compare formulae, we obtain Theorem 2.2.

Put $U \equiv n^{\frac{1}{2}}\{\hat{s}_1(z_{1-\alpha}) - s_1(z_{1-\alpha})\}$. Then $\tilde{T} = T + n^{-1}U$. Let $u_1 = u_1(\alpha)$ be the constant such that $E(T^3U) = u_1 + O(n^{-1})$. In general, if we are dealing with statistics which are smooth functions of means of independently and identically distributed random variables, it is true that $u_1 = 3u$. To see this, write $T = n^{-\frac{1}{2}} \sum_{i=1}^n Z_i + O_p(n^{-\frac{1}{2}})$ and $U = n^{-\frac{1}{2}} \sum_{i=1}^n W_i + O_p(n^{-\frac{1}{2}})$, where the pairs (Z_i, W_i) , for $i = 1, \dots, n$, are independent, zero-mean random variables and $E(Z_i^2) = 1$ for each i . Then

$$u = n^{-1}E\left\{\left(\sum_{i=1}^n Z_i\right)\left(\sum_{i=1}^n W_i\right)\right\} = E(Z_1 W_1),$$

while

$$\begin{aligned} u_1 &= n^{-2}E\left\{\left(\sum_{i=1}^n Z_i\right)^3\left(\sum_{i=1}^n W_i\right)\right\} + O(n^{-1}) \\ &= 3E(Z_1^2)E(Z_1 W_1) + O(n^{-1}) = 3u + O(n^{-1}). \end{aligned}$$

Now, noting that (T, U) has a joint Normal distribution with zero mean, and that $\text{var}(T) = 1 + O(n^{-1})$, it follows that $|ET| + |EU| + |ET^2U| = O(n^{-\frac{1}{2}})$, and we have $E(T^3U) = 3E(TU) + O(n^{-1})$. Therefore,

$$\begin{aligned} k_1(\tilde{T}) &= E(\tilde{T}) = k_1(T) + n^{-1}E(U) = k_1(T) + O(n^{-3/2}), \\ k_2(\tilde{T}) &= E(\tilde{T}^2) - (E\tilde{T})^2 = E(T^2 + 2n^{-1}TU) - (ET)^2 + O(n^{-2}) \\ &= k_2(T) + 2n^{-1}E(TU) + O(n^{-2}) = k_2(T) + 2n^{-1}u + O(n^{-2}), \\ k_3(\tilde{T}) &= E(\tilde{T}^3) - 3E(\tilde{T}^2)E(\tilde{T}) + 2(E\tilde{T})^3 \\ &= E(T^3 + 3n^{-1}T^2U) - 3E(T^2 + 2n^{-1}TU)E(T) + O(n^{-3/2}) \\ &= E(T^3) - 3E(T^2)E(T) + 2(ET)^3 + O(n^{-3/2}) \\ &= k_3(T) + O(n^{-3/2}), \end{aligned}$$

and

$$\begin{aligned} k_4(\tilde{T}) &= E(\tilde{T}^4) - 4E(\tilde{T}^3)E(\tilde{T}) - 3(E\tilde{T}^2)^2 + 12E(\tilde{T}^2)(E\tilde{T})^2 - 6(E\tilde{T})^4 \\ &= E(T^4 + 4n^{-1}T^3U) - 4E(T^3)E(T) - 3\{E(T^2) + 2n^{-1}u\}^2 \\ &\quad + 12E(T^2)(ET)^2 - 6(ET)^4 + O(n^{-3/2}) \\ &= k_4(T) + 4n^{-1}\{E(T^3U) - 3u\} + O(n^{-3/2}) = k_4(T) + O(n^{-3/2}), \end{aligned}$$

as had to be shown.

LEMMA 2.4. Under the assumption that x - and y -populations are Normal, the probability that the percentile- t interval $(-\infty, \hat{b}]$ covers the mean difference is

$$P(\mu_X - \mu_Y \leq \hat{b}) = 1 - \alpha + O(n^{-2}).$$

Proof of Lemma 2.4. Under the assumption of Normality, X and Y have infinitely many finite moments, so we may generalize (2.7) to

$$P(T \leq x) = \Phi(x) + n^{-\frac{1}{2}}q_1(x)\phi(x) + n^{-1}q_2(x)\phi(x) + n^{-3/2}q_3(x)\phi(x) + O(n^{-2}), \quad (2.45)$$

where q_j , for $j = 1, 2, 3$, are polynomials of degree $3j - 1$, odd for even j and even for odd j . Inversion of this expansion gives an extended version of (2.9) for the α 'th quantile of the distribution of T ,

$$y_\alpha = z_\alpha + n^{-\frac{1}{2}}q_{11}(z_\alpha) + n^{-1}q_{21}(z_\alpha) + n^{-3/2}q_{31}(z_\alpha) + O(n^{-2}), \quad (2.46)$$

where q_{11} and q_{21} are given by (2.10) and (2.11) and q_{31} , derived in Appendix A2.3, is

$$\begin{aligned} q_{31}(x) \equiv & -\frac{1}{3}\{q_1(x)\}^3(x^2 - 1) + \frac{3}{2}x\{q_1(x)\}^2q_1'(x) \\ & - q_1(x)\{q_1'(x)\}^2 - \frac{1}{2}q_1''(x)\{q_1(x)\}^2 \\ & - xq_1(x)q_2(x) + q_1'(x)q_2(x) + q_1(x)q_2'(x) - q_3(x). \end{aligned} \quad (2.47)$$

Consequently, the endpoint $\hat{b}$, of a percentile- t confidence interval $(-\infty, \hat{b}]$ admits the expansion

$$\hat{b} = \bar{X} - \bar{Y} + \hat{\sigma}\{z_{1-\alpha} + n^{-\frac{1}{2}}\hat{q}_1(z_{1-\alpha}) + n^{-1}\hat{q}_{21}(z_{1-\alpha}) - n^{-3/2}\hat{q}_{31}(z_{1-\alpha}) + O_p(n^{-2})\}. \quad (2.48)$$

Therefore, the interval $(-\infty, \hat{b}]$ has coverage

$$\begin{aligned} P(\mu_X - \mu_Y \leq \hat{b}) &= P[\mu_X - \mu_Y \leq \bar{X} - \bar{Y} + \hat{\sigma}\{z_{1-\alpha} + n^{-\frac{1}{2}}\hat{q}_1(z_{1-\alpha}) \\ &\quad + n^{-1}\hat{q}_{21}(z_{1-\alpha}) - n^{-3/2}\hat{q}_{31}(z_{1-\alpha}) + O_p(n^{-2})\}] \\ &= P[T + n^{-\frac{1}{2}}\{\hat{q}_1(z_{1-\alpha}) - q_1(z_{1-\alpha})\} \\ &\quad + n^{-1}\{\hat{q}_{21}(z_{1-\alpha}) - q_{21}(z_{1-\alpha})\} \geq -z_{1-\alpha} \\ &\quad - n^{-\frac{1}{2}}q_1(z_{1-\alpha}) - n^{-1}q_{21}(z_{1-\alpha}) + n^{-3/2}q_{31}(z_{1-\alpha})] \\ &\quad + O(n^{-2}) \\ &= P\{T + n^{-1}V \geq -z_{1-\alpha} - n^{-\frac{1}{2}}q_1(z_{1-\alpha}) - n^{-1}q_{21}(z_{1-\alpha}) \\ &\quad + n^{-3/2}q_{31}(z_{1-\alpha})\} + O(n^{-2}), \end{aligned}$$

where $V = V(\alpha) \equiv n^{\frac{1}{2}}\{\hat{q}_1(z_{1-\alpha}) - q_1(z_{1-\alpha})\} + \{\hat{q}_{21}(z_{1-\alpha}) - q_{21}(z_{1-\alpha})\}$. Put $T^\dagger \equiv T + n^{-1}V$. We now proceed as in the proof of Theorem 2.2, by relating the cumulants of T and $T^\dagger$, to develop an Edgeworth expansion for the distribution of $T^\dagger$. We show that $P(T^\dagger \leq x) = P(T \leq x) + O(n^{-2})$, by demonstrating that the first five cumulants of T and $T^\dagger$ agree up to (but not necessarily including) terms of $O(n^{-2})$, and arguing as in the proof of Theorem 2.2. In fact, we shall show that $E\{(T^\dagger)^j\} - E(T^j) = O(n^{-2})$, from which it follows that $k_j(T^\dagger) = k_j(T) + O(n^{-2})$. Notice that

$$\begin{aligned} E\{(T^\dagger)^j\} - E(T^j) &= \sum_{i=1}^j \binom{j}{i} E\{(n^{-1}V)^i T^{j-i}\} \\ &= \sum_{i=1}^j \binom{j}{i} n^{-i} E(V^i T^{j-i}) \\ &= n^{-1} j E(T^{j-1}V) + O(n^{-2}). \end{aligned}$$

Thus, it is sufficient to show that $E(T^{j-1}V) = O(n^{-1})$ for $j = 1, \dots, 5$.

We may write $V = n^{\frac{1}{2}}W_1 + W_2$, where $W_1 \equiv \{\hat{q}_1(z_{1-\alpha}) - q_1(z_{1-\alpha})\}$ and $W_2 \equiv \{\hat{q}_{21}(z_{1-\alpha}) - q_{21}(z_{1-\alpha})\}$ and q_1 and q_{21} are given in Theorem 2.1. From (2.40), we have $T = Z + O_p(n^{-\frac{1}{2}})$, where Z is a Normal $N(0, 1)$ random variable. Since the x - and y -populations are Normal, W_1 has a distribution symmetric about zero with all moments finite, while $Z^{2j-1}W_2$ has a distribution symmetric about zero for $j \geq 1$ with all moments finite. Bearing in mind that $E(\hat{\gamma}_X) = O(n^{-1})$, $E(\hat{\gamma}_X^2) = O(n^{-1})$, $E(\hat{\kappa}_X) = O(n^{-1})$ and $E(\hat{\sigma}_X^2) = \sigma_X^2 + O(n^{-1})$, with similar results for $\hat{\gamma}_Y$, $\hat{\gamma}_Y^2$, $\hat{\kappa}_Y$ and $\hat{\sigma}_Y^2$, then $E(Z^{2j}W_2) = O(n^{-1})$, for $j \geq 0$. Thus,

$$E(V) = n^{\frac{1}{2}}E(W_1) + E(W_2) = O(n^{-1}),$$

since W_1 has a symmetric distribution about zero. Notice that $n^{\frac{1}{2}}E(TW_1) = u + O(n^{-1})$, where u , given by (2.27), is zero in Normal populations, and, as in the proof of Theorem 2.2, $n^{\frac{1}{2}}E(T^3W_1) = 3n^{\frac{1}{2}}E(TW_1) + O(n^{-1}) = O(n^{-1})$. By symmetry, $E(T^2W_1) = E(T^4W_1) = 0$. Therefore,

$$\begin{aligned} E(TV) &\sim n^{\frac{1}{2}}E(TW_1) + E(ZW_2) = O(n^{-1}), \\ E(T^2V) &\sim n^{\frac{1}{2}}E(T^2W_1) + E(Z^2W_2) = O(n^{-1}), \\ E(T^3V) &\sim 3n^{\frac{1}{2}}E(TW_1) + E(Z^3W_2) = O(n^{-1}), \end{aligned}$$

and

$$E(T^4 V) \sim n^{\frac{1}{2}} E(T^4 W_1) + E(Z^4 W_2) = O(n^{-1}),$$

as had to be shown. Consequently, $P(T^\dagger \leq x) = P(T \leq x) + O(n^{-2})$. The coverage probability of $(-\infty, \hat{b}]$ is then

$$\begin{aligned} P(\mu_X - \mu_Y \leq \hat{b}) &= P\{T \geq -z_{1-\alpha} - n^{-\frac{1}{2}}q_1(z_{1-\alpha}) - n^{-1}q_{21}(z_{1-\alpha}) \\ &\quad + n^{-3/2}q_{31}(z_{1-\alpha})\} + O(n^{-2}) \\ &= 1 - \alpha + n^{-\frac{1}{2}}\{q_1(z_{1-\alpha}) - q_1(z_{1-\alpha})\}\phi(z_{1-\alpha}) \\ &\quad + n^{-1}[q_{21}(z_{1-\alpha}) + \frac{1}{2}z_{1-\alpha}\{q_1(z_{1-\alpha})\}^2 \\ &\quad - q_1'(z_{1-\alpha})q_1(z_{1-\alpha}) + q_2(z_{1-\alpha})]\phi(z_{1-\alpha}) \\ &\quad + n^{-3/2}[-q_{31}(z_{1-\alpha}) - \frac{1}{3}\{q_1(z_{1-\alpha})\}^3(z_{1-\alpha}^2 - 1) \\ &\quad + \frac{3}{2}z_{1-\alpha}\{q_1(z_{1-\alpha})\}^2q_1'(z_{1-\alpha}) - q_1(z_{1-\alpha})\{q_1'(z_{1-\alpha})\}^2 \\ &\quad - \frac{1}{2}q_1''(z_{1-\alpha})\{q_1(z_{1-\alpha})\}^2 \\ &\quad - z_{1-\alpha}q_1(z_{1-\alpha})q_2(z_{1-\alpha}) + q_1'(z_{1-\alpha})q_2(z_{1-\alpha}) \\ &\quad + q_1(z_{1-\alpha})q_2'(z_{1-\alpha}) - q_3(z_{1-\alpha})]\phi(z_{1-\alpha}) \\ &\quad + O(n^{-2}) \\ &= 1 - \alpha + O(n^{-2}), \end{aligned} \tag{2.49}$$

on noticing that the coefficients of $n^{-\frac{1}{2}}$, n^{-1} , and $n^{-3/2}$ in (2.49) are identically zero, from (2.11) and (2.47).

APPENDIX A2.1. Proof of Lemma 2.3.

We derive here the first four moments of T and S , treating T first. We use notation from the proof of Theorem 2.1. There, we showed that

$$T = Z\{1 - \frac{1}{2}\sigma R_1 + \sigma^2(\frac{3}{8}R_1^2 + \frac{1}{2}R_2^2) + O_p(n^{-3/2})\}. \quad (\text{A2.1.1})$$

Thus,

$$\begin{aligned} E(T) &= E(Z) - \frac{1}{2}\sigma E(ZR_1) + \frac{3}{8}\sigma^2 E(ZR_1^2) + \frac{1}{2}\sigma^2 E(ZR_2^2) + O(n^{-3/2}) \\ &= -\frac{1}{2}\sigma E(ZR_1) + O(n^{-3/2}) \\ &= -\frac{1}{2}\sigma^{-3} E\{(m^{-\frac{1}{2}}U_2 - n^{-\frac{1}{2}}V_2)(m^{-3/2}U_1 + n^{-3/2}V_1)\} + O(n^{-3/2}) \\ &= -\frac{1}{2}\sigma^{-3} m^{-3/2} \{(r^{-\frac{1}{2}}E(U_1U_2) - r^{3/2}E(V_1V_2))n^{-\frac{1}{2}} + O(n^{-3/2})\} \\ &= -n^{-\frac{1}{2}}\frac{1}{2}s^{-3}g + O(n^{-3/2}), \end{aligned}$$

since $E(U_1U_2) = \gamma_X$ and $E(V_1V_2) = \gamma_Y$, and $s = m^{\frac{1}{2}}\sigma$. From (A2.1.1) we deduce that

$$T^2 = Z^2\{1 - \sigma R_1 + \sigma^2(R_1^2 + R_2^2) + O_p(n^{-3/2})\}.$$

Therefore,

$$\begin{aligned} E(T^2) &= E(Z^2) - \sigma E(Z^2 R_1) + \sigma^2 E(Z^2 R_1^2) + \sigma^2 E(Z^2 R_2^2) + O(n^{-2}) \\ &= 1 - \sigma^{-4} E\{(m^{-1}U_2^2 - 2m^{-\frac{1}{2}}n^{-\frac{1}{2}}U_2V_2 + n^{-1}V_2^2) \\ &\quad \times (m^{-3/2}U_1 + n^{-3/2}V_1)\} \\ &\quad + \sigma^{-6} E\{(m^{-1}U_2^2 - 2m^{-\frac{1}{2}}n^{-\frac{1}{2}}U_2V_2 + n^{-1}V_2^2) \\ &\quad \times (m^{-3}U_1^2 + 2m^{-3/2}n^{-3/2}U_1V_1 + n^{-3}V_1^2)\} \\ &\quad + \sigma^{-4} E\{(m^{-1}U_2^2 - 2m^{-\frac{1}{2}}n^{-\frac{1}{2}}U_2V_2 + n^{-1}V_2^2) \\ &\quad \times (m^{-2}U_2^2 + n^{-2}V_2^2)\} + O(n^{-2}) \\ &= 1 - \sigma^{-4} \{m^{-5/2}E(U_1U_2^2) + n^{-5/2}E(V_1V_2^2)\} \\ &\quad + \sigma^{-6} \{m^{-4}E(U_1^2U_2^2) + m^{-3}n^{-1}E(U_1^2)E(V_2^2) \\ &\quad - 4m^{-2}n^{-2}E(U_1U_2)E(V_1V_2) + m^{-1}n^{-3}E(V_1^2)E(U_2^2) + n^{-4}E(V_1^2V_2^2)\} \\ &\quad + \sigma^{-4} \{m^{-3}E(U_2^4) + n^{-1}m^{-2}E(U_2^2)E(V_2^2) \\ &\quad + n^{-2}m^{-1}E(U_2^2)E(V_2^2) + n^{-3}E(V_2^4)\} + O(n^{-2}) \\ &= 1 - \sigma^{-4} [m^{-3}\{E(X^4) - \sigma_X^4\} + n^{-3}\{E(Y^4) - \sigma_Y^4\}] \end{aligned}$$

$$\begin{aligned}
& + \sigma^{-6}(m^{-4}[\sigma_X^2\{E(X^4) - \sigma_X^4\} + 2\gamma_X^2] \\
& + m^{-3}n^{-1}\sigma_Y^2\{E(X^4) - \sigma_X^4\} - 4m^{-2}n^{-2}\gamma_X\gamma_Y \\
& + m^{-1}n^{-3}\sigma_X^2\{E(Y^4) - \sigma_Y^4\} + n^{-4}[\sigma_Y^2\{E(Y^4) - \sigma_Y^4\} + 2\gamma_Y^2]) \\
& + \sigma^{-4}(3\sigma_X^4m^{-3} + n^{-1}m^{-2}\sigma_Y^2\sigma_X^2 + n^{-2}m^{-1}\sigma_Y^2\sigma_X^2 + 3\sigma_Y^4n^{-3}) \\
& + O(n^{-2}) \\
& = 1 + s^{-4}n^{-1}(r^{-1}\sigma_X^4 + r^2\sigma_Y^4) + s^{-6}n^{-1}\{2(r^{-1}\gamma_X - 2r\gamma_X\gamma_Y + r^3\gamma_Y^2) \\
& \quad - r^{-1}\sigma_X^6 - \sigma_Y^2\sigma_X^4 - r^2\sigma_X^2\sigma_Y^4 - r^3\sigma_Y^6\} \\
& \quad + s^{-4}n^{-1}\{3r^{-1}\sigma_X^4 + \sigma_Y^2\sigma_X^2(1+r) + 3r^2\sigma_Y^4\} + O(n^{-2}) \\
& = 1 + s^{-6}n^{-1}\{2(r^{-1}\gamma_X^2 - 2r\gamma_X\gamma_Y + r^3\gamma_Y^2)\} \\
& \quad + s^{-4}n^{-1}\{3r^{-1}\sigma_X^4 + (1+r)\sigma_X^2\sigma_Y^2 + 3r^2\sigma_Y^4\} + O(n^{-2}) \\
& = 1 + n^{-1}(2s^{-6}g^2 + s^{-4}h^2) + O(n^{-2}).
\end{aligned}$$

To compute the third moment of T , observe from (A2.1.1) that

$$T^3 = Z^3\{1 - \frac{3}{2}\sigma R_1 + O_p(n^{-3/2})\}.$$

Thus,

$$\begin{aligned}
E(T^3) &= E(Z^3) - \frac{3}{2}\sigma E(Z^3 R_1) + O(n^{-3/2}) \\
&= \sigma^{-3}\{m^{-3/2}E(U_2^3) - n^{-3/2}E(V_2^3)\} \\
&\quad - \frac{3}{2}\sigma^{-5}E\{(m^{-3/2}U_2^3 - 3m^{-1}n^{-\frac{1}{2}}U_2^2V_2 + 3m^{-\frac{1}{2}}n^{-1}U_2V_2^2 \\
&\quad - n^{-3/2}V_2^3)(m^{-3/2}U_1 + n^{-3/2}V_1)\} + O(n^{-3/2}) \\
&= \sigma^{-3}(m^{-2}\gamma_X - n^{-2}\gamma_Y) - \frac{3}{2}\sigma^{-5}\{m^{-3}E(U_1U_2^3) \\
&\quad - 3m^{-1}n^{-2}E(U_2^2)E(V_1V_2) + 3m^{-2}n^{-1}E(V_2^2)E(U_1U_2) \\
&\quad - n^{-3}E(V_1V_2^3)\} + O(n^{-3/2}) \\
&= s^{-3}gn^{-\frac{1}{2}} - \frac{9}{2}\sigma^{-5}(m^{-3}\sigma_X^2\gamma_X - m^{-1}n^{-2}\sigma_X^2\gamma_Y \\
&\quad + m^{-2}n^{-1}\sigma_Y^2\gamma_X - n^{-3}\sigma_Y^2\gamma_Y) + O(n^{-3/2}) \\
&= -n^{-\frac{1}{2}}\frac{7}{2}s^{-3}g + O(n^{-3/2}).
\end{aligned}$$

Finally, we may write

$$T^4 = Z^4\{1 - 2\sigma R_1 + \sigma^2(3R_1^2 + 2R_2^2) + O_p(n^{-3/2})\},$$

so that

$$E(T^4) = E(Z^4) - 2\sigma E(Z^4 R_1) + 3\sigma^2 E(Z^4 R_1^2) + 2\sigma^2 E(Z^4 R_2^2) + O(n^{-2})$$

$$\begin{aligned}
&= \sigma^{-4} \{ m^{-2} \mathbb{E}(U_2^4) - 4m^{-3/2} n^{-\frac{1}{2}} \mathbb{E}(U_2^3 V_2) + 6m^{-1} n^{-1} \mathbb{E}(U_2^2 V_2^2) \\
&\quad - 4m^{-\frac{1}{2}} n^{-3/2} \mathbb{E}(U_2 V_2^3) + n^{-2} \mathbb{E}(V_2^4) \} \\
&\quad - 2\sigma^{-6} \mathbb{E} \{ (m^{-2} U_2^4 - 4m^{-3/2} n^{-\frac{1}{2}} U_2^3 V_2 + 6m^{-1} n^{-1} U_2^2 V_2^2 \\
&\quad - 4m^{-\frac{1}{2}} n^{-3/2} U_2 V_2^3 + n^{-2} V_2^4) (m^{-3/2} U_1 + n^{-3/2} V_1) \} \\
&\quad + 3\sigma^{-8} \mathbb{E} \{ (m^{-2} U_2^4 - 4m^{-3/2} n^{-\frac{1}{2}} U_2^3 V_2 + 6m^{-1} n^{-1} U_2^2 V_2^2 \\
&\quad - 4m^{-\frac{1}{2}} n^{-3/2} U_2 V_2^3 + n^{-2} V_2^4) (m^{-3} U_1^2 \\
&\quad + 2m^{-3/2} n^{-3/2} U_1 V_1 + n^{-3} V_1^2) \} \\
&\quad + 2\sigma^{-6} \mathbb{E} \{ (m^{-2} U_2^4 - 4m^{-3/2} n^{-\frac{1}{2}} U_2^3 V_2 + 6m^{-1} n^{-1} U_2^2 V_2^2 \\
&\quad - 4m^{-\frac{1}{2}} n^{-3/2} U_2 V_2^3 + n^{-2} V_2^4) (m^{-2} U_2^2 + n^{-2} V_2^2) \} + O(n^{-2}) \\
&= \sigma^{-4} \{ m^{-2} (3\sigma_X^4 + m^{-1} \kappa_X) + 6m^{-1} n^{-1} \sigma_X^2 \sigma_Y^2 + n^{-2} (3\sigma_Y^4 + n^{-1} \kappa_Y) \} \\
&\quad + 2\sigma^{-6} \{ m^{-4} (15\sigma_X^6 - [4\gamma_X^2 + 6\{\mathbb{E}(X^4) - \sigma_X^4\} \sigma_X^2]) \\
&\quad + m^{-3} n^{-1} [18\sigma_Y^2 \sigma_X^4 - 6\{\mathbb{E}(X^4) - \sigma_X^4\} \sigma_Y^2] \\
&\quad + m^{-2} n^{-2} (3\sigma_Y^4 \sigma_X^2 + 3\sigma_X^4 \sigma_Y^2 + 8\gamma_X \gamma_Y) \\
&\quad + m^{-1} n^{-3} [18\sigma_X^2 \sigma_Y^4 - 6\{\mathbb{E}(Y^4) - \sigma_Y^4\} \sigma_X^2] \\
&\quad + n^{-4} (15\sigma_Y^6 - [4\gamma_Y^2 + 6\{\mathbb{E}(Y^4) - \sigma_Y^4\} \sigma_Y^2]) \} \\
&\quad + 3\sigma^{-8} \{ m^{-5} [3\sigma_X^4 \{\mathbb{E}(X^4) - \sigma_X^4\} + 12\sigma_X^2 \gamma_X^2] \\
&\quad + 6m^{-4} n^{-1} \sigma_Y^2 [2\gamma_X^2 + \sigma_X^2 \{\mathbb{E}(X^4) - \sigma_X^4\}] \\
&\quad + m^{-3} n^{-2} [3\sigma_Y^4 \{\mathbb{E}(X^4) - \sigma_X^4\} - 24\gamma_X \gamma_Y \sigma_X^2] \\
&\quad + n^{-3} m^{-2} [3\sigma_X^4 \{\mathbb{E}(Y^4) - \sigma_Y^4\} - 24\gamma_X \gamma_Y \sigma_Y^2] \\
&\quad + 6m^{-1} n^{-4} \sigma_X^2 [2\gamma_Y^2 + \sigma_Y^2 \{\mathbb{E}(Y^4) - \sigma_Y^4\}] \\
&\quad + n^{-5} [3\sigma_Y^4 \{\mathbb{E}(Y^4) - \sigma_Y^4\} + 12\sigma_Y^2 \gamma_Y^2] \} + O(n^{-2}) \\
&= 3 + n^{-1} (28s^{-6} g^2 + 6s^{-4} k^2 - 2s^{-4} l^2) + O(n^{-2}),
\end{aligned}$$

on collecting terms, and noting the definitions of s, g, k^2 and l^2 .

The moments of S are much easier to calculate than the moments of T . Recall from the proof of Theorem 2.1 that

$$S = Z \equiv \sigma^{-1} (m^{-\frac{1}{2}} U_2 - n^{-\frac{1}{2}} V_2).$$

As a result,

$$\mathbb{E}(S) = \mathbb{E}(Z) = 0,$$

$$\mathbb{E}(S^2) = \mathbb{E}(Z^2) = 1,$$

$$\mathbb{E}(S^3) = \mathbb{E}(Z^3) = \sigma^{-3} \{ m^{-3/2} \mathbb{E}(U_2^3) - n^{-3/2} \mathbb{E}(V_2^3) \} = n^{-\frac{1}{2}} s^{-3} g,$$

and

$$\begin{aligned}
 E(S^4) &= E(Z^4) \\
 &= \sigma^{-4} \{ m^{-2} E(U_2^4) - 4m^{-3/2} n^{-1/2} E(U_2^3 V_2) + 6m^{-1} n^{-1} E(U_2^2 V_2^2) \\
 &\quad - 4m^{-1/2} n^{-3/2} E(U_2 V_2^3) + n^{-2} E(V_2^4) \} \\
 &= \sigma^{-4} \{ m^{-2} (3\sigma_X^4 + m^{-1} \kappa_X) + 6m^{-1} n^{-1} \sigma_X^2 \sigma_Y^2 + n^{-2} (3\sigma_Y^4 + n^{-1} \kappa_Y) \} \\
 &= 3 + n^{-1} s^{-4} b.
 \end{aligned}$$

APPENDIX A2.2. Derivation of the constants u and v in the percentile- t case.

Define U_1, U_2, V_1 and V_2 as in the proof of Theorem 2.1 and put $U_3 \equiv m^{-\frac{1}{2}} \sum_i (X_i^3 - \gamma_X)$ and $V_3 \equiv n^{-\frac{1}{2}} \sum_j (Y_j^3 - \gamma_Y)$, and write $z \equiv z_{1-\alpha}$. In the percentile- t case, u and v are constants such that $n^{\frac{1}{2}} E\{T\{\hat{q}_1(z) - q_1(z)\}\} = u + O(n^{-1})$ and $n^{\frac{1}{2}} E\{T(a_1 - \hat{a}_1)\} = v + O(n^{-1})$, respectively, where $q_1(x) \equiv \frac{1}{6}(2x^2 + 1)s^{-3}g$ and $a_1 \equiv -\frac{1}{6}(2z^2 - 3)s^{-3}g$, s and g being defined as in the proof of Theorem 2.1. Therefore, to evaluate u and v , it is sufficient to consider

$$n^{\frac{1}{2}} E\{T(\hat{s}^{-3}\hat{g} - s^{-3}g)\}, \quad (A2.2.1)$$

where $\hat{s}$ and $\hat{g}$ are obtained from s and g respectively on replacing population moments by sample moments. Observe that

$$\hat{g} = g + r^{-\frac{1}{2}}m^{-\frac{1}{2}}(U_3 - 3\sigma_X^2 U_2) - r^{3/2}n^{-\frac{1}{2}}(V_3 - 3\sigma_Y^2 V_2) + O_p(n^{-1}),$$

and, from the proof of Theorem 2.1, $\hat{s}^{-3} = m^{-3/2}\hat{\sigma}^{-3} = s^{-3}\{1 - \frac{3}{2}\sigma R_1 + O_p(n^{-1})\}$, and $T = Z\{1 - \frac{1}{2}\sigma R_1 + O_p(n^{-1})\}$. Therefore, from the definitions of Z and R_1 , and using the fact that $s = m^{\frac{1}{2}}\sigma$, we obtain

$$\begin{aligned} & n^{\frac{1}{2}} T(\hat{s}^{-3}\hat{g} - s^{-3}g) \\ &= n^{\frac{1}{2}} s^{-1}(U_2 - r^{\frac{1}{2}}V_2)\{1 - \frac{1}{2}s^{-2}(U_1 + r^{-3/2}V_1)r^{-\frac{1}{2}}n^{-\frac{1}{2}}\} \\ & \quad \times \{r^{-1}n^{-\frac{1}{2}}s^{-3}(U_3 - 3\sigma_X^2 U_2) - r^{3/2}n^{-\frac{1}{2}}s^{-3}(V_3 - 3\sigma_Y^2 V_2) \\ & \quad - n^{-\frac{1}{2}}r^{-\frac{1}{2}}\frac{3}{2}s^{-5}g(U_1 + r^{3/2}V_1)\} + O_p(n^{-1}) \\ &= s^{-1}(U_2 - r^{\frac{1}{2}}V_2)\{r^{-1}s^{-3}(U_3 - 3\sigma_X^2 U_2) - r^{3/2}s^{-3}(V_3 - 3\sigma_Y^2 V_2) \\ & \quad - \frac{3}{2}r^{-\frac{1}{2}}s^{-5}g(U_1 + r^{3/2}V_1) - \frac{1}{2}s^{-5}r^{-3/2}n^{-\frac{1}{2}}(U_1 + r^{3/2}V_1)(U_3 - 3\sigma_X^2 U_2) \\ & \quad + \frac{1}{2}s^{-5}rn^{-\frac{1}{2}}(U_1 + r^{3/2}V_1)(V_3 - 3\sigma_Y^2 V_2) \\ & \quad + \frac{3}{4}s^{-7}gr^{-1}n^{-\frac{1}{2}}(U_1 + r^{3/2}V_1)^2\} + O_p(n^{-1}). \end{aligned}$$

Thus, taking expectations,

$$\begin{aligned} & E\{n^{\frac{1}{2}} T(\hat{s}^{-3}\hat{g} - s^{-3}g)\} \\ &= s^{-4}r^{-1}\{E(U_2 U_3) - 3\sigma_X^4 E(U_2^2) + r^3 E(V_2 V_3) - 3r^3 \sigma_Y^4 E(V_2^2)\} \\ & \quad - \frac{3}{2}s^{-6}r^{-\frac{1}{2}}g\{E(U_1 U_2) - r^2 E(V_1 V_2)\} + O(n^{-1}). \end{aligned}$$

Since $E(U_1U_2) = \gamma_X$, $E(V_1V_2) = \gamma_Y$, $E(U_2U_3) = E(X^4)$, $E(V_2V_3) = E(Y^4)$ and $E(U_2^2) = E(V_2^2) = 1$, we may write

$$\begin{aligned} E\{n^{\frac{1}{2}}T(\hat{s}^{-3}\hat{g} - s^{-3}g)\} &= s^{-4}(r^{-1}\kappa_X + r^2\kappa_Y) - \frac{3}{2}s^{-6}g^2 + O(n^{-1}) \\ &= s^{-4}b - \frac{3}{2}s^{-6}g^2 + O(n^{-1}), \end{aligned}$$

where $b \equiv r^{-1}\kappa_X + r^2\kappa_Y$. Consequently,

$$\begin{aligned} u &= \frac{1}{6}(2z^2 + 1)\{(\sigma_X^2 + r\sigma_Y^2)^{-2}(r^{-1}\kappa_X + r^2\kappa_Y) \\ &\quad - \frac{1}{2}(\sigma_X^2 + r\sigma_Y^2)^{-3}(r^{-\frac{1}{2}}\gamma_X - r^{3/2}\gamma_Y)^2\}, \end{aligned}$$

and

$$\begin{aligned} v &= -\frac{1}{6}(2z^2 - 3)\{(\sigma_X^2 + r\sigma_Y^2)^{-2}(r^{-1}\kappa_X + r^2\kappa_Y) \\ &\quad - \frac{3}{2}(\sigma_X^2 + r\sigma_Y^2)^{-3}(r^{-\frac{1}{2}}\gamma_X - r^{3/2}\gamma_Y)^2\}. \end{aligned}$$

APPENDIX A2.3. Derivation of the polynomial q_{31} in the Cornish-Fisher expansion of percentile- t critical point.

The distribution function of T , defined at (2.2), admits expansion (2.45). Cornish-Fisher inversion of (2.45) yields an expansion for the α 'th quantile of the distribution of T given by (2.46). Given expansion (2.46), we find explicit formulae for the polynomials q_{11} , q_{21} and q_{31} in terms of q_1 , q_2 and q_3 by observing that

$$\alpha = P(T \leq y_\alpha).$$

That is,

$$\begin{aligned} \alpha &= P\{T \leq z_\alpha + n^{-\frac{1}{2}}q_{11}(z_\alpha) + n^{-1}q_{21}(z_\alpha) + n^{-3/2}q_{31}(z_\alpha)\} + O(n^{-2}) \\ &= \Phi\{z_\alpha + n^{-\frac{1}{2}}q_{11}(z_\alpha) + n^{-1}q_{21}(z_\alpha) + n^{-3/2}q_{31}(z_\alpha)\} \\ &\quad + n^{-\frac{1}{2}}q_1\{z_\alpha + n^{-\frac{1}{2}}q_{11}(z_\alpha) + n^{-1}q_{21}(z_\alpha)\} \\ &\quad \times \phi\{z_\alpha + n^{-\frac{1}{2}}q_{11}(z_\alpha) + n^{-1}q_{21}(z_\alpha)\} \\ &\quad + n^{-1}q_2\{z_\alpha + n^{-\frac{1}{2}}q_{11}(z_\alpha)\}\phi\{z_\alpha + n^{-\frac{1}{2}}q_{11}(z_\alpha)\} \\ &\quad + n^{-3/2}q_3(z_\alpha)\phi(z_\alpha) + O(n^{-2}), \end{aligned} \tag{A2.3.1}$$

using (2.45). By Taylor expansion of the right-hand side of (A2.3.1),

$$\begin{aligned} \alpha &= \Phi(z_\alpha) + n^{-\frac{1}{2}}\{q_1(z_\alpha) + q_{11}(z_\alpha)\}\phi(z_\alpha) + n^{-1}[q_{21}(z_\alpha) - \frac{1}{2}\{q_{11}(z_\alpha)\}^2 z_\alpha \\ &\quad + q_1'(z_\alpha)q_{11}(z_\alpha) - q_1(z_\alpha)q_{11}(z_\alpha)z_\alpha + q_2(z_\alpha)]\phi(z_\alpha) \\ &\quad + n^{-3/2}[q_{31}(z_\alpha) - q_{11}(z_\alpha)q_{21}(z_\alpha)z_\alpha + \frac{1}{6}\{q_{11}(z_\alpha)\}^3(z_\alpha^2 - 1) \\ &\quad + q_1'(z_\alpha)q_{21}(z_\alpha) + \frac{1}{2}q_1''(z_\alpha)\{q_{11}(z_\alpha)\}^2 \\ &\quad + \frac{1}{2}q_1(z_\alpha)\{q_{11}(z_\alpha)\}^2(z_\alpha^2 - 1) - z_\alpha q_1(z_\alpha)q_{21}(z_\alpha) \\ &\quad - z_\alpha q_1'(z_\alpha)\{q_{11}(z_\alpha)\}^2 + q_2'(z_\alpha)\{q_{11}(z_\alpha)\}^2 - q_2(z_\alpha)q_{11}(z_\alpha)z_\alpha \\ &\quad + q_3(z_\alpha)]\phi(z_\alpha) + O(n^{-2}). \end{aligned}$$

Equating coefficients of $n^{-\frac{1}{2}}$, n^{-1} and $n^{-3/2}$ to zero, and solving, we obtain the polynomials

$$q_{11}(z_\alpha) = -q_1(z_\alpha),$$

$$q_{21}(z_\alpha) = q_1'(z_\alpha)q_1(z_\alpha) - \frac{1}{2}z_\alpha\{q_1(z_\alpha)\}^2 - q_2(z_\alpha),$$

and

$$\begin{aligned} q_{31}(z_\alpha) &= -\frac{1}{3}\{q_1(z_\alpha)\}^3(z_\alpha^2 - 1) + \frac{3}{2}z_\alpha\{q_1(z_\alpha)\}^2 q_1'(z_\alpha) \\ &\quad - q_1(z_\alpha)\{q_1'(z_\alpha)\}^2 - \frac{1}{2}q_1''(z_\alpha)\{q_1(z_\alpha)\}^2 \\ &\quad - z_\alpha q_1(z_\alpha)q_2(z_\alpha) + q_1'(z_\alpha)q_2(z_\alpha) + q_1(z_\alpha)q_2'(z_\alpha) - q_3(z_\alpha). \end{aligned}$$

CHAPTER 3

BOOTSTRAP RESAMPLING AND ITERATION

1. INTRODUCTION

We propose a single, unifying principle guiding the operation of bootstrap reampling, applicable to a wide variety of statistical problems, such as bias reduction, shrinkage, hypothesis testing and confidence interval construction. Our principle differs from other approaches in that it focusses attention directly on a measure of quality or accuracy, expressed in the form of an equation whose solution is sought. It enables us to centre our attention on one or more characteristics which are of major importance in any particular problem — such as coverage error or length for confidence intervals, or bias for point estimation. The aim of the principle is to get those characteristics in which we are interested right. An immediate consequence of our approach is a very general form of bootstrap iteration, directed at improving a particular aspect of a solution to a statistical problem, and unifying and generalizing present disparate accounts of that subject. When employed for bias reduction, iteration of the resampling principle yields a competitor of the generalized jackknife, enabling bias to be reduced to arbitrarily low levels. When applied to coverage error reduction in confidence intervals it produces the techniques of Hall (1986a) and Beran (1987a). The resampling principle leads easily to solutions of new, complex problems, such as empirical versions of confidence intervals proposed by Lehmann (1986). Lehmann argued that an “ideal” confidence interval is one which is short when it covers the true parameter, but not necessarily otherwise. The resampling principle suggests a simple empirical means of constructing such intervals.

In Section 2 we describe the general principle, and in Section 3 we show how it leads naturally to bootstrap iteration. There we demonstrate that in many problems of practical interest, such as bias reduction in point estimation and coverage error reduction in two-sided confidence intervals, each iteration of

the principle reduces error by a factor of order n^{-1} , where n is sample size. In the case of confidence intervals, our result sharpens one of Beran (1987a), who showed that coverage error is reduced by a factor of order $n^{-\frac{1}{2}}$ in two-sided intervals. The main exception to our “ n^{-1} rule” is coverage error of *one*-sided confidence intervals, where error is reduced by a factor of order $n^{-\frac{1}{2}}$ at each iteration.

Our approach to bootstrap iteration serves to unify not just the philosophy of iteration for different statistical problems, but also different techniques of iteration for the same problem. Cases in point are the different methods proposed by Hall (1986a) and Beran (1987a) for correcting coverage error of confidence intervals. Both of these methods fall under our general scheme. We also show that in some circumstances the *infinitely* iterated bootstrap converges. An example where this occurs is iteration to reduce the bias of any polynomial function of population moments. It is easy to find examples where convergence fails to occur. In particular, high-order bootstrap iteration to reduce coverage error in nonparametric percentile- t confidence intervals ultimately produces constant resamples with zero variance.

In Sections 4, 5 and 6 we apply the resampling principle to reduce bias, to construct Lehmann-type confidence intervals, and to shrink parameter estimates in the L^1 metric, respectively.

2. THE PRINCIPLE

We present a general structure for statistical problems in Subsection 2.1, and describe a unified resampling solution to all problems which admit the general structure in Subsection 2.2.

2.1. General formulation of statistical problems. Let F_0 be the true distribution, F_1 the empirical distribution based on a random sample, $f(F_0, F_1)$ a known functional of F_0 and F_1 , and $E(\cdot \mid F_0)$ expectation given that the data came from F_0 . A typical statistical problem has the form: choose f from amongst a designated class of functionals so that

$$E\{f(F_0, F_1) \mid F_0\} = 0. \quad (3.1)$$

The advantage of posing problems in this form is that it focusses attention on a particular characteristic, $E(f \mid F_0)$, which is of major importance. The extent to which we get that characteristic right is measured by the accuracy with which

we solve equation (3.1). The characteristic is usually coverage error, derivative of mean interval length, or level error if the problem is one of constructing confidence intervals or hypothesis tests; see Examples 2.1 and 2.2 below. If we are concerned with reducing bias in point estimation, the characteristic is bias; see Example 2.3.

Example 2.1. Simple confidence intervals. We treat only one-sided intervals as two-sided analogues are similar. Suppose $\hat{\theta} \equiv \psi(F_1)$ estimates the unknown parameter $\theta \equiv \psi(F_0)$. For example, $\psi(F_1) \equiv \int x dF_1(x)$ is a sample mean. Let $\hat{s} \equiv \sigma(F_1)$ estimate the standard deviation of $\hat{\theta}$, and put

$$f_t(F_0, F_1) \equiv I\{\psi(F_0) \leq \psi(F_1) + t\sigma(F_1)\} - (1 - \alpha), \quad (3.2)$$

where t is an arbitrary real number, I is an indicator function and $0 < \alpha < 1$. In many circumstances, the problem of constructing a $(1 - \alpha)$ -level one-sided confidence interval for θ is that of selecting $f = f_t$, of the form (3.2), such that (3.1) holds. Notice that when f is of the form (3.2), the left-hand side of (3.1) equals coverage error, so that this formulation of the problem focusses our attention on a solution with good coverage accuracy.

Example 2.2. Lehmann-type confidence intervals. Lehmann (1986), p.223, argued that “short intervals are desirable when they cover the true parameter value, but not necessarily otherwise”. Arguing in this vein, an ideal $(1 - \alpha)$ -level, two-sided interval for θ would be $[\hat{\theta} - \hat{s}x_{1-\alpha+t}, \hat{\theta} - \hat{s}x_t]$, where $x_\eta \equiv x(F_0)_\eta$ is given by $P\{(\hat{\theta} - \theta)/\hat{s} \leq x_\eta\} = \eta$ and where t is chosen to minimize mean length *conditional on coverage*,

$$\ell(t) \equiv E\{\hat{s}(x_{1-\alpha+t} - x_t)I(\theta \in [\hat{\theta} - \hat{s}x_{1-\alpha+t}, \hat{\theta} - \hat{s}x_t])\}/(1 - \alpha). \quad (3.3)$$

Finding the turning point of ℓ is equivalent to solving $\ell'(t) = 0$, and differentiating under the expectation sign, we see that Lehmann’s problem becomes one of solving an equation like (3.1). A crucial difference between Examples 2.1 and 2.2 is that in the former, the left-hand side of (3.1) equals coverage error, whereas in the latter it describes a derivative of mean interval length.

Example 2.3. Bias correction. Using the notation of Example 2.1, correcting $\hat{\theta} \equiv \psi(F_1)$ for bias is equivalent to solving (3.1) with

$$f_t(F_0, F_1) \equiv \psi(F_1) + t - \psi(F_0).$$

Here, the left-hand side of (3.1) equals the bias of $\hat{\theta}$.

Example 2.4. L^p shrinkage. Consider the problem of estimating the value of some unknown (possibly vector-valued) parameter θ . Often we are able to make an initial estimate, c , of the value of the parameter, either from knowledge of the experimental conditions, from past experience, or from some extraneous source. In such cases it may be reasonable to take the usual estimator for θ , say $\hat{\theta}$, and shrink it towards our initial estimate c , by multiplying the difference $\hat{\theta} - c$ by a shrinkage factor $1 - t$ and adding it to c . That is,

$$\hat{\theta}_s \equiv (1 - t)(\hat{\theta} - c) + c = (1 - t)\hat{\theta} + tc, \quad 0 \leq t \leq 1.$$

The shrunk estimator $\hat{\theta}_s$, though it may be biased, might be closer in some sense than $\hat{\theta}$ to θ . For example, it may have smaller mean squared error than $\hat{\theta}$. Indeed it is widely accepted that minimum mean squared error is a desirable property (see James and Stein, 1961). For this reason, and because mean squared error can be evaluated explicitly, shrinkage is usually studied in the L^2 metric. In that case, the shrinkage parameters admit simple formulae in terms of estimable quantities; see, for example Thompson (1968), Mehta and Srinivasan (1971), and Lawless (1981). The resampling principle allows shrinkage problems to be solved in L^p metrics, for $1 \leq p < \infty$. The case $p = 1$ is of particular interest because of its robustness qualities. In particular, there is a resampling version of L^1 ridge regression.

We shall confine our discussion to L^p shrinkage ($1 \leq p < \infty$) towards a fixed constant c , for scalar θ . There, the problem is that of choosing $t \in [0, 1]$ to minimize $E|(1 - t)\hat{\theta} + tc - \theta|^p$, or equivalently, $E|(1 - t)\psi(F_1) + tc - \psi(F_0)|^p$. Differentiating with respect to t under the expectation sign and equating to zero, we see that we must solve an equation like (3.1).

In more general statistical problems there is no requirement that f and t be scalars. Should we wish to correct two or more estimators for bias simultaneously, or shrink several estimators towards fixed points simultaneously, then f and t would be vector-valued. However, in this chapter, all of our examples will be of scalar f 's and t 's.

2.2. Resampling solutions to statistical problems. The resampling solution to the problem of solving (3.1) is as follows. Draw a same-size sample (called a "resample") at random, with replacement, from the original sample. Let F_2

denote its empiric, conditional on F_1 . Arguing that the relationship between F_1 and F_2 should be similar to that between F_0 and F_1 , choose f from the class of eligible functionals so that

$$E\{f(F_1, F_2) \mid F_1\} = 0. \quad (3.4)$$

This may be done numerically. The resulting f , depending on F_1 , approximates the unobservable f which solves (3.1).

When applied to the confidence interval problem in Example 2.1, this principle yields the familiar percentile- t bootstrap interval (see, for example, Beran, 1987a). Application to Lehmann-type confidence intervals in Example 2.2 is straightforward, although computations require multiple use of resamples; see Section 5. When used to solve the bias-correction problem in Example 2.3, the resampling principle produces $t = \psi(F_1) - E\{\psi(F_2) \mid F_1\}$, and hence the bias-corrected estimate

$$\hat{\theta} \equiv 2\psi(F_1) - E\{\psi(F_2) \mid F_1\}. \quad (3.5)$$

In some point estimation problems, the contribution of F_1 to $f(F_0, F_1)$ is degenerate. For example, the location, t , of the distribution represented by F_0 might be defined as the solution of (3.1) when

$$f(F_0, F_1) = f_t(F_0, F_1) \equiv \int w(x - t) dF_0(x),$$

where w is a known function, such as $w(u) \equiv u$. In this instance, $f(F_0, F_1)$ does not depend on F_1 . The resampling solution to the problem of finding t is to solve equation (3.4), and that produces a general M -estimate of location.

3. ITERATING THE PRINCIPLE

We showed in Section 2 that many statistical problems have the form: find an empirical rule for choosing a functional f from amongst a given class of functionals so that (3.1) holds. Index the class by vector t , writing f as f_t . Take $T(F_0)$ to be the "ideal" t solving (3.1). The resampling principle produces $T(F_1)$, the solution of (3.4), as our bootstrap estimate of $T(F_0)$. We argued that $f_{T(F_1)}$ should approximate $f_{T(F_0)}$, in the sense that

$$E\{f_{T(F_1)}(F_0, F_1) \mid F_0\} \approx 0. \quad (3.6)$$

In this section, we show how to amend $T(F_1)$ so that equation (3.6) is solved more accurately. Introduce an adjustment term u to the functional T so that

$T(\cdot)$ becomes $T(\cdot, u)$, with the property that $T(\cdot, 0) \equiv T(\cdot)$. The idea is that by using an appropriate non-zero value of u , we may reduce the error in (3.6). The error is reduced to zero if we solve

$$E\{f_{T(F_1, u)}(F_0, F_1) \mid F_0\} = 0. \quad (3.7)$$

Put $g_u(F, G) \equiv f_{T(G, u)}(F, G)$. Then equation (3.7) is equivalent to $E\{g_u(F_0, F_1) \mid F_0\} = 0$, which is in exactly the same form as (3.1). The resampling principle therefore argues that u be chosen to solve $E\{g_u(F_1, F_2) \mid F_1\} = 0$; that is,

$$E\{f_{T(F_2, u)}(F_1, F_2) \mid F_1\} = 0. \quad (3.8)$$

Having selected $u = U(F_1)$ in this manner, the once-iterated bootstrap approximation to $f_{T(F_0)}$ becomes $f_{T(F_1, U(F_1))}$. This operation may be repeated several times provided our computational resources can cope.

Should it be not possible to solve (3.8) algebraically, we may derive a numerical solution as follows. Resample B times from F_1 , and let F_2 be the empiric of any one of these resamples. Resample B times from this particular resample, and let F_3 be the empiric of any one of these resamples. Average $f_t(F_2, F_3)$ over these B values of F_3 , for the given F_2 , to obtain first an approximation to $E\{f_t(F_2, F_3) \mid F_2\}$, and then an approximation to the solution $t = T(F_2)$ of $E\{f_t(F_2, F_3) \mid F_2\} = 0$. Repeat this operation for each of the B F_2 's, averaging the resulting values of $f_{T(F_2, u)}(F_1, F_2)$ to obtain an estimate of $E\{f_{T(F_2, u)}(F_1, F_2) \mid F_1\}$. Then solve (3.8) approximately.

There are several ways of introducing the adjustment u to the functional T . We may amend T additively to $T(\cdot, u) \equiv T(\cdot) + u$, or multiplicatively to $T(\cdot, u) \equiv T(\cdot)(1 + u)$, where u is a scalar. In confidence interval problems, adjustments may be made directly to nominal coverage. For instance, in Example 2.1, $t = x(F_0)_\beta$ is the β 'th quantile of the distribution of $-\{\psi(F_1) - \psi(F_0)\}/\sigma(F_1)$, its "ideal" value is $T(F_0) = x(F_0)_{1-\alpha}$, and we might take $T(\cdot, u) \equiv x(\cdot)_{1-\alpha+u}$. Different methods of adjustment usually give slightly different numerical solutions, but they produce the same order of adjustment, as we show two paragraphs below.

It is natural to incorporate further adjustments as additive corrections of the first. Thus, a second iteration could use $T_1(\cdot) \equiv T\{\cdot, U(\cdot)\}$ in place of $T(\cdot)$, and $T_1(\cdot, u) \equiv T\{\cdot, U(\cdot) + u\}$ in place of $T(\cdot, u)$, and so on.

We now give an heuristic argument showing that in many cases of interest, each iteration of the resampling principle reduces the order of magnitude of error

in (3.6) by a factor of order n^{-1} . Suppose we have reduced error to order n^{-j} , for some $j \geq 1$. That is,

$$E\{f_{T(F_1)}(F_0, F_1) \mid F_0\} = c(F_0)n^{-j} + O(n^{-(j+1)}), \quad (3.9)$$

for a smooth functional c . Put

$$d(F_0) \equiv (\partial/\partial u)E\{f_{T(F_1, u)}(F_0, F_1) \mid F_0\} \Big|_{u=0},$$

and assume $d(F_0)$ converges to a non-zero constant as $n \rightarrow \infty$. Then,

$$\Delta \equiv n^{\frac{1}{2}}\{c(F_1)d(F_1)^{-1} - c(F_0)d(F_0)^{-1}\} = O_p(1),$$

as $n \rightarrow \infty$. Put $e(\Delta) \equiv (\partial/\partial u)E\{f_{T(F_1, u)}(F_0, F_1) \mid \Delta\} \Big|_{u=0}$, so that $d(F_0) = E\{e(\Delta)\}$. By Taylor expansion,

$$E\{f_{T(F_1, u)}(F_0, F_1) \mid F_0\} = c(F_0)n^{-j} + d(F_0)u + O(n^{-(j+1)} + u^2).$$

Therefore, equation (3.7) has solution $u = U(F_0) = -\{c(F_0)d(F_0)^{-1}\}n^{-j} + O(n^{-(j+1)})$, and similarly, equation (3.8) has solution

$$\begin{aligned} u &= U(F_1) = -\{c(F_1)d(F_1)^{-1}\}n^{-j} + O_p(n^{-(j+1)}) \\ &= U(F_0) - n^{-j-\frac{1}{2}}\Delta + O_p(n^{-(j+1)}). \end{aligned}$$

Consequently,

$$\begin{aligned} E\{f_{T(F_1, U(F_1))}(F_0, F_1) \mid \Delta\} &= E\{f_{T(F_1, U(F_0))}(F_0, F_1) \mid \Delta\} - n^{-j-\frac{1}{2}}\Delta e(\Delta) \\ &\quad + O_p(n^{-(j+1)}), \end{aligned}$$

and taking expectations,

$$E\{f_{T(F_1, U(F_1))}(F_0, F_1) \mid F_0\} = -n^{-j-\frac{1}{2}}E\{\Delta e(\Delta)\} + O(n^{-(j+1)}). \quad (3.10)$$

As a result, if

$$E\{\Delta e(\Delta)\} = O(n^{-\frac{1}{2}}), \quad (3.11)$$

then

$$E\{f_{T(F_1, U(F_1))}(F_0, F_1) \mid F_0\} = O(n^{-(j+1)}). \quad (3.12)$$

Therefore, a single bootstrap iteration has reduced error from order n^{-j} to order $n^{-(j+1)}$; compare (3.9) and (3.12). We claim that (3.11) holds true in a

great many cases of interest, including bias correction and coverage error reduction for two-sided confidence intervals. To see why (3.11) holds in the case of bias correction, observe that there, the functional $f_t(\cdot)$ is usually smooth in t . As a result, $e(\Delta)$ is bounded in probability, and noting that, in general, $E\{c(F_1)d(F_1)^{-1}\} = c(F_0)d(F_0)^{-1} + O(n^{-1})$, so that $E(\Delta) = O(n^{-\frac{1}{2}})$, we have $E\{\Delta e(\Delta)\} = O(n^{-\frac{1}{2}})$.

The argument for the case of reducing coverage error in two-sided confidence intervals is different, since the functional $f_t(\cdot)$ is no longer smooth in t . We illustrate the argument only in the case of symmetric intervals, other cases being treated formally in Chapter 4. In this instance, we may take

$$f_{T(F_1)}(F_0, F_1) \equiv I\{n^{\frac{1}{2}}|\theta(F_1) - \theta(F_0)| \leq T(F_1)\} - (1 - \alpha),$$

and

$$e(\Delta) \equiv (\partial/\partial u)P\{n^{\frac{1}{2}}|\theta(F_1) - \theta(F_0)| \leq T(F_1, u) \mid \Delta\}\big|_{u=0}.$$

Using Normal approximations to the joint distributions of $n^{\frac{1}{2}}\{\theta(F_1) - \theta(F_0)\}$ and Δ , we see that there exist bivariate Normal random variables (N_1, N_2) with zero means and such that

$$E\{\Delta e(\Delta)\} = E[N_2(\partial/\partial u)P\{|N_1| \leq T(F_0, u) \mid N_2\}] + O(n^{-\frac{1}{2}}).$$

The expectation on the right-hand side is zero. To appreciate why, take the differential operator $\partial/\partial u$ outside the expectation, and note that for any $x > 0$, $E\{N_2P(|N_1| \leq x \mid N_2)\} = 0$. This proves (3.11).

For the case of one-sided intervals, this argument should be modified as follows. Suppose we have reduced error in (3.6) to order $n^{-j/2}$. Equation (3.9) changes to

$$E\{f_{T(F_1)}(F_0, F_1) \mid F_0\} = c(F_0)n^{-j/2} + O(n^{-(j+1)/2}),$$

and equation (3.10) to

$$E\{f_{T(F_1, U(F_1))}(F_0, F_1) \mid F_0\} = -n^{-(j+1)/2}E\{\Delta e(\Delta)\} + O(n^{-(j+1)/2}) \quad (3.13)$$

for the same definition of $e(\Delta)$ as above. Now we may take

$$f_{T(F_1)}(F_0, F_1) \equiv I[n^{\frac{1}{2}}\{\theta(F_1) - \theta(F_0)\} \leq T(F_1)] - (1 - \alpha),$$

and

$$e(\Delta) = (\partial/\partial u)P[n^{\frac{1}{2}}\{\theta(F_1) - \theta(F_0)\} \leq T(F_1, u) \mid \Delta]\big|_{u=0}.$$

Applying the argument used in the case of symmetric, two-sided intervals, we see that there exist bivariate Normal random variables (N_1, N_2) with zero means such that

$$E\{\Delta e(\Delta)\} = E[N_2(\partial/\partial u)P\{|N_1| \leq T(F_0, u) \mid N_2\}] + O(n^{-\frac{1}{2}}) = O(1).$$

Therefore, we see from (3.13), without any subsidiary conditions such as (3.6), that coverage error has been reduced from order $n^{-j/2}$ to order $n^{-(j+1)/2}$.

It should be stressed that iteration of the resampling principle does *not* improve the accuracy with which $T(F_1)$ approximates $T(F_0)$. The ordinary, non-iterated approximation is in error by $O_p(n^{-\frac{1}{2}})$, which is best possible. Instead, successive iterations “tweak” the term of order $n^{-\frac{1}{2}}$ so as to reduce the size of *expected* errors in (3.6).

If the iterated resampling principle is used to construct confidence intervals and if adjustments are made as additive corrections to quantiles, then we get exactly Hall’s (1986a) iterated bootstrap intervals. Alternatively, note that any ordinary bootstrap $(1 - \alpha)$ -level confidence interval, such as a percentile-method or percentile- t interval, employs a bootstrap estimate $T(F_1) \equiv x(F_1)_{1-\alpha}$ of a “true” $(1 - \alpha)$ -quantile $T(F_0) \equiv x(F_0)_{1-\alpha}$. If adjustments are made at the level of nominal coverage, so that $T(\cdot, u) \equiv x(\cdot)_{1-\alpha+u}$, then we obtain precisely Beran’s (1987a) iterated prepivoted intervals. Thus our methodology serves to unify these two approaches.

We conclude this section with a small simulation study of coverage correction in two-sided percentile-method bootstrap confidence intervals for the mean. A more extensive simulation study of bootstrap coverage correction in two-sided percentile-method confidence intervals for the correlation coefficient is given in Chapter 4. The results of our study are summarized in Table 3.1. It demonstrates clearly that one bootstrap iteration does in fact improve coverage accuracy. In our example, the parent population was folded Normal, $|N(0, 1)|$, adjustment was made at the level of nominal coverage, $B = 99$ resamples were drawn in each resampling operation, and each entry was based on 1,000 samples for each value of n . The folded Normal pseudo-random values were generated using NAG Library subroutine G05DDF, and all computations were carried out on a VAX 8700 computer.

TABLE 3.1. *Characteristics of percentile and once-iterated percentile-method confidence intervals for the mean.*

Distribution was folded Normal, and nominal coverage was 0.90. Tabulated characteristics are coverage, mean length, standard deviation of length, and mean value of upper endpoint.

n		Ordinary Percentile	Iterated Percentile
8	cov.	0.831	0.890
	$E(L)$	0.633	0.730
	s.d. (L)	0.211	0.324
	upper	0.330	0.375
9	cov.	0.813	0.877
	$E(L)$	0.609	0.702
	s.d. (L)	0.195	0.286
	upper	0.319	0.364
10	cov.	0.846	0.892
	$E(L)$	0.590	0.685
	s.d. (L)	0.171	0.257
	upper	0.318	0.363
11	cov.	0.847	0.890
	$E(L)$	0.557	0.644
	s.d. (L)	0.156	0.217
	upper	0.286	0.326
12	cov.	0.847	0.897
	$E(L)$	0.547	0.630
	s.d. (L)	0.137	0.192
	upper	0.287	0.324

4. BIAS CORRECTION

We showed in Section 2 that one application of the resampling principle to the problem of reducing the bias of $\hat{\theta}_0 \equiv \psi(F_1)$ for $\theta \equiv \psi(F_0)$ produces the estimate $\hat{\theta}_1$, given by (3.5). Denote by $\hat{\theta}_j$ the estimator obtained from j iterations of the resampling principle. An explicit formula for $\hat{\theta}_j$ is given in Theorem 3.1, proved in Subsection 7.1.

THEOREM 3.1. *The estimator $\hat{\theta}_j$ is given by*

$$\hat{\theta}_j \equiv \sum_{i=1}^{j+1} \binom{j+1}{i} (-1)^{i+1} E\{\psi(F_i) \mid F_1\}, \quad (3.14)$$

where F_{i+1} , conditional on F_i , denotes the empiric of a same-size sample drawn with replacement from F_i . Estimator $\hat{\theta}_j$ has variance of order n^{-1} and bias of order $n^{-(j+1)}$, and the variance of $\hat{\theta}_j$ differs from that of $\hat{\theta}_0$ by a term of order n^{-2} .

The j -times iterated bootstrap estimator $\hat{\theta}_j$ has variance and bias of the same orders as its j -fold jackknife competitor. Parr (1983) and Beran (1984) have given detailed formulae comparing bootstrap and jackknife variance and bias when $j = 1$, but analogues for general j are very complex. See also Efron (1981a, 1981b, 1982), Beran (1982), Efron and Gong (1983) and Wu (1987).

In some circumstances, the limit, $\hat{\theta}_\infty = \lim_j \hat{\theta}_j$, of an infinite number of bias corrections is well-defined. For example, if $\theta \equiv \mu_1^2$ denotes squared population mean, if $\hat{\theta} = \hat{\theta}_0 \equiv \bar{X}^2$ denotes squared sample mean, and if $\hat{s}^2 \equiv n^{-1} \sum_i (X_i - \bar{X})^2$ is sample variance, then, as we show in Lemma 3.4 of Subsection 7.1, j iterations of the resampling principle give $\hat{\theta}_j \equiv \bar{X}^2 - (n-1)^{-1}(1-n^{-j})\hat{s}^2$. As $j \rightarrow \infty$, $\hat{\theta}_j \rightarrow \hat{\theta}_\infty \equiv \bar{X}^2 - (n-1)^{-1}\hat{s}^2$, which is unbiased for θ , and on this occasion is the simple jackknife estimator. A more general result is given in the following lemma.

LEMMA 3.2. *If θ is a polynomial in population moments with known, bounded coefficients, and if $\hat{\theta}_0$ is the same polynomial in sample moments, then $\hat{\theta}_\infty$ exists and is unbiased for θ with variance $O(n^{-1})$. Furthermore, $\hat{\theta}_\infty$ is a polynomial in sample moments.*

We prove Lemma 3.2 in Subsection 7.2.

In complex problems, one may calculate $E\{\psi(F_i) \mid F_1\}$ in (3.14) as follows. Resample B times from F_1 ; resample B times from each of the previous samples, each of the new samples having distribution F_2 ; and so on, until B^{i-1} resamples, that is, B resamples from each of B^{i-2} independent copies of F_{i-1} , have been obtained. Each of these B^{i-1} resamples has distribution F_i . Compute the value of $\psi(\cdot)$ for each of these B^{i-1} resamples, and calculate the arithmetic mean of all B^{i-1} values. This number converges to $E\{\psi(F_i) \mid F_1\}$ as $B \rightarrow \infty$.

5. LEHMANN-TYPE CONFIDENCE INTERVALS

We first apply the resampling principle to construct empirical Lehmann-type intervals. That is, we detail the calculation of the function of which $E\{f(F_1, F_2) \mid F_1\}$ is a formal derivative, in this instance. Adopt notation from Example 2.2, and let $T \equiv (\hat{\theta} - \theta)/\hat{s}$. Conditional on our original n -sample $\mathcal{X}$,

draw B independent n -resamples $\mathcal{X}_1^*, \dots, \mathcal{X}_B^*$ from $\mathcal{X}$, with replacement. Let $(\hat{\theta}_i^*, \hat{s}_i^*)$, for $i = 1, \dots, B$, be versions of $\hat{\theta}$, $\hat{s}$ computed for $\mathcal{X}_i^*$ rather than $\mathcal{X}$. Write $(\hat{\theta}^*, \hat{s}^*)$ for a generic $(\hat{\theta}_i^*, \hat{s}_i^*)$, and put $T^* \equiv (\hat{\theta}^* - \hat{\theta})/\hat{s}^*$. Recall that $x_\eta = x(F_0)_\eta$ is the quantile such that $P(T \leq x_\eta | F_0) = \eta$. First, use the resamples to compute $\hat{x}_\eta = x(F_1)_\eta$, the quantile such that $P(T^* \leq \hat{x}_\eta | F_1) = \eta$, and then use them a second time to calculate

$$\hat{l}_B(t) = \left\{ \sum_{i=1}^B \hat{s}_i^* (\hat{x}_{1-\alpha+t} - \hat{x}_t) I(\hat{\theta} \in [\hat{\theta}_i^* - \hat{s}_i^* \hat{x}_{1-\alpha+t}, \hat{\theta}_i^* - \hat{s}_i^* \hat{x}_t]) \right\} \\ \times \left\{ \sum_{i=1}^B I(\hat{\theta} \in [\hat{\theta}_i^* - \hat{s}_i^* \hat{x}_{1-\alpha+t}, \hat{\theta}_i^* - \hat{s}_i^* \hat{x}_t]) \right\}^{-1}.$$

As $B \rightarrow \infty$, $\hat{l}_B(t)$ converges almost surely to $\hat{l}(t)$, the empirical version of $l(t)$ defined at (3.3):

$$\hat{l}(t) \equiv E\{\hat{s}^*(\hat{x}_{1-\alpha+t} - \hat{x}_t) I(\hat{\theta} \in [\hat{\theta}^* - \hat{s}^* \hat{x}_{1-\alpha+t}, \hat{\theta}^* - \hat{s}^* \hat{x}_t]) | F_1\} / (1 - \alpha).$$

Values $t_0, \hat{t}_0$ minimizing $l(t), \hat{l}(t)$ respectively satisfy $\hat{t}_0 - t_0 \rightarrow 0$ almost surely as $n \rightarrow \infty$. The empirical Lehmann-type interval is $J \equiv [\hat{\theta} - \hat{s} \hat{x}_{1-\alpha+\hat{t}_0}, \hat{\theta} - \hat{s} \hat{x}_{\hat{t}_0}]$.

To describe asymptotic theory, assume the existence of the Edgeworth expansion for the distribution of T ,

$$P(n^{\frac{1}{2}}T \leq x) \sim \Phi(x) + \sum_{i \geq 1} q_i(F_0, x) \phi(x), \quad (3.15)$$

where Φ, ϕ are Standard Normal distribution and density functions respectively. Introduce the notation $z \equiv z_{1-(\alpha/2)}$ where $\Phi(z_{1-(\alpha/2)}) = 1 - (\alpha/2)$, $\psi_{ik} = \psi_{ik}(F_0) \equiv (\partial/\partial x)^k \{q_i(F_0, x) \phi(x)\}|_{x=z}$ for $i \geq 1$, $\psi_{0k} \equiv \Phi^{(k)}(z)$, $\tau^2 = \tau^2(F_0) \equiv \lim_{n \rightarrow \infty} n \text{var}(\hat{s})$, and $\rho = \rho(F_0) \equiv 2z\tau(F_0)/\{\sigma(F_0)(1 - \alpha)\}$. Let r denote the asymptotic correlation between $Z \equiv n^{\frac{1}{2}}(\hat{s} - s)/\tau$ and $n^{\frac{1}{2}}T$, and put $\psi \equiv rz\phi(z)$,

$$u_1 \equiv \lim_{n \rightarrow \infty} nE[T\{\psi_{11}(F_1) - \psi_{11}(F_0)\}]/\psi_{02}, \\ u_2 \equiv \lim_{n \rightarrow \infty} nE[T\{\rho(F_1)\psi(F_1) - \rho(F_0)\psi(F_0)\}]/z,$$

and

$$u_3 \equiv \lim_{n \rightarrow \infty} nE[T\{q_1(F_1, z) - q_1(F_0, z)\}].$$

For our purposes, likelihood-based confidence intervals are intervals of the type $[\hat{\theta} - \hat{s}x_{1-\alpha+u_0}, \hat{\theta} - \hat{s}x_{u_0}]$ where u_0 is chosen to minimize $x_{1-\alpha+u} - x_u$.

They are precisely the ideal “short” intervals discussed earlier; see Chapter 2, Section 9 for a discussion of such intervals and their bootstrap versions. See also Hall (1987a, 1988a). Write $\lambda_L, \lambda_{ET}, \lambda_{LB}$ and $\epsilon_L, \epsilon_{ET}, \epsilon_{LB}$ for mean lengths and coverage errors of Lehmann-type, equal-tailed and likelihood-based bootstrap confidence intervals, respectively. We compare the mean length and coverage error of Lehmann-type intervals with those of equal-tailed and likelihood-based intervals using the following theorem, which is proved in Subsection 7.3.

THEOREM 3.3. *Using the notation introduced above, the mean length of the Lehmann-type interval satisfies*

$$n^{3/2}(\lambda_L - \lambda_{LB}) = \sigma(\rho\psi)^2 z^{-1} + O(n^{-1}), \quad (3.16)$$

and

$$n^{3/2}(\lambda_L - \lambda_{ET}) = \sigma[(\rho\psi)^2 - \{q'_1(z)\}^2] z^{-1} + O(n^{-1}), \quad (3.17)$$

where $q_1(\cdot) = q_1(F_0, \cdot)$ and $\sigma = \sigma(F_0)$. The coverage error of the Lehmann-type interval satisfies

$$|\epsilon_L| + |\epsilon_{ET}| + |\epsilon_{LB}| = O(n^{-1}), \quad (3.18)$$

and

$$\epsilon_L/\epsilon_{LB} = (u_1 + u_2)/u_1 + O(n^{-1}), \quad \epsilon_L/\epsilon_{ET} = (u_1 + u_2)/u_3 + O(n^{-1}). \quad (3.19)$$

The results given in Theorem 3.3 are for bootstrap intervals, but (3.16) and (3.17) continue to hold for the respective “ideal” intervals which could be constructed were the distribution of T known. In this sense, length properties of bootstrap intervals accurately reflect those of their unattainable “ideal” counterparts. Lehmann-type intervals are never shorter than likelihood-based intervals, which are constructed to minimize length regardless of whether θ is covered. This is evidenced by the fact that the right-hand side of (3.16) is non-negative. The situation is not so clear in the case of (3.17), although as we show below, the right-hand side is nonpositive in the circumstance of the Studentized mean for $0 < \alpha < 1$, where $\hat{\theta} \equiv n^{-1} \sum_i X_i$ denotes sample mean and $\hat{s}^2 \equiv n^{-1} \sum_i (X_i - \bar{X})^2$ denotes sample variance. We assume, without loss of generality, that the true mean is 0 and the true variance 1. Hall (1988a), Section 4, Example 1, gives $q_1(x) \equiv \frac{1}{6}\gamma(2x^2 + 1)$, where γ represents standardized population skewness. Since $E\{n^{\frac{1}{2}}(\hat{\theta} - \theta)/\hat{s}\} = -\frac{1}{2}\gamma n^{-\frac{1}{2}} + O(n^{-1})$ (see Appendix A3.1),

we have

$$\begin{aligned} r &\equiv \lim_{n \rightarrow \infty} \text{corr}(Z, n^{\frac{1}{2}}T) = \lim_{n \rightarrow \infty} n\tau^{-1} E\{(\hat{s} - 1)(\hat{\theta} - \theta)/\hat{s}\} \\ &= \lim_{n \rightarrow \infty} n\tau^{-1} E\{(\hat{\theta} - \theta) - (\hat{\theta} - \theta)/\hat{s}\} = \lim_{n \rightarrow \infty} -n^{\frac{1}{2}}\tau^{-1} E(n^{\frac{1}{2}}T) = \frac{1}{2}\gamma\tau^{-1}. \end{aligned}$$

Therefore, $r\tau = \frac{1}{2}\gamma$, and so

$$\rho\psi = 2z^2\phi(z)r\tau(1 - \alpha)^{-1} = \gamma z^2\phi(z)(1 - \alpha)^{-1}.$$

Consequently, the first term on the right-hand side of (3.17) equals

$$z^{-1}[(\rho\psi)^2 - \{q'_1(z)\}^2] = z\gamma^2[z^2\{\phi(z)\}^2(1 - \alpha)^{-2} - \frac{4}{9}].$$

The function $g(z) \equiv z\phi(z)(1 - \alpha)^{-1} = z\phi(z)\{2\Phi(z) - 1\}^{-1}$ is decreasing for $z > 0$, and $0 \leq g(z) < \frac{1}{2}$, when $z > 0$. Thus, for $0 < \alpha < 1$, the first term on the right-hand side of (3.17) is bounded above by $-\frac{7}{36}z\gamma^2$, which is nonpositive.

TABLE 3.2. Asymptotic comparison of length and coverage error, in circumstances of the Studentized mean. Here σ^2 equals population variance and γ equals standardized population skewness.

	$1 - \alpha = 0.90$	$1 - \alpha = 0.95$
$\lim n^{3/2}(\lambda_L - \lambda_{LB})/(\sigma\gamma^2)$	0.052	0.029
$\lim n^{3/2}(\lambda_L - \lambda_{ET})/(\sigma\gamma^2)$	-0.67	-0.84
$\lim \epsilon_L/\epsilon_{LB}$	1.47	1.15
$\lim \epsilon_L/\epsilon_{ET}$	0.55	0.62

Table 3.2 summarizes asymptotics for 90% and 95% intervals in the case of the Studentized mean. These results indicate that Lehmann-type intervals are only marginally longer than likelihood-based intervals, but noticeably shorter than equal-tailed intervals. This property was confirmed by a simulation study summarized in Table 3.3, where all three interval types were compared for $1 - \alpha = 0.95$. Here the parent population was folded Normal, $|N(0, 1)|$, $B = 399$ resamples were drawn from each sample, and each entry was based on 4,000 samples for each value of n . Notice that in this example, all three types of interval exhibit good coverage accuracy.

In this application of the resampling principle, the characteristic of interest has been the derivative of interval length. We could iterate the principle to get

TABLE 3.3. Characteristics of Lehmann-type, likelihood-based and equal-tailed percentile-t bootstrap confidence intervals for the mean. Parent distribution was folded normal, and nominal coverage was 95%. Tabulated characteristics are coverage, mean length, standard deviation of length and mean value of upper endpoint.

n		Lehmann-type	Likelihood-based	Equal-tailed
8	cov.	0.943	0.943	0.951
	length	1.119	1.118	1.235
	std.dev.	0.523	0.523	0.693
	upper	0.667	0.662	0.795
9	cov.	0.949	0.949	0.957
	length	0.997	0.997	1.089
	std.dev.	0.390	0.390	0.505
	upper	0.583	0.578	0.689
10	cov.	0.953	0.953	0.960
	length	0.926	0.926	0.997
	std.dev.	0.334	0.333	0.409
	upper	0.532	0.528	0.622
11	cov.	0.946	0.946	0.952
	length	0.852	0.852	0.906
	std.dev.	0.290	0.290	0.341
	upper	0.489	0.486	0.559
12	cov.	0.946	0.947	0.955
	length	0.798	0.797	0.840
	std.dev.	0.244	0.244	0.277
	upper	0.452	0.449	0.509
13	cov.	0.945	0.945	0.952
	length	0.764	0.764	0.804
	std.dev.	0.226	0.226	0.257
	upper	0.437	0.433	0.490
14	cov.	0.948	0.949	0.955
	length	0.717	0.717	0.751
	std.dev.	0.198	0.197	0.220
	upper	0.404	0.401	0.450
15	cov.	0.944	0.944	0.951
	length	0.682	0.682	0.712
	std.dev.	0.180	0.180	0.200
	upper	0.378	0.375	0.420

that characteristic right to a high order of accuracy. However, it would seem to be of more practical value to focus on coverage error in later resampling, and iterate to reduce the size of that quantity. A virtue of the resampling principle is that it enables a wide variety of characteristics to be corrected for error.

6. L^1 SHRINKAGE

Let $\hat{\theta}$ be an estimator of a parameter θ . The problem of L^1 shrinkage towards a fixed point c is to choose the shrinkage parameter t to minimize $D(t) \equiv E|(1-t)\hat{\theta} + tc - \theta|$. We show next how to derive an empirical solution to this problem using the resampling principle. Conditional on our original n -sample $\mathcal{X}$, draw B independent n -resamples $\mathcal{X}_1^*, \dots, \mathcal{X}_B^*$ from $\mathcal{X}$ with replacement. Let $\hat{\theta}_1^*, \dots, \hat{\theta}_B^*$ denote the respective versions of $\hat{\theta}$ computed using $\mathcal{X}_1^*, \dots, \mathcal{X}_B^*$ instead of $\mathcal{X}$, and put

$$\hat{D}_B(t) \equiv B^{-1} \sum_{i=1}^B |(1-t)\hat{\theta}_i^* + tc - \hat{\theta}|.$$

As $B \rightarrow \infty$, $\hat{D}_B(t)$ converges almost surely to $\hat{D}(t) \equiv E\{|(1-t)\psi(F_2) + tc - \psi(F_1)| \mid F_1\}$. The value $\hat{t}_0$ minimizing $\hat{D}(t)$ is an estimate of the value t_0 minimizing $D(t)$, and an empirical version of the shrunken estimator is $\hat{\theta}_s \equiv (1-\hat{t}_0)\hat{\theta} + \hat{t}_0 c$.

To describe asymptotic theory, let $n^{-1}\sigma^2$ denote the asymptotic variance of $\hat{\theta}$, and suppose that the density of $S \equiv n^{\frac{1}{2}}(\hat{\theta} - \theta)/\sigma$ admits the Edgeworth expansion

$$f(z) = \phi(z)\{1 + n^{-\frac{1}{2}}p(z) + \dots\}, \quad (3.20)$$

where ϕ is the Standard Normal density and p is a polynomial. Put $c_1 \equiv 2 \int_0^\infty p(z)\phi(z) dz$ and $\rho \equiv \sigma/(\theta - c)$. We show in Subsection 7.4 that if $c \neq \theta$ then

$$t_0 = n^{-1}\{\rho^2 + (\pi/2)^{\frac{1}{2}}\rho c_1\} + O(n^{-2}) = O(n^{-1}),$$

and that $\hat{t}_0 \hat{t}_0^{-1} \rightarrow 1$ almost surely as $n \rightarrow \infty$.

7. PROOFS

7.1. Proof of Theorem 3.1. We adopt notation from Section 4, and write $\hat{\theta}_j = \hat{\theta}_j(F_1)$, so that $\psi(F_1) = \hat{\theta}_0(F_1)$. After the j 'th iteration of the resampling principle, we have

$$\hat{\theta}_j = \hat{\theta}_j(F_1) = \hat{\theta}_{j-1}(F_1) + \hat{\theta}_0(F_1) - E\{\hat{\theta}_{j-1}(F_2) \mid F_1\}. \quad (3.21)$$

By writing $\hat{\theta}_{j-1}(F_i)$ in terms of $\hat{\theta}_{j-2}(F_i)$ in (3.21) and $\hat{\theta}_{j-2}(F_i)$ in terms of $\hat{\theta}_{j-3}(F_i)$, and so on, it follows that $\hat{\theta}_j = \sum_{i=1}^{j+1} a_{j,i} E\{\psi(F_i) \mid F_1\}$ for constants $a_{j,i}$, for $i = 1, \dots, j+1$. Observing that

$$\begin{aligned} E\{\hat{\theta}_{j-1}(F_2) \mid F_1\} &= E\left[\sum_{i=1}^j a_{j-1,i} E\{\psi(F_{i+1}) \mid F_2\} \mid F_1\right] \\ &= \sum_{i=1}^j a_{j-1,i} E\{\psi(F_{i+1}) \mid F_1\}, \end{aligned}$$

it follows from (3.21) that

$$\hat{\theta}_j = \sum_{i=1}^{j+1} a_{j,i} E\{\psi(F_i) \mid F_1\} = \psi(F_1) + \sum_{i=1}^j (a_{j-1,i} - a_{j-1,i-1}) E\{\psi(F_{i+1}) \mid F_1\},$$

so that, for $i = 1, \dots, j+1$, the constants $a_{j,i}$ satisfy

$$a_{j,i} = \delta_{1i} + a_{j-1,i} - a_{j-1,i-1}, \quad a_{0,i} = \delta_{1i},$$

where δ_{1i} is 1 if $i = 1$ and zero otherwise. Solving this difference equation, we obtain $a_{j,i} = (-1)^{i+1} \binom{j+1}{i}$, for $i = 1, \dots, j+1$, establishing (3.14).

We next show that the bias of $\hat{\theta}_j$ is of order $n^{-(j+1)}$. We first derive a formula for $E\{\psi(F_i) \mid F_0\}$, for $i = 1, \dots, j+1$. Suppose that for some functionals a_{k_1} , for $k_1 = 1, \dots, j$, we may write

$$E\{\psi(F_1) \mid F_0\} = \psi(F_0) + \sum_{k_1=1}^j a_{k_1}(F_0) n^{-k_1} + O(n^{-(j+1)}),$$

and further assume that we have expansions

$$\begin{aligned} E\{a_{k_1}(F_1) \mid F_0\} &= a_{k_1}(F_0) + \sum_{k_2=1}^{j-k_1} a_{k_1 k_2}(F_0) n^{-k_2} + O(n^{-(j-k_1+1)}), \\ E\{a_{k_1 k_2}(F_1) \mid F_0\} &= a_{k_1 k_2}(F_0) + \sum_{k_3=1}^{j-k_1-k_2} a_{k_1 k_2 k_3}(F_0) n^{-k_3} + O(n^{-(j-k_1-k_2+1)}), \end{aligned}$$

and so on, up to

$$\begin{aligned} E\{a_{k_1 \dots k_{j-1}}(F_1) \mid F_0\} &= a_{k_1 \dots k_{j-1}}(F_0) + \sum_{k_j=1}^{j-k_1-\dots-k_{j-1}} a_{k_1 \dots k_j}(F_0) n^{-k_j} \\ &\quad + O(n^{-(j-k_1-\dots-k_{j-1}+1)}), \end{aligned}$$

for functionals $a_{k_1 k_2}, \dots, a_{k_1 \dots k_j}$. Put $b_{r,i} \equiv E\{b_r(F_i)\}$ where

$$b_r(F_i) \equiv \sum_{k_1=1}^j \sum_{k_2=1}^{j-k_1} \dots \sum_{k_r=1}^{j-k_1-\dots-k_{r-1}} a_{k_1 \dots k_r}(F_i) n^{-(k_1+\dots+k_r)},$$

and $b_r \equiv b_r(F_0)$, for $r = 1, \dots, j$. The $b_{r,i}$ satisfy

$$\begin{aligned} b_{r,i} &= E\{b_r(F_i)\} = E[E\{b_r(F_i) \mid F_{i-1}\}] = E\{b_r(F_{i-1}) + b_{r+1}(F_{i-1})\} \\ &= b_{r,i-1} + b_{r+1,i-1}, \end{aligned}$$

for $i = 1, \dots, j+1$ and $r = 1, \dots, j$, with $b_{r,0} = b_r$. Solving this difference equation yields

$$b_{r,i} = \sum_{l=0}^i \binom{i}{l} b_{r+l}. \quad (3.22)$$

Notice that $b_r = O(n^{-r})$, and for constants $c_{k,i}$, $k = 1, \dots, j$, with $c_{k,0} = 0$, we have

$$E\{\psi(F_i) \mid F_0\} = \psi(F_0) + \sum_{k=1}^j b_k c_{k,i} + O(n^{-(j+1)}). \quad (3.23)$$

To find the $c_{k,i}$'s notice that

$$\begin{aligned} E\{\psi(F_i) \mid F_0\} &= E[E\{\psi(F_i) \mid F_{i-1}\}] \\ &= E\left\{\psi(F_{i-1}) + \sum_{k_1=1}^j a_{k_1}(F_{i-1}) n^{-k_1} + O_p(n^{-(j+1)})\right\} \\ &= E\{\psi(F_{i-1}) \mid F_0\} + b_{1,i-1} + O(n^{-(j+1)}). \end{aligned} \quad (3.24)$$

Thus, substituting formulae for $E\{\psi(F_i) \mid F_0\}$, $E\{\psi(F_{i-1}) \mid F_0\}$ and $b_{1,i-1}$ from (3.22) and (3.23) into (3.24), we get

$$\sum_{k=1}^j b_k c_{k,i} = \sum_{k=1}^j b_k c_{k,i-1} + \sum_{k=1}^i \binom{i-1}{k-1} b_k.$$

Therefore, the $c_{k,i}$'s satisfy $c_{k,i} = c_{k,i-1} + \binom{i-1}{k-1} b_k$, for $k = 1, \dots, i$, with $c_{k,0} = 0$. Hence $c_{k,i} = \binom{i}{k} b_k$, for $k = 1, \dots, i$ and $c_{k,i} = 0$ for $k = i+1, \dots, j$. Thus, using (3.23), we have

$$E\{\psi(F_i) \mid F_0\} = \psi(F_0) + \sum_{k=1}^i \binom{i}{k} b_k + O(n^{-(j+1)}).$$

Consequently,

$$\begin{aligned}
E(\hat{\theta}_j) &= \sum_{i=1}^{j+1} \binom{j+1}{i} (-1)^{i+1} E\{\psi(F_i) \mid F_0\} \\
&= \psi(F_0) \sum_{i=1}^{j+1} \binom{j+1}{i} (-1)^{i+1} + \sum_{i=1}^{j+1} \sum_{k=1}^i \binom{j+1}{i} (-1)^{i+1} \binom{i}{k} b_k \\
&\quad + O(n^{-(j+1)}) \\
&= \psi(F_0) + \sum_{k=1}^j b_k \binom{j+1}{k} (-1)^{k+1} \sum_{m=0}^{j+1-k} \binom{j+1-k}{m} (-1)^m \\
&\quad + O(n^{-(j+1)}) \\
&= \psi(F_0) + O(n^{-(j+1)}),
\end{aligned}$$

since $\sum_{i=1}^{j+1} \binom{j+1}{i} (-1)^{i+1} = 1$ and $\sum_{m=0}^{j+1-k} \binom{j+1-k}{m} (-1)^m = 0$. This proves the result.

To see that $\text{var}(\hat{\theta}_j) = O(n^{-1})$, it suffices to show that $E(\hat{\theta}_j^2) = \theta^2 + O(n^{-1})$. Assume that we have expansion

$$E\{\psi(F_1)^2 \mid F_0\} = \psi(F_0)^2 + n^{-1}b(F_0) + O(n^{-2}),$$

for some functional b , and notice that, from the above argument,

$$E\{\psi(F_i) \mid F_1\} = \psi(F_1) + n^{-1} \binom{i-1}{1} a_1(F_1) + O_p(n^{-2}),$$

for a functional a_1 . Thus,

$$\begin{aligned}
E(\hat{\theta}_j^2) &= E\left(\left[\sum_{i=1}^{j+1} \binom{j+1}{i} (-1)^{i+1} \{\psi(F_1) + n^{-1}(i-1)a_1(F_1)\}\right]^2\right) + O(n^{-2}) \\
&= E\left[\sum_{i=1}^{j+1} \binom{j+1}{i}^2 \{\psi(F_1)^2 + 2n^{-1}(i-1)a_1(F_1)\psi(F_1)\}\right] \\
&\quad + 2E\left[\sum_{k>i}^{j+1} \sum_{i=1}^{j+1} \binom{j+1}{i} \binom{j+1}{k} (-1)^{i+k} \{\psi(F_1)^2\right. \\
&\quad \left.+ n^{-1}a_1(F_1)\psi(F_1)(i+k-2)\}\right] + O(n^{-2}) \\
&= \left[\sum_{i=1}^{j+1} \binom{j+1}{i} (-1)^{i+1} E\{\psi(F_1)^2 \mid F_0\}\right]^2
\end{aligned}$$

$$\begin{aligned}
& + n^{-1} a_1(F_0) \psi(F_0) \sum_{i=1}^{j+1} \sum_{k=1}^{j+1} \binom{j+1}{i} \binom{j+1}{k} (-1)^{i+k} (i+k-2) + O(n^{-2}) \\
& = \psi(F_0)^2 + n^{-1} \psi(F_0) \left\{ b(F_0) \right. \\
& \quad \left. + a_1(F_0) \sum_{i=1}^{j+1} \sum_{k=1}^{j+1} \binom{j+1}{i} \binom{j+1}{k} (-1)^{i+k} (i+k-2) \right\} + O(n^{-2}) \\
& = \psi(F_0)^2 + n^{-1} 2\psi(F_0) \{b(F_0) - a_1(F_0)\} + O(n^{-2}) \tag{3.25} \\
& = \psi(F_0)^2 + O(n^{-1}),
\end{aligned}$$

on noting that $\sum_{i=1}^{j+1} \sum_{k=1}^{j+1} \binom{j+1}{i} \binom{j+1}{k} (-1)^{i+k} (i+k-2) = -2$, since $\sum_{i=1}^{j+1} \binom{j+1}{i} \times (-1)^{i+1} = 1$. This proves the result. Observe that the term of order n^{-1} in (3.25) does not depend on j so we see that $\text{var}(\hat{\theta}_j) = \text{var}(\hat{\theta}_0) + O(n^{-2})$. This happens because successive bias corrections alter only higher-order terms in the variance expansion.

LEMMA 3.4. *If $\theta = \mu_1^2$ denotes squared population mean, $\hat{\theta} = \hat{\theta}_0 \equiv \bar{X}^2$ denotes squared sample mean, and $\hat{s}^2 \equiv n^{-1} \sum_i (X_i - \bar{X})^2$ is sample variance, then j iterations of the resampling principle yield the estimator*

$$\hat{\theta}_j \equiv \bar{X}^2 - (n-1)^{-1} (1 - n^{-j}) \hat{s}^2.$$

Proof. Let $Y_1, \dots, Y_n$ represent a generic resample from F_{i-1} , and $Y_1^*, \dots, Y_n^*$ a generic resample from F_i . Put $\psi(F_i) \equiv (\frac{1}{n} \sum_i Y_i)^2 = \bar{Y}^2$ and $s^2(F_i) = \frac{1}{n} \sum_i (Y_i - \bar{Y})^2$. Since $E(Y_j^* | F_i) = \bar{Y}$ and $E(Y_j^{*2} | F_i) = \frac{1}{n} \sum_i Y_i^2 = s^2(F_i) + \psi(F_i)$, it follows that

$$\begin{aligned}
E\{\psi(F_{i+1}) | F_i\} &= E(\bar{Y}^{*2} | F_i) = n^{-2} [n\{s^2(F_i) + \psi(F_i)\} + n(n-1)\psi(F_i)] \\
&= \psi(F_i) + n^{-1} s^2(F_i),
\end{aligned}$$

and

$$\begin{aligned}
E\{s^2(F_{i+1}) | F_i\} &= E\left(\frac{1}{n} \sum_i Y_i^{*2} - \bar{Y}^{*2} | F_i\right) \\
&= \{s^2(F_i) + \psi(F_i)\} - \{\psi(F_i) + n^{-1} s^2(F_i)\} \\
&= s^2(F_i)(1 - n^{-1}),
\end{aligned}$$

for $i \geq 1$. Therefore, taking expectations conditional on F_1 ,

$$E\{\psi(F_{i+1}) | F_1\} = E\{\psi(F_i) | F_1\} + n^{-1} E\{s^2(F_i) | F_1\}, \tag{3.26}$$

and

$$E\{s^2(F_{i+1}) \mid F_1\} = E\{s^2(F_i) \mid F_1\}(1 - n^{-1}), \quad (3.27)$$

for $i \geq 1$, with $\psi(F_1) = \bar{X}^2$ and $s^2(F_1) = n^{-1} \sum_i (X_i - \bar{X})^2 = \hat{s}^2$. From (3.27) we deduce

$$E\{s^2(F_i) \mid F_1\} = \hat{s}^2(1 - n^{-1})^{i-1}, \quad (3.28)$$

and substituting (3.28) into (3.26) and solving, we get

$$\begin{aligned} E\{\psi(F_{i+1}) \mid F_1\} &= \bar{X}^2 + n^{-1} \hat{s}^2 \sum_{k=0}^i (1 - n^{-1})^k \\ &= \bar{X}^2 + (n-1)^{-1} \hat{s}^2 \sum_{k=1}^{i+1} (1 - n^{-1})^k. \end{aligned} \quad (3.29)$$

Substituting (3.29) into (3.14) we obtain

$$\begin{aligned} \hat{\theta}_j &= \sum_{i=1}^{j+1} \binom{j+1}{i} (-1)^{i+1} \left\{ \bar{X}^2 + (n-1)^{-1} \hat{s}^2 \sum_{k=1}^i (1 - n^{-1})^k \right\} \\ &= \bar{X}^2 + (n-1)^{-1} \hat{s}^2 \sum_{i=1}^{j+1} \binom{j+1}{i} (-1)^{i+1} \sum_{k=1}^i (1 - n^{-1})^k, \end{aligned}$$

since $\sum_{i=1}^{j+1} \binom{j+1}{i} (-1)^{i+1} = 1$. Observing that

$$\begin{aligned} &\sum_{i=1}^{j+1} \binom{j+1}{i} (-1)^{i+1} \sum_{k=1}^i (1 - n^{-1})^k \\ &= \sum_{i=1}^{j+1} \binom{j+1}{i} (-1)^{i+1} (n-1) \left\{ 1 - (1 - \frac{1}{n})^{i-1} \right\} \\ &= (n-1) \left\{ 1 + (1-n)^{-1} \sum_{i=1}^{j+1} \binom{j+1}{i} (-1)^{i+1} (1 - \frac{1}{n})^i \right\} \\ &= n^{-j} - 1, \end{aligned}$$

we arrive at the desired formula.

7.2. Proof of Lemma 3.2. Note that if μ_j denotes the j 'th population moment, then $(\mu_{j_1})^{i_1} \cdots (\mu_{j_r})^{i_r}$ has a unique, unbiased, nonparametric estimator expressible as a polynomial in sample moments up to the $(\sum_k i_k j_k)$ 'th, with coefficients equal to inverses of polynomials in n^{-1} , and hence also to power series in n^{-1} . Successive iterations do no more than generate higher-order terms in these series.

7.3. Asymptotics for Lehmann-type intervals: Proof of Theorem 3.3.

Notation is from Section 5. Suppose that $J \equiv [\hat{\theta} - \hat{s}a, \hat{\theta} + \hat{s}b]$ has coverage $(1 - \alpha)$. Then $n^{\frac{1}{2}}a$ and $n^{\frac{1}{2}}b$ converge to $z \equiv z_{1-(\alpha/2)}$ and admit expansions $n^{\frac{1}{2}}a \sim z + \sum_{i \geq 1} n^{-i/2}a_i$ and $n^{\frac{1}{2}}b \sim z + \sum_{i \geq 1} n^{-\frac{1}{2}}b_i$. Also, we have $P(-b \leq T \leq a) = 1 - \alpha$. If J is Lehmann-type for θ then $J' \equiv [-\hat{\theta} - \hat{s}b, -\hat{\theta} + \hat{s}a]$ is Lehmann-type for $-\theta$. From this, and the fact that the polynomials q_i in (3.15) are even/odd functions for odd/even j , follows $b_i = (-1)^i a_i$, for $i \geq 1$. Substituting expansions for a and b into the formula $P(-b \leq T \leq a) = 1 - \alpha$, and using (3.15), we see, as in Section 9 of Chapter 2, that formula (2.33) of that section holds for a_i . Taking $l = 1$ in (2.33), we get

$$a_2 = -(\psi_{20} + \psi_{11}a_1 + \frac{1}{2}\psi_{02}a_1^2)/\psi_{01}. \quad (3.30)$$

Furthermore, a and b minimize

$$\begin{aligned} (1 - \alpha)l &\equiv (a + b)\mathbf{E}\{\hat{s}I(\theta \in J)\} \\ &= (a + b)[(1 - \alpha)\sigma + n^{-\frac{1}{2}}\tau\mathbf{E}\{ZI(-b \leq T \leq a)\}]. \end{aligned} \quad (3.31)$$

Suppose the joint density of Z and T admits expansion

$$f(x, t) = \phi(x, t) + n^{-\frac{1}{2}}r_1(x, t)\phi(x, t) + O(n^{-1}),$$

for a polynomial r_1 . Then,

$$\begin{aligned} &\mathbf{E}\{ZI(-b \leq T \leq a)\} \\ &= \int_{-b}^a dt \int_{-\infty}^{\infty} x f(x, t) dx \\ &= \int_{-b}^a dt \int_{-\infty}^{\infty} x \phi(x, t) dx + n^{-\frac{1}{2}} \int_{-z}^z dt \int_{-\infty}^{\infty} x r_1(x, t) \phi(x, t) dx + O(n^{-1}) \\ &= \int_{-z}^z dt \int_{-\infty}^{\infty} x \phi(x, t) dx + n^{-\frac{1}{2}} \left\{ a_1 \int_{-\infty}^{\infty} x \phi(x, z) dx + b_1 \int_{-\infty}^{\infty} x \phi(x, -z) dx \right. \\ &\quad \left. + \int_{-z}^z dt \int_{-\infty}^{\infty} x r_1(x, t) \phi(x, t) dx \right\} + O(n^{-1}). \end{aligned}$$

Therefore, by (3.31), and noting that $a_1 = -b_1$ and $a_2 = b_2$,

$$\begin{aligned} (1 - \alpha)l &= 2n^{-\frac{1}{2}}z\sigma(1 - \alpha) + 2n^{-1}\tau z \int_{-z}^z dt \int_{-\infty}^{\infty} x \phi(x, t) dx \\ &\quad + 2n^{-3/2} \left(a_2\sigma(1 - \alpha) + z \left[a_1\tau \int_{-\infty}^{\infty} x \{\phi(x, z) - \phi(x, -z)\} dx \right. \right. \end{aligned}$$

$$\begin{aligned}
& + \tau \int_{-z}^z dt \int_{-\infty}^{\infty} x r_1(x, t) \phi(x, t) dx \Big] \Big) + O(n^{-2}) \\
& = v_1 + 2n^{-3/2} \left\{ 2\tau a_1 z \int_{-\infty}^{\infty} x \phi(x, z) dx \right. \\
& \quad \left. - (1 - \alpha) \sigma(\psi_{11} a_1 + \tfrac{1}{2} \psi_{02} a_1^2) \psi_{01}^{-1} \right\} + O(n^{-2}),
\end{aligned}$$

the last line using (3.30), where v_1 denotes a term not depending on a_1, a_2 , etc. Differentiating the $n^{-3/2}$ term with respect to a_1 and equating to zero, we conclude that a_1 minimizing $(1 - \alpha)l$ satisfies

$$4z\tau \int_{-\infty}^{\infty} x \phi(x, z) dx - 2\sigma(1 - \alpha)(\psi_{11} + \psi_{02} a_1) \psi_{01}^{-1} = 0. \quad (3.32)$$

Since the asymptotic joint distribution of Z and T is bivariate Normal $N(0, 0; 1, 1, r)$, it follows that

$$\begin{aligned}
& \int_{-\infty}^{\infty} x \phi(x, z) dx \\
& = \int_{-\infty}^{\infty} (x - rz) \{2\pi(1 - r^2)\}^{-\frac{1}{2}} \exp\{-\tfrac{1}{2}(1 - r^2)^{-1}(x - rz)^2\} \phi(z) dx \\
& \quad + rz \phi(z) \int_{-\infty}^{\infty} \{2\pi(1 - r^2)\}^{-\frac{1}{2}} \exp\{-\tfrac{1}{2}(1 - r^2)^{-1}(x - rz)^2\} dx \\
& = rz \phi(z) = \psi.
\end{aligned}$$

Therefore, solving (3.32), we have

$$a_1 = (\rho \psi \psi_{01} - \psi_{11}) \psi_{02}^{-1}, \quad (3.33)$$

and, substituting (3.33) into (3.30),

$$a_2 = (\psi_{11}^2 - \rho^2 \psi^2 \psi_{01}^2 - 2\psi_{02} \psi_{20}) (2\psi_{01} \psi_{02})^{-1}. \quad (3.34)$$

The above calculations are for an “ideal” Lehmann-type interval J . By replacing a and b there by their empirical counterparts $\hat{a}$ and $\hat{b}$, we may repeat the argument for $\hat{J} \equiv [\hat{\theta} - \hat{s}\hat{a}, \hat{\theta} + \hat{s}\hat{b}]$ to obtain that $n^{\frac{1}{2}}\hat{a} = z + n^{-\frac{1}{2}}\hat{a}_1 + n^{-1}\hat{a}_2 + O_p(n^{-3/2})$ and $n^{\frac{1}{2}}\hat{b} = z - n^{-\frac{1}{2}}\hat{a}_1 + n^{-1}\hat{a}_2 + O_p(n^{-3/2})$, where $\hat{a}_1, \hat{a}_2$ admit the formulae for a_1 and a_2 , but with F_1 replacing F_0 . The mean length of $\hat{J}$ is

$$\begin{aligned}
\lambda_L & = E\{\hat{s}(\hat{a} + \hat{b})\} = n^{-\frac{1}{2}} 2z E(\hat{s}) + n^{-3/2} 2E(\hat{s}\hat{a}_2) + O(n^{-5/2}) \\
& = n^{-\frac{1}{2}} 2z E(\hat{s}) + n^{-3/2} 2\sigma a_2 + O(n^{-5/2}).
\end{aligned} \quad (3.35)$$

To obtain similar expressions for the mean lengths of equal-tailed and likelihood-based intervals, we take expectations of formula (4.11) of Hall (1988a), where in the case of equal-tailed intervals, $\hat{s}_2(z) \equiv \hat{q}_1(z)\hat{q}'_1(z) - \frac{1}{2}z\{\hat{q}_1(z)\}^2 - \hat{q}_2(z)$ (see formula (1.5) of Hall (1988a) with p 's replaced by q 's), and in the case of likelihood-based intervals, $\hat{s}_2(z) \equiv (\frac{1}{2}\hat{\psi}_{11}^2\hat{\psi}_{02}^{-1} - \hat{\psi}_{20})\hat{\psi}_{01}^{-1}$ (see formula (2.10) of Hall (1988a)). This gives

$$\lambda_{ET} = n^{-\frac{1}{2}}2zE(\hat{s}) + n^{-3/2}2\sigma[q_1(z)q'_1(z) - \frac{1}{2}z\{q_1(z)\}^2 - q_2(z)] + O(n^{-5/2}), \quad (3.36)$$

and

$$\lambda_{LB} = n^{-\frac{1}{2}}2zE(\hat{s}) + n^{-3/2}2\sigma(\psi_{11}^2 - 2\psi_{02}\psi_{20})(2\psi_{01}\psi_{02})^{-1} + O(n^{-5/2}). \quad (3.37)$$

From the definition of ψ_{ik} , for $i \geq 0$, we deduce

$$\begin{aligned} q_1(z) &= \psi_{10}\psi_{01}^{-1}, \quad q_2(z) = \psi_{20}\psi_{01}^{-1}, \\ q'_1(z) &= (\psi_{11}\psi_{01} - \psi_{10}\psi_{02})\psi_{01}^{-2}, \quad z = -\psi_{02}\psi_{01}^{-1}, \end{aligned} \quad (3.38)$$

so that $q_1(z)q'_1(z) - \frac{1}{2}z\{q_1(z)\}^2 - q_2(z) = \psi_{10}\psi_{01}^{-3}(\psi_{01}\psi_{11} - \psi_{10}\psi_{02}) + \frac{1}{2}\psi_{02}\psi_{10}^2\psi_{01}^{-3} - \psi_{20}\psi_{01}^{-1}$. Thus,

$$\begin{aligned} n^{3/2}(\lambda_L - \lambda_{ET}) &= 2\sigma(\frac{1}{2}\psi_{11}^2\psi_{01}^{-1}\psi_{02}^{-1} - \frac{1}{2}\rho^2\psi^2\psi_{01}\psi_{02}^{-1} - \psi_{20}\psi_{01}^{-1} - \psi_{10}\psi_{11}\psi_{01}^{-2} \\ &\quad + \psi_{10}^2\psi_{02}\psi_{01}^{-3} - \frac{1}{2}\psi_{02}\psi_{10}^2\psi_{01}^{-3} + \psi_{20}\psi_{01}^{-1}) + O(n^{-1}) \\ &= \sigma(\psi_{11}^2\psi_{01}^{-1}\psi_{02}^{-1} - 2\psi_{10}\psi_{11}\psi_{01}^{-2} + \psi_{02}\psi_{10}^2\psi_{01}^{-3} \\ &\quad - \rho^2\psi^2\psi_{01}\psi_{02}^{-1}) + O(n^{-1}) \\ &= \sigma\psi_{01}\psi_{02}^{-1}\{\psi_{01}^{-2}(\psi_{11} - \psi_{02}\psi_{10}\psi_{01}^{-1})^2 - \rho^2\psi^2\} + O(n^{-1}) \\ &= \sigma[(\rho\psi)^2 - \{q'_1(z)\}^2]z^{-1} + O(n^{-1}), \end{aligned}$$

on noting (3.33), (3.34) and (3.38). Formula (3.16) follows directly from (3.35) and (3.37).

To derive (3.18) and (3.19) we deduce coverage errors of the three types of interval using formula (4.14) of Hall (1988a). In the case of equal-tailed intervals, substituting $s_1(z) \equiv q_1(z)$ and $s_2(z) \equiv q_1(z)q'_1(z) - \frac{1}{2}z\{q_1(z)\}^2 - q_2(z)$ into that formula yields

$$\begin{aligned} \epsilon_{ET} &= -2n^{-1}[\frac{1}{2}\{q_1(z)\}^2z + q_1(z)\{q'_1(z) - q_1(z)z\} - q_2(z) - q_1(z)q'_1(z) \\ &\quad + \frac{1}{2}z\{q_1(z)\}^2 + q_2(z) + u_3z]\phi(z) + O(n^{-2}) \\ &= -n^{-1}2u_3z\phi(z) + O(n^{-2}), \end{aligned}$$

and for likelihood-based intervals, Hall (1988a), Section 4.6, gives

$$\epsilon_{LB} = -n^{-1}2u_1z\phi(z) + O(n^{-2}).$$

For Lehmann-type intervals, we use Hall's (1988a) formula (4.14) with $s_1 \equiv a_1$ and $s_2 \equiv a_2$. Thus,

$$\begin{aligned} \epsilon_L = & -2n^{-1} \left[\frac{1}{2}a_1^2z + a_1\{q'_1(z) - q_1(z)z\} - q_2(z) - a_2 + uz \right] \phi(z) \\ & + O(n^{-2}), \end{aligned} \quad (3.39)$$

where $u \equiv E\{n^{\frac{1}{2}}T(\hat{a}_1 - a_1)\} = u_1 + u_2$. Substituting formulae (3.33), (3.34) and (3.38) for a_1, a_2, q_1, q_2, q'_1 and z in terms of ψ_{ik} , for $i \geq 0$, into (3.39), we conclude that

$$\begin{aligned} \epsilon_L = & -2n^{-1} \left\{ -\frac{1}{2}(\rho\psi\psi_{01}\psi_{02}^{-1} - \psi_{11}\psi_{02}^{-1})^2\psi_{02}\psi_{01}^{-1} - (\rho\psi\psi_{01}\psi_{02}^{-1} - \psi_{11}\psi_{02}^{-1}) \right. \\ & \times (\psi_{11}\psi_{01}^{-1} - \psi_{10}\psi_{02}\psi_{01}^{-1} + \psi_{10}\psi_{02}\psi_{01}^{-1}) - \psi_{20}\psi_{01}^{-1} \\ & \left. - \frac{1}{2}\psi_{11}^2\psi_{02}^{-1}\psi_{01}^{-1} + \frac{1}{2}\rho^2\psi^2\psi_{01}\psi_{02}^{-1} + \psi_{20}\psi_{01}^{-1} + (u_1 + u_2)z \right\} \phi(z) + O(n^{-2}) \\ = & -2n^{-1} \left\{ -\frac{1}{2}\rho^2\psi^2\psi_{01}\psi_{02}^{-1} + \rho\psi\psi_{11}\psi_{02}^{-1} - \frac{1}{2}\psi_{11}^2\psi_{01}^{-1}\psi_{02}^{-1} - \rho\psi\psi_{11}\psi_{02}^{-1} \right. \\ & \left. + \psi_{11}^2\psi_{02}^{-1}\psi_{01}^{-1} - \frac{1}{2}\psi_{11}^2\psi_{02}^{-1}\psi_{01}^{-1} + \frac{1}{2}\rho^2\psi^2\psi_{01}\psi_{02}^{-1} + (u_1 + u_2)z \right\} \phi(z) \\ & + O(n^{-2}) \\ = & -n^{-1}2(u_1 + u_2)z\phi(z) + O(n^{-2}). \end{aligned}$$

Formulae (3.18) and (3.19) follow from the above expansions for ϵ_L , ϵ_{ET} and ϵ_{LB} .

7.4. Asymptotics for L^1 shrinkage. Minimizing $D(t)$ is equivalent to minimizing $d(t) \equiv (1-t)E|Z - n^{-\frac{1}{2}}a|$, where $a \equiv nt(\theta - c)\{\sigma(1-t)\}^{-1}$. If $\delta = O(n^{-\frac{1}{2}})$, then

$$\begin{aligned} E|Z - \delta| &= \int_{-\infty}^{\delta} (\delta - z)f(z) dz + \int_{\delta}^{\infty} (z - \delta)f(z) dz \\ &= \int_{-\infty}^0 (\delta - z)f(z) dz - \int_0^{\delta} (z - \delta)f(z) dz + \int_0^{\infty} (z - \delta)f(z) dz \\ &\quad - \int_0^{\delta} (z - \delta)f(z) dz \\ &= E|Z| + \delta \left(\int_{-\infty}^0 - \int_0^{\infty} \right) f(z) dz - 2 \int_0^{\delta} (z - \delta)f(z) dz \end{aligned}$$

$$\begin{aligned}
&= E|Z| - \delta n^{-\frac{1}{2}} 2 \int_0^\infty p(z) \phi(z) dz - 2 \int_0^\delta z f(z) dz \\
&\quad + 2\delta \int_0^\delta f(z) dz + O(n^{-2}) \\
&= E|Z| - \delta n^{-\frac{1}{2}} c_1 + \delta^2 \phi(0) + O(n^{-2}) \\
&= E|Z| - \delta n^{-\frac{1}{2}} c_1 + (2\pi)^{-\frac{1}{2}} \delta^2 + O(n^{-2}),
\end{aligned}$$

on noting $\phi(0) = (2\pi)^{-\frac{1}{2}}$, and the coefficient of $n^{-j/2}$ in expansion (3.20) is an odd/even polynomial according as j is odd/even. Hence,

$$\begin{aligned}
d(t) &= (1-t)E|Z| - c_1 \rho^{-1} t + (2\pi)^{-\frac{1}{2}} \rho^{-2} n t^2 (1-t)^{-1} + O(n^{-2}) \\
&= (1-t)E|Z| - c_1 \rho^{-1} t + (2\pi)^{-\frac{1}{2}} \rho^{-2} n t^2 + O(n^{-2}), \tag{3.40}
\end{aligned}$$

if $t = O(n^{-1})$. The minimizing value, t_0 , satisfies $2nt_0 = (2\pi)^{\frac{1}{2}}(\rho^2 E|Z| + \rho c_1) + O(n^{-1})$, on differentiating (3.40) with respect to t and equating to zero. Since $E|Z| = (2\pi^{-1})^{\frac{1}{2}} + O(n^{-1})$, it follows that $t_0 = n^{-1}\{\rho^2 + (\pi/2)^{\frac{1}{2}} \rho c_1\} + O(n^{-2})$. In the empirical case, we proceed in an almost identical manner, except that it is convenient to integrate by parts to express integrals involving f in terms of the distribution function. Noting that empirical versions of ρ and c_1 converge almost surely to their true values, we have $\hat{t}_0 \hat{t}_0^{-1} \rightarrow 1$ almost surely as $n \rightarrow \infty$.

APPENDIX A3.1. Proof of Lemma A3.1.

LEMMA A3.1. *If θ and s^2 denote population mean and variance respectively, and $\hat{\theta}$ and $\hat{s}^2$ are sample mean and sample variance based on a sample of size n , then*

$$E\{n^{\frac{1}{2}}(\hat{\theta} - \theta)/\hat{s}\} = -\frac{1}{2}n^{-\frac{1}{2}}\gamma + O(n^{-1}),$$

where γ is standardized population skewness.

Proof. Assume, without loss of generality, that the population has zero mean and unit variance, and write $U_1 \equiv n^{-\frac{1}{2}} \sum_i X_i$ and $U_2 \equiv n^{-\frac{1}{2}} \sum_i (X_i^2 - 1)$. The random variables U_1 and U_2 are $O_p(1)$. Notice that $\bar{X} = n^{-\frac{1}{2}}U_1$, and

$$\begin{aligned}\hat{s}^2 &\equiv n^{-1} \sum_i (X_i - \bar{X})^2 \\ &= 1 + n^{-1} \sum_i (X_i^2 - 1) - \bar{X}^2 \\ &= 1 + n^{-\frac{1}{2}}U_2 + O_p(1).\end{aligned}$$

Thus,

$$\begin{aligned}n^{\frac{1}{2}}(\hat{\theta} - \theta)/\hat{s} &= n^{\frac{1}{2}}\bar{X}\{1 + n^{-\frac{1}{2}}U_2 + O_p(n^{-1})\}^{-\frac{1}{2}} \\ &= U_1\{1 - \frac{1}{2}n^{-\frac{1}{2}}U_2 + O_p(n^{-1})\} \\ &= U_1 - \frac{1}{2}n^{-\frac{1}{2}}U_1U_2 + O_p(n^{-1}).\end{aligned}$$

The desired result follows on noting that $E(U_1) = 0$ and $E(U_1U_2) = \gamma$.

CHAPTER 4

ON BOOTSTRAP ITERATION FOR COVERAGE CORRECTION IN CONFIDENCE INTERVALS

1. INTRODUCTION, DEFINITIONS AND NOTATION

1.1. Introduction. We focus on the use of the iterated resampling principle introduced in Chapter 3 to reduce coverage error in confidence intervals. When applied in this context, the iterated resampling principle encompasses the iterated bootstrap coverage-correction techniques of Hall (1986a) and Beran (1987a), although in this chapter we shall confine our attention to Beran's method of iteration. This method can be viewed as a calibration technique using the bootstrap; see Loh (1987). We give direct proofs that in the case of one-sided confidence intervals, each iteration of the coverage-correction algorithm reduces coverage error by a factor of order $n^{-\frac{1}{2}}$, while for two-sided intervals, the reduction in error is by a factor of order n^{-1} . We also provide explicit formulae for dominant remainder terms in expansions of coverage error after bootstrap iteration in both one- and two-sided cases. Our result in the case of two-sided intervals sharpens one of Beran (1987a), who showed that each iteration of the principle reduced error in two-sided intervals by a factor of order $n^{-\frac{1}{2}}$.

Section 2 contains our main results about bootstrap iteration for coverage error reduction. Section 3 describes the effect of bootstrap coverage correction on individual critical points. For one-sided intervals, we show that one bootstrap iteration forces the critical points of the iterated interval to be second-order correct. This is not true for two-sided intervals, where second-order correctness does not play such a crucial role in attaining a high order of coverage accuracy. In Section 4, we show how bootstrap iteration affects the mean length of confidence intervals. It turns out that the change in mean length between original and iterated intervals is proportional to the coverage error of the original interval. This means that a coverage-corrected interval tends to be longer than the original

interval when the original interval undercovers, but tends to be shorter than the original interval when it overcovers. The results of a simulation study of iterated bootstrap confidence intervals for the mean are examined in Section 5.

We illustrate the use of the coverage-correction algorithm in Section 6, where we discuss the application of bootstrap iteration to two complex interval construction problems. In Subsection 6.1, we demonstrate that a single iteration of the resampling principle reduces the level error of a Bartlett-adjusted likelihood ratio test from $O(n^{-2})$ to $O(n^{-3})$, where n is sample size, and in Subsection 6.2 we discuss the problem of constructing accurate bootstrap confidence intervals for the correlation coefficient. For the latter, we propose two practical and accurate methods to overcome a problem often encountered in complex situations: that percentile-method intervals have poor coverage accuracy (see Rasmussen, 1987), and percentile- t intervals have erratically varying lengths and endpoints (see Efron, 1981a). The first method is based on coverage-correction of percentile-method intervals, and the second is the application of percentile- t methods to a transformed correlation coefficient. A simulation study supports the conclusion that both methods yield intervals with good coverage properties (unlike ordinary percentile intervals) and stable endpoints and length (unlike ordinary percentile- t intervals).

1.2. Parametric and nonparametric bootstraps. Suppose we are interested in constructing confidence intervals for a univariate parameter θ . Let $\hat{\theta}$ be our estimate of θ , based on a sample of size n , say $\mathcal{X} \equiv \{X_1, \dots, X_n\}$, and with asymptotic variance $n^{-1}\sigma^2$. Let $\hat{\sigma}^2$ be an estimate of σ^2 . In the nonparametric case, a “resample”, $\mathcal{X}^* \equiv \{X_1^*, \dots, X_n^*\}$, is drawn at random, with replacement from $\mathcal{X}$. In the parametric case, suppose the density f_λ of the sampling distribution is completely determined except for a vector λ of unknown parameters. Use $\mathcal{X}$ to estimate λ by $\hat{\lambda} = \hat{\lambda}(\mathcal{X})$, and write $\mathcal{X}^*$ for a random sample drawn from the population with density $f_{\hat{\lambda}}$. In both parametric and nonparametric cases, let $\hat{\theta}^*$ and $\hat{\sigma}^*$ be versions of $\hat{\theta}$ and $\hat{\sigma}$ computed using the resample $\mathcal{X}^*$ rather than the sample $\mathcal{X}$.

We shall use two examples throughout this chapter to illustrate our arguments and conclusions in nonparametric and parametric circumstances. First, suppose we are interested in estimating the mean θ of a continuous distribution without making parametric assumptions. Let $\hat{\theta}$ denote sample mean and $\hat{\sigma}^2$ sample variance, the latter having divisor n rather than $n - 1$. Then the

distributions of $n^{\frac{1}{2}}(\hat{\theta}^* - \hat{\theta})/\hat{\sigma}$ and $n^{\frac{1}{2}}(\hat{\theta}^* - \hat{\theta})/\hat{\sigma}^*$, conditional on $\mathcal{X}$, approximate the unconditional distributions of $n^{\frac{1}{2}}(\hat{\theta} - \theta)/\sigma$ and $n^{\frac{1}{2}}(\hat{\theta} - \theta)/\hat{\sigma}$ respectively. In a parametric framework, let $\hat{\theta}$ and $\hat{\sigma}$ be “bootstrap estimates” (that is, obtained by replacing functionals of a distribution function by functionals of the empiric) of θ and σ , and assume that the unknown parameters λ are functions of location and scale of $\hat{\theta}$, and that the statistics $n^{\frac{1}{2}}(\hat{\theta} - \theta)/\sigma$ and $n^{\frac{1}{2}}(\hat{\theta} - \theta)/\hat{\sigma}$ are location and scale invariant. Examples of this include inference about the mean in a Normal $N(\theta, \sigma^2)$ population, and about the mean θ of an exponential distribution. We shall focus on the latter example. For such cases, the distributions of $n^{\frac{1}{2}}(\hat{\theta}^* - \theta)/\hat{\sigma}$ and $n^{\frac{1}{2}}(\hat{\theta}^* - \theta)/\hat{\sigma}^*$, either conditional on $\mathcal{X}$, or unconditionally, are identical to those of $n^{\frac{1}{2}}(\hat{\theta} - \theta)/\sigma$ and $n^{\frac{1}{2}}(\hat{\theta} - \theta)/\hat{\sigma}$, respectively.

Note that “bootstrap population moments”, used throughout this chapter, have differing interpretations in nonparametric and parametric contexts. In the nonparametric case, bootstrap population moments are moments of the sample $\mathcal{X}$; in the parametric case, they are moments with respect to the density $f_{\hat{\lambda}}$. In the parametric case, we assume that $\int x f_{\hat{\lambda}}(x) dx$ equals the mean $\bar{X}$ of $\mathcal{X}$. For example, this is true if $f_{\hat{\lambda}}$ is from an exponential family and $\hat{\lambda}$ is the maximum likelihood estimate of λ .

1.3. Definitions and Notation. We close this section by defining the types of intervals discussed in this chapter, namely two types of percentile-method interval, percentile- t , and accelerated bias-corrected intervals.

The percentile- t argument for constructing confidence intervals is founded on the idea that the distribution of $T^* \equiv n^{\frac{1}{2}}(\hat{\theta}^* - \hat{\theta})/\hat{\sigma}^*$, conditional on $\mathcal{X}$, should approximate the unconditional distribution of $T \equiv n^{\frac{1}{2}}(\hat{\theta} - \theta)/\hat{\sigma}$. Thus, defining

$$\hat{t}_{\alpha} \equiv \sup\{t : P(T^* \leq t \mid \mathcal{X}) \leq \alpha\},$$

each of $J_1 \equiv (-\infty, \hat{\theta} - n^{-\frac{1}{2}}\hat{t}_{\alpha}\hat{\sigma}]$, $J_2 \equiv [\hat{\theta} - n^{-\frac{1}{2}}\hat{t}_{1-\alpha}\hat{\sigma}, \infty)$ and $J_3 \equiv [\hat{\theta} - n^{-\frac{1}{2}}\hat{t}_{1-(\alpha/2)}\hat{\sigma}, \hat{\theta} - n^{-\frac{1}{2}}\hat{t}_{\alpha/2}\hat{\sigma}]$ is a nominal $(1 - \alpha)$ -level percentile- t interval for θ .

There are two main types of percentile-method interval, the first based on the argument that the bootstrap distribution of $\hat{\theta}^*$ should be close to the unconditional distribution of $\hat{\theta}$, and the second obtained by arguing that the distribution of $\hat{\theta}^* - \hat{\theta}$, conditional on $\mathcal{X}$, should approximate the unconditional distribution of $\hat{\theta} - \theta$. Consequently, if we define

$$\hat{u}_{\alpha} \equiv \sup\{u : P(\hat{\theta}^* \leq u \mid \mathcal{X}) \leq \alpha\}, \quad \hat{v}_{\alpha} \equiv \sup\{v : P(\hat{\theta}^* - \hat{\theta} \leq v \mid \mathcal{X}) \leq \alpha\},$$

then each of $J_{11} \equiv (-\infty, \hat{u}_{1-\alpha}]$, $J_{12} \equiv [\hat{u}_\alpha, \infty)$ and $J_{13} \equiv [\hat{u}_{\alpha/2}, \hat{u}_{1-(\alpha/2)}]$ is a percentile-method interval of the first type, and each of $J_{21} \equiv (-\infty, \hat{\theta} - \hat{v}_\alpha]$, $J_{22} \equiv [\hat{\theta} - \hat{v}_{1-\alpha}, \infty)$ and $J_{23} \equiv [\hat{\theta} - \hat{v}_{1-(\alpha/2)}, \hat{\theta} - \hat{v}_{\alpha/2}]$ is a percentile-method interval of the second type. Each of the above intervals has nominal coverage $1 - \alpha$. Hall (1988a) termed the first type of percentile-method interval the “backwards” interval, and the second-type the “hybrid” interval. We shall retain Hall’s terminology for convenience.

Efron (1987) proposed his accelerated bias-corrected procedure as an improvement on the “backwards” percentile method, which, although invariant under monotone transformations, has poor coverage properties. Efron’s (1987) technique retains this invariance, but reduces coverage error. To describe the construction of an accelerated bias-corrected (ABC) interval, put

$$\begin{aligned}\hat{d} &\equiv \Phi^{-1}\{P(\hat{\theta}^* - \hat{\theta} \leq 0 \mid \mathcal{X})\}, \\ \hat{\delta}_a &= \hat{\delta}_a(\alpha) \equiv \Phi[\hat{d} + (\hat{d} + z_\alpha)\{1 - \hat{a}(\hat{d} + z_\alpha)\}^{-1}],\end{aligned}$$

where $\hat{a}$ is the bootstrap estimate of the “acceleration constant” and z_α is the solution of $\Phi(z_\alpha) = \alpha$. Let K_j , for $j = 1, 2, 3$, be the same as interval J_{1j} , except that percentiles $\hat{u}_\alpha$, $\hat{u}_{1-\alpha}$ etc. are replaced by $\hat{u}_{\hat{\delta}_a(\alpha)}$, $\hat{u}_{\hat{\delta}_a(1-\alpha)}$ etc. The intervals K_j , for $j = 1, 2, 3$, are accelerated bias-corrected intervals for θ with nominal coverage $1 - \alpha$. Hall (1988a) shows that in many important cases, the acceleration constant is simply one-sixth of the third moment of the first-order approximation to $n^{\frac{1}{2}}(\hat{\theta} - \theta)/\sigma$.

Recall from Section 4 of Chapter 2 that a critical point is said to be *second-order correct* if it agrees with the critical point of a theoretical “ideal” interval based on the distribution of $n^{\frac{1}{2}}(\hat{\theta} - \theta)/\hat{\sigma}$ to order $n^{-1} = (n^{-\frac{1}{2}})^2$; that is, to second order. We shall use the notion of second-order correctness in Section 3. Many of the results in this chapter follow on from results in Hall (1988a), and where practicable we have retained Hall’s notation. Further notation will be introduced as it is needed.

2. EFFECT OF BOOTSTRAP ITERATION ON COVERAGE ERROR

We describe in Subsection 2.1 the coverage-correction algorithm. In Subsections 2.2 and 2.3 we discuss coverage correction in one- and two-sided intervals, respectively, and in Subsection 2.4 we study bootstrap coverage correction in the context of the parametric bootstrap.

2.1. The Algorithm. The coverage correction algorithm runs as follows. Suppose we have a sample $\mathcal{X} \equiv \{X_1, \dots, X_n\}$ drawn from the distribution F_0 , and denote a resample from $\mathcal{X}$ by $\mathcal{X}^* \equiv \{X_1^*, \dots, X_n^*\}$. (Recall that the resample $\mathcal{X}^*$ is drawn differently in parametric and nonparametric cases.) Let $I_0 = I_0(\alpha)$ be a nominal $(1 - \alpha)$ -level confidence interval for the unknown parameter θ . To indicate the dependence of I_0 on $\mathcal{X}$, we write it as $I_0(\alpha | \mathcal{X})$. Denote the coverage of $I_0(\alpha | \mathcal{X})$ by $\pi_0(\alpha)$; that is, $\pi_0(\alpha) \equiv P\{\theta \in I_0(\alpha | \mathcal{X})\}$. Adopting a slightly different notation to that introduced in Chapter 3, index the class of functionals f by the adjusted nominal coverage by putting

$$f_{\beta_\alpha}(F_0, F_1) \equiv I\{\theta \in I_0(\beta_\alpha | \mathcal{X})\} - (1 - \alpha),$$

where I is an indicator function. To obtain an interval $I_0(\beta_\alpha | \mathcal{X})$ with exact coverage $1 - \alpha$, we must solve the equation $E\{f_{\beta_\alpha}(F_0, F_1) | F_0\} = 0$ for β_α ; that is, we must solve for β_α the equation

$$\pi_0(\beta_\alpha) \equiv P\{\theta \in I_0(\beta_\alpha | \mathcal{X}) | F_0\} = 1 - \alpha. \quad (4.1)$$

Now, let $I_0(\alpha | \mathcal{X}^*)$ denote the version of $I_0(\alpha | \mathcal{X})$ computed using the resample $\mathcal{X}^*$ rather than the sample $\mathcal{X}$. The resampling principle argues that $\hat{\beta}_\alpha = \beta_\alpha(F_1)$ be chosen to solve the equation $E\{f_{\hat{\beta}_\alpha}(F_1, F_2) | F_1\} = 0$; that is

$$\hat{\pi}_0(\hat{\beta}_\alpha) = P\{\hat{\theta} \in I_0(\hat{\beta}_\alpha | \mathcal{X}^*) | \mathcal{X}\} = 1 - \alpha, \quad (4.2)$$

and the bootstrap coverage-corrected interval is $I_1(\alpha) \equiv I_0(\hat{\beta}_\alpha | \mathcal{X})$. Our method of iteration can be interpreted as a calibration technique; see Loh (1987). Clearly, by replacing I_0 by I_1 in the preceding algorithm, we may iterate the procedure to further correct for coverage.

2.2. One-sided intervals. Let $I_0 = I_0(\alpha | \mathcal{X}) \equiv (-\infty, \hat{\theta}(\alpha)]$ denote a one-sided, $(1 - \alpha)$ -level confidence interval for θ , for $0 < \alpha < 1$, and suppose that the coverage error of I_0 is $O(n^{-j/2})$ for some $j \geq 1$. Typically, for percentile-method and Normal theory intervals we have $j = 1$, while for percentile- t and accelerated bias-corrected intervals, $j = 2$. Assume that the coverage probability of I_0 admits the expansion

$$\pi_0(\alpha) \equiv P\{\theta \in I_0(\alpha | \mathcal{X})\} = (1 - \alpha) + n^{-j/2} r_1(z_{1-\alpha}) \phi(z_{1-\alpha}) + O(n^{-(j+1)/2}), \quad (4.3)$$

where $z_{1-\alpha}$ is the solution of $\Phi(z_{1-\alpha}) = 1 - \alpha$, and r_1 is a polynomial with coefficients depending on the moments of X ; see Hall (1988a), Section 4.5. Then, the bootstrap estimate of π_0 , $\hat{\pi}_0$, satisfies

$$\hat{\pi}_0(\alpha) \equiv P\{\hat{\theta} \in I_0(\alpha | \mathcal{X}^*) | \mathcal{X}\} = (1 - \alpha) + n^{-j/2} \hat{r}_1(z_{1-\alpha}) \phi(z_{1-\alpha}) + O_p(n^{-(j+1)/2}), \quad (4.4)$$

where $\hat{r}_1$ differs from r_1 in that population moments appearing in coefficients of the latter are replaced by their respective bootstrap versions in the former. Let $\hat{\beta}_\alpha$ be the solution of equation (4.2), and put $I_1 = I_1(\alpha) \equiv I_0(\hat{\beta}_\alpha | \mathcal{X})$. Let v_1 be the polynomial such that

$$nE[(\hat{\theta} - \theta)\sigma^{-1}\{\hat{r}_1(z_{1-\alpha}) - r_1(z_{1-\alpha})\}] = v_1(z_{1-\alpha}) + O(n^{-1}), \quad (4.5)$$

and put $r_2(z_{1-\alpha}) \equiv z_{1-\alpha}v_1(z_{1-\alpha})$. Theorem 4.1, proved in Subsection 7.1, gives an expression for the coverage probability of I_1 .

THEOREM 4.1. *Using the above notation, the interval I_1 has coverage*

$$\pi_1(\alpha) \equiv P\{\theta \in I_1(\alpha | \mathcal{X})\} = (1 - \alpha) + n^{-(j+1)/2} r_2(z_{1-\alpha}) \phi(z_{1-\alpha}) + O(n^{-(j+2)/2}). \quad (4.6)$$

Therefore, the coverage-correction algorithm has reduced error from $O(n^{-j/2})$ to $O(n^{-(j+1)/2})$.

2.3. Two-sided intervals. Let $I_0 = I_0(\alpha | \mathcal{X}) \equiv [\hat{\theta}_1(\alpha), \hat{\theta}_2(\alpha)]$ denote a two-sided $(1 - \alpha)$ -level confidence interval for θ , for $0 < \alpha < 1$, and assume that the coverage error of I_0 is $O(n^{-j})$ for some $j \geq 1$. All of the common two-sided bootstrap intervals typically have coverage error of order n^{-1} , as does the Normal theory interval. Suppose that the coverage probability of I_0 admits expansion

$$\pi_0(\alpha) \equiv P\{\theta \in I_0(\alpha | \mathcal{X})\} = (1 - \alpha) + n^{-j} t_1(z) \phi(z) + O(n^{-(j+1)}), \quad (4.7)$$

where $z \equiv z_{1-(\alpha/2)}$ is the solution of $\Phi(z_{1-(\alpha/2)}) = 1 - (\alpha/2)$, and t_1 is an odd polynomial with coefficients depending on the moments of X . Oddness is a characteristic property of t_1 , not a condition; see Hall (1988a), Section 4.5, in the case $j = 2$. The bootstrap estimate of π_0 satisfies

$$\hat{\pi}_0(\alpha) \equiv P\{\hat{\theta} \in I_0(\alpha | \mathcal{X}^*) | \mathcal{X}\} = (1 - \alpha) + n^{-j} \hat{t}_1(z) \phi(z) + O_p(n^{-(j+1)}), \quad (4.8)$$

where $\hat{t}_1$ is obtained from t_1 by replacing population moments there by their bootstrap estimates. Let $\hat{\beta}_\alpha$ be the solution of equation (4.2), and put $I_1 = I_1(\alpha | \mathcal{X}) \equiv I_0(\hat{\beta}_\alpha | \mathcal{X})$. Let v_2 be the polynomial such that

$$nE[(\hat{\theta} - \theta)\sigma^{-1}\{\hat{t}_1(z) - t_1(z)\}] = v_2(z) + O(n^{-1}), \quad (4.9)$$

and $g(x, y) \equiv \sum_i \sum_j a_{ij} x^i y^j$ the odd, bivariate cubic polynomial such that the joint density function h_n of $[n^{\frac{1}{2}}(\hat{\theta} - \theta)\hat{\sigma}^{-1}, \frac{1}{2}n^{\frac{1}{2}}\{\hat{t}_1(z) - t_1(z)\}]$ admits expansion

$$h_n(x, y) = \phi(x, y) + n^{-\frac{1}{2}}g(x, y)\phi(x, y) + O(n^{-1}),$$

where $\phi(x, y)$ is the joint density of the bivariate Normal limit. Assume that for polynomials $\hat{s}_{1i}$ and $\hat{s}_{2i}$, the critical points $\hat{\theta}_i(\alpha)$, for $i = 1, 2$, admit Cornish-Fisher expansions

$$\hat{\theta}_i(\alpha) = \hat{\theta} + n^{-\frac{1}{2}}\hat{\sigma}\{(-1)^i z + n^{-\frac{1}{2}}\hat{s}_{1i}(z) + n^{-1}\hat{s}_{2i}(z)\} + O_p(n^{-2}). \quad (4.10)$$

Both $\hat{s}_{1i}$ and $\hat{s}_{2i}$ are sample versions of nonrandom polynomials s_{1i} and s_{2i} in which population moments have been replaced by sample moments. It turns out that to ensure that $\pi_0(\alpha) = 1 - \alpha + O(n^{-1})$, it is necessary to have $\hat{s}_{11} \equiv \hat{s}_{12}$. We also need $\hat{s}_{21} \equiv -\hat{s}_{22}$. The latter would normally hold in practice, since $\hat{s}_{21}(x) = \hat{s}_{22}(-x)$, and $\hat{s}_{22}$ is an odd polynomial. We impose it as a condition.

Let $D_1 \equiv n^{\frac{1}{2}}(\hat{\theta} - \theta)\hat{\sigma}^{-1}$ and $D_2 \equiv \frac{1}{2}n^{\frac{1}{2}}\{\hat{t}_1(z) - t_1(z)\}$, let ρ , σ_2^2 be the asymptotic correlation between D_1 and D_2 and the asymptotic variance of D_2 respectively, and N be a Standard Normal variate. Put

$$\begin{aligned} \frac{1}{2}w(\theta) &\equiv \theta^2 v_2(z)\{\theta s_{11}(z) + s'_{11}(z)\} + \theta \sum_i \sum_j \sum_l (-1)^{i+l} a_{ij} \binom{j+1}{l} \\ &\times E(N^{j+1-l}) \sigma_2^{j+1} \rho^l (1 - \rho^2)^{\frac{1}{2}j+1-l} He_{i+l-1}(\theta), \end{aligned}$$

where $He_j(x)$ is the j 'th Hermite polynomial (see, for example, Petrov (1975), p.137), and let η be the polynomial such that

$$\int e^{itx} d\{\eta(x)\phi(x)\} = w(it)e^{-\frac{1}{2}t^2},$$

and put $t_2(z) \equiv \eta(-z)$. We give in Theorem 4.2 an expression for the coverage probability of the coverage-corrected interval, I_1 .

THEOREM 4.2. Assume $\hat{s}_{21} = -\hat{s}_{22}$. Using the notation introduced above, the interval I_1 has coverage probability

$$\pi_1(\alpha) \equiv P\{\theta \in I_1(\alpha \mid \mathcal{X})\} = (1 - \alpha) + n^{-(j+1)}t_2(z)\phi(z) + O(n^{-(j+2)}). \quad (4.11)$$

Theorem 4.2 is proved in Subsection 7.2, and confirms that in two-sided intervals, bootstrap iteration reduces coverage error from $O(n^{-j})$ to $O(n^{-(j+1)})$.

2.4. Coverage correction and the parametric bootstrap. We show that, provided $\hat{\theta} - \theta$ can be made exactly pivotal by rescaling, a single iteration of the coverage-correction algorithm produces an interval with perfect coverage accuracy. Recall from Subsection 1.2 that in certain parametric problems, the distribution of $n^{\frac{1}{2}}(\hat{\theta}^* - \hat{\theta})/\hat{\sigma}^*$, either conditional on $\mathcal{X}$, or unconditionally, is identical to the unconditional distribution of $n^{\frac{1}{2}}(\hat{\theta} - \theta)/\hat{\sigma}$. Consequently, an interval based on the bootstrap distribution of $n^{\frac{1}{2}}(\hat{\theta} - \theta)/\hat{\sigma}$, namely a percentile- t interval, has perfect coverage accuracy. Bootstrap coverage correction in that case merely reproduces the same interval. Therefore, we focus on non-pivotal intervals, such as those derived by “backwards” or “hybrid” percentile methods. It turns out that a single bootstrap iteration of any one of the intervals discussed in Subsection 1.3 produces a percentile- t interval with exact coverage, with further iterations reproducing the same interval. We illustrate this in the case of a one-sided “backwards” percentile interval $(-\infty, \hat{u}_{1-\alpha}]$ for θ ; see Subsection 1.3. Equation (4.1) may be written, using functional notation,

$$\begin{aligned} 1 - \alpha &= \pi_0(\beta_\alpha) = P\{\theta(F_0) \leq \hat{u}_{1-\beta_\alpha}(F_1)\} \\ &= P[\theta(F_0) \leq \sup\{u : P[\theta(F_2) \leq u \mid F_1] \leq 1 - \beta_\alpha\}] \\ &= P[\theta(F_0) \leq \sup\{\theta(F_1) + n^{-\frac{1}{2}}\sigma(F_1)t : P[n^{\frac{1}{2}}\{\theta(F_2) \\ &\quad - \theta(F_1)\}/\sigma(F_1) \leq t \mid F_1] \leq 1 - \beta_\alpha\}] \\ &= P[-n^{\frac{1}{2}}\{\theta(F_1) - \theta(F_0)\}/\sigma(F_1) \leq \sup\{t : P[n^{\frac{1}{2}}\{\theta(F_2) \\ &\quad - \theta(F_1)\}/\sigma(F_1) \leq t \mid F_1] \leq 1 - \beta_\alpha\}]. \end{aligned} \quad (4.12)$$

On replacing F_0 by F_1 , F_1 by F_2 and F_2 by F_3 in (4.12), equation (4.2) becomes

$$\begin{aligned} 1 - \alpha &= \hat{\pi}_0(\hat{\beta}_\alpha) \\ &= P[-n^{\frac{1}{2}}\{\theta(F_2) - \theta(F_1)\}/\sigma(F_2) \\ &\quad \leq \sup\{t : P[n^{\frac{1}{2}}\{\theta(F_3) - \theta(F_2)\}/\sigma(F_2) \leq t \mid F_2] \leq 1 - \hat{\beta}_\alpha\} \mid F_1] \end{aligned}$$

$$\begin{aligned}
&= P \left[-n^{\frac{1}{2}} \{ \theta(F_2) - \theta(F_1) \} / \sigma(F_2) \right. \\
&\quad \leq n^{\frac{1}{2}} \sigma(F_2)^{-1} \sup \{ u : P[\theta(F_3) \leq u \mid F_2] \leq 1 - \hat{\beta}_\alpha \} \\
&\quad \left. - n^{\frac{1}{2}} \theta(F_2) \sigma(F_2)^{-1} \mid F_1 \right] \\
&= P \left[-n^{\frac{1}{2}} \{ \theta(F_2) - \theta(F_1) \} / \sigma(F_2) \right. \\
&\quad \left. \leq n^{\frac{1}{2}} \sigma(F_2)^{-1} \hat{u}_{1-\hat{\beta}_\alpha}(F_2) - n^{\frac{1}{2}} \theta(F_2) \sigma(F_2)^{-1} \mid F_1 \right] \\
&= 1 - P \left[n^{\frac{1}{2}} \{ \theta(F_2) - \theta(F_1) \} / \sigma(F_2) \right. \\
&\quad \left. \leq n^{\frac{1}{2}} \theta(F_2) \sigma(F_2)^{-1} - n^{\frac{1}{2}} \sigma(F_2)^{-1} \hat{u}_{1-\hat{\beta}_\alpha}(F_2) \mid F_1 \right]. \tag{4.13}
\end{aligned}$$

However, the distribution of $n^{\frac{1}{2}} \{ \theta(F_2) - \theta(F_1) \} / \sigma(F_2)$, conditional on F_1 , is identical to the unconditional distribution of $n^{\frac{1}{2}} \{ \theta(F_1) - \theta(F_0) \} / \sigma(F_1)$, and the percentile- t interval $(-\infty, \hat{\theta} - n^{-\frac{1}{2}} \hat{t}_\alpha \hat{\sigma})$ is exact so that $P\{n^{\frac{1}{2}}(\hat{\theta} - \theta)/\hat{\sigma} \leq \hat{t}_\alpha\} = \alpha$. Therefore, it follows from (4.13) that

$$\hat{t}_\alpha = n^{\frac{1}{2}} \theta(F_1) \sigma(F_1)^{-1} - n^{\frac{1}{2}} \sigma(F_1)^{-1} \hat{u}_{1-\hat{\beta}_\alpha}(F_1);$$

that is,

$$\hat{u}_{1-\hat{\beta}_\alpha} = \hat{\theta} - n^{-\frac{1}{2}} \hat{t}_\alpha \hat{\sigma}.$$

Thus, the one-sided, iterated interval is $(-\infty, \hat{\theta} - n^{-\frac{1}{2}} \hat{t}_\alpha \hat{\sigma}]$, which is just an exact $(1 - \alpha)$ -level percentile- t interval for θ . A similar argument shows that a single iteration of a “hybrid” percentile interval also produces an exact percentile- t interval. The general result does require the property that $\hat{\theta} - \theta$ can be made exactly pivotal by rescaling, and so does not hold universally in parametric problems.

3. EFFECT OF BOOTSTRAP COVERAGE CORRECTION ON CRITICAL POINTS

We derive expansions for critical points of once-iterated bootstrap confidence intervals in both one- and two-sided cases. Behaviour differs between one- and two-sided intervals, so we treat them separately, in Subsections 3.1 and 3.2 respectively.

3.1. One-sided intervals. Let $I_0 = I_0(\alpha) \equiv (-\infty, \hat{\theta}(\alpha)]$ denote a $(1 - \alpha)$ -level one-sided confidence interval for θ , and assume that the critical point $\hat{\theta}(\alpha)$ admits Cornish-Fisher expansion

$$\hat{\theta}(\alpha) = \hat{\theta} + n^{-\frac{1}{2}} \hat{\sigma} \{ z_{1-\alpha} + n^{-\frac{1}{2}} \hat{s}_1(z_{1-\alpha}) + n^{-1} \hat{s}_2(z_{1-\alpha}) \} + O_p(n^{-2}), \tag{4.14}$$

where $\hat{s}_1$ and $\hat{s}_2$ are sample versions of theoretical polynomials s_1 and s_2 , and $z_{1-\alpha}$ is the solution of $\Phi(z_{1-\alpha}) = 1 - \alpha$. Hall (1988a), Section 4.2, gives Cornish-Fisher expansions of the common bootstrap critical points in terms of the polynomials p_1 , q_1 , p_{21} , and q_{21} defined in Section 1.4 of that paper. In particular, percentile- t , “backwards” percentile, “hybrid” percentile and accelerated bias-corrected critical points admit expansions

$$\begin{aligned}\hat{\theta}_{STUD}(\alpha) &= \hat{\theta} + n^{-\frac{1}{2}}\hat{\sigma}\{z_{1-\alpha} + n^{-\frac{1}{2}}\hat{q}_1(z_{1-\alpha}) + n^{-1}\hat{q}_{21}(z_{1-\alpha})\} \\ &\quad + O_p(n^{-2}),\end{aligned}\tag{4.15}$$

$$\begin{aligned}\hat{\theta}_{BACK}(\alpha) &= \hat{\theta} + n^{-\frac{1}{2}}\hat{\sigma}\{z_{1-\alpha} - n^{-\frac{1}{2}}\hat{p}_1(z_{1-\alpha}) + n^{-1}\hat{p}_{21}(z_{1-\alpha})\} \\ &\quad + O_p(n^{-2}),\end{aligned}\tag{4.16}$$

$$\begin{aligned}\hat{\theta}_{HYB}(\alpha) &= \hat{\theta} + n^{-\frac{1}{2}}\hat{\sigma}\{z_{1-\alpha} + n^{-\frac{1}{2}}\hat{p}_1(z_{1-\alpha}) + n^{-1}\hat{p}_{21}(z_{1-\alpha})\} \\ &\quad + O_p(n^{-2}),\end{aligned}\tag{4.17}$$

and

$$\begin{aligned}\hat{\theta}_{ABC}(\alpha) &= \hat{\theta} + n^{-\frac{1}{2}}\hat{\sigma}\{z_{1-\alpha} + n^{-\frac{1}{2}}\hat{q}_1(z_{1-\alpha}) + n^{-1}(\{\hat{p}_1(z_{1-\alpha}) + \hat{q}_1(z_{1-\alpha})\} \\ &\quad \times [\{\hat{p}_1(z_{1-\alpha}) + \hat{q}_1(z_{1-\alpha}) - 2\hat{p}_1(0)\}z_{1-\alpha}^{-1} - \hat{p}'_1(z_{1-\alpha})] \\ &\quad + \hat{p}_{21}(z_{1-\alpha}))\} + O_p(n^{-2}),\end{aligned}\tag{4.18}$$

respectively, where we have retained Hall’s notation. The theoretical “ideal” interval based on the distribution of $n^{\frac{1}{2}}(\hat{\theta} - \theta)/\hat{\sigma}$ has critical point $\hat{\theta}_{ideal}(\alpha)$ satisfying

$$\hat{\theta}_{ideal}(\alpha) = \hat{\theta} + n^{-\frac{1}{2}}\hat{\sigma}\{z_{1-\alpha} + n^{-\frac{1}{2}}q_1(z_{1-\alpha}) + n^{-1}q_{21}(z_{1-\alpha})\} + O(n^{-2}),$$

and the interval $(-\infty, \hat{\theta}_{ideal}(\alpha)]$ has coverage probability exactly $1 - \alpha$. Therefore, a critical point admitting expansion (4.14) is second-order correct if $s_1(z) = q_1(z)$, so that percentile- t and accelerated bias-corrected critical points are second-order correct, while “backwards” and “hybrid” percentile-method critical points are not.

Suppose that for some $j \geq 1$, the interval I_0 has coverage probability $\pi_0(\alpha)$ which admits expansion

$$\pi_0(\alpha) = 1 - \alpha + n^{-j/2}r_1(z_{1-\alpha})\phi(z_{1-\alpha}) + n^{-(j+1)/2}r(z_{1-\alpha})\phi(z_{1-\alpha}) + O(n^{-(j+2)/2}),\tag{4.19}$$

for polynomials r_1 and r . Denote the iterated interval I_1 by $(-\infty, \hat{\theta}^{(1)}(\alpha)]$. We show in Theorem 4.3, proved in Subsection 7.3, that bootstrap coverage correction produces an interval with second-order correct critical point.

THEOREM 4.3. *Bootstrap coverage correction of $I_0 \equiv (-\infty, \hat{\theta}(\alpha)]$ produces $I_1 \equiv (-\infty, \hat{\theta}^{(1)}(\alpha)]$ where $\hat{\theta}^{(1)}(\alpha)$ is a second-order correct critical point admitting expansion*

$$\begin{aligned} \hat{\theta}^{(1)}(\alpha) = \hat{\theta} + n^{-\frac{1}{2}}\hat{\sigma} [z_{1-\alpha} + n^{-\frac{1}{2}}\hat{q}_1(z_{1-\alpha}) + n^{-1}\{q_{21}(z_{1-\alpha}) + uz_{1-\alpha}\}] \\ + O_p(n^{-2}), \end{aligned} \quad (4.20)$$

in which q_1 and q_{21} are polynomials occurring in the expansion of the “ideal” $(1 - \alpha)$ -level critical point based on the distribution of $n^{\frac{1}{2}}(\hat{\theta} - \theta)/\hat{\sigma}$, and u is the constant such that

$$nE[(\hat{\theta} - \theta)\hat{\sigma}^{-1}\{\hat{s}_1(z_{1-\alpha}) - s_1(z_{1-\alpha})\}] = u + O(n^{-1}).$$

Applying Theorem 4.3 to percentile- t , “backwards” and “hybrid” percentile, and accelerated bias-corrected critical points (4.14)–(4.18), we obtain expressions for the critical points of the corresponding iterated intervals,

$$\begin{aligned} \hat{\theta}_{STUD}^{(1)}(\alpha) = \hat{\theta} + n^{-\frac{1}{2}}\hat{\sigma}[z_{1-\alpha} + n^{-\frac{1}{2}}\hat{q}_1(z_{1-\alpha}) + n^{-1}\{q_{21}(z_{1-\alpha}) + u_1z_{1-\alpha}\}] \\ + O_p(n^{-2}), \end{aligned} \quad (4.21)$$

$$\begin{aligned} \hat{\theta}_{BACK}^{(1)}(\alpha) = \hat{\theta} + n^{-\frac{1}{2}}\hat{\sigma}[z_{1-\alpha} + n^{-\frac{1}{2}}\hat{q}_1(z_{1-\alpha}) + n^{-1}\{q_{21}(z_{1-\alpha}) - u_2z_{1-\alpha}\}] \\ + O_p(n^{-2}), \end{aligned} \quad (4.22)$$

$$\begin{aligned} \hat{\theta}_{HYB}^{(1)}(\alpha) = \hat{\theta} + n^{-\frac{1}{2}}\hat{\sigma}[z_{1-\alpha} + n^{-\frac{1}{2}}\hat{q}_1(z_{1-\alpha}) + n^{-1}\{q_{21}(z_{1-\alpha}) + u_2z_{1-\alpha}\}] \\ + O_p(n^{-2}), \end{aligned} \quad (4.23)$$

and $\hat{\theta}_{ABC}^{(1)}(\alpha)$ agrees with $\hat{\theta}_{STUD}^{(1)}(\alpha)$ up to terms of $O(n^{-2})$, where u_1 and u_2 are constants such that

$$nE[(\hat{\theta} - \theta)\hat{\sigma}^{-1}\{\hat{q}_1(z_{1-\alpha}) - q_1(z_{1-\alpha})\}] = u_1 + O(n^{-1}), \quad (4.24)$$

and

$$nE[(\hat{\theta} - \theta)\hat{\sigma}^{-1}\{\hat{p}_1(z_{1-\alpha}) - p_1(z_{1-\alpha})\}] = u_2 + O(n^{-1}), \quad (4.25)$$

respectively. A second iteration of the coverage-correction algorithm to the “backwards” and “hybrid” percentile intervals renders the twice-iterated critical points $\hat{\theta}_{BACK}^{(2)}(\alpha)$ and $\hat{\theta}_{HYB}^{(2)}(\alpha)$ *third-order equivalent* to the critical points $\hat{\theta}_{STUD}^{(1)}(\alpha)$ and $\hat{\theta}_{ABC}^{(1)}(\alpha)$; that is, the critical points agree up to but not including terms of order n^{-2} .

Hall (1988a), equation (4.14), gives the following formula for coverage probability of one-sided intervals with critical points admitting expansion (4.14),

$$\begin{aligned}\pi_0(\alpha) = & 1 - \alpha + n^{-\frac{1}{2}}\{s_1(z_{1-\alpha}) - q_1(z_{1-\alpha})\} \\ & - n^{-1}[\frac{1}{2}s_1^2(z_{1-\alpha})z_{1-\alpha} + s_1(z_{1-\alpha})\{q_1'(z_{1-\alpha}) - q_1(z_{1-\alpha})z_{1-\alpha}\} \\ & - q_2(z_{1-\alpha}) - s_2(z_{1-\alpha}) + uz_{1-\alpha}]\phi(z_{1-\alpha}) + O(n^{-3/2}),\end{aligned}\quad (4.26)$$

where u is defined in Theorem 4.3. Putting $j = 1$ in Theorem 4.1, it follows from (4.26) that $r_1(z_{1-\alpha}) = s_1(z_{1-\alpha}) - q_1(z_{1-\alpha})$. Noting (4.25) and applying Theorem 4.1 with $j = 1$, we deduce coverage probability expansions for once-iterated “backwards” and “hybrid” percentile intervals:

$$\pi_{1,BACK}(\alpha) = 1 - \alpha - n^{-1}(u_1 + u_2)z_{1-\alpha}\phi(z_{1-\alpha}) + O(n^{-3/2}), \quad (4.27)$$

and

$$\pi_{1,HYB}(\alpha) = 1 - \alpha + n^{-1}(u_2 - u_1)z_{1-\alpha}\phi(z_{1-\alpha}) + O(n^{-3/2}), \quad (4.28)$$

respectively, since $s_1 = -p_1$ for “backwards” intervals and $s_1 = p_1$ for “hybrid” intervals. For percentile- t and accelerated bias-corrected intervals, $s_1 = q_1$, so both $\pi_{1,STUD}(\alpha)$ and $\pi_{1,ABC}(\alpha)$ are equal to $1 - \alpha + O(n^{-3/2})$, by Theorem 4.1. Similarly, the coverage probabilities of twice-iterated “backwards” and “hybrid” intervals are in error by $O(n^{-3/2})$. We illustrate our arguments and conclusions through two examples.

Example 1: Nonparametric estimation of mean. Let $X_1, \dots, X_n$ be independent and identically distributed observations from a continuous, univariate population with zero mean and unit variance. Define standardized skewness, kurtosis and fifth cumulant by $\gamma \equiv E(X_1^3)$, $\kappa_4 \equiv E(X_1^4) - 3$, and $\kappa_5 \equiv E(X_1^5) - 10E(X_1^3)$, respectively. Sample versions of the mean, variance and standardized third, fourth, and fifth cumulants are $\hat{\theta} \equiv n^{-1} \sum_i X_i$, $\hat{\sigma}^2 \equiv n^{-1} \sum_i (X_i - \bar{X})^2$, $\hat{\gamma} \equiv \hat{\sigma}^{-3} n^{-1} \sum_i (X_i - \bar{X})^3$, $\hat{\kappa}_4 \equiv \hat{\sigma}^{-4} n^{-1} \sum_i (X_i - \bar{X})^4 - 3$, and $\hat{\kappa}_5 \equiv \hat{\sigma}^{-5} \sum_i (X_i - \bar{X})^5 - 10\hat{\sigma}^{-3} \sum_i (X_i - \bar{X})^3$, respectively. Hall (1988a), Section 4.2, Example 1, gives Cornish-Fisher expansions for percentile- t , “backwards” and “hybrid” percentile, and accelerated bias-corrected critical points. The polynomials p_1 , q_1 , p_{21} , and q_{21} are there shown to be $p_1(x) \equiv -\frac{1}{6}\gamma(x^2 - 1)$, $q_1(x) \equiv \frac{1}{6}\gamma(2x^2 + 1)$, $p_{21}(x) \equiv x\{\frac{1}{24}\kappa_4(x^2 - 3) - \frac{1}{36}\gamma^2(2x^2 - 5)\}$ and $q_{21}(x) \equiv x\{-\frac{1}{12}\kappa_4(x^2 - 3) + \frac{5}{72}\gamma^2(4x^2 - 1) + \frac{1}{4}(x^2 + 3)\}$, respectively. Consequently,

$$nE[(\hat{\theta} - \theta)\hat{\sigma}^{-1}\{\hat{p}_1(z_{1-\alpha}) - p_1(z_{1-\alpha})\}] = -\frac{1}{6}(z_{1-\alpha}^2 - 1)nE\{(\hat{\theta} - \theta)\hat{\sigma}^{-1}(\hat{\gamma} - \gamma)\},$$

and

$$nE[(\hat{\theta} - \theta)\hat{\sigma}^{-1}\{\hat{q}_1(z_{1-\alpha}) - q_1(z_{1-\alpha})\}] = \frac{1}{6}(2z_{1-\alpha}^2 + 1)nE\{(\hat{\theta} - \theta)\hat{\sigma}^{-1}(\hat{\gamma} - \gamma)\}.$$

We show in Appendix A4.1 that

$$\lim_{n \rightarrow \infty} nE\{(\hat{\theta} - \theta)\hat{\sigma}^{-1}(\hat{\gamma} - \gamma)\} = \kappa_4 - \frac{3}{2}\gamma^2,$$

and thus,

$$u_1 = \frac{1}{6}(2z_{1-\alpha}^2 + 1)(\kappa_4 - \frac{3}{2}\gamma^2), \quad u_2 = -\frac{1}{6}(z_{1-\alpha}^2 - 1)(\kappa_4 - \frac{3}{2}\gamma^2).$$

Therefore, it follows from (4.22) and (4.23) that

$$\begin{aligned} \hat{\theta}_{BACK}^{(1)}(\alpha) &= \hat{\theta} + n^{-\frac{1}{2}}\hat{\sigma}(z_{1-\alpha} + n^{-\frac{1}{2}}\frac{1}{6}\hat{\gamma}(2z_{1-\alpha}^2 + 1) \\ &\quad + n^{-1}[z_{1-\alpha}\{\frac{1}{12}\kappa_4(z_{1-\alpha}^2 + 1) + \frac{1}{72}\gamma^2(2z_{1-\alpha}^2 + 13) \\ &\quad + \frac{1}{4}(z_{1-\alpha}^2 + 3)\}]) + O_p(n^{-2}), \end{aligned}$$

and

$$\begin{aligned} \hat{\theta}_{HYB}^{(1)}(\alpha) &= \hat{\theta} + n^{-\frac{1}{2}}\hat{\sigma}(z_{1-\alpha} + n^{-\frac{1}{2}}\frac{1}{6}\hat{\gamma}(2z_{1-\alpha}^2 + 1) \\ &\quad + n^{-1}[z_{1-\alpha}\{-\frac{1}{12}\kappa_4(3z_{1-\alpha}^2 - 5) + \frac{1}{72}\gamma^2(38z_{1-\alpha}^2 - 23) \\ &\quad + \frac{1}{4}(z_{1-\alpha}^2 + 3)\}]) + O_p(n^{-2}). \end{aligned}$$

From (4.26) and (4.27), the coverage probabilities of once-iterated “backwards” and “hybrid” percentile intervals are

$$\begin{aligned} \pi_{1,BACK}(\alpha) &= 1 - \alpha - \frac{1}{6}n^{-1}z_{1-\alpha}(z_{1-\alpha}^2 + 2)(\kappa_4 - \frac{3}{2}\gamma^2)\phi(z_{1-\alpha}) \\ &\quad + O(n^{-3/2}), \end{aligned} \quad (4.29)$$

and

$$\pi_{1,HYB}(\alpha) = 1 - \alpha - \frac{1}{2}n^{-1}z_{1-\alpha}^3(\kappa_4 - \frac{3}{2}\gamma^2)\phi(z_{1-\alpha}) + O(n^{-3/2}), \quad (4.30)$$

respectively.

Once-iterated “backwards” and “hybrid” percentile intervals are competitors of ordinary percentile- t and accelerated bias-corrected intervals which have coverage error of order n^{-1} without iteration. Hall (1988a), Section 4.5, gives the coverage probabilities of these intervals:

$$\begin{aligned} \pi_{0,STUD}(\alpha) &= 1 - \alpha - \frac{1}{6}n^{-1}z_{1-\alpha}(2z_{1-\alpha}^2 + 1)(\kappa_4 - \frac{3}{2}\gamma^2)\phi(z_{1-\alpha}) \\ &\quad + O(n^{-3/2}), \end{aligned} \quad (4.31)$$

and

$$\begin{aligned}\pi_{0,ABC}(\alpha) = 1 - \alpha - n^{-1}z_{1-\alpha}\left\{\frac{1}{24}\kappa_4(5z_{1-\alpha}^2 + 13) - \frac{1}{8}\gamma^2(2z_{1-\alpha}^2 + 5) \right. \\ \left. + \frac{1}{4}(z_{1-\alpha}^2 + 3)\right\}\phi(z_{1-\alpha}) + O(n^{-3/2}).\end{aligned}\quad (4.32)$$

Direct comparison between (4.29) and (4.30) indicates that we would prefer the once-iterated “backwards” percentile interval over the once-iterated “hybrid” percentile interval on the basis of asymptotic coverage accuracy if $\frac{1}{6}(z_{1-\alpha}^2 + 2) < \frac{1}{2}z_{1-\alpha}^2$; that is, if $z_{1-\alpha}^2 > 1$, or, alternatively, if $1 - \alpha > 0.84$ or $1 - \alpha < 0.16$. Similarly, comparing (4.29) and (4.30), and (4.30) and (4.31) respectively, we would prefer the once-iterated “backwards” percentile interval over the ordinary percentile- t interval, and the ordinary percentile- t interval over the once-iterated “hybrid” interval if $z_{1-\alpha}^2 > 1$; that is, if $1 - \alpha > 0.84$ or $1 - \alpha < 0.16$. In particular, for 90% and 95% intervals, the once-iterated “backwards” percentile interval has the highest coverage accuracy of all three intervals.

Using expansions (4.29)–(4.32) in Theorem 4.1, with $j = 2$, and the facts that

$$\lim_{n \rightarrow \infty} nE\{(\hat{\theta} - \theta)\hat{\sigma}^{-1}(\hat{\kappa}_4 - \kappa_4)\} = \kappa_5 - 2\gamma\kappa_4,$$

and

$$\lim_{n \rightarrow \infty} nE\{(\hat{\theta} - \theta)\hat{\sigma}^{-1}(\hat{\gamma}^2 - \gamma^2)\} = 2\gamma(\kappa_4 - \frac{3}{2}\gamma^2),$$

which are proved in Appendix A4.1, we may write down expansions for coverage probabilities of once-iterated percentile- t and accelerated bias-corrected intervals and twice-iterated “backwards” and “hybrid” percentile intervals, respectively:

$$\begin{aligned}\pi_{1,STUD}(\alpha) = 1 - \alpha - n^{-3/2}\frac{1}{6}z_{1-\alpha}^2(2z_{1-\alpha}^2 + 1)(\kappa_5 - 5\gamma\kappa_4 + \frac{9}{2}\gamma^3)\phi(z_{1-\alpha}) \\ + O(n^{-2}),\end{aligned}\quad (4.33)$$

$$\begin{aligned}\pi_{1,ABC}(\alpha) = 1 - \alpha - n^{-3/2}z_{1-\alpha}^2\left\{\frac{1}{24}\kappa_5(5z_{1-\alpha}^2 + 13) - \frac{1}{12}\gamma\kappa_4(11z_{1-\alpha}^2 + 28) \right. \\ \left. + \frac{3}{8}\gamma^3(2z_{1-\alpha}^2 + 5)\right\}\phi(z_{1-\alpha}) + O(n^{-2}),\end{aligned}$$

$$\begin{aligned}\pi_{2,BACK}(\alpha) = 1 - \alpha - n^{-3/2}\frac{1}{6}z_{1-\alpha}^2(z_{1-\alpha}^2 + 2)(\kappa_5 - 5\gamma\kappa_4 + \frac{9}{2}\gamma^3)\phi(z_{1-\alpha}) \\ + O(n^{-2}),\end{aligned}\quad (4.34)$$

and

$$\begin{aligned}\pi_{2,HYB}(\alpha) = 1 - \alpha - n^{-3/2}\frac{1}{2}z_{1-\alpha}^4(\kappa_5 - 5\gamma\kappa_4 + \frac{9}{2}\gamma^3)\phi(z_{1-\alpha}) \\ + O(n^{-2}),\end{aligned}\quad (4.35)$$

From (4.20), the once-iterated critical points $\hat{\theta}_{STUD}^{(1)}(\alpha)$ and $\hat{\theta}_{ABC}^{(1)}(\alpha)$ both have

the form

$$\begin{aligned} \hat{\theta} + n^{-\frac{1}{2}}\hat{\sigma} \left[z_{1-\alpha} + n^{-\frac{1}{2}}\frac{1}{6}\hat{\gamma}(2z_{1-\alpha}^2 + 1) + n^{-1}z_{1-\alpha} \left\{ \frac{1}{12}\kappa_4(3z_{1-\alpha}^2 + 5) \right. \right. \\ \left. \left. - \frac{1}{72}\gamma^2(16z_{1-\alpha}^2 + 23) + \frac{1}{4}(z_{1-\alpha}^2 + 3) \right\} \right] + O(n^{-2}). \end{aligned}$$

Comparing expansions (4.33), (4.34) and (4.35) in pairs, it follows that we would prefer, on the basis of asymptotic coverage accuracy, a twice-iterated “backwards” percentile interval over either a once-iterated percentile- t interval or a twice-iterated “hybrid” percentile interval, provided $1 - \alpha > 0.84$ or $1 - \alpha < 0.16$. In this instance, we see by repeated iteration that $m + 1$ iterations of a “backwards” percentile interval yield an interval with asymptotically higher coverage accuracy than that produced by m iterations of a percentile- t interval, or $m + 1$ iterations of a “hybrid” percentile interval, if $1 - \alpha > 0.84$, for $m \geq 1$. In particular, this is true for nominal 90% and 95% intervals.

Example 2: Estimation of exponential mean. Let $X_1, \dots, X_n$ be independent and identically distributed observations from a distribution with density $f_\theta(y) \equiv \theta^{-1} \exp(-\theta^{-1}y)$, for $y > 0$. The maximum likelihood estimate of θ is the sample mean $\hat{\theta} \equiv n^{-1} \sum_i X_i$, which is also the maximum likelihood estimator of $\sigma (= \theta)$. In this case, $\hat{\theta} - \theta$ can be made exactly pivotal by rescaling, so one iteration of the coverage-correction algorithm produces a percentile- t interval with perfect coverage accuracy. Hence, from Hall (1988a), Section 4.2, Example 2,

$$\begin{aligned} \hat{\theta}_{BACK}^{(1)}(\alpha) &= \hat{\theta}_{HYB}^{(1)}(\alpha) = \hat{\theta}_{STUD}^{(1)}(\alpha) = \hat{\theta}_{ABC}^{(1)}(\alpha) \\ &= \hat{\theta} + n^{-\frac{1}{2}}\hat{\sigma} \left\{ z_{1-\alpha} + n^{-\frac{1}{2}}\frac{1}{3}(2z_{1-\alpha}^2 + 1) + n^{-1}\frac{1}{36}z_{1-\alpha}(13z_{1-\alpha}^2 + 17) \right\} \\ &\quad + O(n^{-2}). \end{aligned}$$

As stated in Chapter 3, the bootstrap coverage-correction algorithm is directed at constructing intervals with accurate coverage. However, as we now show, this does not necessarily mean that by applying our algorithm, we get the endpoint of the interval right. To illustrate this point, let $I_0 \equiv (-\infty, \hat{\theta}(\alpha)]$ be a $(1 - \alpha)$ -level, one-sided interval for θ , where $\hat{\theta}(\alpha)$ admits Cornish-Fisher expansion (4.14) and $\pi_0(\alpha)$, the coverage probability of I_0 , admits expansion (4.15) with $j = 1$. A theoretical “ideal” exact interval for θ is $(-\infty, \hat{\theta}_{ideal}(\alpha)]$, where $\hat{\theta}_{ideal}(\alpha)$ admits expansion

$$\hat{\theta}_{ideal}(\alpha) = \hat{\theta} + n^{-\frac{1}{2}}\hat{\sigma} \{ z_{1-\alpha} + n^{-\frac{1}{2}}q_1(z_{1-\alpha}) + n^{-1}q_{21}(z_{1-\alpha}) + \dots \},$$

and its bootstrap version, the percentile- t interval, is $(-\infty, \hat{\theta}_{STUD}(\alpha)]$, where $\hat{\theta}_{STUD}(\alpha)$ is given by (4.16). Let $\hat{\beta}_\alpha$ be the solution of the equation $\hat{\pi}_0(\hat{\beta}_\alpha) = 1 - \alpha$, and define $\tilde{\beta}_\alpha$ and $\bar{\beta}_\alpha$ to be random variables such that $\hat{\theta}(\tilde{\beta}_\alpha) = \hat{\theta}_{STUD}(\alpha)$ and $\hat{\theta}(\bar{\beta}_\alpha) = \hat{\theta}_{ideal}(\alpha)$, respectively. Thus, $\hat{\beta}_\alpha$ is the bootstrap coverage correction, $\tilde{\beta}_\alpha$ is the “optimal” correction if we are trying to get $\hat{\theta}(\alpha)$ to look like $\hat{\theta}_{STUD}(\alpha)$, and $\bar{\beta}_\alpha$ is the unobservable “optimal” correction if we are trying to get $\hat{\theta}(\alpha)$ to look like $\hat{\theta}_{ideal}(\alpha)$. It can be shown that $\lim_{n \rightarrow \infty} \text{corr}(\hat{\beta}_\alpha, \tilde{\beta}_\alpha) = 1$ and

$$\begin{aligned} & \lim_{n \rightarrow \infty} \text{corr}(\hat{\beta}_\alpha, \bar{\beta}_\alpha) \\ &= \lim_{n \rightarrow \infty} \left(\frac{\text{E}[\{\hat{s}_1(z_{1-\alpha}) - s_1(z_{1-\alpha})\} \{\hat{s}_1(z_{1-\alpha}) - s_1(z_{1-\alpha}) - \hat{q}_1(z_{1-\alpha}) + q_1(z_{1-\alpha})\}]}{\text{E}[\{\hat{s}_1(z_{1-\alpha}) - s_1(z_{1-\alpha})\}^2]} \right) \\ & \quad \times \left(\frac{\text{E}[\{\hat{s}_1(z_{1-\alpha}) - s_1(z_{1-\alpha}) - \hat{q}_1(z_{1-\alpha}) + q_1(z_{1-\alpha})\}^2]}{\text{E}[\{\hat{s}_1(z_{1-\alpha}) - s_1(z_{1-\alpha})\}^2]} \right)^{-\frac{1}{2}}. \end{aligned}$$

We do not give a proof of the above formula since the proof is neither instructive, nor central to our argument. Now consider the example of constructing nonparametric percentile intervals for the mean. In the case of “backwards” percentile intervals, $s_1(x) = -p_1(x) = \frac{1}{6}\gamma(x^2 - 1)$, and so $\text{corr}(\hat{\beta}_\alpha, \tilde{\beta}_\alpha)$ converges to one, while $\text{corr}(\hat{\beta}_\alpha, \bar{\beta}_\alpha)$ converges to the sign of $(1 - z_{1-\alpha}^2)$. In the same way, for “hybrid” percentile intervals, $\text{corr}(\hat{\beta}_\alpha, \tilde{\beta}_\alpha)$ converges to one, while $\text{corr}(\hat{\beta}_\alpha, \bar{\beta}_\alpha)$ converges to the sign of $(z_{1-\alpha}^2 - 1)$. As a result, if $z_{1-\alpha}^2 > 1$ — that is, if $1 - \alpha > 0.84$ or $1 - \alpha < 0.16$ — bootstrap coverage correction pulls the once-iterated “hybrid” percentile interval *towards* the “ideal” interval, but pulls the once-iterated “backwards” percentile interval *away* from the “ideal” interval. On the other hand, both once-iterated percentile intervals are pulled *towards* the percentile- t interval. This result is striking because we showed earlier that, if $\alpha > 0.84$, the once-iterated “backwards” interval has *better* asymptotic coverage accuracy than either the once-iterated “hybrid” interval or the percentile- t interval, despite the fact that iteration pulls the “backwards” interval away from the “ideal” interval, towards the percentile- t interval. Therefore, we do not need to get closer to the “ideal” interval to obtain good coverage accuracy. In fact, no matter how many times we iterate our algorithm, we can get no closer than $O(n^{-1})$ to the “ideal” interval, since $\hat{q}_1(z_{1-\alpha}) - q_1(z_{1-\alpha}) = O_p(n^{-\frac{1}{2}})$.

3.2. Two-sided intervals. Let $I_0 = I_0(\alpha) \equiv [\hat{\theta}_1(\alpha), \hat{\theta}_2(\alpha)]$ be a $(1 - \alpha)$ -level, two-sided confidence interval for θ , and assume that the critical points $\hat{\theta}_i$ admit

Cornish-Fisher expansions

$$\hat{\theta}_i(\alpha) = \hat{\theta} + n^{-\frac{1}{2}}\hat{\sigma}\{(-1)^i z + n^{-\frac{1}{2}}\hat{s}_{1i}(z) + n^{-1}\hat{s}_{2i}(z)\} + O_p(n^{-2}), \quad (4.36)$$

for $i = 1, 2$, and polynomials $\hat{s}_{1i}$ and $\hat{s}_{2i}$, where $z \equiv z_{1-(\alpha/2)}$ is the solution of $\Phi(z_{1-(\alpha/2)}) = 1 - (\alpha/2)$. As in Section 2, $\hat{s}_{1i}$ and $\hat{s}_{2i}$ are sample versions of nonrandom polynomials s_{1i} and s_{2i} , in which population moments have been replaced by sample moments. We assume as in Subsection 2.2 that $\hat{s}_{21} = -\hat{s}_{22}$. Formulae for the upper critical point $\hat{\theta}_2(\alpha)$ in the cases of percentile- t , “backwards” and “hybrid” percentile, and accelerated bias-corrected intervals are given by (4.15)–(4.18) with $z_{1-\alpha}$ replaced by z .

Suppose, as in Section 2, that the coverage probability of I_0 admits expansion (4.7) for some $j \geq 1$, where t_1 is an odd polynomial. As shown in the Lemma following the proof of Theorem 4.2, it is necessary to have $\hat{s}_{11} = \hat{s}_{12}$ for there to be no term of order $n^{-\frac{1}{2}}$ in (4.7). Theorem 4.4 provides a formula for the endpoints of the once-iterated interval $[\hat{\theta}_1^{(1)}(\alpha), \hat{\theta}_2^{(1)}(\alpha)]$, when $j = 1$ in (4.7).

THEOREM 4.4. *Using the above notation, and assuming that $s_{21} = -s_{22}$, if $j = 1$ in (4.7) then the endpoints $\hat{\theta}_i^{(1)}(\alpha)$, for $i = 1, 2$, of the bootstrap coverage-corrected interval I_1 have expansions*

$$\begin{aligned} \hat{\theta}_i^{(1)}(\alpha) = \hat{\theta} + n^{-\frac{1}{2}}\hat{\sigma}\{ & (-1)^i z + n^{-\frac{1}{2}}\hat{s}_{1i}(z) + n^{-1}(-1)^i [\tfrac{1}{2}s_{1i}^2(z)z \\ & + s_{1i}(z)\{q_1'(z) - zq_1(z)\} - q_2(z) + uz] \\ & + O_p(n^{-2}), \end{aligned} \quad (4.37)$$

for $i = 1, 2$, where u is the constant such that

$$nE[(\hat{\theta} - \theta)\hat{\sigma}^{-1}\{\hat{s}_{1i}(z) - s_{1i}(z)\}] = u + O(n^{-1}). \quad (4.38)$$

Theorem 4.4 is proved in Subsection 7.4.

Notice that, for two-sided intervals, bootstrap coverage correction does not force critical points to be second-order correct. This reinforces Hall's (1988a) point that second-order correctness is not an important issue as far as coverage accuracy in two-sided confidence intervals is concerned.

Using (4.37) and (4.38), we may write general formulae for once-iterated percentile- t , “backwards” and “hybrid” percentile, and accelerated bias-corrected critical points:

$$\hat{\theta}_{STUD,i}^{(1)}(\alpha) = \hat{\theta} + n^{-\frac{1}{2}}\hat{\sigma}\{(-1)^i z + n^{-\frac{1}{2}}\hat{q}_1(z) + n^{-1}(-1)^i\{q_{21}(z) + u_1 z\}\}$$

$$\begin{aligned}
& + O_p(n^{-2}), \\
\hat{\theta}_{BACK,i}^{(1)}(\alpha) &= \hat{\theta} + n^{-\frac{1}{2}} \hat{\sigma}((-1)^i z - n^{-\frac{1}{2}} \hat{p}_1(z) + n^{-1}(-1)^i [\frac{1}{2} p_1^2(z) z \\
& \quad - p_1(z) \{q_1'(z) - z q_1(z)\} - q_2(z) - u_2 z]) + O_p(n^{-2}), \\
\hat{\theta}_{HYB,i}^{(1)}(\alpha) &= \hat{\theta} + n^{-\frac{1}{2}} \hat{\sigma}((-1)^i z + n^{-\frac{1}{2}} \hat{p}_1(z) + n^{-1}(-1)^i [\frac{1}{2} p_1^2(z) z \\
& \quad + p_1(z) \{q_1'(z) - z q_1(z)\} - q_2(z) + u_2 z]) + O_p(n^{-2}),
\end{aligned}$$

where u_1 and u_2 are defined at (4.24) and (4.25) respectively, and $\hat{\theta}_{ABC,i}^{(1)}(\alpha)$ agrees with $\hat{\theta}_{STUD,i}^{(1)}(\alpha)$ up to $O(n^{-2})$. We illustrate Theorem 4.4 in our two examples.

Example 1: Nonparametric estimation of mean. See Subsection 3.1 for notation and other details. Here, we merely give explicit formulae for critical points after iteration. As in Subsection 3.1, $u_1 = \frac{1}{6}(2z^2 + 1)(\kappa_4 - \frac{3}{2}\gamma^2)$ and $u_2 = -\frac{1}{6}(z^2 - 1)(\kappa_4 - \frac{3}{2}\gamma^2)$, so that

$$\begin{aligned}
\hat{\theta}_{STUD,i}^{(1)}(\alpha) &= \hat{\theta} + n^{-\frac{1}{2}} \hat{\sigma} [(-1)^i z + n^{-\frac{1}{2}} \frac{1}{6} \hat{\gamma}(2z^2 + 1) \\
& \quad + n^{-1}(-1)^i z \{ \frac{1}{12} \kappa_4(3z^2 + 5) - \frac{1}{72} \gamma^2(16z^2 + 23) + \frac{1}{4}(z^2 + 3) \}] \\
& \quad + O_p(n^{-2}), \\
\hat{\theta}_{BACK,i}^{(1)}(\alpha) &= \hat{\theta} + n^{-\frac{1}{2}} \hat{\sigma} [(-1)^i z + n^{-\frac{1}{2}} \frac{1}{6} \hat{\gamma}(z^2 - 1) \\
& \quad + n^{-1}(-1)^i z \{ \frac{1}{12} \kappa_4(z^2 + 1) + \frac{1}{72} \gamma^2(z^2 - 1)^2 + \frac{1}{4}(z^2 + 3) \}] \\
& \quad + O_p(n^{-2}),
\end{aligned}$$

and

$$\begin{aligned}
\hat{\theta}_{HYB,i}^{(1)}(\alpha) &= \hat{\theta} + n^{-\frac{1}{2}} \hat{\sigma} [(-1)^i z - n^{-\frac{1}{2}} \frac{1}{6} \hat{\gamma}(z^2 - 1) \\
& \quad + n^{-1}(-1)^i z \{ -\frac{1}{12} \kappa_4(3z^2 - 5) + \frac{1}{72} \gamma^2(9z^4 + 14z^2 - 23) \\
& \quad + \frac{1}{4}(z^2 + 3) \}] + O_p(n^{-2}),
\end{aligned}$$

and $\hat{\theta}_{ABC,i}^{(1)}(\alpha)$ admits the same expansion as $\hat{\theta}_{STUD,i}^{(1)}(\alpha)$ up to terms of order n^{-2} .

Example 2: Estimation of exponential mean. See Subsection 3.1 for notation. Recall from Subsection 2.3 that, provided $\hat{\theta} - \theta$ can be made exactly pivotal by rescaling, one iteration of the coverage-correction algorithm produces a percentile- t interval with perfect coverage accuracy. Substituting formulae for p_1 , q_1 , p_{21} , and q_{21} , given in Section 4.3 of Hall (1988a), and noting that $u_1 = u_2 = 0$ in this case, we obtain the following formulae for iterated critical

points, using the usual notation:

$$\begin{aligned}\hat{\theta}_{STUD,i}^{(1)}(\alpha) &= \hat{\theta}_{ABC,i}^{(1)}(\alpha) = \hat{\theta}_{STUD,i}(\alpha) = \hat{\theta}_{ABC,i}(\alpha) \\ &= \hat{\theta} + n^{-\frac{1}{2}}\hat{\sigma}\{(-1)^i z + n^{-\frac{1}{2}}\frac{1}{3}(2z^2 + 1) \\ &\quad + n^{-1}(-1)^i\frac{1}{36}z(13z^2 + 17)\} + O_p(n^{-2}), \\ \hat{\theta}_{BACK,i}^{(1)}(\alpha) &= \hat{\theta} + n^{-\frac{1}{2}}\hat{\sigma}\{(-1)^i z + n^{-\frac{1}{2}}\frac{1}{3}(z^2 - 1) \\ &\quad + n^{-1}(-1)^i\frac{1}{36}z(2z^4 + 5z^2 - 7)\} + O_p(n^{-2}),\end{aligned}$$

and

$$\begin{aligned}\hat{\theta}_{HYB,i}^{(1)}(\alpha) &= \hat{\theta} + n^{-\frac{1}{2}}\hat{\sigma}\{(-1)^i z - n^{-\frac{1}{2}}\frac{1}{3}(z^2 - 1) \\ &\quad + n^{-1}(-1)^i\frac{1}{36}z(18z^4 - 35z^2 + 17)\} + O_p(n^{-2}).\end{aligned}$$

Each of these critical points results in exact $(1 - \alpha)$ -level percentile- t intervals.

4. EFFECT OF BOOTSTRAP COVERAGE CORRECTION ON INTERVAL LENGTH

In this section we derive expressions for the change in mean length of interval and the change in the variance of length of interval after bootstrap coverage correction. By way of notation, let $I_0 = I_0(\alpha) \equiv [\hat{\theta}_1(\alpha), \hat{\theta}_2(\alpha)]$ be a $(1 - \alpha)$ -level, two-sided confidence interval for θ , with length denoted by L , and let $I_1 = I_1(\alpha) \equiv [\hat{\theta}_1^{(1)}(\alpha), \hat{\theta}_2^{(1)}(\alpha)]$ be the interval obtained by coverage correction of I_0 , with length denoted by $L^{(1)}$. Let $z \equiv z_{1-(\alpha/2)}$ be the solution of $\Phi(z_{1-(\alpha/2)}) = 1 - (\alpha/2)$. Assume that the coverage probability, $\pi_0(\alpha)$, of I_0 admits expansion (4.7) for some $j \geq 1$, and that the endpoints $\hat{\theta}_i(\alpha)$, for $i = 1, 2$, of I_0 admit Cornish-Fisher expansions

$$\begin{aligned}\hat{\theta}_i(\alpha) &= \hat{\theta} + n^{-\frac{1}{2}}\hat{\sigma}\{(-1)^i z + n^{-\frac{1}{2}}\hat{s}_{1i}(z) + n^{-1}\hat{s}_{2i}(z) + \dots \\ &\quad + n^{-j}\hat{s}_{2j,i}(z)\} + O_p(n^{-(j+1)}),\end{aligned}\tag{4.39}$$

for polynomials $\hat{s}_{1i}, \dots, \hat{s}_{2j,i}$ for $i = 1, 2$. We give, in Theorem 4.5, an expression for the change in mean length of interval arising from coverage correction of I_0 .

THEOREM 4.5. *Assuming expansions (4.7) and (4.39) for the coverage probability and endpoints of I_0 respectively, the mean difference in length between the original and iterated intervals is*

$$E(L^{(1)} - L) = -n^{-(2j+1)/2}\sigma t_1(z) + O(n^{-(j+1)}).\tag{4.40}$$

Theorem 4.5 is proved in Subsection 7.5.

Clearly (4.40) indicates that if $t_1(z) < 0$, then there is an asymptotic mean increase in length between I_0 and I_1 , while if $t_1(z) > 0$, there is an asymptotic mean decrease in length between I_0 and I_1 . This means that if the original interval undercovers, the effect of coverage correction is to produce an interval with greater asymptotic mean length, while if the original interval overcovers, then coverage correction yields an interval with shorter asymptotic mean length. If $j = 1$ in (4.7), we deduce from Hall (1988a), equation (4.15), that

$$t_1(z) = -2[\frac{1}{2}s_1^2(z)z + s_1(z)\{q_1'(z) - zq_1(z)\} - s_2(z) + uz],$$

where

$$\hat{\theta}_i(\alpha) = \hat{\theta} + n^{-\frac{1}{2}}\{(-1)^i z + n^{-\frac{1}{2}}\hat{s}_1(z) + n^{-1}(-1)^i\hat{s}_2(z)\} + O_p(n^{-2}),$$

and u is given in Theorem 4.3. Therefore, in that case, we have

$$\begin{aligned} E(L^{(1)} - L) &= 2n^{-3/2}\sigma z[\frac{1}{2}s_1^2(z)z + s_1(z)\{q_1'(z) - zq_1(z)\} \\ &\quad - q_2(z) - s_2(z) + uz] + O(n^{-2}). \end{aligned} \quad (4.41)$$

Example 1: Nonparametric estimation of mean. See Subsection 3.1 for notation. Hall (1988a), Section 4.5, gives expressions for coverage probabilities of one-sided percentile- t , “backwards” and “hybrid” percentile, and accelerated bias-corrected intervals. Substituting Hall’s formulae for s_1 , s_2 , q_1 , q_2 , and u into (4.41) in each of these cases gives

$$\begin{aligned} E(L_{STUD}^{(1)} - L_{STUD}) &= n^{-3/2}\sigma z(2z^2 + 1)\frac{1}{3}(\kappa_4 - \frac{3}{2}\gamma^2) + O(n^{-2}), \\ E(L_{BACK}^{(1)} - L_{BACK}) &= n^{-3/2}\sigma z\{\frac{1}{12}\kappa_4(z^2 + 5) + \frac{1}{36}\gamma^2(z^4 + 2z^2 - 9) \\ &\quad + \frac{1}{2}(z^2 + 3)\} + O(n^{-2}), \\ E(L_{HYB}^{(1)} - L_{HYB}) &= n^{-3/2}\sigma z\{-\frac{1}{12}\kappa_4(7z^2 - 13) + \frac{1}{12}\gamma^2(3z^4 + 6z^2 - 11) \\ &\quad + \frac{1}{2}(z^2 + 3)\} + O(n^{-2}), \end{aligned}$$

and

$$\begin{aligned} E(L_{ABC}^{(1)} - L_{ABC}) &= n^{-3/2}\sigma z\{\frac{1}{12}\kappa_4(5z^2 + 13) - \frac{1}{4}\gamma^2(2z^2 + 5) \\ &\quad + \frac{1}{2}(z^2 + 3)\} + O(n^{-2}). \end{aligned}$$

If $\gamma = \kappa_4 = 0$ (for example, when the underlying population is Normal), the change in mean length for percentile- t intervals is $O(n^{-2})$, while for percentile

and accelerated bias-corrected intervals, it is $\frac{1}{2}n^{-3/2}\sigma z(z^2 + 3) + O(n^{-2})$, this expression arising from the difference between the Standard Normal distribution and an expansion of Student's- t distribution function with $n - 1$ degrees of freedom. In the case of distributions with non-zero skewness and kurtosis, the mean length for "backwards" intervals asymptotically increases when $\kappa_4 \geq 0$, while for "hybrid" intervals, asymptotic mean length increases when $\kappa_4 \leq 0$.

Example 2: Estimation of exponential mean. Since in this example iteration of "backwards" and "hybrid" percentile intervals results in percentile- t intervals with perfect coverage accuracy, and iteration of percentile- t and accelerated bias-corrected intervals merely reproduces the same intervals, we focus on *actual* interval length rather than *mean* interval length. Using formulae for endpoints of percentile- t , accelerated bias-corrected and iterated "backwards" and "hybrid" percentile intervals derived in Example 2 of Subsection 3.2, we have

$$\begin{aligned} L_{STUD} &= L_{ABC} = L_{STUD}^{(1)} = L_{ABC}^{(1)} \\ &= 2n^{-\frac{1}{2}}\hat{\sigma}\{z + n^{-1}\frac{1}{36}z(13z^2 + 17)\} + O_p(n^{-2}), \\ L_{BACK}^{(1)} &= 2n^{-\frac{1}{2}}\hat{\sigma}\{z + n^{-1}\frac{1}{36}z(2z^4 + 5z^2 - 7)\} + O_p(n^{-2}), \end{aligned}$$

and

$$L_{HYB}^{(1)} = 2n^{-\frac{1}{2}}\hat{\sigma}\{z + n^{-1}\frac{1}{36}z(18z^4 - 35z^2 + 17)\} + O_p(n^{-2}).$$

Table 4.1 relates lengths of percentile- t , accelerated bias-corrected and iterated "backwards" and "hybrid" percentile intervals when $1 - \alpha = 0.90$ and $1 - \alpha = 0.95$.

TABLE 4.1. Lengths of 90% and 95% confidence intervals in exponential case. Values tabulated are proportional to the amount by which length exceeds $2n^{-\frac{1}{2}}\sigma z$.

Type of interval	length, 90%	length 95%
Iterated "backwards"	0.97	2.27
Iterated "hybrid"	2.47	8.07
Percentile- t , ABC	2.38	3.64

Amongst the intervals considered in Table 4.1, the iterated "backwards" interval is shortest, while the iterated "hybrid" interval is longest. Since each

interval has coverage exactly $1 - \alpha$, the iterated “backwards” interval would be preferred on the basis of length.

To close this section, we give an expression for the difference between the variance of the length of the original interval I_0 and the variance of the length of the iterated interval I_1 . Suppose that I_0 has endpoints satisfying (4.39), and coverage probability satisfying (4.7).

THEOREM 4.6. *Under the above assumptions, the variance of the length of the iterated interval I_1 equals*

$$\text{var}(L^{(1)}) = \text{var}(L) - 4n^{-(j+1)} \text{cov}\{\hat{\sigma}, \hat{\sigma}\hat{t}_1(z_{1-(\alpha/2)})\}z_{1-(\alpha/2)} + O(n^{-(j+2)}). \quad (4.41a)$$

Therefore, bootstrap coverage correction does not always result in an increase in the variance of the interval length, as the covariance term in (4.41a) may be either positive or negative. Theorem 4.6 is proved in Subsection 7.6.

5. SIMULATION STUDY

We give in this section the results of a small simulation study of iterated two-sided bootstrap confidence intervals for the Studentized mean. The results of our study are summarized in Table 4.2. In our example, the parent population was folded Normal, samples were of size 10, $B = 299$ resamples were drawn in each resampling operation, and each entry was based on 1000 samples. The folded Normal pseudo-random numbers were generated using NAG Library subroutine G05DDF, and all computations were carried out on a VAX 8700 computer.

The results are in general agreement with the conclusions of our asymptotic analysis. One iteration of the coverage-correction algorithm does improve coverage accuracy markedly in each case. The iterated “backwards” percentile method performs best, having higher coverage accuracy and shorter mean length than any of the other intervals. For each type of interval, mean length increased or decreased according as to whether the original interval undercovered or overcovered.

6. APPLICATIONS OF COVERAGE CORRECTION

We consider two applications of the coverage correction algorithm. In Subsection 6.1, we examine the application of bootstrap coverage correction to Bartlett-corrected tests, and in Subsection 6.2, we discuss the problem of constructing

TABLE 4.2. Characteristics of ordinary percentile, percentile- t , iterated percentile and iterated percentile- t intervals for the mean. Nominal coverage was 90%. Tabulated characteristics are coverage, mean length, standard deviation of length, and mean value of upper endpoint.

	Type of interval					
	"Backwards"		"Hybrid"		Percentile- t	
	BACK	Iterated BACK	HYB	Iterated HYB	STUD	Iterated STUD
cov.	.846	.892	.837	.842	.924	.891
len.	.590	.685	.587	.701	.895	.855
s.d.	.171	.257	.164	.289	.457	.580
upper	.318	.363	.285	.345	.605	.582

accurate confidence intervals for the correlation coefficient, especially in nonnormal populations.

6.1. Bootstrap coverage correction for Bartlett-corrected likelihood-ratio tests. Let $L(\omega)$ denote the likelihood of a random sample drawn from a distribution depending on a vector ω of unknown parameters. A common test of the simple null hypothesis $H_0 : \omega = \omega_0$ is to calculate the maximum likelihood estimator $\hat{\omega}$ of ω , and assess the truth of H_0 by studying the likelihood ratio $L(\hat{\omega})/L(\omega)$, or equivalently, by examining

$$w \equiv 2 \log \{L(\hat{\omega})/L(\omega)\} = 2 \{l(\hat{\omega}) - l(\omega)\},$$

where $l = \log L$. In regular cases, the asymptotic distribution of w is chi-squared with d degrees of freedom, where d is the dimension of w .

Bartlett (1937) introduced an adjustment which improves the quality of this chi-squared approximation. Bartlett's argument runs as follows. Suppose that under the null hypothesis we may write

$$E(w) = d \{1 + n^{-1}b_1 + O(n^{-2})\},$$

for some b_1 , and let $\hat{b}_1$ equal b_1 if b_1 is known, or a $\sqrt{n}$ -consistent estimator of b_1 , otherwise. Then the distribution of

$$w' \equiv (1 + n^{-1}\hat{b}_1)^{-1}w \tag{4.42}$$

is, in general, closer to a chi-squared distribution with d degrees of freedom than is the distribution of w . The factor $(1 + n^{-1}\hat{b}_1)^{-1}$ in (4.42) is called a Bartlett adjustment factor. Several authors note that while in general,

$$P(w \leq x) = P(X_d^2 \leq x) + O(n^{-1}),$$

for real x , the distribution of w' satisfies

$$P(w' \leq x) = P(X_d^2 \leq x) + O(n^{-3/2}), \quad (4.43)$$

where X_d^2 denotes a random variable having a chi-squared distribution with d degrees of freedom; see for example, Lawley (1956), Hayakawa (1976), and Barndorff-Nielsen and Cox (1984). Barndorff-Nielsen and Cox (1984) give a particularly simple derivation of (4.43). Barndorff-Nielsen and Hall (1988) have refined (4.43) to show that, in fact,

$$P(w' \leq x) = P(X_d^2 \leq x) + O(n^{-2}). \quad (4.44)$$

We sketch here an argument showing that one application of the resampling principle reduces level error in the Bartlett-corrected test from $O(n^{-2})$ to $O(n^{-3})$. A more detailed argument, along similar lines to that of Barndorff-Nielsen and Hall (1988), is given in Subsection 7.7 to elucidate the finer points of our argument. Suppose that

$$P(w' \leq x) = P(X_d^2 \leq x) + R(x)n^{-2} + O(n^{-5/2}),$$

for a smooth, bounded function R of x , depending on the unknown parameters ω . Let $x_{1-\alpha}$ denote the $100(1-\alpha)$ 'th percentage point of a chi-squared distribution with d degrees of freedom. Then,

$$P(w' \leq x) = 1 - \alpha + R(x_{1-\alpha})n^{-2} + O(n^{-5/2}). \quad (4.45)$$

Put $f_t(F_0, F_1) \equiv I\{w'(F_1) \leq x_{1-t}\} - (1 - \alpha)$, where I is an indicator function, and let $t = \beta_\alpha \equiv \beta_\alpha(F_0)$ be the solution of $E\{f_t(F_0, F_1) \mid F_0\} = 0$; that is,

$$P\{w'(F_1) \leq x_{1-t} \mid F_0\} = 1 - \alpha. \quad (4.46)$$

It follows from (4.45) that $P(w' \leq x_{1-\beta_\alpha}) = 1 - \beta_\alpha + R(x_{1-\beta_\alpha})n^{-2} + O(n^{-5/2})$, and, noting (4.46),

$$\begin{aligned} \beta_\alpha &= \alpha + R(x_{1-\beta_\alpha})n^{-2} + O(n^{-5/2}) \\ &= \alpha + R(x_{1-\alpha})n^{-2} + O(n^{-5/2}), \end{aligned}$$

the last line using the fact that $\beta_\alpha = \alpha + O(n^{-2})$, and so $x_{1-\beta_\alpha} = x_{1-\alpha} + O(n^{-2})$. The resampling principle argues that we estimate $\beta_\alpha(F_0)$ by the solution $t = \hat{\beta}_\alpha \equiv \beta_\alpha(F_1)$ of the equation $E\{f_t(F_1, F_2) \mid F_1\} = 0$; that is, $P\{w'(F_2) \leq x_{1-t} \mid F_1\} = 1 - \alpha$. The solution $\hat{\beta}_\alpha$ admits expansion

$$\begin{aligned}\hat{\beta}_\alpha &= \alpha + \hat{R}(x_{1-\hat{\beta}_\alpha})n^{-2} + O_p(n^{-5/2}) \\ &= \alpha + \hat{R}(x_{1-\alpha})n^{-2} + O_p(n^{-5/2}) \\ &= \alpha + R(x_{1-\alpha})n^{-2} + O_p(n^{-5/2})\end{aligned}\tag{4.47}$$

$$= \beta_\alpha + O_p(n^{-5/2}),\tag{4.48}$$

where $\hat{R}$ is obtained from R by replacing unknown parameters in R by their bootstrap estimates. Equation (4.47) follows on noting that $x_{1-\hat{\beta}_\alpha} = x_{1-\alpha} + O_p(n^{-2})$, and $\hat{R}(x) = R(x) + O_p(n^{-1/2})$. Inverting the chi-squared distribution with d degrees of freedom on both sides of (4.48), and Taylor-expanding, it follows that

$$x_{1-\hat{\beta}_\alpha} = x_{1-\beta_\alpha} + O_p(n^{-5/2}),$$

and thus,

$$P(w' \leq x_{1-\hat{\beta}_\alpha}) = P(w' \leq x_{1-\beta_\alpha}) + O(n^{-5/2}),\tag{4.49}$$

since β_α is chosen so that (4.46) holds. It is an extremely tedious calculation to show that the term of order $n^{-5/2}$ vanishes in (4.49). Barndorff-Nielsen and Hall (1988) show that terms of order $n^{-3/2}$ vanish in an expansion of the distribution of w' about a chi-squared distribution with d degrees of freedom. We expand their argument in Subsection 7.7 to show that terms of order $n^{-5/2}$ also vanish in an expansion of the distribution of w' . Therefore,

$$P(w' \leq x_{1-\hat{\beta}_\alpha}) = 1 - \alpha + O(n^{-3});$$

that is, level error is reduced to $O(n^{-3})$. Our result sharpens one of Beran (1988), who showed that bootstrap iteration reduces level error in Bartlett-corrected tests from $O(n^{-3/2})$ to $O(n^{-2})$.

6.2. Better nonparametric bootstrap confidence intervals for the correlation coefficient. The problem of constructing bootstrap confidence intervals for the correlation coefficient has attracted considerable recent attention. The very popular bootstrap algorithm, the “backwards” percentile method, has been criticized because of coverage inaccuracy (see Lunneborg, 1985, and Rasmussen, 1987). Efron (1981a) has pointed out that the percentile- t method fails

strikingly when used for the correlation coefficient. Efron employed a jackknife estimate to standardize for scale. We have repeated Efron's experiment, using both the jackknife and nonparametric delta method estimates of variance to Studentize, with almost identical results. The percentile- t bootstrap method founders in both cases because the lengths of these confidence intervals fluctuate erratically, their large expectations and standard deviations reflecting the inaccuracy of these particular variance estimates in small samples.

In this subsection, we argue that the correlation coefficient is better treated by specialized techniques. The percentile- t method, which is becoming accepted as an accurate and standard tool for simpler problems (see for example, Bickel and Freedman, 1981; Beran, 1987a, 1987b, 1988; Schenker, 1985; Hall, 1987b, 1988a; Singh, 1987), is apparently less suitable for more complex statistics such as the correlation coefficient. We propose two practical and accurate alternative methods. The first is based on coverage correction of the "backwards" percentile method, and the other is the use of percentile- t methods on a transformed correlation coefficient. Both promise to be useful in other complex interval construction problems.

We begin by considering coverage correction of the "backwards" percentile-method interval for the correlation coefficient. Acknowledging that "backwards" percentile intervals have good stability properties — that is, successive bootstrap estimates of percentile-method critical points do not fluctuate erratically — we adjust the interval so as to improve its much-criticized coverage accuracy. By so doing, we manage to retain the interval's stability, yet greatly enhance its coverage performance.

The basic "backwards" percentile interval has coverage error of order n^{-1} , where n is sample size (see Hall, 1988a), and, as we have shown in Section 2, the coverage-corrected "backwards" interval has coverage error of order n^{-2} . However, it could be unwise to rely too heavily on this type of asymptotic result when dealing with the correlation coefficient in small samples, especially given that the highly erratic behaviour of percentile- t intervals is certainly not apparent from Edgeworth expansions.

In practice, the iterated percentile method for constructing confidence intervals is relatively computationally expensive, when compared with ordinary percentile or percentile- t methods. In particular, if at each resampling operation, B resamples are drawn, then the time taken to construct an iterated percentile method interval is roughly proportional to B^2 compared with B for an ordinary

percentile or percentile- t interval. Depending on the coverage accuracy required, it is our experience that one iteration is usually enough to substantially improve coverage accuracy to a reasonable level.

We conclude our discussion of iterated percentile-method intervals for the correlation coefficient, ρ , with a simulation study. Table 4.3 contains a summary of simulations from three parent populations, using sample sizes $n = 8, 10$, and 12 . For simulation 1, the sample $\mathcal{X} \equiv \{(X_1, Y_1), \dots, (X_n, Y_n)\}$, is drawn from a bivariate folded Normal population, denoted by $|N(\mathbf{0}, \mathbf{I})|$, with $\rho = 0$; for simulation 2, $\rho = 0.5$, with the sample $\mathcal{X}$ being drawn such that $X_i = Z_{1i} + W_i$, and $Y_i = Z_{2i} + W_i$, where Z_{1i} , Z_{2i} , and W_i were distributed as independent folded Normal variates, for $i = 1, \dots, n$; and for simulation 3, the sample $\mathcal{X}$ is drawn from a bivariate lognormal population with $\rho = (1 + e^{\frac{1}{2}})^{-1} \approx 0.37754$. In each case, adjustment was made directly to nominal coverage, $B = 299$ resamples were drawn in each resampling operation, and each entry was based on 1000 samples for each value of n . The simulation study indicates that one bootstrap iteration does improve coverage accuracy dramatically. Especially significant is the performance of the iterated percentile method when the parent population is bivariate lognormal. In this instance, the iterated percentile method interval has much better coverage accuracy than the percentile- t methods considered next. In fact, coverage of the iterated percentile method intervals is correct; that is, not significantly different from the nominal coverage in each of nine combinations of sample size and the parent populations considered. What is more, this coverage correction is realized without unreasonable increase in average interval length over the ordinary percentile method.

We now consider the use of percentile- t methods for constructing confidence intervals for the correlation coefficient. This method is usually applied to statistics for which there is available a reasonably stable variance estimate; see for example, Bickel and Freedman (1981), Hinkley and Wei (1984), Beran (1987a, 1987b, 1988), Hall (1987b, 1988a), and Singh (1987). For instance, Hall (1988a) shows that the percentile- t method works well when constructing confidence intervals for the Studentized mean. The performance of percentile- t intervals for a parameter θ depends largely on how well we are able to estimate the variance of the estimator $\hat{\theta}$. For the correlation coefficient, use of standard estimates of variance, such as the jackknife estimate (Miller, 1974), or the nonparametric delta method estimate (Efron and Gong, 1983, p.40), to Studentize, results in percentile- t intervals with erratically varying lengths and endpoints. We pro-

TABLE 4.3. Characteristics of ordinary and iterated percentile method bootstrap confidence intervals for the correlation coefficient. Nominal coverage is 90% throughout. Tabulated characteristics are estimated coverage (nominal SE is .01), average length, standard error of length, and average value of upper endpoint.

n	Simulation 1 ($\rho = 0$)		Simulation 2 ($\rho = 0.5$)		Simulation 3 ($\rho = 0.3775$)	
	Ordinary percentile method	Iterated percentile method	Ordinary percentile method	Iterated percentile method	Ordinary percentile method	Iterated percentile method
8 cov.	.847	.908	.837	.905	.840	.903
len.	1.16	1.41	.97	1.25	1.02	1.29
s.d.	.29	.37	.37	.46	.36	.44
upper	.55	.68	.84	.89	.86	.91
10 cov.	.844	.905	.846	.899	.840	.896
len.	1.02	1.21	.84	1.05	.89	1.10
s.d.	.24	.31	.31	.40	.28	.36
upper	.51	.60	.82	.86	.83	.87
12 cov.	.863	.910	.841	.893	.848	.898
len.	.92	1.07	.74	.89	.82	.99
s.d.	.20	.28	.26	.34	.24	.30
upper	.46	.53	.79	.82	.80	.84

pose to 1) transform the correlation coefficient to a statistic for which there is available a reasonably stable variance estimate, 2) use the percentile- t method to construct a $(1 - \alpha)$ -level confidence interval for the transformed correlation coefficient, and then 3) invert this interval into a $(1 - \alpha)$ -level confidence interval for the correlation coefficient.

First consider the transformation of the correlation coefficient, ρ . Fisher's (1921) transformation of the sample correlation coefficient $\hat{\rho}$, assuming the underlying population is bivariate Normal, has variance almost independent of ρ . Let $Z \equiv \frac{1}{2} \log\{(1 + \hat{\rho})/(1 - \hat{\rho})\}$. Then, for even fairly small samples from a bivariate Normal population, Z is approximately Normally distributed with mean $\xi \equiv \frac{1}{2} \log\{(1 + \rho)/(1 - \rho)\}$ and variance $(n - 3)^{-1}$; see Kendall and Stuart (1977), vol.2, p.312. Efron (1982) and Efron and Gong (1983) suggest that, under the assumption that samples are drawn from a bivariate Normal population, the percentile method can be used to construct good confidence intervals for ρ since it is invariant under monotone transformations, and the z -transformed correlation

coefficient has almost constant variance. However, this does not resolve the problem of what to do when the underlying population is not bivariate Normal. We suggest that Fisher's z -transformation can still be used when the parent population is nonnormal, that a percentile- t method be used to construct a $(1 - \alpha)$ -level confidence interval for ξ , and this interval transformed back into a $(1 - \alpha)$ -level confidence interval for ρ .

To construct a percentile- t interval for ξ , we need an estimate of the variance of Z . It follows from Kendall and Stuart (1977), vol.1, p.251, that, asymptotically, $n^{\frac{1}{2}}(Z - \xi)$ is distributed with zero mean and variance

$$\rho^2(1 - \rho^2)^{-2} \left\{ \mu_{22}\mu_{11}^{-2} + \frac{1}{4}(\mu_{40}\mu_{20}^{-2} + \mu_{04}\mu_{02}^{-2} + 2\mu_{22}\mu_{20}^{-1}\mu_{02}^{-1}) - (\mu_{31}\mu_{20}^{-1}\mu_{11}^{-1} + \mu_{13}\mu_{02}^{-1}\mu_{11}^{-1}) \right\}, \quad (4.50)$$

provided the bivariate moments μ_{ij} , for $i, j = 0, \dots, 4$, exist. (When the underlying population is bivariate Normal, the expression (4.50) equals one.) The obvious plug-in estimate of (4.50) is just the nonparametric delta method estimate of the variance of the transformed correlation coefficient. Unfortunately, the nonparametric delta method does not provide a very stable estimate of (4.50) in small samples, because of the difficulty of estimating high-order moments accurately. Miller (1974), Section 5.3, advocates using the jackknife estimate of the variance of the transformed statistic. He notes that "transformations are needed to keep the jackknife on scale and thus prevent distortion of the results". Therefore, we also investigate the use of the jackknife estimate of the variance of Z in constructing percentile- t intervals for the transformed correlation coefficient.

A particularly attractive feature of Fisher's z -transformation is that, since it is one-to-one, strictly increasing, and maps $[-1, 1]$ onto the real line, all the intervals obtained for ρ are constrained to lie within $[-1, 1]$. We have found that constructing percentile- t intervals for ρ without first transforming the correlation coefficient results in occasional extremely long intervals that *contain* $[-1, 1]$. In principle, rather than using Fisher's z -transformation, we could have used almost any one-to-one, strictly increasing function mapping $[-1, 1]$ onto the real line.

We conclude this section with a simulation study comparing the performance of four percentile- t methods. Four sets of simulations were carried out using different parent populations. In the first, the sample $\mathcal{X} \equiv \{(X_1, Y_1), \dots, (X_n, Y_n)\}$ was drawn from a bivariate Normal population with $\rho = 0$; in the second, $\mathcal{X}$ was drawn from a bivariate folded Normal population with $\rho = 0$; in the third, $\mathcal{X}$ was

TABLE 4.4. Characteristics of four percentile-t method bootstrap confidence intervals for the correlation coefficient. The parent population is bivariate Normal with $\rho = 0$. Nominal coverages are 90%, 95% and 99%. Tabulated characteristics are estimated coverage (nominal SE is .01 for 90% intervals, .007 for 95% intervals, and .003 for 99% intervals), average length, standard error of length and average value of upper endpoint. Estimated coverages that are more than two standard errors away from the nominal coverage are marked with a *, and those that are more than ten standard errors from the nominal coverage are marked with a †.

n		Untransformed percentile-t						Ordinary			Transformed percentile-t					
		Delta method			Jackknife			Percentile			Delta method			Jackknife		
	α	.90	.95	.99	.90	.95	.99	.90	.95	.99	.90	.95	.99	.90	.95	.99
10	cov.	.94*	.98*	1.00*	.94*	.98*	1.00*	.85*	.90*	.98*	.89	.96	.99	.89	.93*	.99
	len.	1.99	3.25	14.7	2.13	3.49	16.8	1.00	1.20	1.57	1.09	1.33	1.70	1.09	1.31	1.68
	st.err.	1.64	2.78	30.9	1.92	3.27	42.3	.24	.26	.25	.33	.33	.25	.36	.38	.27
	upper	.99	1.73	10.4	1.07	1.87	12.1	.50	.58	.75	.55	.68	.87	.54	.65	.86
15	cov.	.93*	.97*	.99	.94*	.97*	.99	.86*	.93*	.98*	.90	.95	.99	.89	.94*	.99
	len.	1.20	1.63	3.37	1.25	1.72	3.65	.82	.97	1.28	.90	1.07	1.41	.87	1.08	1.44
	st.err.	.50	.78	3.10	.57	.92	3.31	.17	.18	.20	.24	.26	.28	.28	.32	.31
	upper	.59	.84	1.92	.62	.89	2.08	.41	.47	.63	.45	.54	.73	.43	.55	.75
20	cov.	.92*	.97*	1.00*	.93*	.98*	.99	.88*	.92*	.98*	.90	.95	.99	.89	.94	.98
	len.	.92	1.20	2.12	.95	1.24	2.27	.71	.84	1.12	.76	.92	1.23	.77	.92	1.25
	st.err.	.32	.43	1.01	.36	.48	1.18	.13	.14	.17	.19	.21	.26	.23	.27	.29
	upper	.46	.60	1.18	.48	.62	1.27	.36	.40	.54	.38	.45	.64	.40	.47	.65

drawn such that $X_i = Z_{1i} + W_i$, and $Y_i = Z_{2i} + W_i$, where Z_{1i} , Z_{2i} , and W_i are distributed as independent folded Normal variates, for $i = 1, \dots, n$, resulting in $\rho = 0.5$; and in the fourth, $\mathcal{X}$ was drawn from a bivariate lognormal distribution with $\rho = (1 + e^{\frac{1}{2}})^{-1} \approx 0.37754$. The five types of interval constructed were as follows: (i) Ordinary percentile-t interval, Studentizing with the nonparametric delta method estimate of the variance of $\hat{\rho}$; (ii) As in (i), except Studentizing with the jackknife estimate of the variance of $\hat{\rho}$; (iii) Ordinary percentile-method interval; (iv) Transformed percentile-t interval, involving Fisher's z , Studentizing with the nonparametric delta method estimate of the variance of Z (sample

TABLE 4.5. Characteristics of four percentile-t method bootstrap confidence intervals for the correlation coefficient. The parent populations are independent folded Normals. Nominal coverages are 90%, 95% and 99%. Tabulated characteristics are estimated coverage (nominal SE is .01 for 90% intervals, .007 for 95% intervals, and .003 for 99% intervals), average length, standard error of length and average value of upper endpoint. Estimated coverages that are more than two standard errors away from the nominal coverage are marked with a *, and those that are more than ten standard errors from the nominal coverage are marked with a †.

n		Untransformed percentile-t						Ordinary			Transformed percentile-t					
		Delta method			Jackknife			Percentile			Delta method			Jackknife		
	α	.90	.95	.99	.90	.95	.99	.90	.95	.99	.90	.95	.99	.90	.95	.99
10	cov.	.95*	.98*	1.00*	.95*	.97*	1.00*	.85*	.92*	.98*	.92	.95	.99	.89	.95	.99
	len.	1.96	3.08	15.7	2.10	3.31	16.5	1.01	1.21	1.56	1.10	1.30	1.70	1.11	1.34	1.71
	st.err.	1.36	2.54	39.1	1.57	2.79	37.0	.23	.25	.25	.32	.32	.26	.36	.36	.27
	upper	.92	1.55	11.5	1.00	1.68	11.9	.53	.61	.74	.54	.65	.88	.54	.68	.89
15	cov.	.92*	.96	.99	.92*	.96	.99	.86*	.90*	.98*	.90	.94	.99	.90	.94	.99
	len.	1.16	1.64	3.39	1.24	1.79	3.61	.81	.97	1.28	.87	1.05	1.43	.90	1.08	1.46
	st.err.	.60	1.00	2.40	.78	1.43	2.74	.17	.19	.21	.25	.28	.27	.26	.30	.29
	upper	.59	.84	1.96	.64	.93	2.09	.42	.46	.61	.44	.51	.74	.44	.55	.76
20	cov.	.93*	.96	.99	.93*	.97*	1.00*	.88	.91*	.98*	.90	.93*	.99	.90	.93*	.99
	len.	.89	1.20	2.19	.93	1.27	2.30	.70	.84	1.12	.74	.91	1.26	.78	.93	1.27
	st.err.	.32	.49	1.04	.40	.61	1.18	.13	.15	.18	.19	.23	.26	.21	.25	.26
	upper	.46	.63	1.23	.48	.67	1.30	.35	.42	.53	.37	.46	.66	.39	.47	.67

moments in (4.50)); and (v) As in (iv), except Studentizing with the jackknife estimate of the variance of Z . Throughout, each entry is based on 1000 samples from each of the sample sizes $n = 10, 15$, and 20 , with $B = 299$ resamples being drawn at each resampling operation. Results are given for nominal coverages 90%, 95%, and 99%.

Throughout Tables 4.4–4.7, estimated coverages that are greater than two standard errors from the nominal coverage are marked with an asterisk (*), and those that are greater than ten standard errors from the nominal coverage are marked with a dagger (†). It is clear from Tables 4.4–4.7 that in all the cases

TABLE 4.6. Characteristics of four percentile- t method bootstrap confidence intervals for the correlation coefficient. The samples $\mathcal{X} \equiv \{X_1, \dots, X_n\}$ and $\mathcal{Y} = \{Y_1, \dots, Y_n\}$ are drawn so that $X_i = Z_{1i} + W_i$ and $Y_i = Z_{2i} + W_i$ where Z_{1i} , Z_{2i} and W_i are distributed as independent folded Normal variates for $i = 1, \dots, n$ and $\rho = 0.5$. Nominal coverages are 90%, 95% and 99%. Tabulated characteristics are estimated coverage (nominal SE is .01 for 90% intervals, .007 for 95% intervals, and .003 for 99% intervals), average length, standard error of length and average value of upper endpoint. Estimated coverages that are more than two standard errors away from the nominal coverage are marked with a *, and those that are more than ten standard errors from the nominal coverage are marked with a †.

n		Untransformed percentile- t						Ordinary			Transformed percentile- t					
		Delta method			Jackknife			Percentile			Delta method			Jackknife		
	α	.90	.95	.99	.90	.95	.99	.90	.95	.99	.90	.95	.99	.90	.95	.99
10	cov.	.91	.98*	.99	.91	.98*	.99	.83*	.92*	.97*	.87*	.95	.99	.90	.94	.99
	len.	1.76	2.81	9.93	1.90	3.05	10.5	.82	1.01	1.41	.99	1.22	1.58	1.01	1.22	1.60
	st.err.	2.44	3.17	18.6	2.42	3.78	19.6	.32	.33	.37	.39	.40	.35	.42	.42	.38
	upper	1.04	1.35	5.35	1.10	1.44	5.61	.82	.86	.92	.80	.86	.95	.80	.86	.95
15	cov.	.92*	.97*	.99	.92*	.96	.99	.85*	.92*	.97*	.90	.94	.98*	.90	.94	.98*
	len.	.99	1.34	2.65	1.06	1.43	2.87	.66	.79	1.11	.76	.92	1.27	.78	.96	1.30
	st.err.	.57	.75	2.51	.67	.88	3.37	.21	.24	.29	.28	.32	.35	.31	.36	.38
	upper	.85	.96	1.53	.87	1.00	1.64	.77	.80	.86	.76	.81	.90	.77	.81	.91
20	cov.	.91	.96	.99	.91	.96	1.00*	.86*	.93*	.99	.88	.94	.99	.88	.94	.98
	len.	.77	1.00	1.76	.81	1.06	1.88	.57	.69	.96	.65	.78	1.09	.67	.82	1.12
	st.err.	.33	.44	.90	.38	.51	1.04	.16	.18	.23	.22	.24	.31	.25	.28	.34
	upper	.78	.88	1.16	.80	.90	1.21	.73	.77	.83	.73	.78	.87	.74	.79	.87

considered, the ordinary percentile- t intervals perform poorly. Even though they have reasonable coverage, often overcovering, they tend to be very long — some of them contain $[-1, 1]$ — and the estimated standard error of interval length is large. In comparison, the percentile-method intervals, although undercovering rather badly (statistically significantly in all cases), tend to be shorter, with fairly stable length. Except when the underlying population is bivariate lognormal, the transformed percentile- t intervals perform very well, having significantly better

TABLE 4.7. Characteristics of four percentile- t method bootstrap confidence intervals for the correlation coefficient. The parent population is bivariate log-normal with $\rho \approx 0.3775$. Nominal coverages are 90%, 95% and 99%. Tabulated characteristics are estimated coverage (nominal SE is .01 for 90% intervals, .007 for 95% intervals, and .003 for 99% intervals), average length, standard error of length and average value of upper endpoint. Estimated coverages that are more than two standard errors away from the nominal coverage are marked with a *, and those that are more than ten standard errors from the nominal coverage are marked with a †.

n		Untransformed percentile- t						Ordinary			Transformed percentile- t					
		Delta method			Jackknife			Percentile			Delta method			Jackknife		
	α	.90	.95	.99	.90	.95	.99	.90	.95	.99	.90	.95	.99	.90	.95	.99
10	cov.	.82*	.91*	.97*	.87*	.94*	.99	.83*	.92*	.98*	.77†	.88*	.96*	.85*	.90*	.97*
	len.	2.24	3.58	12.1	2.82	4.65	15.9	.90	1.08	1.45	.99	1.20	1.56	1.13	1.30	1.65
	st.err.	2.50	4.17	21.1	3.16	5.75	28.8	.28	.30	.33	.40	.40	.36	.44	.43	.36
	upper	.95	1.20	4.16	1.13	1.49	5.33	.83	.89	.94	.72	.78	.87	.77	.80	.90
15	cov.	.80*	.87†	.97*	.87*	.93*	.98*	.85*	.92*	.98*	.77†	.84†	.94†	.83*	.90*	.97*
	len.	1.28	1.89	4.10	1.71	2.56	5.70	.75	.89	1.20	.81	.98	1.33	.97	1.16	1.48
	st.err.	.90	1.68	4.39	1.41	2.74	6.47	.21	.22	.24	.32	.36	.39	.43	.42	.39
	upper	.85	.99	1.43	1.02	1.22	1.88	.77	.82	.89	.69	.73	.82	.75	.78	.86
20	cov.	.82*	.88†	.95†	.87*	.92*	.98*	.85*	.93*	.98*	.80†	.86†	.94†	.86*	.90*	.96*
	len.	1.08	1.44	2.50	1.40	1.91	3.50	.66	.78	1.03	.74	.87	1.16	.86	1.03	1.33
	st.err.	.69	1.18	1.86	1.12	1.87	3.15	.18	.18	.20	.30	.34	.36	.38	.42	.40
	upper	.84	.96	1.24	1.00	1.17	1.59	.74	.78	.85	.69	.73	.80	.73	.78	.85

coverage accuracy than the percentile-method intervals, while being only slightly longer than them, in general.

When the underlying population is bivariate lognormal (see Table 4.7), the transformed percentile- t intervals that use the nonparametric delta method to estimate the variance of Z fail badly. They suffer from severe undercoverage and are still longer than the percentile-method intervals. This is probably due to the instability in small samples of the plug-in estimate of the asymptotic variance, (4.50). The transformed percentile- t intervals that employ a jackknife estimate of the variance of Z have better coverage accuracy than the other transformed

percentile- t intervals, although they still undercover moderately, having coverage accuracy about the same as the corresponding percentile-method intervals. This is a fairly encouraging result in favour of the jackknife since that method is fairly general, making no explicit use of (4.50). None of the four percentile- t methods yielded coverage accuracy as high as the iterated "backwards" percentile method discussed earlier.

Overall, the transformed percentile- t interval which used the jackknife estimate of the variance of Z performed consistently best amongst the percentile- t intervals considered. However, none of the percentile- t intervals discussed performed particularly well when the underlying population was heavy tailed. In that case, if high coverage accuracy is desired, it would be better to use the iterated "backwards" percentile method.

7. PROOFS

7.1. Proof of Theorem 4.1. Write the coverage probability

$$\pi_0(\alpha) = 1 - \alpha + \{n^{-j/2}r_1(z_{1-\alpha}) + n^{-(j+1)/2}r(z_{1-\alpha})\}\phi(z_{1-\alpha}) + O(n^{-(j+2)/2})$$

for a polynomial r . Since $\hat{r}(z_{1-\alpha}) = r(z_{1-\alpha}) + O_p(n^{-\frac{1}{2}})$, we have

$$\hat{\pi}_0(\alpha) = \pi_0(\alpha) + n^{-(j+1)/2}U\phi(z_{1-\alpha}) + O_p(n^{-(j+2)/2}),$$

where $U \equiv n^{\frac{1}{2}}\{\hat{r}_1(z_{1-\alpha}) - r_1(z_{1-\alpha})\}$. Let β_α be the solution of $\pi_0(\beta_\alpha) = 1 - \alpha$, and define the $O_p(1)$ random variable b by $z_{1-\hat{\beta}_\alpha} = z_{1-\beta_\alpha} + n^{-(j+1)/2}b$. Then,

$$\begin{aligned} 1 - \alpha &= \hat{\pi}_0(\hat{\beta}_\alpha) = \pi_0(\hat{\beta}_\alpha) + n^{-(j+1)/2}U\phi(z_{1-\alpha}) + O_p(n^{-(j+2)/2}) \\ &= \pi_0(\beta_\alpha) + \Phi(z_{1-\hat{\beta}_\alpha}) - 1 + \alpha + n^{-(j+1)/2}U\phi(z_{1-\alpha}) + O_p(n^{-(j+2)/2}) \\ &= 1 - \alpha + n^{-(j+1)/2}(b + U)\phi(z_{1-\alpha}) + O_p(n^{-(j+2)/2}), \end{aligned}$$

whence

$$b = -U + O_p(n^{-\frac{1}{2}}).$$

Write the critical point $\hat{\theta}(\alpha)$ as $\hat{\theta} + n^{-\frac{1}{2}}\hat{\sigma}\{z_{1-\alpha} + n^{-\frac{1}{2}}\hat{f}(z_{1-\alpha})\}$, where $\hat{f}$ is $O_p(1)$. Then,

$$\hat{\theta}(\hat{\beta}_\alpha) = \hat{\theta}(\beta_\alpha) + n^{-\frac{1}{2}}\hat{\sigma}\{n^{-(j+1)/2}b + O_p(n^{-(j+2)/2})\},$$

and so,

$$\begin{aligned}\pi_1(\alpha) &\equiv P\{\theta \leq \hat{\theta}(\hat{\beta}_\alpha)\} \\ &= P\{\theta \leq \hat{\theta}(\beta_\alpha) + n^{-(j+2)/2}\hat{\sigma}b\} + O(n^{-(j+2)/2}) \\ &= P(T \geq -z_{1-\beta_\alpha}) + O(n^{-(j+2)/2}),\end{aligned}$$

where $T \equiv S + n^{-(j+1)/2}b$ and $S \equiv n^{\frac{1}{2}}(\hat{\theta} - \theta)/\hat{\sigma} + n^{-\frac{1}{2}}\hat{f}(z_{1-\beta_\alpha})$. We shall prove that for each x ,

$$P(T \leq x) = P(S \leq x) + n^{-(j+1)/2}v_1(z_{1-\alpha})x\phi(x) + O(n^{-(j+2)/2}), \quad (4.51)$$

from which it follows on taking $x = -z_{1-\beta_\alpha}$, that

$$1 - \pi_1(\alpha) = 1 - \pi_0(\beta_\alpha) - n^{-(j+1)/2}v_1(z_{1-\alpha})z_{1-\alpha}\phi(z_{1-\alpha}) + O(n^{-(j+2)/2}),$$

giving the desired result.

To derive (4.51), observe that

$$\begin{aligned}E(T^k) &= E\{(S + n^{-(j+1)/2}b)^k\} \\ &= E(S^k) + n^{-(j+1)/2}kE(S^{k-1}b) + O(n^{-(j+2)/2}).\end{aligned}$$

Let (N_1, N_2) have the joint limiting Normal distribution of (S, b) . Then both N_1 and N_2 have zero mean, N_1 has unit variance and $E(S^{k-1}b) = E(N_1^{k-1}N_2) + O(n^{-\frac{1}{2}})$. But

$$\begin{aligned}E(N_1^{k-1}N_2) &= E\{N_1^{k-1}E(N_2 | N_1)\} \\ &= E(N_1^k) \text{corr}(N_1, N_2) \{\text{var}(N_2)/\text{var}(N_1)\}^{\frac{1}{2}} \\ &= E(N_1^k)E(N_1N_2) = -E(N_1^k)v_1(z_{1-\alpha}) + O(n^{-\frac{1}{2}}),\end{aligned}$$

since $E(N_1N_2) = E(Sb) + O(n^{-\frac{1}{2}}) = -E(SU) + O(n^{-\frac{1}{2}}) = -v_1(z_{1-\alpha}) + O(n^{-\frac{1}{2}})$. Consequently,

$$E(T^k) = E(S^k) - n^{-(j+1)/2}kv_1(z_{1-\alpha})E(N_1^k) + O(n^{-(j+2)/2}),$$

and so the moment generating functions ψ_S and ψ_T of S and T respectively satisfy

$$\psi_T(\theta) = \psi_S(\theta) - n^{-(j+1)/2}v_1(z_{1-\alpha})\theta^2 \exp(\theta^2/2) + O(n^{-(j+2)/2}),$$

on noting that $\psi_S(\theta) = E(e^{\theta S}) = \sum_k E(S^k) \theta^k / k!$ and $\sum_k k E(N_1^k) \theta^k / k! = \theta^2 \exp(\theta^2/2)$. But S is asymptotically Normal $N(0, 1)$, so that $\psi_S(\theta) = e^{\frac{1}{2}\theta^2} + O(n^{-\frac{1}{2}})$ and

$$\psi_T(\theta) = \psi_S(\theta) \{1 - n^{-(j+1)/2} v_1(z_{1-\alpha}) \theta^2 + O(n^{-(j+2)/2})\}.$$

Taking logarithms of both sides and equating coefficients of $\theta^j / j!$, we deduce that all cumulants of T , except for the second, agree with those of S up to terms of order $n^{-(j+1)/2}$, and

$$k_2(T) = k_2(S) - n^{-(j+1)/2} 2v_1(z_{1-\alpha}) + O(n^{-(j+2)/2}).$$

Result (4.51) follows on noting the contributions which cumulants make to an Edgeworth expansion; see Section 3 of Chapter 1.

7.2. Proof of Theorem 4.2. Write the coverage probability

$$\pi_0(\alpha) = 1 - \alpha + \{n^{-j} t_1(z) + n^{-(j+1)} t(z)\} \phi(z) + O(n^{-(j+2)}),$$

where $z \equiv z_{1-(\alpha/2)}$, for a polynomial t . Then,

$$\hat{\pi}_0(\alpha) = \pi_0(\alpha) + (n^{-(2j+1)/2} V_1 + n^{-(2j+3)/2} V) \phi(z) + O_p(n^{-(j+2)}),$$

where $V_1 \equiv n^{\frac{1}{2}} \{\hat{t}_1(z) - t_1(z)\}$ and $V \equiv n^{\frac{1}{2}} \{\hat{t}(z) - t(z)\}$. Let β_α be the solution of $\pi_0(\beta_\alpha) = 1 - \alpha$ and let b_1, b be $O_p(1)$ random variables such that $z_{1-(\hat{\beta}_\alpha/2)} = z_{1-(\beta_\alpha/2)} + n^{-(2j+1)/2} b_1 + n^{-(2j+3)/2} b + O_p(n^{-(j+2)})$. Henceforth, we write z' for $z_{1-(\beta_\alpha/2)}$. Then,

$$\begin{aligned} 1 - \alpha &= \hat{\pi}_0(\hat{\beta}_\alpha) = \pi_0(\hat{\beta}_\alpha) + (n^{-(2j+1)/2} V_1 + n^{-(2j+3)/2} V) \phi(z) + O_p(n^{-(j+2)}) \\ &= 2\Phi(z_{1-(\hat{\beta}_\alpha/2)}) - 1 + \{n^{-j} t_1(z_{1-(\hat{\beta}_\alpha/2)}) + n^{-(j+1)} t(z_{1-(\hat{\beta}_\alpha/2)})\} \\ &\quad \times \phi(z_{1-(\hat{\beta}_\alpha/2)}) + (n^{-(2j+1)/2} V_1 + n^{-(2j+3)/2} V) \phi(z) + O_p(n^{-(j+2)}) \\ &= 1 - \beta_\alpha + (n^{-(2j+1)/2} b_1 + n^{-(2j+3)/2} b) 2\phi(z') \\ &\quad + \{n^{-j} t_1(z') + n^{-(j+1)} t(z')\} \phi(z') \\ &\quad + n^{-j} \cdot n^{-(2j+1)/2} b_1 \{t'_1(z') - z' t_1(z')\} \phi(z') \\ &\quad + (n^{-(2j+1)/2} V_1 + n^{-(2j+3)/2} V) \phi(z) + O_p(n^{-(j+2)}) \\ &= \pi_0(\beta_\alpha) + n^{-(2j+1)/2} (2b_1 + V_1) \phi(z) + n^{-(2j+3)/2} (2b + V \\ &\quad + n^{1-j} [b_1 z t_1(z) + b_1 \{t'_1(z) - z t_1(z)\}]) \phi(z) + O_p(n^{-(j+2)}) \\ &= 1 - \alpha + n^{-(2j+1)/2} (2b_1 + V_1) \phi(z) + n^{-(2j+3)/2} \{2b + V \\ &\quad + n^{1-j} b_1 t'_1(z)\} \phi(z) + O_p(n^{-(j+2)}), \end{aligned}$$

where the second-last identity is obtained by noticing that $z' = z - n^{-j} \frac{1}{2} t_1(z) + O(n^{-(j+1)})$. Therefore, we may take $b_1 \equiv -\frac{1}{2} V_1$ and $b \equiv -\frac{1}{2} V + n^{1-j} \frac{1}{4} V_1 t_1'(z)$. Write

$$\hat{\theta}_i(\alpha) = \hat{\theta} + n^{-\frac{1}{2}} \hat{\sigma} \{(-1)^i z + n^{-\frac{1}{2}} \hat{s}_{1i}(z) + n^{-1} \hat{s}_{2i}(z) + n^{-3/2} \hat{f}_i(z)\}$$

for polynomials $\hat{s}_{1i}$ and $\hat{s}_{2i}$, for $i = 1, 2$, and an $O_p(1)$ function $\hat{f}_i$. Then,

$$\begin{aligned} \hat{\theta}_i(\hat{\beta}_\alpha) &= \hat{\theta}_i(\beta_\alpha) + n^{-\frac{1}{2}} \hat{\sigma} [(-1)^i (n^{-(2j+1)/2} b_1 + n^{-(2j+3)/2} b) \\ &\quad + n^{-(2j+1)/2} b_1 \{n^{-\frac{1}{2}} \hat{s}'_{1i}(z) + n^{-1} \hat{s}'_{2i}(z)\} + O_p(n^{-(j+2)})] \\ &= \hat{\theta}_i(\beta_\alpha) + n^{-\frac{1}{2}} \hat{\sigma} [n^{-(2j+1)/2} (-1)^i b_1 + n^{-(2j+2)/2} b_1 \hat{s}'_{1i}(z) \\ &\quad + n^{-(2j+3)/2} \{(-1)^i b + b_1 \hat{s}'_{2i}(z)\} + O_p(n^{-(j+2)})]. \end{aligned}$$

Hence,

$$\begin{aligned} P\{\hat{\theta}_1(\hat{\beta}_\alpha) \leq \theta \leq \hat{\theta}_2(\hat{\beta}_\alpha)\} &= P\{\theta \leq \hat{\theta}_2(\hat{\beta}_\alpha)\} - P\{\theta \leq \hat{\theta}_1(\hat{\beta}_\alpha)\} \\ &= P(T_2 \geq -z') - P(T_1 \geq z') + O(n^{-(j+2)}), \end{aligned} \quad (4.52)$$

where $T_i = S_i + W_i$, $S_i \equiv n^{\frac{1}{2}} (\hat{\theta} - \theta) \hat{\sigma}^{-1} + n^{-\frac{1}{2}} \hat{s}_{1i}(z') + n^{-1} \hat{s}_{2i}(z') + n^{-3/2} \hat{f}_i(z')$, and $W_i \equiv n^{-(2j+1)/2} (-1)^i b_1 + n^{-(2j+2)/2} b_1 \hat{s}'_{1i}(z) + n^{-(2j+3)/2} \{(-1)^i b + b_1 \hat{s}'_{2i}(z)\}$.

In view of (4.52), we must calculate

$$P(T_2 \geq x) - P(T_1 \geq -x) = 1 - \{P(T_2 \leq x) + P(-T_1 \leq x)\},$$

for $x = -z'$. That is, we must calculate $P(T_2 \leq x) + P(-T_1 \leq x)$. This may be done by inverting the moment generating function

$$\psi_T(\theta) \equiv E(e^{\theta T_2}) + E(e^{-\theta T_1}) = \sum_{k=0}^{\infty} \{E(T_2^k) + E(-T_1)^k\} \theta^k / k!, \quad (4.53)$$

and so we compute $E(T_2^k) + (-1)^k E(T_1^k)$. Observe that

$$E(T_i^k) = E(S_i + W_i)^k = E(S_i^k) + k E(S_i^{k-1} W_i) + O(n^{-(j+2)}). \quad (4.54)$$

Put $S \equiv n^{\frac{1}{2}} (\hat{\theta} - \theta) \hat{\sigma}^{-1}$. Then,

$$\begin{aligned} S_i^{k-1} &= S^{k-1} + (k-1) S^{k-2} \{n^{-\frac{1}{2}} \hat{s}_{1i}(z') + n^{-1} \hat{s}_{2i}(z')\} \\ &\quad + \frac{1}{2} (k-1)(k-2) S^{k-3} n^{-1} \hat{s}_{1i}^2(z') + O_p(n^{-3/2}) \\ &= S^{k-1} + (k-1) S^{k-2} \{n^{-\frac{1}{2}} \hat{s}_{1i}(z) + n^{-1} \hat{s}_{2i}(z)\} \\ &\quad + \frac{1}{2} (k-1)(k-2) S^{k-3} n^{-1} \hat{s}_{1i}^2(z) + O_p(n^{-3/2}), \end{aligned} \quad (4.55)$$

and

$$\begin{aligned}
S_i^{k-1} W_i &= n^{-(2j+1)/2} (-1)^i S^{k-1} b_1 \\
&\quad + n^{-(2j+2)/2} \{ (-1)^i (k-1) S^{k-2} b_1 \hat{s}_{1i}(z) + S^{k-1} b_1 \hat{s}'_{1i}(z) \} \\
&\quad + n^{-(2j+3)/2} [S^{k-1} \{ (-1)^i b + b_1 \hat{s}'_{2i}(z) \} + (k-1) S^{k-2} b_1 \hat{s}_{1i}(z) \hat{s}'_{1i}(z) \\
&\quad + (-1)^i b_1 \{ (k-1) S^{k-2} \hat{s}_{2i}(z) + \tfrac{1}{2} (k-1)(k-2) S^{k-3} \hat{s}_{1i}^2(z) \}] \\
&\quad + O_p(n^{-(j+2)}).
\end{aligned}$$

It follows from Lemma 4.7 at the end of the proof that $\hat{s}_{11} = \hat{s}_{12}$. Recall that $\hat{s}_{21} = -\hat{s}_{22}$. Therefore, if k is odd,

$$\begin{aligned}
S(k) &\equiv S_2^{k-1} W_2 + (-1)^k S_1^{k-1} W_1 \\
&= n^{-(2j+1)/2} 2 S^{k-1} b_1 + n^{-(2j+2)/2} 2(k-1) S^{k-2} b_1 \hat{s}_{11}(z) \\
&\quad + n^{-(2j+3)/2} [2 S^{k-1} \{ b + b_1 \hat{s}'_{22}(z) \} \\
&\quad + b_1 (k-1)(k-2) S^{k-3} \hat{s}_{11}^2(z)] + O_p(n^{-(j+2)}),
\end{aligned}$$

while if k is even,

$$\begin{aligned}
S(k) &= n^{-(2j+2)/2} 2 S^{k-1} b_1 \hat{s}'_{11}(z) + n^{-(2j+3)/2} \{ 2(k-1) S^{k-2} b_1 \hat{s}_{11}(z) \hat{s}'_{11}(z) \\
&\quad + 2 b_1 (k-1) S^{k-2} \hat{s}_{22}(z) \} + O_p(n^{-(j+2)}).
\end{aligned}$$

The expected value of the $n^{-(2j+3)/2}$ term is $O(n^{-(j+2)})$ in both cases. In fact, for odd k ,

$$\begin{aligned}
E\{S(k)\} &= n^{-(2j+1)/2} 2E(S^{k-1} b_1) + n^{-(2j+2)/2} 2(k-1)E(S^{k-2} b_1) s_{11}(z) \\
&\quad + O(n^{-(j+2)}),
\end{aligned}$$

while for even k ,

$$E\{S(k)\} = n^{-(2j+2)/2} 2E(S^{k-1} b_1) s'_{11}(z) + O(n^{-(j+2)}).$$

Letting $\psi_S(\theta) \equiv E(e^{\theta S_2}) + E(e^{-\theta S_1})$, noting formula (4.54), and recalling (4.53), we obtain

$$\begin{aligned}
\psi_T(\theta) &= \psi_S(\theta) + \sum_{k=0}^{\infty} k E\{S(k)\} \theta^k / k! + O(n^{-(j+2)}) \\
&= \psi_S(\theta) + n^{-(j+1)} 2 \left[\sum_{k=0}^{\infty} \frac{\theta^{2k+1}}{(2k)!} \{ n^{\frac{1}{2}} E(S^{2k} b_1) + 2k E(S^{2k-1} b_1) s_{11}(z) \} \right. \\
&\quad \left. + \sum_{k=1}^{\infty} \frac{\theta^{2k}}{(2k-1)!} E(S^{2k-1} b_1) s'_{11}(z) \right] + O(n^{-(j+2)}).
\end{aligned}$$

If the polynomial v_2 is as defined at (4.9), then $E(S^{2k-1}b_1) = -E(N^{2k})v_2(z) + O(n^{-1})$, where N is Normal $N(0, 1)$. Therefore,

$$\begin{aligned}
 & - \sum_{k=1}^{\infty} \left\{ \frac{\theta^{2k+1}}{(2k)!} 2k E(S^{2k-1}b_1) s_{11}(z) + \frac{\theta^{2k}}{(2k-1)!} E(S^{2k-1}b_1) s'_{11}(z) \right\} + O(n^{-1}) \\
 &= v_2(z) \{ \theta s_{11}(z) + s'_{11}(z) \} \sum_{k=1}^{\infty} \frac{\theta^{2k}}{(2k-1)!} E(N^{2k}) \\
 &= v_2(z) \{ \theta s_{11}(z) + s'_{11}(z) \} \theta \sum_{k=1}^{\infty} \frac{\theta^{k-1}}{(k-1)!} E(N^k) \\
 &= v_2(z) \{ \theta s_{11}(z) + s'_{11}(z) \} \theta \frac{d}{d\theta} \{ \exp(\theta^2/2) \} \\
 &= v_2(z) \{ \theta s_{11}(z) + s'_{11}(z) \} \theta^2 \exp(\theta^2/2). \tag{4.57}
 \end{aligned}$$

Let $g(x, y) = \sum_i \sum_j a_{ij} x^i y^j$ be the polynomial introduced prior to Theorem 4.2. The joint density h_n , of $(S, -b_1)$ admits expansion

$$h_n(x, y) = \phi(x, y) + n^{-\frac{1}{2}} g(x, y) \phi(x, y) + n^{-1} g_1(x, y) \phi(x, y) + O(n^{-3/2}),$$

where g_1 is an even polynomial of degree 6 and ϕ is the density of the joint Normal limit (N_1, N_2) of $(S, -b_1)$. Since $x^{2k}y$ is an odd function,

$$\begin{aligned}
 -n^{\frac{1}{2}} E(S^{2k}b_1) &= n^{\frac{1}{2}} \iint x^{2k} y h_n(x, y) dx dy \\
 &= \iint x^{2k} y g(x, y) \phi(x, y) dx dy + O(n^{-1}). \tag{4.58}
 \end{aligned}$$

Let N_1, N_3 be independent Normal $N(0, 1)$ random variables, and set $\rho \equiv \text{corr}(N_1, N_2)$ and $\sigma_2^2 \equiv \text{var}(N_2)$. Then

$$\begin{aligned}
 & \iint x^{2k} y g(x, y) \phi(x, y) dx dy \\
 &= \sum_i \sum_j a_{ij} E(N_1^{2k+i} N_2^{j+1}) \\
 &= \sum_i \sum_j a_{ij} E\{N_1^{2k+i} E(N_2^{j+1} | N_1)\} \\
 &= \sum_i \sum_j a_{ij} E[N_1^{2k+i} \{N_3 \sigma_2 (1 - \rho^2)^{\frac{1}{2}} + \rho \sigma_2 N_1\}^{j+1}] \\
 &= \sum_i \sum_j \sum_l a_{ij} \binom{j+1}{l} E[N_1^{2k+i+l} N_3^{j+1-l} (\rho \sigma_2)^l \{\sigma_2 (1 - \rho^2)^{\frac{1}{2}}\}^{j+1-l}] \\
 &= \sum_i \sum_j \sum_l a_{ij} \binom{j+1}{l} E(N_1^{2k+i+l}) E(N_3^{j+1-l}) \sigma_2^{j+1} \rho^l (1 - \rho^2)^{\frac{1}{2}(j+1-l)},
 \end{aligned}$$

and so, by (4.58),

$$\begin{aligned}
 & - \sum_{k=0}^{\infty} \frac{\theta^{2k+1}}{(2k)!} n^{\frac{1}{2}} E(S^{2k} b_1) + O(n^{-1}) \\
 & = \theta \sum_i \sum_j \sum_l a_{ij} \binom{j+1}{l} E(N^{j+1-l}) \sigma_2^{j+1} \rho^l (1 - \rho^2)^{\frac{1}{2}(j+1-l)} \\
 & \quad \times \sum_{k=0}^{\infty} \frac{\theta^{2k}}{(2k)!} E(N^{2k+i+l}) .
 \end{aligned}$$

The general term in the triple series over i, j, l is zero unless $j+1-l$ is even, which is equivalent to saying that $i+l$ is even (since $a_{ij} = 0$ unless $i+j$ is odd). For even m ,

$$\begin{aligned}
 \sum_{k=0}^{\infty} \frac{\theta^{2k}}{(2k)!} E(N^{2k+m}) & = \sum_{k=0}^{\infty} \frac{\theta^k}{k!} E(N^{k+m}) \\
 & = \left(\frac{\partial}{\partial \theta} \right)^m E(e^{\theta N}) = (-1)^m H e_{m-1}(\theta) \exp(\theta^2/2) .
 \end{aligned}$$

Hence,

$$\begin{aligned}
 & - \sum_{k=0}^{\infty} \frac{\theta^{2k+1}}{(2k)!} n^{\frac{1}{2}} E(S^{2k} b_1) + O(n^{-1}) \\
 & = \theta \exp(\theta^2/2) \sum_i \sum_j \sum_l a_{ij} \binom{j+1}{l} E(N^{j+1-l}) \sigma_2^{j+1} \\
 & \quad \times \rho^l (1 - \rho^2)^{\frac{1}{2}(j+1-l)} (-1)^{i+l} H e_{i+l-1}(\theta) . \tag{4.59}
 \end{aligned}$$

Combining (4.56), (4.57) and (4.59), we conclude that

$$\psi_T(\theta) = \psi_S(\theta) - n^{-(j+1)} w(\theta) \exp(\theta^2/2) + O(n^{-(j+2)}) ,$$

where

$$\begin{aligned}
 w(\theta) & \equiv 2\theta^2 v_2(z) \{ \theta s_{11}(z) + s'_{11}(z) \} + 2\theta \sum_i \sum_j \sum_l (-1)^{i+l} a_{ij} \binom{j+1}{l} \\
 & \quad \times E(N^{j+1-l}) \sigma_2^{j+1} \rho^l (1 - \rho^2)^{\frac{1}{2}(j+1-l)} H e_{i+l-1}(\theta) .
 \end{aligned}$$

Consequently,

$$\begin{aligned}
 & P(T_2 \leq x) + P(-T_1 \leq x) \\
 & = P(S_2 \leq x) + P(-S_1 \leq x) + n^{-(j+1)} \xi(x) \phi(x) + O(n^{-(j+2)}) ,
 \end{aligned}$$

where ξ is the polynomial such that

$$\int e^{itx} d\{\xi(x)\phi(x)\} = -w(it) \exp(-t^2/2).$$

Taking $x = -z'$ in this expansion and noting (4.52), we deduce that

$$\begin{aligned} \pi_1(\alpha) &\equiv P\{\hat{\theta}_1(\hat{\beta}_\alpha) \leq \theta \leq \hat{\theta}_2(\hat{\beta}_\alpha)\} \\ &= 1 - \{P(T_2 \leq -z') + P(-T_1 \leq -z')\} + O(n^{-(j+2)}) \\ &= 1 - \{P(S_2 \leq -z') + P(-S_1 \leq -z')\} - n^{-(j+1)}\xi(-z)\phi(-z) \\ &\quad + O(n^{-(j+2)}) \\ &= 1 - \alpha - n^{-(j+1)}\xi(-z)\phi(z) + O(n^{-(j+2)}), \end{aligned}$$

since $1 - \alpha = \pi_1(\beta_\alpha) = P(S_2 \geq -z') - P(S_1 \geq z') = 1 - \{P(S_2 \leq -z') + P(-S_1 \leq -z')\}$. This completes the proof.

LEMMA 4.7. *In the above proof, $\hat{s}_{11} \equiv \hat{s}_{12}$.*

Proof. Observe that

$$1 - \alpha = P\{\hat{\theta}_1(\beta_\alpha) \leq \theta \leq \hat{\theta}_2(\beta_\alpha)\} = 1 - \{P(S_2 \leq -z') + P(-S_1 \leq -z')\}.$$

Also, by the first line of (4.55), and since $z' = z + O(n^{-1})$,

$$S_2^k + (-1)^k S_1^k = S^k \{1 + (-1)^k\} + n^{-\frac{1}{2}} k S^{k-1} \{s_{11}(z) + (-1)^k s_{12}(z)\} + O_p(n^{-1}).$$

Hence,

$$\begin{aligned} \psi_S(\theta) &\equiv \sum_{k=0}^{\infty} E\{S_2^k + (-1)^k S_1^k\} \theta^k / k! \\ &= E(e^{\theta S}) + E(e^{-\theta S}) + n^{-\frac{1}{2}} \{s_{11}(z) - s_{12}(z)\} \sum_{k=0}^{\infty} E(S^{2k}) \frac{\theta^{2k+1}}{(2k)!} + O(n^{-1}) \\ &= \exp(\theta^2/2) [2 + n^{-\frac{1}{2}} \{s_{11}(z) - s_{12}(z)\} \theta] + O(n^{-1}). \end{aligned}$$

Inverting this moment generating function to obtain $P(S_2 \leq x) + P(-S_1 \leq x)$ and taking $x = -z' = -z + O(n^{-1})$, we deduce that a term of order $n^{-\frac{1}{2}}$ is present in our expression unless $s_{11}(z) = s_{12}(z)$.

7.3. Proof of Theorem 4.3. Write the coverage probability

$$\pi_0(\alpha) = 1 - \alpha + n^{-\frac{1}{2}} r_1(z_{1-\alpha}) \phi(z_{1-\alpha}) + n^{-1} r(z_{1-\alpha}) \phi(z_{1-\alpha}) + O(n^{-3/2}),$$

for polynomials r_1 and r . If $j \neq 1$ in (4.15), then $r_1 \equiv 0$. Choosing $\hat{\beta}_\alpha$ so that $\hat{\pi}_0(\hat{\beta}_\alpha) = 1 - \alpha$, we have

$$\begin{aligned} 1 - \alpha &= \hat{\pi}_0(\hat{\beta}_\alpha) \\ &= 1 - \hat{\beta}_\alpha + n^{-\frac{1}{2}} \hat{r}_1(z_{1-\hat{\beta}_\alpha}) \phi(z_{1-\hat{\beta}_\alpha}) + n^{-1} r(z_{1-\hat{\beta}_\alpha}) \phi(z_{1-\hat{\beta}_\alpha}) + O_p(n^{-3/2}). \end{aligned}$$

Thus,

$$\begin{aligned} 1 - \hat{\beta}_\alpha &= 1 - \alpha - n^{-\frac{1}{2}} \hat{r}_1(z_{1-\hat{\beta}_\alpha}) \phi(z_{1-\hat{\beta}_\alpha}) - n^{-1} r(z_{1-\hat{\beta}_\alpha}) \phi(z_{1-\hat{\beta}_\alpha}) \\ &\quad + O_p(n^{-3/2}). \end{aligned} \quad (4.60)$$

From (4.60), it follows that $1 - \hat{\beta}_\alpha = 1 - \alpha - n^{-\frac{1}{2}} \hat{r}_1(z_{1-\alpha}) \phi(z_{1-\alpha}) + O_p(n^{-1})$, since $z_{1-\hat{\beta}_\alpha} = z_{1-\alpha} + O_p(n^{-1})$. As a result,

$$z_{1-\hat{\beta}_\alpha} = z_{1-\alpha} - n^{-\frac{1}{2}} \hat{r}_1(z_{1-\alpha}) + O_p(n^{-1}). \quad (4.61)$$

Substituting (4.61) into (4.60), and Taylor-expanding about $z_{1-\alpha}$, we obtain

$$\begin{aligned} 1 - \hat{\beta}_\alpha &= 1 - \alpha - n^{-\frac{1}{2}} \{ \hat{r}_1(z_{1-\alpha}) - n^{-\frac{1}{2}} \hat{r}_1(z_{1-\alpha}) \hat{r}'_1(z_{1-\alpha}) \} \{ \phi(z_{1-\alpha}) \\ &\quad - n^{-\frac{1}{2}} \hat{r}_1(z_{1-\alpha}) \phi'(z_{1-\alpha}) \} - n^{-1} r(z_{1-\alpha}) \phi(z_{1-\alpha}) + O_p(n^{-3/2}) \\ &= 1 - \alpha - n^{-\frac{1}{2}} \hat{r}_1(z_{1-\alpha}) \phi(z_{1-\alpha}) + n^{-1} \{ \hat{r}_1(z_{1-\alpha}) \hat{r}'_1(z_{1-\alpha}) \\ &\quad - z_{1-\alpha} \hat{r}_1^2(z_{1-\alpha}) - r(z_{1-\alpha}) \} \phi(z_{1-\alpha}) + O_p(n^{-3/2}). \end{aligned}$$

Inverting the Standard Normal distribution function and Taylor-expanding on both sides of this equation, we get

$$\begin{aligned} z_{1-\hat{\beta}_\alpha} &= z_{1-\alpha} - n^{-\frac{1}{2}} \hat{r}_1(z_{1-\alpha}) + n^{-1} \{ r_1(z_{1-\alpha}) r'_1(z_{1-\alpha}) \\ &\quad - \frac{1}{2} z_{1-\alpha} r_1^2(z_{1-\alpha}) - r(z_{1-\alpha}) \} + O_p(n^{-3/2}). \end{aligned} \quad (4.62)$$

From Hall (1988a), equation (4.14), we have the formulae $r_1(z_{1-\alpha}) \equiv s_1(z_{1-\alpha}) - q_1(z_{1-\alpha})$ and $r(z_{1-\alpha}) \equiv -[\frac{1}{2} s_1^2(z_{1-\alpha}) z_{1-\alpha} + s_1(z_{1-\alpha}) \{ q'_1(z_{1-\alpha}) - q_1(z_{1-\alpha}) z_{1-\alpha} \} - q_2(z_{1-\alpha}) + u z_{1-\alpha}]$, where u is defined in Theorem 4.3, for the polynomials appearing in an expansion of coverage probability. Therefore,

$$\begin{aligned} \hat{\theta}^{(1)}(\alpha) &\equiv \hat{\theta}(\hat{\beta}_\alpha) \\ &= \hat{\theta} + n^{-\frac{1}{2}} \hat{\sigma} \{ z_{1-\hat{\beta}_\alpha} + n^{-\frac{1}{2}} \hat{s}_1(z_{1-\hat{\beta}_\alpha}) + n^{-1} \hat{s}_2(z_{1-\hat{\beta}_\alpha}) \} + O_p(n^{-2}) \\ &= \hat{\theta} + n^{-\frac{1}{2}} \hat{\sigma} [z_{1-\alpha} + n^{-\frac{1}{2}} \{ \hat{s}_1(z_{1-\alpha}) - \hat{r}_1(z_{1-\alpha}) \} + n^{-1} \{ r_1(z_{1-\alpha}) r'_1(z_{1-\alpha}) \\ &\quad - \frac{1}{2} z_{1-\alpha} r_1^2(z_{1-\alpha}) - r(z_{1-\alpha}) - s'_1(z_{1-\alpha}) r_1(z_{1-\alpha}) + s_2(z_{1-\alpha}) \}] \\ &\quad + O_p(n^{-2}), \end{aligned}$$

the last line following on noting expansion (4.62) and Taylor-expanding about $z_{1-\alpha}$. Substituting our formulae for r_1 and r into the above expression for $\hat{\theta}^{(1)}(\alpha)$, we conclude that

$$\begin{aligned}\hat{\theta}^{(1)}(\alpha) &= \hat{\theta} + n^{-\frac{1}{2}}\hat{\sigma}[z_{1-\alpha} + n^{-\frac{1}{2}}\hat{q}_1(z_{1-\alpha}) + n^{-1}\{q_1'(z_{1-\alpha})q_1(z_{1-\alpha}) \\ &\quad - \frac{1}{2}z_{1-\alpha}q_1^2(z_{1-\alpha}) - q_2(z_{1-\alpha}) + uz_{1-\alpha}\}] + O_p(n^{-2}).\end{aligned}$$

The desired result follows on observing that $q_{21}(x) \equiv q_1'(x)q_1(x) - \frac{1}{2}xq_1^2(x) - q_2(x)$.

7.4. Proof of Theorem 4.4. Write $\pi_0(\alpha) = 1 - \alpha + n^{-1}t_1(z)\phi(z) + O(n^{-2})$, where $z \equiv z_{1-(\alpha/2)}$, for a polynomial t_1 . If $\hat{\beta}_\alpha$ is the solution of $\hat{\pi}_0(\hat{\beta}_\alpha) = 1 - \alpha$, then

$$1 - \alpha = \hat{\pi}_0(\hat{\beta}_\alpha) = 1 - \hat{\beta}_\alpha + n^{-1}\hat{t}_1(z_{1-(\hat{\beta}_\alpha/2)})\phi(z_{1-(\hat{\beta}_\alpha/2)}) + O_p(n^{-2}).$$

Thus,

$$\begin{aligned}1 - (\hat{\beta}_\alpha/2) &= 1 - (\alpha/2) - n^{-1}\frac{1}{2}\hat{t}_1(z_{1-(\hat{\beta}_\alpha/2)})\phi(z_{1-(\hat{\beta}_\alpha/2)}) + O_p(n^{-2}) \\ &= 1 - (\alpha/2) - n^{-1}\frac{1}{2}\hat{t}_1(z)\phi(z) + O_p(n^{-2}),\end{aligned}\tag{4.63}$$

since $z_{1-(\hat{\beta}_\alpha/2)} = z + O_p(n^{-1})$. It follows from (4.63) that

$$z_{1-(\hat{\beta}_\alpha/2)} = z - n^{-1}\frac{1}{2}\hat{t}_1(z) + O_p(n^{-2}).$$

Hence,

$$\begin{aligned}\hat{\theta}_i^{(1)}(\alpha) &\equiv \hat{\theta}_i(\hat{\beta}_\alpha) \\ &= \hat{\theta} + n^{-\frac{1}{2}}\hat{\sigma}\{(-1)^i z_{1-(\hat{\beta}_\alpha/2)} + n^{-\frac{1}{2}}\hat{s}_{1i}(z_{1-(\hat{\beta}_\alpha/2)}) + n^{-1}\hat{s}_{2i}(z_{1-(\hat{\beta}_\alpha/2)})\} \\ &\quad + O_p(n^{-2}) \\ &= \hat{\theta} + n^{-\frac{1}{2}}\hat{\sigma}\{(-1)^i z + n^{-\frac{1}{2}}\hat{s}_{1i}(z) + n^{-1}\{s_{2i}(z) - (-1)^i \frac{1}{2}t_1(z)\}\} \\ &\quad + O_p(n^{-2}).\end{aligned}\tag{4.64}$$

We deduce from Hall (1988a), equation (4.15), that

$$t_1(z) = -2[\frac{1}{2}s_{11}^2(z)z + s_{11}(z)\{q_1'(z) - zq_1(z)\} - q_2(z) - s_{22}(z) + uz],$$

where u is defined in Theorem 4.4. Using this formula for $t_1(z)$ and the fact that $s_{11} = s_{12}$ and $s_{21} = -s_{22}$, the term of order n^{-1} in (4.64) is

$$\begin{aligned}s_{2i}(z) - (-1)^i \frac{1}{2}t_1(z) \\ = (-1)^i [\frac{1}{2}s_{1i}^2(z)z + s_{1i}(z)\{q_1'(z) - zq_1(z)\} - q_2(z) + uz].\end{aligned}\tag{4.65}$$

Substituting (4.65) into (4.64) completes the proof.

7.5. Proof of Theorem 4.5. The length of I_0 is

$$L = n^{-\frac{1}{2}}\hat{\sigma}[2z + n^{-\frac{1}{2}}\{\hat{s}_{12}(z) - \hat{s}_{11}(z)\} + n^{-1}\{\hat{s}_{22}(z) - \hat{s}_{21}(z)\} + \cdots \\ + n^{-j}\{\hat{s}_{2j,2}(z) - \hat{s}_{2j,1}(z)\}] + O_p(n^{-(j+1)}), \quad (4.66)$$

where $z \equiv z_{1-(\alpha/2)}$. Since π_0 admits expansion (4.39), and $\hat{\beta}_\alpha$ is the solution of $\hat{\pi}_0(\hat{\beta}_\alpha) = 1 - \alpha$, we have

$$1 - \alpha = 1 - \hat{\beta}_\alpha + n^{-j}\hat{t}_1(z_{1-(\hat{\beta}_\alpha/2)})\phi(z_{1-(\hat{\beta}_\alpha/2)}) + O_p(n^{-(j+1)}),$$

so that

$$1 - (\hat{\beta}_\alpha/2) = 1 - (\alpha/2) - n^{-j}\frac{1}{2}\hat{t}_1(z)\phi(z) + O_p(n^{-(j+1)}),$$

since $z_{1-(\hat{\beta}_\alpha/2)} = z + O_p(n^{-j})$. Thus,

$$z_{1-(\hat{\beta}_\alpha/2)} = z - n^{-j}\frac{1}{2}\hat{t}_1(z) + O_p(n^{-(j+1)}).$$

Consequently,

$$\begin{aligned} \hat{\theta}_i^{(1)}(\alpha) &\equiv \hat{\theta}_i(\hat{\beta}_\alpha) \\ &= \hat{\theta} + n^{-\frac{1}{2}}\hat{\sigma}[(-1)^i z + n^{-\frac{1}{2}}\hat{s}_{1i}(z) + \cdots + n^{-(2j-1)/2}\hat{s}_{2j-1,i}(z) \\ &\quad + n^{-j}\{\hat{s}_{2j,i}(z) - (-1)^i\frac{1}{2}\hat{t}_1(z)\}] + O_p(n^{-(j+1)}). \end{aligned}$$

Therefore, the length of the iterated interval I_1 is

$$\begin{aligned} L^{(1)} &= n^{-\frac{1}{2}}\hat{\sigma}[2z + n^{-\frac{1}{2}}\{\hat{s}_{12}(z) - \hat{s}_{11}(z)\} + \cdots \\ &\quad + n^{-(2j-1)/2}\{\hat{s}_{2j-1,2}(z) - \hat{s}_{2j-1,1}(z)\} \\ &\quad + n^{-j}\{\hat{s}_{2j,2}(z) - \hat{s}_{2j,1}(z) - \hat{t}_1(z)\}] + O_p(n^{-(j+1)}) \\ &= L - n^{-(2j+1)/2}\hat{\sigma}\hat{t}_1(z) + O_p(n^{-(j+1)}). \end{aligned} \quad (4.67)$$

Noting that $\hat{\sigma} = \sigma + O_p(n^{-\frac{1}{2}})$ and $\hat{t}_1(z) = t_1(z) + O_p(n^{-\frac{1}{2}})$, and taking expectations of both sides of (4.67), we obtain

$$E(L^{(1)} - L) = -n^{-(2j+1)/2}\sigma t_1(z) + O(n^{-(j+1)}),$$

completing the proof.

7.6. Proof of Theorem 4.6. Using expansion (4.66) for the length of I_0 , we deduce

$$\text{var}(L) = 4n^{-1}z^2\text{var}(\hat{\sigma}) + \sum_{k=3}^{2j+2} n^{-k/2} \sum'_{(m,l,k)} \text{cov}(a_l, a_m) + O(n^{-(j+5/2)}), \quad (4.68)$$

where $\sum'_{(m,l,k)}$ denotes summation over positive integers m, l , such that $m + l = k$, and $a_1 \equiv 2z\hat{\sigma}$, and $a_d \equiv \hat{\sigma}\{\hat{s}_{d2}(z) - \hat{s}_{d1}(z)\}$, for $d = 2, \dots, 2j$, are the coefficients of $n^{-d/2}$ in (4.66). Clearly, since $\hat{\sigma} = \sigma + O_p(n^{-\frac{1}{2}})$ and $\hat{s}_{di}(z) = s_{di}(z) + O_p(n^{-\frac{1}{2}})$, for $d = 1, \dots, 2j$, and $i = 1, 2$, we have $\text{cov}(a_l, a_m) = O(n^{-1})$ for all l, m .

Using expansion (4.67) for the length $L^{(1)}$ of the iterated interval I_1 , we obtain

$$\text{var}(L^{(1)}) = 4n^{-1}z^2\text{var}(\hat{\sigma}) + \sum_{k=3}^{2j+2} n^{-k/2} \sum'_{(m,l,k)} \text{cov}(b_l, b_m) + O(n^{-(j+5/2)}), \quad (4.69)$$

where $b_d \equiv a_d$, for $d = 1, \dots, 2j - 1$, and

$$b_{2j} \equiv \hat{\sigma}\{\hat{s}_{2j,2}(z) - \hat{s}_{2j,1}(z) - \hat{t}_1(z)\} = a_{2j} - \hat{\sigma}\hat{t}_1(z),$$

are the coefficients of $n^{-j/2}$ for $d = 1, \dots, 2j$ in (4.67). Since the only terms of order $n^{-(j+3/2)}$ or greater which involve a_{2j} in (4.68) or b_{2j} in (4.69) have $l = 2j$ and $m = 1$, or $l = 1$ and $m = 2j$, it follows that

$$\begin{aligned} \text{var}(L^{(1)}) &= 4n^{-1}z^2\text{var}(\hat{\sigma}) + \sum_{k=3}^{2j+2} n^{-k/2} \sum'_{(m,l,k)} \text{cov}(a_l, a_m) \\ &\quad - 2n^{-(j+1)}\text{cov}(a_1, a_{2j}) + 2n^{-(j+1)}\text{cov}(b_1, b_{2j}) + O(n^{-(j+5/2)}) \\ &= \text{var}(L) - 2n^{-(j+1)}\{\text{cov}(a_1, a_{2j}) - \text{cov}(b_1, b_{2j})\} \\ &\quad + O(n^{-(j+5/2)}). \end{aligned} \quad (4.70)$$

Substituting formulae for a_1, a_{2j}, b_1 , and b_{2j} into (4.70), we get

$$\text{var}(L^{(1)}) = \text{var}(L) + 2n^{-(j+1)}\text{cov}\{2\hat{\sigma}z, -\hat{\sigma}\hat{t}_1(z)\} + O(n^{-(j+5/2)}),$$

from which the result follows.

7.7. Asymptotics for Bartlett-correction: Proof that terms of order $n^{-5/2}$ vanish in (4.49). By way of notation, we indicate components of vectors and matrices using superscripts, and define

$$l_{r_1 \dots r_j}(\omega) \equiv (\partial^j / \partial \omega^{r_1} \dots \partial \omega^{r_j}) l(\omega).$$

Throughout we shall employ summation notation, so that summation is over repeated indices, except that summation over the special index j will be indicated explicitly.

Our proof follows similar lines to Barndorff-Nielsen and Hall's (1988) proof that terms of order $n^{-\frac{1}{2}}$ and $n^{-3/2}$ vanish in an expansion of the distribution of w' about a chi-squared distribution with d degrees of freedom. As in that paper, we shall deal only with the full likelihood statistic, the argument in the case of the partitioned statistic differing only in minor notational details. We first obtain, by Taylor expansion, an expansion for $w \equiv 2\{l(\hat{\omega}) - l(\omega)\}$:

$$w = 2 \sum_{j=1}^7 (j!)^{-1} (\hat{\omega} - \omega)^{r_1} \dots (\hat{\omega} - \omega)^{r_j} l_{r_1 \dots r_j}(\omega) + O(n^{-3}).$$

Next, expanding $l_r(\hat{\omega})$ about ω , and observing that $l_r(\hat{\omega}) = 0$, we have

$$l_r(\omega) = - \sum_{j=1}^6 (j!)^{-1} (\hat{\omega} - \omega)^{r_1} \dots (\hat{\omega} - \omega)^{r_j} l_{rr_1 \dots r_j}(\omega) + O_p(n^{-5/2}).$$

Combining these two expansions and noting that $(j!)^{-1} - \{(j-1)!\}^{-1} = -\{(j-2)!j\}^{-1}$, we see that

$$w = \sum_{j=2}^7 \{-\tfrac{1}{2}(j-2)!j\}^{-1} (\hat{\omega} - \omega)^{r_1} \dots (\hat{\omega} - \omega)^{r_j} l_{r_1 \dots r_j}(\omega) + O_p(n^{-3}). \quad (4.71)$$

Put $\xi^r \equiv n^{\frac{1}{2}}(\hat{\omega} - \omega)^r$, $\lambda_{r_1 \dots r_j} \equiv n^{-1}E\{l_{r_1 \dots r_j}(\omega)\}$, and $\eta_{r_1 \dots r_j} \equiv n^{-\frac{1}{2}}\{l_{r_1 \dots r_j}(\omega) - n\lambda_{r_1 \dots r_j}\}$. Notice that $\lambda_{r_1 \dots r_j}$ does not depend on n .

The Bartlett correction factor has the form $(1 + \hat{b}_1 n^{-1})^{-1}$, where $\hat{b}_1$ is either known or estimable from the data. Assuming the latter to be the case, suppose that $\hat{b}_1$ is a $\sqrt{n}$ -consistent estimator of b_1 . In general, we may write $\hat{b}_1$ as follows,

$$\hat{b}_1 = b_1 + n^{-\frac{1}{2}}\zeta_1 + n^{-1}\zeta_2\zeta_3 + n^{-3/2}\zeta_4\zeta_5\zeta_6 + O_p(n^{-2}),$$

for zero-mean, $O_p(1)$ random variables $\zeta_1, \zeta_2, \zeta_3, \zeta_4, \zeta_5$, and ζ_6 . For example, if b_1 is a three-times differentiable function θ of a mean μ , then if $\bar{X} \equiv n^{-1} \sum_i X_i$ estimates μ , we have

$$\begin{aligned} \hat{b}_1 &\equiv \theta(\bar{X}) \\ &= \theta(\mu) + (\bar{X} - \mu)\theta'(\mu) + (\bar{X} - \mu)^2 \tfrac{1}{2}\theta''(\mu) + (\bar{X} - \mu)^3 \tfrac{1}{6}\theta'''(\mu) + O_p(n^{-2}). \end{aligned}$$

Then, the estimated Bartlett correction factor is

$$(1 + \hat{b}_1 n^{-1})^{-1} = 1 - b_1 n^{-1} - n^{-3/2} \zeta_1 + n^{-2} (b_1^2 - \zeta_2 \zeta_3) + n^{-5/2} (2b_1 \zeta_1 - \zeta_4 \zeta_5 \zeta_6) + O_p(n^{-3}). \quad (4.72)$$

Now, applying the Bartlett correction, we obtain from (4.71) and (4.72), the following expansion for w' ,

$$\begin{aligned} w' &\equiv (1 + \hat{b}_1 n^{-1})^{-1} w \\ &= \{1 - b_1 n^{-1} - n^{3/2} \zeta_1 + n^{-2} (b_1^2 - \zeta_2 \zeta_3) + n^{-5/2} (2b_1 \zeta_1 - \zeta_4 \zeta_5 \zeta_6)\} \\ &\quad \times \sum_{j=2}^7 n^{-(j-2)/2} \{-\tfrac{1}{2}(j-2)!\}^{-1} \xi^{r_1} \dots \xi^{r_j} (\lambda_{r_1 \dots r_j} + n^{-\frac{1}{2}} \eta_{r_1 \dots r_j}) \\ &\quad + O_p(n^{-3}). \end{aligned} \quad (4.73)$$

Let ν denote the vector comprised of $\zeta_1, \dots, \zeta_6$ and all the ξ 's and η 's appearing in (4.73). In regular cases, ν has an asymptotic multivariate Normal limit, with zero mean and finite variance covariance matrix, V . We consider only the case where V is nonsingular. Of course, the matrix Λ whose elements are λ_{rs} is nonsingular, and $-\Lambda^{-1}$ is the asymptotic variance-covariance matrix of $(\xi^1, \dots, \xi^d)$. Therefore, we may express ν as a linear transform of a vector δ having a limiting Normal $N(0, I)$ distribution and such that $-\xi^r \xi^s \lambda_{rs} = \sum_{1 \leq j \leq d} (\delta^j)^2$. Using this representation in (4.73), it follows that

$$\begin{aligned} w' &= \sum_{j=1}^d (\delta^j)^2 + \sum_{j=3}^7 n^{-(j-2)/2} c_{1, r_1 \dots r_j} \delta^{r_1} \dots \delta^{r_j} \\ &\quad + \sum_{j=2}^5 n^{-j/2} c_{2, r_1 \dots r_j} \delta^{r_1} \dots \delta^{r_j} + O_p(n^{-3}), \end{aligned} \quad (4.74)$$

for constants $c_{1, r_1 \dots r_j}$ and $c_{2, r_1 \dots r_j}$, not depending on n . To see how (4.74) follows from (4.73), merely multiply term by term through the summation in (4.73) and notice that powers of $n^{-k/2}$ always multiply sums of products of k or $k+2$ elements of ν . This is the crucial feature of (4.74): when k is even/odd, the coefficient of $n^{-k/2}$ in (4.74) is the sum of a product of an even/odd number of δ 's.

In regular problems, the vector δ has a density f which admits an expansion of the form

$$f(z) = \phi(z) + \sum_{j=1}^5 n^{-j/2} p_j(z) \phi(z) dz, \quad (4.75)$$

where ϕ is the Standard Normal density and the $p_j(z)$ are polynomials. As Barndorff-Nielsen and Hall (1988) point out, an almost universal property of cumulants results in p_j being an odd function for odd j , and an even function for even j ; see James (1955, 1958), James and Mayne (1962), Bhattacharya and Ghosh (1978) and Withers (1983).

Combining (4.74) and (4.75), we conclude that, for any real x , $P(w' \leq x) = \mathcal{I} + O(n^{-3})$, where

$$\mathcal{I} \equiv \int_{\mathcal{R}(x;n)} \left\{ \phi(z) + \sum_{j=1}^5 n^{-j/2} p_j(z) \phi(z) \right\} dz,$$

and

$$\begin{aligned} \mathcal{R}(x;n) \equiv \left\{ z : \sum_{j=1}^d (z^j)^2 + \sum_{j=3}^7 n^{-(j-2)/2} c_{1,r_1 \dots r_j} z^{r_1} \dots z^{r_j} \right. \\ \left. + \sum_{j=2}^5 n^{-j/2} c_{2,r_1 \dots r_j} z^{r_1} \dots z^{r_j} \leq x \right\}. \end{aligned}$$

If $\mathcal{I}$ is expanded in a power series in powers of $n^{-\frac{1}{2}}$, then odd powers vanish. A general proof of this result is tedious, so we merely establish that terms of order $n^{-\frac{1}{2}}$, $n^{-3/2}$, and $n^{-5/2}$ are zero. We show that if p is an even polynomial then

$$J_E \equiv \int_{\mathcal{R}(x;n)} p(z) \phi(z) dz = a_1 + a_2 n^{-1} + a_3 n^{-2} + O(n^{-3}), \quad (4.76)$$

for constants a_1 , a_2 , and a_3 , and if p is an odd polynomial, then

$$J_O \equiv \int_{\mathcal{R}(x;n)} p(z) \phi(z) dz = a_4 n^{-\frac{1}{2}} + a_5 n^{-3/2} + O(n^{-5/2}), \quad (4.77)$$

for constants a_4 and a_5 . The desired result follows directly from (4.76) and (4.77). We begin by introducing some notation. Put $g(z) \equiv \sum_{1 \leq j \leq d} (z^j)^2$, $g_1(z) \equiv c_{1,r_1 r_2 r_3} z^{r_1} z^{r_2} z^{r_3}$, and $g_j(z) = c_{1,r_1 \dots r_{j+2}} z^{r_1} \dots z^{r_{j+2}} + c_{2,r_1 \dots r_j} z^{r_1} \dots z^{r_j}$, for $j = 2, 3, 4$, and 5. Let y be a fixed real number and define integrals

$$J_i(y) = \int_{\mathcal{A}_i} p(z) \phi(z) dz,$$

for $i = 1, \dots, 5$, and even/odd polynomials p , where the integration is carried out over the sets $\mathcal{A}_i(y) \equiv \{z : g(z) + \sum_{1 \leq j \leq i-1} n^{-j/2} g_j(z) + n^{-i/2} y g_i(z) \leq x\}$

for $i = 1, \dots, 5$, and the null sum $\sum_{j=1}^0$ is interpreted as zero. Notice that $J_{i+1}(0) = J_i(1)$ for $i = 1, \dots, 4$. Also define the integrals

$$J_{i,\pm}(y) = \int_{\mathcal{A}_{i,\pm}} p(z)\phi(z) dz,$$

for $i = 1, \dots, 5$, where the integration is over the sets

$$\mathcal{A}_{i,\pm} \equiv \left\{ z : g(z) + \sum_{j=1}^{i-1} (-1)^j n^{-j/2} g_j(z) + y g_i(z) \leq x \right\}.$$

First assume that p is even. Expanding $J_1(y)$ about $y = 0$, we get

$$\begin{aligned} J_1(y) = J_1(0) + n^{-\frac{1}{2}} y C_{11} + \frac{1}{2} n^{-1} y^2 C_{12} + \frac{1}{6} n^{-3/2} y^3 C_{13} \\ + \frac{1}{24} n^{-2} y^4 C_{14} + \frac{1}{120} n^{-5/2} y^5 C_{15} + O(n^{-3}), \end{aligned} \quad (4.78)$$

where the C 's do not depend on n . Changing variable from z to $-z$ in the integral defining $J_1(y)$, we see that $J_1(-y) = J_1(y)$, so C_{1j} must be zero for odd j . Hence,

$$J_1(1) = J_1(0) + \frac{1}{2} n^{-1} C_{12} + \frac{1}{24} n^{-2} C_{14} + O(n^{-3}). \quad (4.79)$$

Next, expand $J_2(y)$ about $y = 0$, obtaining

$$J_2(y) = J_2(0) + n^{-1} y C_{21} + n^{-2} \frac{1}{2} y^2 C_{22} + O(n^{-3}). \quad (4.80)$$

Recalling $J_2(0) = J_1(1)$, and combining (4.79) and (4.80), we have

$$J_2(1) = J_1(0) + n^{-1} (\frac{1}{2} C_{12} + C_{21}) + n^{-2} (\frac{1}{2} C_{22} + \frac{1}{24} C_{14}). \quad (4.81)$$

Expanding $J_3(y)$ about zero, we get $J_3(y) = J_3(0) + n^{-3/2} y C_{31} + O(n^{-3})$. Barndorff-Nielsen and Hall (1988) have shown that $J_3(1)$ has no term of order $n^{-3/2}$. Consequently, $C_{31} = 0$, and

$$\begin{aligned} J_3(1) &= J_3(0) + O(n^{-3}) = J_2(1) + O(n^{-3}) \\ &= J_1(0) + n^{-1} (\frac{1}{2} C_{12} + C_{21}) + n^{-2} (\frac{1}{2} C_{22} + \frac{1}{24} C_{14}) + O(n^{-3}). \end{aligned}$$

Now, expanding $J_4(y)$ about zero, we find $J_4(y) = J_4(0) + n^{-2} y C_{41} + O(n^{-4})$. Thus,

$$\begin{aligned} J_4(1) &= J_4(0) + n^{-2} C_{41} + O(n^{-4}) = J_3(1) + n^{-2} C_{41} + O(n^{-4}) \\ &= J_1(0) + n^{-1} (\frac{1}{2} C_{12} + C_{21}) + n^{-2} (\frac{1}{2} C_{22} + \frac{1}{24} C_{14} + C_{41}) + O(n^{-3}). \end{aligned}$$

Finally, consider the integral $J_{5,\pm}(y)$. By Taylor expansion, $J_{5,+}$ and its derivatives differ from those of $J_{5,-}$ by $O(n^{-\frac{1}{2}})$. In particular, $J'_{5,+}(0) = J'_{5,-}(0) + O(n^{-\frac{1}{2}})$. Hence,

$$\begin{aligned} J_E &= J_{5,+}(n^{-3/2}) = J_{5,+}(0) + n^{-3/2} J'_{5,+}(0) + O(n^{-3}) \\ &= J_{5,+}(0) + n^{-3/2} J'_{5,-}(0) + O(n^{-2}). \end{aligned} \quad (4.82)$$

Changing variable from z to $-z$ in $J_{5,+}(y)$, it follows that $J_{5,+}(y) = J_{5,-}(-y)$, whence

$$J_E = J_{5,-}(-n^{-3/2}) = J_{5,-}(0) - n^{-3/2} J'_{5,-}(0) + O(n^{-3}). \quad (4.83)$$

Noting that $J_{5,+}(0) = J_{5,-}(0)$, and comparing (4.82) and (4.83), we deduce that $J'_{5,-}(0) = 0$. Therefore,

$$\begin{aligned} J_E &= J_{5,-}(0) + O(n^{-3}) = J_{5,+}(0) + O(n^{-3}) = J_4(1) + O(n^{-3}) \\ &= J_1(0) + n^{-1}(\tfrac{1}{2}C_{12} + C_{21}) + n^{-2}(\tfrac{1}{2}C_{22} + \tfrac{1}{24}C_{14} + C_{41}) + O(n^{-3}). \end{aligned}$$

This proves (4.76). We use a similar argument to prove (4.77). Suppose now that p is odd. Expanding $J_1(y)$ about zero gives (4.78). Changing variable from z to $-z$ in the integral defining $J_1(y)$, we deduce that $J_1(y) = -J_1(-y)$. Hence $J_1(0) = 0$ and $C_{1j} = 0$ for even j , and so

$$J_1(1) = n^{-\frac{1}{2}}C_{11} + n^{-3/2}\tfrac{1}{6}C_{31} + O(n^{-5/2}).$$

Next expand $J_2(y)$ about zero to obtain (4.80). In this case, both C_{21} and C_{22} are zero. To see this, consider the integral $J_{2,\pm}(y)$. By Taylor expansion, $J_{2,+}$ and its derivatives differ from those of $J_{2,-}$ by $O(n^{-\frac{1}{2}})$. In particular, $J'_{2,+}(0) = J'_{2,-}(0) + O(n^{-\frac{1}{2}})$ and $J''_{2,+}(0) = J''_{2,-}(0) + O(n^{-\frac{1}{2}})$. As a result,

$$\begin{aligned} J_2(y) &= J_{2,+}(n^{-1}y) = J_{2,+}(0) + n^{-1}y J'_{2,+}(0) + O(n^{-3}) \\ &= J_{2,+}(0) + n^{-1}y J'_{2,-}(0) + O(n^{-3/2}). \end{aligned} \quad (4.84)$$

Changing variable from z to $-z$ in the integral $J_{2,+}(y)$, we get $J_{2,+}(y) = -J_{2,-}(-y)$, whence

$$J_2(y) = J_{2,-}(-n^{-1}y) = J_{2,-}(0) - n^{-1}y J'_{2,-}(0) + \tfrac{1}{2}n^{-2}y^2 J''_{2,-}(0) + O(n^{-3}). \quad (4.85)$$

Comparing (4.84) and (4.85), and noting that since $J_{2,+}(y) = -J_{2,-}(-y)$, we have $J_{2,+}(0) = J_{2,-}(0) = J_{2,+}''(0) = J_{2,-}''(0) = 0$, it follows that $J_{2,-}'(0) = 0$. Consequently, $J_2(y) = J_2(0) + O(n^{-3})$, and

$$\begin{aligned} J_2(1) &= J_2(0) + O(n^{-3}) = J_1(1) + O(n^{-3}) \\ &= n^{-\frac{1}{2}}C_{11} + n^{-3/2}\frac{1}{6}C_{13} + O(n^{-5/2}). \end{aligned}$$

It only remains to expand $J_3(y)$, $J_4(y)$ and $J_5(y)$ about $y = 0$, and note that the term of order n^{-2} in the expansion of $J_4(y)$ is zero in the same way as for $J_2(y)$, to establish (4.77).

APPENDIX A4.1. Proof of Lemma A4.1.

LEMMA A4.1. *Using the notation of Example 1, Subsection 3.1,*

$$\begin{aligned}\lim_{n \rightarrow \infty} nE\{(\hat{\theta} - \theta)\hat{\sigma}^{-1}(\hat{\gamma} - \gamma)\} &= \kappa_4 - \frac{3}{2}\gamma^2, \\ \lim_{n \rightarrow \infty} nE\{(\hat{\theta} - \theta)\hat{\sigma}^{-1}(\hat{\gamma}^2 - \gamma^2)\} &= 2\gamma(\kappa_4 - \frac{3}{2}\gamma^2),\end{aligned}$$

and

$$\lim_{n \rightarrow \infty} nE\{(\hat{\theta} - \theta)\hat{\sigma}^{-1}(\hat{\kappa}_4 - \kappa_4)\} = \kappa_5 - 2\gamma\kappa_4.$$

Proof. Firstly, define the zero-mean, $O_p(1)$ random variables $U_1 \equiv n^{-\frac{1}{2}} \sum_i X_i$, $U_2 \equiv n^{-\frac{1}{2}} \sum_i (X_i^2 - 1)$, $U_3 \equiv n^{-\frac{1}{2}} \sum_i (X_i^3 - \gamma)$, and $U_4 \equiv n^{-\frac{1}{2}} \sum_i (X_i^4 - \kappa_4 + 3)$. Then, $\hat{\sigma}^2 = 1 + n^{-\frac{1}{2}}U_2 + O_p(n^{-1})$, and

$$n^{\frac{1}{2}}(\hat{\theta} - \theta)/\hat{\sigma} = U_1(1 - \frac{1}{2}n^{-\frac{1}{2}}U_2) + O_p(n^{-1}) = U_1 + O_p(n^{-\frac{1}{2}}). \quad (A4.1)$$

We may write the sample standardized skewness as

$$\begin{aligned}\hat{\gamma} &\equiv \hat{\sigma}^{-3}n^{-1} \sum_i (X_i - \bar{X})^3 \\ &= (1 - \frac{1}{2}n^{-\frac{1}{2}}U_2)^3 \{\gamma + n^{-\frac{1}{2}}(U_3 - 3U_1)\} + O_p(n^{-\frac{1}{2}}) \\ &= \gamma + n^{-\frac{1}{2}}(U_3 - 3U_1 - \frac{3}{2}U_2\gamma) + O_p(n^{-1}),\end{aligned}$$

and thus,

$$\hat{\gamma}^2 = \gamma^2 + n^{-\frac{1}{2}}2\gamma(U_3 - 3U_1 - \frac{3}{2}U_2\gamma) + O_p(n^{-1}).$$

Therefore,

$$\begin{aligned}n^{\frac{1}{2}}(\hat{\gamma} - \gamma) &= U_3 - 3U_1 - \frac{3}{2}U_2\gamma + O_p(n^{-\frac{1}{2}}), \\ n^{\frac{1}{2}}(\hat{\gamma}^2 - \gamma^2) &= 2\gamma(U_3 - 3U_1 - \frac{3}{2}U_2\gamma) + O_p(n^{-\frac{1}{2}}).\end{aligned} \quad (A4.2)$$

Now, the sample standardized kurtosis satisfies

$$\begin{aligned}\hat{\kappa}_4 &\equiv \hat{\sigma}^{-4}n^{-1} \sum_i (X_i - \bar{X})^4 - 3 \\ &= (1 - 2n^{-\frac{1}{2}}U_2)\{\kappa_4 + 3 + n^{-\frac{1}{2}}(U_4 - 4U_1\gamma)\} - 3 + O_p(n^{-1}) \\ &= \kappa_4 + n^{-\frac{1}{2}}\{U_4 - 4U_1\gamma - 2(\kappa_4 + 3)U_2\} + O_p(n^{-1}).\end{aligned}$$

Thus,

$$n^{\frac{1}{2}}(\hat{\kappa}_4 - \kappa_4) = U_4 - 4U_1\gamma - 2(\kappa_4 + 3)U_2 + O_p(n^{-\frac{1}{2}}). \quad (A4.3)$$

Combining expansions (A4.1)–(A4.3), we obtain

$$\begin{aligned}
 n(\hat{\theta} - \theta)\hat{\sigma}^{-1}(\hat{\gamma} - \gamma) &= U_1U_3 - 3U_1^2 - \frac{3}{2}U_1U_2\gamma + O_p(n^{-\frac{1}{2}}), \\
 n(\hat{\theta} - \theta)\hat{\sigma}^{-1}(\hat{\gamma}^2 - \gamma^2) &= 2\gamma(U_1U_3 - 3U_1^2 - \frac{3}{2}U_1U_2\gamma) + O_p(n^{-\frac{1}{2}}), \\
 n(\hat{\theta} - \theta)\hat{\sigma}^{-1}(\hat{\kappa}_4 - \kappa_4) &= U_1U_4 - 4U_1^2\gamma - 2(\kappa_4 + 3)U_1U_2 + O_p(n^{-\frac{1}{2}}).
 \end{aligned} \tag{A4.4}$$

Noting that $E(U_1^2) = 1$, $E(U_1U_2) = \gamma$, $E(U_1U_3) = E(X_1^4) = \kappa_4 + 3$, and $E(U_1U_4) = E(X_1^5) = \kappa_5 + 10\gamma$, and taking expectations of (A4.4) as $n \rightarrow \infty$, we deduce the required result.

CHAPTER 5

BOOTSTRAP METHODS FOR CONSTRUCTING CONFIDENCE INTERVALS FOR QUANTILES

1. INTRODUCTION

The problem of constructing confidence intervals and hypothesis tests for quantiles, particularly the median, has received considerable attention for many years. The so-called sign test for a quantile of a continuous distribution is thought to be the first distribution-free test ever used, its simplest form dating back to John Arbuthnot, who used it to support "An Argument for Divine Providence taken from the constant Regularity observ'd in the Births of Both Sexes" (1710–1712). Recently, computer-intensive statistical methods such as the bootstrap and jackknife techniques, which are known to perform well in simple problems, have been investigated as possible solutions to more complex problems, such as the quantile problem, with mixed success; recent references include Miller (1974), Maritz and Jarrett (1978), Efron (1979a, 1982), Kaigh (1983), McKean and Schrader (1984), Ghosh, Parr, Singh and Babu (1985), Sheather (1987), and Sheather and McKean (1987).

In this chapter, we discuss the use of bootstrap and jackknife methods in the problem of constructing accurate confidence intervals for quantiles. In Section 2, we question the efficacy of standard bootstrap techniques, such as the percentile method, bias corrections, bootstrap coverage correction and percentile-t methods, in the quantile problem. We show that if the sample quantile is defined appropriately, then a "backwards" percentile-method interval for a quantile is exactly equivalent to an interval derived via a sign test; see Kendall and Stuart (1977), p.546, for the latter. This fact is implicit in some earlier work (for example, Efron (1979a), p.5), but the explicit relationship between percentile-method and sign test-method critical points has not been given before. One consequence of this result is that refinements of percentile-method intervals, such as Efron's

(1981a, 1987) bias-correction and accelerated bias-correction, do no more than adjust sign test intervals. As a result, and in contrast to many other problems, percentile-method intervals have coverage error of precise order $n^{-\frac{1}{2}}$ in both one- and two-sided cases, and bias corrections fail to improve the order of coverage accuracy beyond that level.

Of at least equal significance is the fact that bootstrap iteration directed at improving coverage accuracy (see Chapters 3 and 4) has no role to play in the quantile problem and, if used, has absolutely no effect on the order of coverage accuracy of percentile-method confidence intervals for quantiles. This is in marked distinction to more classical problems, where each iteration reduces the order of coverage error by the factor $n^{-\frac{1}{2}}$ for one-sided intervals and n^{-1} for two-sided intervals. It thus emerges that standard bootstrap techniques perform poorly in the problem of constructing confidence intervals for quantiles: the percentile method is inaccurate for given levels, and produces nothing new; iteration of the percentile method fails completely; bias correction and accelerated bias correction are merely adjustments to sign-test intervals and therefore fail; and percentile-t is hardly a practicable alternative because there is not a suitable variance estimate.

In view of the poor performance of standard bootstrap methods in the quantile problem, we direct our attention in Sections 3 and 4 to the question of using the bootstrap estimate of sample quantile variance, introduced by Maritz and Jarrett (1978) and Efron (1979a), to "Studentize" the sample quantile in order to construct Normal theory confidence intervals for the quantile. In Section 3, we examine the rate at which the bootstrap quantile variance estimator converges to the true variance, and show that its relative error is of precise order $n^{-1/4}$. Thus, considered as a point estimator, the bootstrap variance estimator converges more slowly than the Siddiqui-Bloch-Gastwirth estimator and kernel estimators, which typically have smaller errors, of order at most $n^{-2/5}$. Since the asymptotic variance of a sample quantile depends on the value of the population density at the quantile, estimating the sample quantile variance is tantamount to density estimation. We also study the convergence properties of a bootstrap density estimate based on the bootstrap quantile variance estimator in Section 3.

In Section 4, we show that the level error of confidence intervals and hypothesis tests for population quantiles constructed using the bootstrap estimate of sample quantile variance is of precise order $n^{-\frac{1}{2}}$ in both one- and two-sided cases. This contrasts markedly with more classical problems, where the error

is of order $n^{-\frac{1}{2}}$ in the one-sided case, but n^{-1} in the two-sided case. It arises from an unusual feature of the Edgeworth expansion, in that the leading term, of order $n^{-\frac{1}{2}}$, is proportional to a polynomial containing both odd and even powers of the argument. We compare, both theoretically and by simulation, the performance of bootstrap-Studentized intervals and intervals constructed using Siddiqui-Bloch-Gastwirth (SBG) variance estimates. Our results show that for two-sided confidence intervals and hypothesis tests, and in large samples, the bootstrap variance estimate is inferior to the SBG variance estimate provided the smoothing parameter in the latter is chosen to minimize level error. A simulation study suggests that for nominal 90% and 95% two-sided confidence intervals for quantiles other than the median, and in small samples, bootstrap-Studentized intervals have better coverage accuracy than SBG-Studentized intervals.

Section 5 treats the problem of using the jackknife method to estimate sample quantile variance. We show that the jackknife estimate fails to be consistent, and that the jackknife-Studentized quantile has a non-normal limiting distribution. Proofs of theorems stated in Sections 1–5 are given in Section 6.

We close this section with some notation. Let $X, X_1, X_2, \dots$ denote independent and identically distributed observations from a population with distribution function F , and write $X_{n1} \leq \dots \leq X_{nn}$ for the order statistics of the sample $\mathcal{X} \equiv \{X_1, \dots, X_n\}$. Let $\hat{F}$ be the empiric distribution function corresponding to $\mathcal{X}$. Given $0 < p < 1$, put $r \equiv [np] + 1$, where $[\cdot]$ is the integer part function. Define

$$F^{-1}(p) \equiv \sup\{x : F(x) \leq p\},$$

and $\hat{F}^{-1}(p)$ similarly. If the p 'th quantile, ξ_p , of the distribution of X is uniquely defined then it is given by $\xi_p \equiv F^{-1}(p)$. We take the p 'th sample quantile to be $X_{nr} = \hat{F}^{-1}(p)$, although the results in Sections 3, 4 and 5 continue to hold if r is defined as $[np]$ or $[np] + 1$, indeed if $r = np + o(n^{\frac{1}{2}})$. Put $F_{nr}(x) \equiv P(X_{nr} \leq x)$ and $x_\alpha \equiv F_{nr}^{-1}(1 - \alpha)$ for $0 < \alpha < 1$, and define Φ and ϕ to be Standard Normal distribution and density functions, respectively.

2. ON THE ACCURACY OF BOOTSTRAP PERCENTILE METHOD CONFIDENCE INTERVALS FOR A QUANTILE

The “backwards” percentile method of constructing bootstrap confidence intervals, described in Section 1 of Chapter 4, is the oldest and probably the most popular of several approaches, dating from Efron's (1979a) seminal paper. However, as we have noted in Chapter 2 and Subsection 6.2 of Chapter 4,

percentile-method intervals can suffer from poor coverage accuracy. Several numerical and theoretical studies (for example, Efron, 1981a, 1987; Hall, 1988a) show that in many other cases, the percentile method performs poorly. For example, when used to construct confidence intervals for a population mean using a sample of size n , it produces confidence intervals with coverage error of order $n^{-\frac{1}{2}}$, rather than the order n^{-1} associated with percentile-t.

Techniques such as percentile-t, which rely on standardization for scale, are not really a practicable alternative in the case of confidence intervals for quantiles, since there is no natural estimate of the variance of a sample quantile. We show in Section 3 that the bootstrap variance estimate converges very slowly, at rate only $n^{-1/4}$, and Hall and Sheather (1988) show that Edgeworth expansions for the distribution of the Studentized quantile are substantially different from those for, say, the Studentized mean. Thus, in the case of quantiles, we are led to reconsider the percentile method, or refinements of the percentile method, such as Efron's (1981a, 1987) bias correction or accelerated bias correction.

To describe our results in detail, first suppose that the distribution function F is continuous. Now we introduce the bootstrap. Draw a sample $\mathcal{X}^* \equiv \{X_1^*, \dots, X_n^*\}$ from $\mathcal{X}$, with replacement. Let its order statistics be $X_{n1}^* \leq \dots \leq X_{nn}^*$. We call X_{nr}^* the p 'th resample quantile. Put $\hat{F}_{nr}(x) \equiv P(X_{nr}^* \leq x \mid \mathcal{X})$. We claim that

$$\hat{F}_{nr}^{-1}(1 - \alpha) \equiv \hat{F}^{-1}\{F(x_\alpha)\}, \quad (5.1)$$

for $0 < \alpha < 1$. To see why, let $Bi(n, q)$ denote a binomial random variable with parameters n and q , where $0 < q < 1$. Then,

$$\begin{aligned} \hat{F}_{nr}^{-1}(1 - \alpha) &= \sup\{x : \hat{F}_{nr}(x) \leq 1 - \alpha\} \\ &= \sup\{x : P[Bi(n, \hat{F}(x)) \geq r \mid \mathcal{X}] \leq 1 - \alpha\} \\ &= \sup\{x : P[Bi(n, \hat{F}(x)) \geq r \mid \mathcal{X}] \leq P[Bi(n, F(x_\alpha)) \geq r]\} \\ &= \sup\{x : \hat{F}(x) \leq F(x_\alpha)\} \\ &= \hat{F}^{-1}\{F(x_\alpha)\}, \end{aligned}$$

since $P\{Bi(n, q) \geq r\}$ is increasing in q .

Recall from Subsection 1.3 of Chapter 4 that if we are interested in a parameter θ , estimated by $\hat{\theta}$, and if $\hat{\theta}^*$ denotes the version of $\hat{\theta}$ computed using the resample $\mathcal{X}^*$ rather than the sample $\mathcal{X}$, then defining $\hat{u}_\alpha \equiv \{u : P(\hat{\theta}^* \leq u \mid \mathcal{X}) \leq \alpha\}$, the intervals $(-\infty, \hat{u}_{1-\alpha}]$, $[\hat{u}_\alpha, \infty)$ and $[\hat{u}_{\alpha/2}, \hat{u}_{1-(\alpha/2)}]$ are "backwards" percentile-method intervals for θ with nominal coverage $1 - \alpha$. In the

case of quantiles, we take $\theta = \xi_p$, $\hat{\theta} = X_{nr}$ and $\hat{\theta}^* = X_{nr}^*$, and so $\hat{u}_\alpha = \hat{F}_{nr}^{-1}(\alpha)$ for $0 < \alpha < 1$. Hence, by (5.1), percentile-method critical points are obtained from inversion of the empiric distribution function, exactly as are sign test or binomial-method critical points.

In the case of confidence intervals for quantiles, no amount of adjustment of the nominal level, $1 - \alpha$, will produce percentile-method intervals with exact coverage. This is because, being sign test intervals, their coverage probabilities are sums of atoms of a binomial distribution, and only in extremely lucky and unrealistic circumstances can the sum possibly add to exactly $1 - \alpha$. This contrasts with the behaviour of percentile-method intervals in most other circumstances, where optimal adjustment can yield perfect coverage accuracy. Indeed Efron's (1981a, 1987) bias-correction and accelerated bias-correction methods exploit this fact to give improved coverage accuracy. An even more annoying feature of percentile-method intervals in the case of quantiles is that for a given value of $1 - \alpha$, coverage accuracy cannot, in general, be reduced below order $n^{-\frac{1}{2}}$. This happens because $P\{\xi_p \leq \hat{F}^{-1}\{F(x_\alpha)\}\}$ is a binomial probability, and atoms of the $Bi(n, p)$ distribution are of order $n^{-\frac{1}{2}}$ except in the extreme tails:

$$\begin{aligned} P\{\xi_p \leq \hat{F}^{-1}\{F(x_\alpha)\}\} &= P\{Bi(n, p) \leq nF(x_\alpha)\}, \\ \max_{0 \leq s \leq n} \left| \binom{n}{s} p^s (1-p)^{n-s} - \{np(1-p)\}^{-\frac{1}{2}} \phi(y) \right| &= O(n^{-1}), \end{aligned} \quad (5.2)$$

where $y \equiv (s - np)\{np(1-p)\}^{-\frac{1}{2}}$. (See Petrov (1975), p. 197). For example, if $p = \frac{1}{2}$, if $\alpha = 1 - \Phi(2)$, and if the sequence $\{n_k, k \geq 1\}$ is defined by $n_k = 4k^2$, then, along that sequence,

$$\inf_{0 < \beta < 1} |P\{\xi_p \leq \hat{F}^{-1}\{F(x_\beta)\}\} - (1 - \alpha)| \sim n_k^{-\frac{1}{2}} \phi(2). \quad (5.3)$$

See Appendix A5.1 for a proof. Therefore, in this instance and in many others, coverage accuracy of the bootstrap confidence interval $(-\infty, \hat{F}_{nr}^{-1}(1 - \alpha)]$ cannot be reduced below order $n^{-\frac{1}{2}}$ by adjusting the nominal level. The same is true in the case of two-sided intervals, since the dominant error in coverage accuracy arises from rounding errors which compound rather than cancel in the two-sided case. It follows that correction methods such as Efron's bias correction and accelerated bias correction, which merely adjust the level of percentile-method intervals, also have coverage error of size $n^{-\frac{1}{2}}$.

Bootstrap iteration to improve coverage accuracy of confidence intervals has been studied in Chapter 4. Recall from Subsection 2.1 of Chapter 4, that for the

one-sided “backwards” percentile-method interval $(-\infty, \hat{u}_{1-\alpha}]$, Beran’s (1987a) method of bootstrap iteration may be employed as follows. Estimate the true coverage, $\pi(\alpha) \equiv P(\theta \leq \hat{u}_{1-\alpha})$, by

$$\hat{\pi}(\alpha) \equiv P(\hat{\theta} \leq \hat{u}_{1-\alpha}^* \mid \mathcal{X}),$$

where $\hat{u}_{1-\alpha}^* \equiv \sup\{u : P(\hat{\theta}^{**} \leq u \mid \mathcal{X}^*) \leq 1 - \alpha\}$, $\mathcal{X}^*$ is a resample from $\mathcal{X}$, and $\hat{\theta}^{**}$ is the version of $\hat{\theta}$ computed using $\mathcal{X}^{**}$, a resample from $\mathcal{X}^*$, rather than $\mathcal{X}$. The once-iterated “backwards” percentile interval is $(-\infty, \hat{u}_{1-\hat{\beta}_\alpha}]$, where $\hat{\beta}_\alpha$ is the solution of the equation $\hat{\pi}(\hat{\beta}_\alpha) = 1 - \alpha$. Beran’s method may be thought of as a calibration technique. Loh (1987) gives an excellent, detailed account of calibration-related methods. This interpretation, which seems natural in this context, was given by DiCiccio and Romano (1988).

In the case where θ is the quantile ξ_p , the true coverage $\pi(\alpha)$ is known anyway. This follows from the fact that

$$F(x_\alpha) = \sup\{F(x) : F_{nr}(x) \leq 1 - \alpha\} = \sup\{u : P[Bi(n, u) \geq r] \leq 1 - \alpha\},$$

which is known, and by (5.2), $\pi(\alpha) = P\{Bi(n, p) \leq nF(x_\alpha)\}$, which must also be known. Hence bootstrap iteration has no role to play in this problem, for it serves merely to estimate the *known* quantity $\pi(\alpha)$. In particular, bootstrap iteration cannot be used to reduce coverage error below the order $n^{-\frac{1}{2}}$ associated with sign test intervals.

3. THE EXACT CONVERGENCE RATE OF THE BOOTSTRAP QUANTILE VARIANCE ESTIMATOR

The bootstrap variance estimator for a sample quantile was introduced independently by Maritz and Jarrett (1978) and Efron (1979a). In this section, we show that it converges to the true variance quite slowly, having a relative error of order $n^{-1/4}$, where n is sample size. In contrast, the earlier Siddiqui-Bloch-Gastwirth (SBG) estimator (Siddiqui, 1960; Bloch and Gastwirth, 1968) has a smaller relative error of size $n^{-2/5}$. Kernel methods developed by Parzen (1979), Csörgö (1983), Falk (1986) and Welsh (1987, 1988) may be employed to achieve even smaller errors of size $n^{-\frac{1}{2}+\epsilon}$ for any $\epsilon > 0$. In this sense, the bootstrap variance estimator is inferior to more classical estimators. The bootstrap variance estimator has been studied numerically by Maritz and Jarrett (1978), McKean

and Schrader (1984) and Sheather (1986a), among others. Consistency has been proved by Babu (1986). See also Ghosh, Parr, Singh and Babu (1985).

Using the notation introduced in Section 1 of this chapter, the bootstrap variance estimator of $\sigma^2 \equiv \text{var}(X_{nr})$ is

$$\hat{\sigma}^2 \equiv \sum_{j=1}^n (X_{nj} - X_{nr})^2 w_j, \quad (5.4)$$

where

$$w_j \equiv r \binom{n}{r} \int_{(j-1)/n}^{j/n} x^{r-1} (1-x)^{n-r} dx.$$

By way of notation, a function g will be said to satisfy a Lipschitz condition of order τ in a neighbourhood of a point z if for some $C > 0$ and $\epsilon > 0$, $|g(x) - g(y)| \leq C|x - y|^\tau$, whenever $x, y \in (z - \epsilon, z + \epsilon)$.

The exact variance σ^2 depends closely on the value at ξ_p of the density $f \equiv F'$. Therefore, estimation of σ^2 involves, either explicitly or implicitly, density estimation. If f has a bounded derivative then

$$n\sigma^2 = p(1-p)f(\xi_p)^{-2} + O(n^{-1}), \quad (5.5)$$

as $n \rightarrow \infty$; see David and Johnson (1954), Pearson and Pearson (1931, 1932) and Van Zwet (1964). We give a refinement of this result in Theorem 5.1, which is proved in Subsection 6.1.

THEOREM 5.1. Assume $E|X|^\eta < \infty$ for some $\eta > 0$, $f \equiv F'$ exists and satisfies a Lipschitz condition of order $\tau \in (0, 1]$ in a neighbourhood of ξ_p , and $f(\xi_p) > 0$. Then,

$$\sigma^2 = n^{-1}p(1-p)f(\xi_p)^{-2} + O(n^{-1-(\tau/2)}),$$

as $n \rightarrow \infty$.

Next, we describe a central limit theorem for $\hat{\sigma}^2$, defined at (5.4).

THEOREM 5.2. Assume the conditions of Theorem 5.1, and in addition that $\tau > \frac{1}{2}$. Then $n^{5/4}(\hat{\sigma}^2 - \sigma^2)$ is asymptotically Normally distributed with zero mean and variance $2\pi^{-1/2}\{p(1-p)\}^{3/2}f(\xi_p)^{-4}$.

A proof of Theorem 5.2 is given in Subsection 6.2. It is trivial to modify the proof of Theorem 5.2 so as to show that with $\tau \geq \frac{1}{2}$ (instead of $\tau > \frac{1}{2}$), we have

$\hat{\sigma}^2 - \sigma^2 = O_p(n^{-5/4})$. However, it does not seem possible to obtain asymptotic Normality without $\tau > \frac{1}{2}$. Of course, when $\tau > \frac{1}{2}$, we have from Theorem 5.1 that $\sigma^2 = n^{-1}p(1-p)f(\xi_p)^{-2} + o(n^{-5/4})$.

In view of (5.5), we are implicitly estimating f with an error of larger order than n^{-1} . Therefore, our conclusions about the quality of $\hat{\sigma}^2$ as an estimator of σ^2 apply equally to $\hat{f}(\xi_p) \equiv \{p(1-p)(n\hat{\sigma}^2)^{-1}\}^{\frac{1}{2}}$ as an estimator of $f(\xi_p)$. Under the conditions of Theorem 5.2, $\hat{f}(\xi_p)$ is a consistent estimator of $f(\xi_p)$ and satisfies the central limit theorem

$$n^{1/4}\{\hat{f}(\xi_p) - f(\xi_p)\} \rightarrow N(0, \lambda^2), \quad (5.6)$$

in distribution, where $\lambda^2 \equiv (2\pi^{\frac{1}{2}})^{-1}\{p(1-p)\}^{-\frac{1}{2}}f(\xi_p)^2$. More generally, if the condition $\tau > \frac{1}{2}$ is weakened to $\tau \geq \frac{1}{2}$ then $\hat{f}(\xi_p) = f(\xi_p) + O_p(n^{-1/4})$.

There is considerable interest in estimation of the so-called *sparsity function*, $\eta(p) \equiv 1/f(\xi_p)$; see McKean and Schrader (1984), Sheather (1986a, 1986b) and Welsh (1987, 1988). An immediate corollary of our work is that $1/\hat{f}(\xi_p)$ is an inferior point estimator of $\eta(p)$.

4. THE ERROR INCURRED USING THE BOOTSTRAP VARIANCE ESTIMATE WHEN CONSTRUCTING CONFIDENCE INTERVALS FOR QUANTILES

The bootstrap estimate of the variance of a sample quantile is frequently used to "Studentize" the sample quantile, either explicitly, in the construction of confidence intervals for population quantiles, or implicitly, in statements of accuracy such as "sample quantile plus or minus two standard errors." Its numerical performance in these contexts, with particular reference to coverage accuracy of two-sided confidence intervals for the median, has been studied in depth (see, for example, McKean and Schrader, 1984; Sheather and McKean, 1987). In this section, we provide theoretical confirmation of the results of these studies, by stating and proving an Edgeworth expansion which describes the precise order of coverage accuracy of bootstrap-Studentized confidence intervals for population quantiles.

A major competitor of the bootstrap variance estimator is the Siddiqui-Bloch-Gastwirth (SBG) estimator, mentioned briefly in Section 3, where the variance estimate is based on a density estimate. It is known that when the smoothing parameter, m , in this quantity is chosen in a manner which is optimal for coverage accuracy (meaning that $m \approx n^{2/3}$), the resulting two-sided

confidence interval has coverage error of precise order $n^{-2/3}$, as sample size, n , increases. On the other hand, when the smoothing parameter is chosen in a manner which optimizes the pointwise convergence rate ($m \approx n^{4/5}$), the coverage error is larger, of precise order $n^{-2/5}$. A third proposal, the so-called "Normal quantile" method, studied by McKean and Schrader (1984) and Sheather and McKean (1987), has $m \approx n^{1/2}$ and results in coverage error of precise order $n^{-1/2}$. All of these conclusions are clear from Edgeworth expansions in Hall and Sheather (1988).

The bootstrap estimate of the variance, σ^2 , of X_{nr} is given at (5.4). When $p = \frac{1}{2}$ and n is even, the sample median is often defined by $\frac{1}{2}(X_{n, \frac{1}{2}n} + X_{n, \frac{1}{2}n+1})$. There, $\hat{\sigma}^2$ admits a more complicated formula given by Maritz and Jarrett (1978).

As expected, the bootstrap-Studentized quantile $(X_{nr} - \xi_p)/\hat{\sigma}$ has an asymptotic Standard Normal distribution. This follows from the usual central limit theorem for X_{nr} (see, for example, David (1981), p.255) and from the fact that $\hat{\sigma}\sigma^{-1}$ tends to unity as $n \rightarrow \infty$ (Ghosh, Parr, Singh and Babu, 1985; Babu 1986). Our interest centres on the rate at which this Normal limit is approached. By way of regularity conditions, assume that $f \equiv F'$ and its first two derivatives exist and are bounded in a neighbourhood of ξ_p , and that $f(\xi_p) > 0$. Suppose also that for some $\eta > 0$, $E|X|^\eta < \infty$. Put $\epsilon_n \equiv [np] + 1 - np$,

$$Q_1(z) \equiv \frac{1}{4}\{\pi p(1-p)\}^{-\frac{1}{2}}z(z^2 + 1 + 2\sqrt{2}),$$

and

$$\begin{aligned} Q_2(z) \equiv & \frac{1}{6}\{p(1-p)\}^{-\frac{1}{2}}(1+p)(z^2 - 1) \\ & - \frac{1}{2}[\{p(1-p)\}^{-\frac{1}{2}}p + \{p(1-p)\}^{\frac{1}{2}}f'(\xi_p)f(\xi_p)^{-2}]z^2 \\ & - \{p(1-p)\}^{-\frac{1}{2}}\{\frac{1}{2}(1-p) + \epsilon_n\}. \end{aligned}$$

Note that $Q_1(z)$ is an odd polynomial whereas $Q_2(z)$ is even. The following theorem allows us to find expansions for the coverage probability of bootstrap-Studentized intervals.

THEOREM 5.3. *Under the above regularity conditions,*

$$\sup_{-\infty < z < \infty} |\mathbb{P}\{(X_{nr} - \xi_p)/\hat{\sigma} \leq z\} - \Phi(z) - n^{-\frac{1}{2}}\{Q_1(z) + Q_2(z)\}\phi(z)| = O(n^{-3/4}),$$

as $n \rightarrow \infty$.

Theorem 5.3 is proved in Subsection 6.3.

Let $z_{1-\alpha}$ denote the solution of $\Phi(z_{1-\alpha}) = 1 - \alpha$ for $0 < \alpha < 1$. The intervals $I_1 \equiv (-\infty, X_{nr} + \hat{\sigma}z_{1-\alpha}]$ and $I_2 \equiv [X_{nr} - \hat{\sigma}z_{1-(\alpha/2)}, X_{nr} + \hat{\sigma}z_{1-(\alpha/2)}]$ are both confidence intervals for ξ_p with nominal coverage $1 - \alpha$. Their actual coverages are, by Theorem 5.3,

$$P(\xi_p \in I_1) = 1 - \alpha + n^{-\frac{1}{2}}\{Q_1(z_{1-\alpha}) - Q_2(z_{1-\alpha})\}\phi(z_{1-\alpha}) + O(n^{-3/4}), \quad (5.7)$$

and

$$P(\xi_p \in I_2) = 1 - \alpha + n^{-\frac{1}{2}}2Q_1(z_{1-(\alpha/2)})\phi(z_{1-(\alpha/2)}) + O(n^{-3/4}). \quad (5.8)$$

Note that the even polynomial Q_2 has cancelled from the term of order $n^{-\frac{1}{2}}$ in formula (5.8). Expansions (5.7) and (5.8) show that coverage error is of precise order $n^{-\frac{1}{2}}$ in both one- and two-sided confidence intervals. In this sense, bootstrap-Studentized intervals perform similarly to sign test intervals; see Section 2.

Exactly the same expansions would be obtained if z_α were a quantile of Student's t distribution with $n - \nu$ degrees of freedom (any fixed integer ν) rather than a quantile of the Normal distribution. The reason is that z_α in the former case differs from z_α in the latter case only by a term of order n^{-1} , which is negligible relative to even the remainder terms in (5.7) and (5.8).

One simple application of Theorem 5.3 is to relate the coverage errors of different types of confidence interval. For example, direct comparison of two-sided Normal quantile method (where $m = \min(p, 1 - p)n^{\frac{1}{2}}$) and bootstrap-Studentized intervals does not involve the value of f at ξ_p . This is the case of most interest when comparing bootstrap-Studentized and SBG-Studentized intervals, as Normal quantile intervals have the same order of coverage accuracy as bootstrap-Studentized intervals, while the SBG intervals which use $m \approx n^{2/3}$ and $m \approx n^{4/5}$ have higher and lower orders of coverage accuracy, respectively. It follows from equation (2.7) of Hall and Sheather (1988) that the two-sided Normal quantile interval, say I_3 , has coverage probability

$$P(\xi_p \in I_3) = 1 - \alpha - \frac{1}{2}n^{-\frac{1}{2}}\{\min(p, 1 - p)\}^{-\frac{1}{2}}z_{1-(\alpha/2)}^3 + o(n^{-\frac{1}{2}}). \quad (5.9)$$

Comparing expansions (5.8) and (5.9) we see that the bootstrap-Studentized interval has better asymptotic coverage accuracy than the Normal quantile interval if and only if

$$z_{1-(\alpha/2)}^2 > ([\pi \max\{p^{-1} - 1, (p^{-1} - 1)^{-1}\}]^{\frac{1}{2}} - 1)^{-1}(1 + 2\sqrt{2}).$$

In particular, for nominal 95% intervals ($\alpha = 0.05$), the bootstrap-Studentized interval has smaller asymptotic coverage error for p outside the interval (0.44,0.55), and for nominal 90% intervals ($\alpha = 0.10$), the bootstrap-Studentized interval is better for p outside the interval (0.35,0.65).

The results of a simulation study comparing bootstrap-Studentized and SBG intervals are given in Table 5.1. In that example, the underlying population was folded Normal $|N(0, 1)|$, $B = 299$ resamples were drawn in each resampling operation and each entry was based on 1000 samples for each of sample sizes $n = 11$, 25, and 99, representing small, moderate and large samples. All calculations were carried out on a VAX 8700 computer.

Table 5.1. Coverage of bootstrap-Studentized and SBG-Studentized intervals. The former is listed as "Boot", the latter as $\gamma = \frac{1}{2}, \frac{2}{3}$ or $\frac{4}{5}$, indicating the value of γ in the formula $m = \min(p, 1 - p)n^\gamma$. Population was folded Normal.

p	type	90% intervals			95% intervals		
		$n = 11$	$n = 25$	$n = 99$	$n = 11$	$n = 25$	$n = 99$
0.05	Boot	.838	.895	.880	.872	.931	.926
	$\gamma = \frac{1}{2}$	.796	.769	.741	.817	.814	.786
	$\gamma = \frac{2}{3}$	.796	.769	.802	.817	.814	.857
	$\gamma = \frac{4}{5}$	.796	.769	.802	.817	.814	.857
0.1	Boot	.863	.903	.881	.910	.942	.931
	$\gamma = \frac{1}{2}$	.723	.754	.736	.782	.797	.782
	$\gamma = \frac{2}{3}$	.723	.754	.836	.782	.797	.886
	$\gamma = \frac{4}{5}$	.723	.842	.845	.782	.894	.892
0.25	Boot	.884	.895	.874	.931	.944	.931
	$\gamma = \frac{1}{2}$	.753	.823	.817	.803	.872	.870
	$\gamma = \frac{2}{3}$	.842	.857	.849	.880	.907	.909
	$\gamma = \frac{4}{5}$	.842	.875	.857	.880	.923	.921
0.5	Boot	.877	.894	.888	.925	.940	.934
	$\gamma = \frac{1}{2}$	.825	.850	.857	.875	.902	.906
	$\gamma = \frac{2}{3}$	.871	.889	.883	.918	.942	.933
	$\gamma = \frac{4}{5}$	.901	.909	.898	.945	.963	.942

The major feature to emerge from Table 5.1 is that the bootstrap method performs comparatively well over the range of n and p values. In particular, it performs significantly better than its SBG competitors when both n and p are small. This is despite the fact that in small samples, the bootstrap-Studentized intervals undercover badly, and suggests that large samples are needed for differences between coverage errors of orders $n^{-2/5}$, $n^{-1/2}$, and $n^{-2/3}$ to become apparent in simulations. The results of our study extend those of McKean and Schrader (1984) and Sheather and McKean (1987) to quantiles other than the median and suggest that for small samples and extreme quantiles, the bootstrap-Studentized approach is a practical alternative to SBG methods.

From a theoretical viewpoint, a fascinating consequence of Theorem 5.3 is that the Edgeworth expansion of coverage accuracy has first term, of order $n^{-1/2}$, proportional to a polynomial containing both odd and even powers of the argument. In more standard problems, such as those discussed in Chapters 2, 3 and 4, the polynomial of order $n^{-1/2}$ is even, and so its influence cancels from two-sided confidence intervals, which therefore have coverage error of order n^{-1} . The fact that the leading polynomial is not even in the case of the bootstrap-Studentized quantile is the reason that the coverage accuracy of both one- and two-sided confidence intervals is of order $n^{-1/2}$.

The curious behaviour of our Edgeworth expansion may be explained as follows. The Edgeworth expansion of a Studentized quantile consists of a "main" series of terms decreasing in powers of $n^{-1/2}$, arising from the numerator in the Studentized ratio, multiplied by a "secondary" series arising from the denominator. As we have shown in Section 3, the bootstrap variance estimate has relative error of precise order $n^{-1/4}$, so we might expect that the secondary series decreases in powers of $n^{-1/4}$. The proof of Theorem 5.3 shows that there is indeed a term of nominal order $n^{-1/4}$ in our expansions. However, its size is proportional to a coefficient of correlation between two asymptotically Normal random variables which are asymptotically independent, with the result that its precise order is $n^{-1/2}$, not $n^{-1/4}$. For both main and secondary series, the j 'th term is even or odd according as j is odd or even, respectively. The term of order $n^{-1/2}$ in the combined series includes the first, even term of the main series and the second, odd term of the secondary series.

In principle, it is possible to improve on the coverage error in expansions (5.7) and (5.8) by a one-term Edgeworth correction. In the case of two-sided intervals, such a correction is explicit, not depending on the value of f at ξ_p .

There, replacing $z_{1-(\alpha/2)}$ by $y_{1-(\alpha/2)} \equiv z_{1-(\alpha/2)} - n^{-\frac{1}{2}}Q_1(z_{1-(\alpha/2)})$ in I_2 , the interval $I_4 \equiv [X_{nr} - \hat{\sigma}y_{1-(\alpha/2)}, X_{nr} + \hat{\sigma}y_{1-(\alpha/2)}]$ has coverage, by Theorem 5.3,

$$P(\xi_p \in I_4) = 1 - \alpha + O(n^{-3/4}).$$

In practice, however, we have found that such corrections are not successful in small to moderate samples, due both to the polynomial behaviour of coefficients in Edgeworth expansions and the large sample sizes required for such improvements in coverage error to become apparent in simulations. It is not as easy to see how to apply a one-term Edgeworth correction in the one-sided case, where it is first necessary to estimate the values of f and f' at ξ_p . In theory, replacing f and f' by consistent estimators in the correction allows coverage error to be reduced, but in practice the results are disappointing.

An immediate corollary of Theorem 5.3 is that a Berry-Esseen theorem holds:

$$\sup_{-\infty < z < \infty} |P\{(X_{nr} - \xi_p)/\hat{\sigma} \leq z\} - \Phi(z)| = O(n^{-\frac{1}{2}}),$$

as $n \rightarrow \infty$. This result is available under somewhat weaker assumptions than those stated prior to Theorem 5.3.

5. ON USING THE JACKKNIFE TO ESTIMATE QUANTILE VARIANCE

As we have seen from Sections 2, 3 and 4, the bootstrap performs relatively poorly in the quantile problem compared with its behaviour in simpler problems. We might therefore consider the use of other techniques, such as the jackknife, which is a major competitor of the bootstrap in simple problems, in the case of quantiles. The problem of estimating sample quantile variance has been viewed as a kind of "smoking gun" against the jackknife technique, mainly because of the jackknife's failure to provide a consistent variance estimate in the case of the median; see Efron (1979a), Ghosh, Parr, Singh and Babu (1985), and Sheather (1987). It might be thought, however, that a Studentized ratio using the jackknife estimate of sample quantile variance has a limiting Normal distribution, at least for certain quantiles, such as the median (see for example, Miller (1974), p.8). In this section, we show that, in fact, the limiting distribution of the jackknife-Studentized quantile is markedly non-normal, having infinite mean. This is in stark contrast to more conventional problems, such as that of Studentizing the

sample mean, where the jackknife-Studentized mean has an exact Student's t distribution with $(n - 1)$ degrees of freedom, where n is sample size.

Using the notation introduced in Section 1, and writing $s \equiv [(n - 1)p] + 1$, where $[\cdot]$ is the integer part function, the delete-one jackknife estimate of ξ_p is $\hat{\xi}_p \equiv sn^{-1}X_{n,s+1} + (1 - sn^{-1})X_{n,s}$, with variance estimate

$$\hat{\sigma}_J^2 \equiv (n - 1)(n - s)sn^{-2}(X_{n,s+1} - X_{n,s})^2; \quad (5.10)$$

see Miller (1974). Theorem 5.4 relates the jackknife variance estimate and the true variance.

THEOREM 5.4. Assume $E|X|^\eta < \infty$ for some $\eta > 0$, $f \equiv F'$ exists and is continuous at ξ_p , and $f(\xi_p) > 0$. Then

$$\sigma^2 = n^{-1}p(1 - p)f(\xi_p)^{-2} + o(n^{-1}), \quad (5.11a)$$

and

$$\hat{\sigma}_J^2 \sigma^{-2} = W + o_p(1), \quad (5.11b)$$

as $n \rightarrow \infty$, where W has a Weibull distribution with parameters 1 and $\frac{1}{2}$.

Theorem 5.4 is proved in Subsection 6.4. An immediate consequence of Theorem 5.4 is that the jackknife variance estimate fails to be consistent. In fact, $\hat{\sigma}_J^2$ is asymptotically unbiased for exactly twice the true variance.

The following theorem, proved in Subsection 6.5, gives the asymptotic distribution of the Studentized ratio $(X_{nr} - \xi_p)/\hat{\sigma}_J$.

THEOREM 5.5. Under the regularity conditions assumed in Theorem 5.4,

$$\sup_{-\infty < z < \infty} |\mathbb{P}\{(X_{nr} - \xi_p)/\hat{\sigma}_J \leq z\} - G(z)| = o(1), \quad (5.12)$$

as $n \rightarrow \infty$, where $G(z) \equiv \frac{1}{2} + \text{sgn}(z)\Phi(-|z|^{-1})\exp\{(2z^2)^{-1}\}$ for $z \neq 0$ and $G(0) = \frac{1}{2}$.

The distribution with distribution function G is that of the ratio of a Standard Normal random variable to an independent, unit-mean, negative exponential random variable. A little calculus shows that the t 'th absolute moment of G is $2^{t/2}\pi^{-\frac{1}{2}}\Gamma(1 - t)\Gamma\{(1 + t)/2\}$, for $0 < t < 1$, but infinite for $t \geq 1$. The asymptotic density of $(X_{nr} - \xi_p)/\hat{\sigma}_J$ is $g(z) \equiv z^{-2}(2\pi)^{-\frac{1}{2}} - \text{sgn}(z)\Phi(-|z|^{-1})\exp\{(2z^2)^{-1}\}$ for $z \neq 0$ with $g(0) = \phi(0)$.

Since the Studentized ratio $(X_{nr} - \xi_p)/\hat{\sigma}_J$ is asymptotically pivotal, the quantiles of G could be used to construct asymptotically exact confidence intervals for ξ_p . However, these intervals would be very long in general, as G is heavy tailed. To illustrate this, if $z_{1-\alpha}$ denotes the solution of $\Phi(z_{1-\alpha}) = 1 - \alpha$, and $y_{1-\alpha}$ the solution of $G(y_{1-\alpha}) = 1 - \alpha$, then, to three decimal places, $y_{0.9} = 7.344$, $y_{0.95} = 15.327$ and $y_{0.975} = 31.287$ compared with $z_{0.9} = 1.28$, $z_{0.95} = 1.645$ and $z_{0.975} = 1.96$. Indeed, if one were to wrongly assume an asymptotic Standard Normal distribution for the jackknife-Studentized quantile in constructing a 95% confidence interval for ξ_p , the coverage error would asymptotically equal 0.123 in the one-sided case and 0.255 in the two-sided case.

The reason for the failure of the jackknife in the quantile problem is that sample quantiles are insufficiently smooth functionals of the empiric distribution function. Parr (1985) discusses smoothness conditions under which the jackknife works reasonably. See also Sen (1988). A different approach to the question of using the jackknife in constructing confidence intervals for quantiles is presented by Kaigh (1983). Kaigh avoids the problem of inconsistency in the variance estimate by using a U-statistic, whose jackknife variance estimator is consistent, to estimate the quantile.

6. PROOFS

Throughout all proofs, symbols $C, C_1, C_2, \dots$ denote generic positive constants.

6.1. Proof of Theorem 5.1. The probability integral transform $u = F(x)$ maps the continuous order statistic X_{nr} into the r 'th order statistic U_{nr} of a sample of size n from a Uniform distribution on $(0, 1)$. If we invert the relation $U_{nr} = F(X_{nr})$, writing $X_{nr} = F^{-1}(U_{nr})$, and expand $F^{-1}(U_{nr})$ in a Taylor series about $q \equiv E(U_{nr}) = (n+1)^{-1}r$, we obtain

$$X_{nr} = F^{-1}(q) + (U_{nr} - q)[f\{F^{-1}(q)\}]^{-1} + R_0, \quad (5.13)$$

where

$$R_0 \equiv (U_{nr} - q) \int_0^1 \left\{ (f[F^{-1}\{q + t(U_{nr} - q)\}])^{-1} - [f\{F^{-1}(q)\}]^{-1} \right\} dt.$$

Recall that for some $\epsilon, \zeta > 0$, $|f(x) - f(y)| \leq C|x - y|^\tau$ whenever $x, y \in I(\zeta) \equiv (\xi_p - \zeta, \xi_p + \zeta)$, and observe that $q = (n+1)^{-1}r = p + O(n^{-1})$. Therefore, for

sufficiently large n , $F^{-1}(q) \in I(\frac{1}{2}\zeta)$ and so,

$$|R_0| \leq |U_{nr} - q| \int_0^1 C_1 |t(U_{nr} - q)|^\tau dt = C_2 |U_{nr} - q|^{1+\tau}. \quad (5.14)$$

Sen (1959) shows that if $E|X|^\eta < \infty$ for some $\eta > 0$, then $E(X_{nr}^2)$ exists for $r_0 \leq r \leq n - r_0 + 1$, where $r_0\eta = 2$. Thus, for n sufficiently large, $\sigma^2 = \text{var}(X_{nr})$ exists. David (1981), p.36, shows that $\text{var}(U_{nr}) = (n+2)^{-1}q(1-q) = n^{-1}p(1-p) + O(n^{-2})$. Hence, noting (5.13) and the fact that $F^{-1}(q) = \xi_p + O(n^{-1})$, we have

$$\begin{aligned} \text{var}(X_{nr}) &= n^{-1}p(1-p)f(\xi_p)^{-2} + 2\text{cov}\{(U_{n,r} - q)f(\xi_p)^{-1}, R_0\} \\ &\quad + \text{var}(R_0) + O(n^{-2}) \\ &= n^{-1}p(1-p)f(\xi_p)^{-2} + O(n^{-1-(\tau/2)}), \end{aligned}$$

the last line following on observing (5.14) and the fact that $E(U_{nr} - q)^{2+\tau} = O(n^{-1-(\tau/2)})$.

6.2. Proof of Theorem 5.2. The proof is divided into seven steps.

Step (i): Preliminaries. We begin by simplifying the weights w_j . Put $\gamma \equiv (j-1)/n$, $\beta \equiv \{n\gamma(1-\gamma)\}^{\frac{1}{2}}$ and $z \equiv (j-r-1)/\beta$, each being functions of j . Interpret $\beta^{-1}\phi(z)$ as zero if $j = 1$.

LEMMA 5.6. *There exist positive constants C_1, C_2 such that*

$$w_j = \beta^{-1}\phi(z) + r_j,$$

where $|r_j| \leq C_1 n^{-1} \exp\{-C_2 n^{-1}(j-r)^2\}$, for $1 \leq j \leq n$.

Proof of Lemma 5.6. Note that $w_j = I_{\gamma+n^{-1}}(r, n-r+1) - I_\gamma(r, n-r+1)$, where

$$\begin{aligned} I_y(a, b) &\equiv B(a, b)^{-1} \int_0^y x^{a-1}(1-x)^{b-1} dx \\ &= \sum_{j=a}^n \binom{n}{j} y^j (1-y)^{n-j}, \quad 0 < y < 1, \end{aligned}$$

is the incomplete beta function, the second equality provided a, b are positive integers and $b = n - a + 1$. Let N be a binomial $Bi(n, 1-\nu)$ random variable. Then

$$I_\nu(r, n-r+1) = P(N \leq n-r) = P(Z \leq z),$$

where $Z = Z(\nu) \equiv \{N - n(1 - \nu)\}\{n\nu(1 - \nu)\}^{-\frac{1}{2}}$ and $z = z(\nu) \equiv \{n - r - n(1 - \nu)\}\{n\nu(1 - \nu)\}^{-\frac{1}{2}}$. If $n\nu$ is an integer then, using an Edgeworth expansion for the binomial distribution function (see for example, Gnedenko and Kolmogorov (1968), p.213), we have

$$\begin{aligned} P(Z \leq z) &= \Phi(z) + n^{-\frac{1}{2}}\left\{\frac{1}{6}u(1 - z^2) + \frac{1}{2}v^{-1}\right\}\phi(z) \\ &\quad + n^{-1}z\left\{\frac{1}{72}u^2(10z^2 - z^4 - 15)\right. \\ &\quad \left.+ \frac{1}{24}\kappa(3 - z^2) - \frac{1}{12}v^{-2} + \frac{1}{12}v^{-1}u(z^2 - 3)\right\}\phi(z) \\ &\quad + \sum_{j=3}^4 n^{-j/2}s_j(z)\phi(z) + O(n^{-5/2}), \end{aligned} \quad (5.15)$$

uniformly in $\epsilon < \nu < 1 - \epsilon$ for any $\epsilon \in (0, \frac{1}{2})$, where s_3, s_4 are polynomials, $v = v(\nu) = \{\nu(1 - \nu)\}^{\frac{1}{2}}$, $u = u(\nu) \equiv v^{-1}(1 - 2\nu)$ and $\kappa = \kappa(\nu) \equiv v^{-2} - 6$. Expansion (5.15) has been given to one more term than we shall need to prove Lemma 5.6, but we will need the extra term in the proof of Theorem 5.3. Notice that $v(\nu + n^{-1})^{-1} = \{\nu(1 - \nu) + n^{-1}(1 - 2\nu) - n^{-2}\}^{-\frac{1}{2}} = v^{-1} - n^{-1}\frac{1}{2}v^{-2}u + O(n^{-2})$, $u(\nu + n^{-1}) = u + O(n^{-1})$, $\kappa(\nu + n^{-1}) = \kappa + O(n^{-1})$, and

$$\begin{aligned} z(\nu + n^{-1}) &= (n\nu - r + 1)n^{-\frac{1}{2}}\{v^{-1} - n^{-1}\frac{1}{2}v^{-2}u + O(n^{-2})\} \\ &= z + n^{-\frac{1}{2}}v^{-1} - n^{-1}\frac{1}{2}v^{-1}uz + O(n^{-3/2}), \end{aligned}$$

where v, u, κ and z denote those functions evaluated at ν . Therefore,

$$\begin{aligned} \Phi\{z(\nu + n^{-1})\} &= \Phi(z) + n^{-\frac{1}{2}}v^{-1}\phi(z) \\ &\quad - n^{-1}\frac{1}{2}v^{-1}z(u + v^{-1})\phi(z) + O(n^{-3/2}), \end{aligned} \quad (5.16)$$

$$\begin{aligned} n^{-\frac{1}{2}}\left[\frac{1}{6}u(\nu + n^{-1})\{1 - z^2(\nu + n^{-1})\} + \frac{1}{2}v(\nu + n^{-1})^{-1}\right]\phi\{z(\nu + n^{-1})\} \\ = n^{-\frac{1}{2}}\left[\frac{1}{6}u\{1 - z^2 - 2n^{-\frac{1}{2}}v^{-1}z\} + \frac{1}{2}v^{-1}\right]\{\phi(z) - n^{-\frac{1}{2}}v^{-1}z\phi(z)\} \\ \quad + O(n^{-3/2}) \\ = n^{-\frac{1}{2}}\left\{\frac{1}{6}u(1 - z^2) + \frac{1}{2}v^{-1}\right\}\phi(z) \\ \quad - n^{-1}v^{-1}z\left\{\frac{1}{6}u(3 - z^2) + \frac{1}{2}v^{-1}\right\}\phi(z) + O(n^{-3/2}), \end{aligned} \quad (5.17)$$

and

$$\begin{aligned} n^{-1}z(\nu + n^{-1})\left[\frac{1}{72}u^2(\nu + n^{-1})\{10z^2(\nu + n^{-1}) - z^4(\nu + n^{-1}) - 15\}\right. \\ \quad \left.+ \frac{1}{24}\kappa(\nu + n^{-1})\{3 - z^2(\nu + n^{-1})\} - \frac{1}{12}v(\nu + n^{-1})^{-2}\right. \\ \quad \left.+ \frac{1}{12}v(\nu + n^{-1})^{-1}u(\nu + n^{-1})\{z^2(\nu + n^{-1}) - 3\}\right]\phi\{z(\nu + n^{-1})\} \\ = n^{-1}z\left\{\frac{1}{72}u^2(10z^2 - z^4 - 15) + \frac{1}{24}\kappa(3 - z^2) - \frac{1}{12}v^{-2}\right. \\ \quad \left.+ \frac{1}{12}v^{-1}u(z^2 - 3)\right\}\phi(z) + O(n^{-3/2}). \end{aligned} \quad (5.18)$$

Adding formulae (5.16)–(5.18) and noting expansion (5.15), we conclude that

$$\begin{aligned} w_j &= P\{Z(\gamma + n^{-1}) \leq z(\gamma + n^{-1})\} - P\{Z(\gamma) \leq z(\gamma)\} \\ &= n^{-\frac{1}{2}}v^{-1}\phi(z) + n^{-1}v^{-1}z\frac{1}{6}\{u(z^2 - 6) - 6v^{-1}\}\phi(z) + O(n^{-3/2}), \end{aligned} \quad (5.19)$$

$$= n^{-\frac{1}{2}}v^{-1}\phi(z) + O(n^{-1}), \quad (5.20)$$

where u , v , and z denote those functions evaluated at $\nu = \gamma$. These expansions hold uniformly in $\epsilon < \gamma < 1 - \epsilon$; that is, in $\epsilon n < j < n - \epsilon n$ for any $\epsilon \in (0, \frac{1}{2})$.

The same argument may be used to generalize (5.19) to

$$w_j = \sum_{i=1}^3 n^{-i/2} t_i(z) \phi(z) + O(n^{-2}), \quad (5.21)$$

for polynomials t_1 , t_2 , and t_3 . We deduce from Bernstein's inequality (Hoeffding (1963), p.17) that

$$\begin{aligned} w_j &\leq \min[P\{Z(\gamma + n^{-1}) \leq z(\gamma + n^{-1})\}, 1 - P\{Z(\gamma) \leq z(\gamma)\}] \\ &\leq \exp\{-2cn^{-1}(j - r)^2\}, \end{aligned}$$

for a positive constant $c < \frac{1}{2}$ and for all j . Also, $(2\pi)^{\frac{1}{2}}\phi(z) \leq \exp\{-n^{-1}(j - r - 1)^2\}$, and so both w_j and $|\sum_{1 \leq i \leq 3} n^{-i/2} t_i(z) \phi(z)|$ are bounded by the quantity $C_1 \exp\{-2cn^{-1}(j - r)^2\}$ uniformly in all j . Therefore,

$$\left| w_j - \sum_{i=1}^3 n^{-i/2} t_i(z) \phi(z) \right| \leq C_2 \exp\{-2cn^{-1}(j - r)^2\},$$

uniformly in j . If $\epsilon < \frac{1}{2} \min(p, 1 - p)$ and either $j \leq \epsilon n$ or $j \geq n - \epsilon n$, then this bound is smaller, $C_3 n^{-2}$. Hence, by (5.21), and for all j ,

$$\begin{aligned} \left| w_j - \sum_{i=1}^3 n^{-i/2} t_i(z) \phi(z) \right| &\leq C_4 \min[n^{-2}, \exp\{-2cn^{-1}(j - r)^2\}] \\ &\leq C_4 n^{-1} \exp\{-cn^{-1}(j - r)^2\}. \end{aligned} \quad (5.22)$$

The bound on the right-hand side also applies to $|n^{-i/2} t_i(z) \phi(z)|$ for $i = 2, 3$, and all j . Lemma 5.6 follows from (5.20) and (5.22) and observing that $\beta = n^{\frac{1}{2}}v$, the function v being evaluated at γ .

Note, for future reference, that we may generalize Lemma 5.6 to: there exist constants C_5 and C_6 such that

$$\begin{aligned} w_j &= \beta^{-1}\phi(z) + \beta^{-2}\frac{1}{6}z\{(1 - 2\gamma)(z^2 - 6) - 6\}\phi(z) + r_{1j} \\ &= n^{-\frac{1}{2}}\{\gamma(1 - \gamma)\}^{-\frac{1}{2}}\phi(z) + n^{-1}\{\gamma(1 - \gamma)\}^{-1}\frac{1}{6}z\{(1 - 2\gamma)(z^2 - 6) - 6\} \\ &\quad + r_{ij}, \end{aligned} \quad (5.23)$$

where $|r_{1j}| \leq C_5 n^{-3/2} \exp\{-C_6 n^{-1}(j-r)^2\}$ uniformly in $1 \leq j \leq n$ and $n \geq 1$. This follows by using the same argument as above to generalize (5.19) to

$$w_j = \sum_{i=1}^5 n^{-i/2} t_i(z) \phi(z) + O(n^{-3}), \quad (5.24)$$

and using Bernstein's inequality and (5.24) to obtain the bound

$$\begin{aligned} \left| w_j - \sum_{i=1}^5 n^{-i/2} t_i(z) \phi(z) \right| &\leq C_7 \min[n^{-3}, \exp\{-2cn^{-1}(j-r)^2\}] \\ &\leq C_7 n^{-3/2} \exp\{-cn^{-1}(j-r)^2\}. \end{aligned} \quad (5.25)$$

Expression (5.23) follows on noting (5.19) and the fact that the bound on the right-hand side of (5.25) also applies to $|n^{-i/2} t_i(z) \phi(z)|$ for $i = 3, 4, 5$.

Put $H(x) \equiv F^{-1}(e^{-x})$, and let $Z_1, \dots, Z_n$ denote independent and identically distributed, unit-mean, negative exponential random variables. Define $s_j \equiv \text{sgn}(r-j)$, $m_{0j} \equiv \min(r, j)$, $m_{1j} \equiv \max(r, j) - 1$, $a \equiv H'(\sum_{r \leq k \leq n} k^{-1})$,

$$A_j \equiv \sum_{k=j}^n k^{-1} Z_k, \quad B_j \equiv \sum_{k=m_{0j}}^{m_{1j}} k^{-1} Z_k, \quad \text{and} \quad b_j \equiv E(B_j) = \sum_{k=m_{0j}}^{m_{1j}} k^{-1},$$

for $1 \leq j \leq n$, the latter two quantities being zero if $j = r$. Using Rényi's representation (David (1981), p.21), the sequence $\{X_{nj}, 1 \leq j \leq n\}$ has the same distribution as $\{H(A_j), 1 \leq j \leq n\}$, and in a slight abuse of notation, we shall write

$$X_{nj} = H\left(\sum_{k=j}^n k^{-1} Z_k\right), \quad 1 \leq j \leq n.$$

By Taylor expansion of $H(\sum_{j \leq k \leq n} k^{-1} Z_k)$ about A_r , we have

$$X_{nj} - X_{nr} = H\left(\sum_{k=j}^n k^{-1} Z_k\right) - H\left(\sum_{k=r}^n k^{-1} Z_k\right) = D_j + R_{1j},$$

where $D_j \equiv s_j a B_j$, $R_{1j} \equiv R_{2j} + R_{3j}$, $R_{2j} \equiv s_j B_j \{H'(A_r) - a\}$, and

$$R_{3j} \equiv s_j B_j \int_0^1 \{H'(A_r + ts_j B_j) - H'(A_r)\} dt.$$

Thus,

$$(X_{nj} - X_{nr})^2 = a^2 b_j^2 + 2a^2 b_j (B_j - b_j) + a^2 (B_j - b_j)^2 + 2D_j R_{1j} + R_{1j}^2. \quad (5.26)$$

Assuming $E|X|^\eta < \infty$, we have, using Markov's inequality,

$$\begin{aligned} P(|X_j| \leq n^{2/\eta}, \text{ all } j \leq n) &= \{1 - P(|X| > n^{2/\eta})\}^n \\ &\geq (1 - n^{-2} E|X|^\eta)^n = 1 + O(n^{-1}). \end{aligned}$$

Therefore, with probability tending to one as $n \rightarrow \infty$, $\max_{j \leq n} (X_{nj} - X_{nr})^2 \leq 4n^{4/\eta}$. Thus, by Lemma 5.6, for any $0 < \delta < \min(p, 1-p)$ and with probability tending to one,

$$\sum_{|j-r| > \delta n} (X_{nj} - X_{nr})^2 w_j \leq C_1 n^{4/\eta} \sum_{|j-r| > \delta n} n^{-\frac{1}{2}} \exp\{-C_2(\delta)n\} = O(n^{-\lambda}),$$

for all $\lambda > 0$. Hence writing $\sum'_j$ for summation over j with $|j-r| \leq \delta n$, we see that it suffices to prove Theorem 5.2 for $\hat{\sigma}^2(\delta) \equiv \sum'_j (X_{nj} - X_{nr})^2 w_j$ instead of $\hat{\sigma}^2$, where δ may be chosen arbitrarily small. Observe from (5.26) that

$$\hat{\sigma}^2(\delta) = S_1 + S_2 + T_1 + T_2 + T_3, \quad (5.27)$$

where $S_1 \equiv \sum'_j a^2 b_j^2 w_j$, $S_2 \equiv \sum'_j 2a^2 b_j (B_j - b_j) w_j$, $T_1 \equiv \sum'_j a^2 (B_j - b_j)^2 w_j$, $T_2 \equiv \sum'_j 2D_j R_{1j} w_j$, and $T_3 \equiv \sum'_j R_{1j}^2 w_j$. These terms are treated in steps (ii)–(vi) below.

Step(ii): S_1 . If f satisfies a Lipschitz condition of order $\tau = \frac{1}{2} + \epsilon$ in a neighbourhood of ξ_p , then $a \equiv H'(\sum_{r \leq k \leq n} k^{-1}) = H'\{\log(n/r)\} + O(n^{-\frac{1}{2}-\epsilon}) = -pf(\xi_p)^{-1} + o(n^{-1/4})$, the last identity following on noting that $H'(x) = -e^{-x} \times f\{F^{-1}(e^{-x})\}^{-1}$ and $r = n\{p + O(n^{-1})\}$. Observe that

$$\begin{aligned} \beta &= \{n\gamma(1-\gamma)\}^{\frac{1}{2}} = \{n^{-1}(j-1)(n-j+1)\}^{\frac{1}{2}} \\ &= [n^{-1}(r-1)(n-r+1)\{1 + O(n^{-1}|j-r|)\}]^{\frac{1}{2}} \\ &= \{np(1-p)\}^{\frac{1}{2}} + O(n^{-\frac{1}{2}}|j-r|). \end{aligned}$$

Then, using expansions for b_j and w_j , and approximating series by integrals, we obtain

$$\begin{aligned} \sum'_j b_j^2 w_j &= \sum'_j \{\log(j/r)\}^2 w_j + O\left\{\sum'_j n^{-1} \log(j/r) w_j\right\} \\ &= \sum'_j r^{-2} (j-r)^2 w_j + O\left(\sum'_j r^{-3} |j-r|^3 w_j + \sum'_j n^{-1} r^{-1} |j-r| w_j\right) \end{aligned}$$

$$\begin{aligned}
&= \sum_j' r^{-2} (j-r)^2 \beta^{-1} \phi\{(j-r)\beta^{-1}\} + O\left(\sum_j' r^{-3} |j-r|^3 \beta^{-1} \phi\{(j-r)\beta^{-1}\}\right. \\
&\quad + \sum_j' n^{-1} r^{-1} |j-r| \beta^{-1} \phi\{(j-r)\beta^{-1}\} \\
&\quad \left.+ \sum_j' r^{-2} (j-r)^2 C_1 n^{-1} \exp\{-C_2 n^{-1} (j-r)^2\}\right) \\
&= \sum_j' r^{-2} (j-r)^2 \{np(1-p)\}^{-\frac{1}{2}} \phi[(j-r)\{np(1-p)\}^{-\frac{1}{2}}] \\
&\quad + O\left(\sum_j' r^{-3} |j-r|^3 \beta^{-1} \phi\{(j-r)\beta^{-1}\}\right. \\
&\quad + \sum_j' n^{-1} r^{-1} |j-r| \beta^{-1} \phi\{(j-r)\beta^{-1}\} \\
&\quad + \sum_j' r^{-2} (j-r)^2 n^{-1} \exp\{-C_2 n^{-1} (j-r)^2\} \\
&\quad + \sum_j' r^{-2} |j-r|^3 n^{-3/2} \phi[(j-r)\{np(1-p)\}^{-\frac{1}{2}}] \\
&\quad \left.+ \sum_j' r^{-2} (j-r)^4 n^{-3/2} \{np(1-p)\}^{-1} \phi[(j-r)\{np(1-p)\}^{-\frac{1}{2}}]\right) \\
&= \int_{r-\delta n}^{r+\delta n} r^{-2} (x-r)^2 \{np(1-p)\}^{-\frac{1}{2}} \phi[(x-r)\{np(1-p)\}^{-\frac{1}{2}}] dx \\
&\quad + O\left(\int_{r-\delta n}^{r+\delta n} r^{-3} |x-r|^3 \beta^{-1} \phi\{(x-r)\beta^{-1}\} dx\right. \\
&\quad + \int_{r-\delta n}^{r+\delta n} n^{-1} r^{-1} |x-r| \beta^{-1} \phi\{(x-r)\beta^{-1}\} dx \\
&\quad + \int_{r-\delta n}^{r+\delta n} r^{-2} (x-r)^2 n^{-1} \exp\{-C_2 n^{-1} (x-r)^2\} dx \\
&\quad + \int_{r-\delta n}^{r+\delta n} r^{-2} |x-r|^3 n^{-3/2} \phi[(x-r)\{np(1-p)\}^{-\frac{1}{2}}] dx \\
&\quad \left.+ \int_{r-\delta n}^{r+\delta n} r^{-2} (x-r)^4 n^{-5/2} \phi[(x-r)\{np(1-p)\}^{-\frac{1}{2}}] dx\right) \\
&= r^{-2} \{np(1-p)\} \int_{-\delta n^{\frac{1}{2}} \{p(1-p)\}^{-\frac{1}{2}}}^{\delta n^{\frac{1}{2}} \{p(1-p)\}^{-\frac{1}{2}}} y^2 \phi(y) dy \\
&\quad + O\left\{r^{-3} \beta^3 \int_{-\delta n/\beta}^{\delta n/\beta} |y|^3 \phi(y) dy + n^{-1} r^{-1} \beta \int_{-\delta n/\beta}^{\delta n/\beta} |y| \phi(y) dy\right. \\
&\quad + r^{-2} n^{\frac{1}{2}} \int_{-\delta n^{\frac{1}{2}}}^{\delta n^{\frac{1}{2}}} y^2 \exp(-C_2 y^2) dy + r^{-2} n^{\frac{1}{2}} \int_{-\delta n^{\frac{1}{2}} \{p(1-p)\}^{-\frac{1}{2}}}^{\delta n^{\frac{1}{2}} \{p(1-p)\}^{-\frac{1}{2}}} |y|^3 \phi(y) dy \\
&\quad \left.+ r^{-2} \int_{-\delta n^{\frac{1}{2}} \{p(1-p)\}^{-\frac{1}{2}}}^{\delta n^{\frac{1}{2}} \{p(1-p)\}^{-\frac{1}{2}}} y^4 \phi(y) dy\right\}
\end{aligned}$$

$$\begin{aligned}
&= r^{-2} \{np(1-p)\} \int_{-\infty}^{\infty} y^2 \phi(y) dy + O \left[r^{-3} \beta^3 \int_{-\infty}^{\infty} |y|^3 \phi(y) dy \right. \\
&\quad \left. + n^{-1} r^{-1} \beta \int_{-\infty}^{\infty} |y| \phi(y) dy + r^{-2} n^{\frac{1}{2}} \left\{ \int_{-\infty}^{\infty} y^2 \exp(-C_2 y^2) dy \right. \right. \\
&\quad \left. \left. + \int_{-\infty}^{\infty} |y|^3 \phi(y) dy \right\} + r^{-2} \int_{-\infty}^{\infty} y^4 \phi(y) dy \right] \\
&= n^{-1} p^{-1} (1-p) + O(n^{-3/2}) \\
&= n^{-1} p^{-1} (1-p) + o(n^{-5/4}).
\end{aligned}$$

Therefore $S_1 = n^{-1} p(1-p) f(\xi_p)^{-2} + o(n^{-5/4})$.

Step (iii): S_2 . Observe that S_2 is a weighted sum of independent, centred exponential random variables. By Lyapounov's Theorem (Chung (1974), p.200), we see that S_2 is asymptotically Normal $N(0, s^2)$, where $s^2 \equiv E(S_2^2)$ and

$$\begin{aligned}
(4a^4)^{-1} s^2 &= E \left\{ \sum_j' b_j w_j (B_j - b_j) \right\}^2 \\
&= E \left\{ \sum_j' b_j w_j \sum_{k=m_{0j}}^{m_{1j}} k^{-1} (Z_k - 1) \right\}^2 \\
&= E \left\{ \sum_{j=r-\delta n}^{r-1} b_j w_j \sum_{k=j}^{r-1} k^{-1} (Z_k - 1) + \sum_{j=r+1}^{r+\delta n} b_j w_j \sum_{k=r}^{j-1} k^{-1} (Z_k - 1) \right\}^2 \\
&= E \left\{ \sum_{k=r-\delta n}^{r-1} k^{-1} (Z_k - 1) \sum_{j=r-\delta n}^k b_j w_j + \sum_{k=r}^{r+\delta n-1} k^{-1} (Z_k - 1) \sum_{j=k+1}^{r+\delta n} b_j w_j \right\}^2 \\
&= \sum_{k=r-\delta n}^{r-1} k^{-2} \left(\sum_{j=r-\delta n}^k b_j w_j \right)^2 + \sum_{k=r}^{r+\delta n-1} k^{-2} \left(\sum_{j=k+1}^{r+\delta n} b_j w_j \right)^2, \quad (5.28)
\end{aligned}$$

since the Z_k 's are independent and $\text{var}(Z_k) = 1$. Approximating sums by integrals in (5.28), and using the facts that $b_j \sim \log(j/r) \sim r^{-1}(j-r)$, and $w_j \sim \beta^{-1} \phi(z)$, we obtain

$$\begin{aligned}
(4a^4)^{-1} s^2 &\sim \int_{r-\delta n}^r x^{-2} \left[\int_{-\infty}^x r^{-1} (y-r) \beta^{-1} \phi\{(y-r)\beta^{-1}\} dy \right]^2 dx \\
&\quad + \int_r^{r+\delta n} x^{-2} \left[\int_x^{\infty} r^{-1} (y-r) \beta^{-1} \phi\{(y-r)\beta^{-1}\} dy \right]^2 dx \\
&= \int_{r-\delta n}^r x^{-2} r^{-2} \beta^2 \left\{ \int_{-\infty}^{(x-r)/\beta} y \phi(y) dy \right\}^2 dx
\end{aligned}$$

$$\begin{aligned}
& + \int_r^{r+\delta n} x^{-2} r^{-2} \beta^2 \left\{ \int_{-\infty}^{(x-r)/\beta} y \phi(y) dy \right\}^2 dx \\
& = \beta^3 r^{-2} \int_{-\delta n/\beta}^{\delta n/\beta} (r + \beta x)^{-2} \left\{ \int_{-\infty}^x y \phi(y) dy \right\}^2 dx \\
& \sim \beta^3 r^{-4} \int_{-\infty}^{\infty} \left\{ \int_{-\infty}^x y \phi(y) dy \right\}^2 dx \\
& = \beta^3 r^{-4} \int_{-\infty}^{\infty} \{\phi(x)\}^2 dx = \beta^3 / (2r^4 \pi^{\frac{1}{2}}).
\end{aligned}$$

Therefore, $s^2 \sim n^{-5/2} 2\pi^{-\frac{1}{2}} \{p(1-p)\}^{-3/2} f(\xi_p)^{-4}$.

Step (iv): T_1 . Observe that $T_1 \geq 0$ and noting, as for Lemma 5.6, that there exist positive constants C_1, C_2 such that $w_j \leq C_1 n^{-\frac{1}{2}} \exp\{-C_2 n^{-1}(j-r)^2\}$,

$$\begin{aligned}
E(T_1) & = a^2 \sum_j' E(B_j - b_j)^2 w_j = a^2 \sum_j' \left(\sum_{k=m_{0j}}^{m_{1j}} k^{-2} w_j \right) \\
& \leq C_1 \sum_j' \left(\sum_{k=m_{0j}}^{m_{1j}} k^{-2} \right) n^{-\frac{1}{2}} \exp\{-C_2 n^{-1}(j-r)^2\} \\
& \leq C_3(\delta) n^{-5/2} \sum_j' |j-r| \exp\{-C_2 n^{-1}(j-r)^2\} \\
& \leq C_4(\delta) n^{-5/2} \int_{-\infty}^{\infty} |x| \exp(-C_2 n^{-1} x^2) dx = O(n^{-3/2}).
\end{aligned}$$

Hence $T_1 = o_p(n^{-5/4})$.

Step (v): T_2 . For some $\epsilon, \zeta > 0$, $|f(x) - f(y)| \leq C|x-y|^{\frac{1}{2}+\epsilon}$ whenever $x, y \in I(\zeta) \equiv (\xi_p - \zeta, \xi_p + \zeta)$. If δ is chosen sufficiently small, then with probability tending to one as $n \rightarrow \infty$, $X_{nj} = H(A_j) \in I(\frac{1}{2}\zeta)$ whenever $|j-r| \leq \delta n$. In this circumstance, since $H'(x) = -e^{-x}/f\{H(x)\}$, we have $|R_{2j}| \leq C|B_j||A_r - EA_r|^{\frac{1}{2}+\epsilon}$ and $|R_{3j}| \leq C|B_j|^{(3/2)+\epsilon}$ for $|j-r| \leq \delta n$. Therefore, with high probability,

$$|T_2| \leq CT_2' \equiv C \sum_j' (B_j^2 |A_r - EA_r|^{\frac{1}{2}+\epsilon} + |B_j|^{(5/2)+\epsilon}) w_j. \quad (5.29)$$

Observe that

$$\begin{aligned}
E|B_j|^t & = E \left| \sum_{k=m_{0j}}^{m_{1j}} k^{-1} Z_k \right|^t \leq E(Z_1^t) \sum_{k=m_{0j}}^{m_{1j}} k^{-t} \\
& \leq C(t) n^{-t} |j-r| \leq C(t) (n^{-1} |j-r|)^t,
\end{aligned} \quad (5.30)$$

for $t \geq 1$. Also, using Jensen's inequality, we have

$$E|A_r - EA_r|^{1+2\epsilon} = E\left|\sum_{k=r}^n k^{-1}(Z_k - 1)\right|^{1+2\epsilon} \leq \left\{E\left|\sum_{k=r}^n k^{-1}(Z_k - 1)\right|^{2+4\epsilon}\right\}^{\frac{1}{2}}. \quad (5.31)$$

Applying Rosenthal's inequality (Hall and Heyde (1980), Theorem 2.12, p.23) to the bound on the right-hand side of (5.31), we get

$$\begin{aligned} E|A_r - EA_r|^{1+2\epsilon} &\leq \left\{C_1\left(\sum_{k=r}^n k^{-2}\right)^{1+2\epsilon} + C_2\sum_{k=r}^n k^{-(2+4\epsilon)}\right\}^{\frac{1}{2}} \\ &\leq \{C_3n^{-(1+2\epsilon)} + C_4n^{-(1+4\epsilon)}\}^{\frac{1}{2}} \leq C_5n^{-\frac{1}{2}-\epsilon}. \end{aligned} \quad (5.32)$$

Therefore, using Cauchy-Schwarz, (5.30), and (5.32), we see from (5.29) that

$$\begin{aligned} E(T'_2) &\leq \sum_j' \{(EB_j^4 E|A_r - EA_r|^{1+2\epsilon})^{\frac{1}{2}} + (E|B_j|^{5+2\epsilon})^{\frac{1}{2}}\} w_j \\ &\leq C_1 \sum_j' \{n^{-2}(j-r)^2 n^{-1/4-\epsilon/2} + (n^{-1}|j-r|)^{(5/2)+\epsilon}\} \\ &\quad \times n^{-\frac{1}{2}} \exp\{-C_2 n^{-1}(j-r)^2\} \\ &\sim C_1 \left[\int_{r-\delta n}^{r+\delta n} n^{-11/4-\epsilon/2} (x-r)^2 \exp\{-C_2 n^{-1}(x-r)^2\} dx \right. \\ &\quad \left. + \int_{r-\delta n}^{r+\delta n} n^{-3-\epsilon} |x-r|^{(5/2)+\epsilon} \exp\{-C_2 n^{-1}(x-r)^2\} dx \right] \\ &\sim C_1 \left\{ n^{-5/4-\epsilon/2} \int_{-\infty}^{\infty} x^2 \exp(-C_2 x^2) dx \right. \\ &\quad \left. + n^{-5/4-\epsilon/2} \int_{-\infty}^{\infty} |x|^{(5/2)+\epsilon} \exp(-C_2 x^2) dx \right\} \\ &= O(n^{-5/4-\epsilon/2}). \end{aligned}$$

Therefore, $T_2 = o_p(n^{-5/4})$.

Step (vi): T_3 . Using the bounds obtained for $|R_{2j}|$ and $|R_{3j}|$ in Step (v), we have

$$\begin{aligned} |T_3| &= \left| \sum_j' R_{1j}^2 w_j \right| \\ &\leq \sum_j' (R_{2j}^2 + 2|R_{2j}||R_{3j}| + R_{3j}^2) w_j \\ &\leq C_1 \sum_j' (B_j^2 |A_r - EA_r|^{1+2\epsilon} + 2|B_j|^{(5/2)+\epsilon} |A_r - EA_r|^{\frac{1}{2}+\epsilon} \\ &\quad + |B_j|^{3+2\epsilon}) w_j. \end{aligned}$$

Using Cauchy-Schwarz and bounds (5.30)–(5.32) for $E|B_j|^t$, $E|A_r - EA_r|^{1+2\epsilon}$, and $E|A_r - EA_r|^{2+4\epsilon}$, respectively, it follows that

$$\begin{aligned}
 E|T_3| &\leq C_1 \sum_j' \{ (EB_j^4 E|A_r - EA_r|^{2+4\epsilon})^{\frac{1}{2}} + 2(E|B_j|^{5+2\epsilon} E|A_r - EA_r|^{1+2\epsilon})^{\frac{1}{2}} \\
 &\quad + (E|B_j|^{6+4\epsilon})^{\frac{1}{2}} \} w_j \\
 &\leq C_2 \sum_j' \{ n^{-2}(j-r)^2 n^{-\frac{1}{2}-\epsilon} + 2(n^{-1}|j-r|)^{(5/2)+\epsilon} n^{-1/4-\epsilon/2} \\
 &\quad + (n^{-1}|j-r|)^{3+2\epsilon} \} n^{-\frac{1}{2}} \exp\{-C_3 n^{-1}(j-r)^2\} \\
 &\sim C_2 \left\{ n^{-3/2-\epsilon} \int_{-\infty}^{\infty} x^2 \exp(-C_3 x^2) dx \right. \\
 &\quad + 2n^{-3/2-\epsilon} \int_{-\infty}^{\infty} |x|^{(5/2)+\epsilon} \exp(-C_3 x^2) dx \\
 &\quad \left. + n^{-3/2-\epsilon} \int_{-\infty}^{\infty} |x|^{3+2\epsilon} \exp(-C_3 x^2) dx \right\} \\
 &= O(n^{-3/2-\epsilon}),
 \end{aligned}$$

and so $T_3 = o_p(n^{-5/4})$.

Step (vii): Completion. Steps (ii) and (iii) show that $S_1 + S_2$ is asymptotically Normally distributed with mean $n^{-1}p(1-p)f(\xi_p)^{-2} + o(n^{-5/4})$ and variance $n^{-5/2}2\pi^{-\frac{1}{2}}\{p(1-p)\}^{3/2}f(\xi_p)^{-4} + o(n^{-5/2})$, and steps (iv)–(vi) demonstrate that $T_1 + T_2 + T_3 = o_p(n^{-5/4})$. Theorem 5.2 now follows from (5.27). Formula (5.6) follows on noting that $n^{\frac{1}{2}}\text{var}\{\hat{f}(\xi_p) - f(\xi_p)\} = n^{-\frac{1}{2}}p(1-p)^{\frac{1}{4}}\sigma^{-3}\text{var}(\hat{\sigma}^2 - \sigma^2)$.

6.3. Proof of Theorem 5.3. We shall use the so-called “delta method” to identify the Edgeworth expansion. Formal verification that the remainder is indeed of the stated order, given in more detail later, consists primarily of truncating the series in the definition of $\hat{\sigma}^2$ to a sum over values j satisfying $|j-r| \leq C_1 n^{\frac{1}{2}}$, for a large positive constant C_1 , and using (5.23) and the fact that $E|X|^\eta < \infty$ (so that $\max_{j \leq n} (X_{nj} - X_{nr})^2 \leq C_2 n^{C_3}$, some $C_2, C_3 > 0$) to show that the remainder can be made $O(n^{-C_4})$ with probability tending to one for any given $C_4 > 0$, by choosing C_1 large. We then proceed in the same way as in steps (ii)–(vi) of the previous proof to calculate the remainder of our truncated estimator. The argument below details the proof from this point on.

Put $H(x) \equiv F^{-1}(e^{-x})$, and let $\beta_1, \dots, \beta_n$ denote independent and identically distributed centred exponential random variables. Define m_{0j} , m_{1j} , and s_j

as in the proof of Theorem 5.2. Changing notation slightly from that proof, put

$$\mu \equiv \sum_{k=r}^n k^{-1}, \quad \mu_j \equiv s_j \sum_{k=m_{0j}}^{m_{1j}} k^{-1}, \quad \delta \equiv \sum_{k=r}^n k^{-1} \beta_k, \quad \delta_j \equiv \sum_{k=m_{0j}}^{m_{1j}} k^{-1} \beta_k,$$

for $1 \leq j \leq n$. For future reference, note that $H'(x) = -e^{-x} f\{F^{-1}(e^{-x})\}^{-1}$ and $H''(x) = -H'(x) - e^{-2x} f'\{F^{-1}(e^{-x})\} f\{F^{-1}(e^{-x})\}^{-3}$. Therefore, since $\mu = \log(n/r) + O(n^{-1}) = -\log p + O(n^{-1})$, it follows that

$$\begin{aligned} H(\mu) &= \xi_p + O(n^{-1}), & H'(\mu) &= -pf(\xi_p)^{-1} + O(n^{-1}), \\ H''(\mu) &= pf(\xi_p)^{-1} - p^2 f'(\xi_p) f(\xi_p)^{-3} + O(n^{-1}). \end{aligned} \quad (5.33)$$

Furthermore,

$$t_j \equiv \sum_{k=j}^n k^{-1} (1 + \beta_k) = \left(\sum_{k=r}^n + s_j \sum_{k=m_{0j}}^{m_{1j}} \right) k^{-1} (1 + \beta_k) = \mu + \mu_j + \delta + \delta_j.$$

Now, $E(\delta^2) = 2 \sum_{r \leq k \leq n} k^{-2} = O(n^{-1})$ and $E(\delta_j^2) = 2 \sum_{m_{0j} \leq k \leq m_{1j}} k^{-2} = O(n^{-2} |j - r|) = O(n^{-3/2})$ uniformly in $|j - r| \leq Cn^{1/2}$, for any $C > 0$. By Rényi's representation (David (1981), p.21), we may write $X_{nj} = H(t_j)$, for $1 \leq j \leq n$. Therefore, without loss of generality, and for $1 \leq j \leq n$,

$$\begin{aligned} X_{nj} &= H(\mu + \mu_j + \delta + \delta_j) \\ &= H(\mu + \mu_j) + (\delta + \delta_j) H'(\mu + \mu_j) + \frac{1}{2} \delta^2 H''(\mu + \mu_j) + S_{1j}, \end{aligned}$$

where S_{ij} denotes a random variable satisfying $E(S_{ij}) = O(n^{-3/2} |j - r|^{1/2}) = O(n^{-5/4})$ uniformly in $|j - r| \leq Cn^{1/2}$, for $i = 1, 2$. Consequently, since $\mu_r = \delta_r = 0$,

$$\begin{aligned} X_{nj} - X_{nr} &= H(\mu + \mu_j) - H(\mu) + \delta \{H'(\mu + \mu_j) - H'(\mu)\} \\ &\quad + \frac{1}{2} \delta^2 \{H''(\mu + \mu_j) - H''(\mu)\} + \delta_j H'(\mu + \mu_j) + S_{2j}. \end{aligned}$$

Now, $|H^{(i)}(\mu + \mu_j) - H^{(i)}(\mu)| = O(n^{-1/2})$ uniformly in $|j - r| \leq Cn^{1/2}$ for $i = 0, 1, 2$, because $|\mu_j| = |\sum_{m_{0j} \leq k \leq m_{1j}} k^{-1}| = O(n^{-1} |j - r|) = O(n^{-1/2})$ uniformly in such j 's. As a result,

$$\begin{aligned} (X_{nj} - X_{nr})^2 &= \{H(\mu + \mu_j) - H(\mu)\}^2 + \delta_j^2 H'(\mu + \mu_j)^2 \\ &\quad + 2\delta \{H(\mu + \mu_j) - H(\mu)\} \{H'(\mu + \mu_j) - H'(\mu)\} \\ &\quad + 2\delta_j \{H(\mu + \mu_j) - H(\mu)\} H'(\mu + \mu_j) + S_{3j}, \end{aligned} \quad (5.34)$$

where S_{3j} denotes a random variable satisfying $E(S_{3j}) = O(n^{-5/2}|j-r|^{3/2}) = O(n^{-7/4})$ uniformly in $|j-r| \leq Cn^{1/2}$.

Next we examine the weights, w_j . Recall from the proof of Lemma 5.6 that there exist constants C_1, C_2 so that (5.23) holds, viz.

$$w_j = n^{-\frac{1}{2}} \{\gamma(1-\gamma)\}^{-\frac{1}{2}} \phi(z) + n^{-1} \{\gamma(1-\gamma)\}^{-\frac{1}{2}} \frac{1}{6} z \{(1-2\gamma)(z^2-6)-6\} \phi(z) + r_{1j},$$

where $|r_{1j}| \leq C_1 n^{-3/2} \exp\{-C_2 n^{-1}(j-r)^2\}$ uniformly in $1 \leq j \leq n$ and $n \geq 1$, $\gamma \equiv (j-1)/n$ and $z \equiv (j-r-1)\{n\gamma(1-\gamma)\}^{-\frac{1}{2}}$. It follows that, given any $C_3 > 0$, there exists $C_4 > 0$ such that $|w_j| = O(n^{-C_3})$ uniformly in $|j-r| > C_4 n^{1/2}$. Since we have assumed $E|X|^\eta < \infty$, for some $\eta > 0$, we have, as in the proof of Theorem 5.2, $P(|X_j| \leq n^{2/\eta}, \text{ all } j \leq n) = 1 + O(n^{-1})$. Therefore, with probability tending to one as $n \rightarrow \infty$, $\max_{j \leq n} (X_{nj} - X_{nr})^2 \leq 4n^{4/\eta}$, and hence

$$\sum_{|j-r| > C_4 n^{1/2}} (X_{nj} - X_{nr})^2 w_j \leq C_1 n^{4/\eta - C_3 + \frac{1}{2}} = O(n^{-\lambda}),$$

for some $\lambda > 0$, depending on C_3 . We may make λ as large as we like by choosing C_3 appropriately. Consequently, given any $\lambda > 0$, for some $C_4 > 0$, we deduce that

$$\hat{\sigma}^2(C_4) \equiv \sum_j'' (X_{nj} - X_{nr})^2 w_j = \sigma^2 + O(n^{-\lambda}),$$

where $\sum_j''$ denotes summation over indices j for which $|j-r| \leq C_4 n^{1/2}$. Therefore, we conclude from (5.23) that, for such C_3, C_4 ,

$$\begin{aligned} n\hat{\sigma}^2 &= n \sum_{j=1}^n (X_{nj} - X_{nr})^2 w_j \\ &= A_0^2(1 + A_1\delta + U_1 + U_2) + n \sum_j'' S_{3j} w_j + n \sum_{|j-r| > C_4 n^{1/2}} S_{3j} w_j \\ &= A_0^2(1 + A_1\delta + U_1 + U_2) \\ &\quad + O_p\left(n^{-3/4} \sum_j'' w_j + n^{-3/2-C_3} \sum_{|j-r| > C_4 n^{1/2}} |j-r|^{3/2}\right) \\ &= A_0^2(1 + A_1\delta + U_1 + U_2) + O_p(n^{-3/4}), \end{aligned} \tag{5.35}$$

where

$$A_0^2 \equiv n \sum_{j=1}^n [\{H(\mu + \mu_j) - H(\mu)\}^2 + E(\delta_j^2) H'(\mu + \mu_j)^2] w_j,$$

$$A_0^2 A_1 \equiv 2n \sum_{j=1}^n \{H(\mu + \mu_j) - H(\mu)\} \{H'(\mu + \mu_j) - H'(\mu)\} w_j,$$

$$A_0^2 U_1 \equiv 2n \sum_{j=1}^n \delta_j \{H(\mu + \mu_j) - H(\mu)\} H'(\mu + \mu_j) w_j,$$

and

$$A_0^2 U_2 \equiv n \sum_{j=1}^n (\delta_j^2 - E\delta_j^2) H'(\mu + \mu_j)^2 w_j.$$

Since for any $C_3 > 0$ there exists $C_4 > 0$ such that $|w_j| = O(n^{-C_3})$ uniformly in $|j - r| > C_4 n^{\frac{1}{2}}$, we could have used $\sum_j''$ rather than $\sum_{1 \leq j \leq n}$ in the above definitions, as the difference $(\sum_{|j-r| > C_4 n^{\frac{1}{2}}})$ can be made of smaller order than $n^{-3/4}$ by appropriate choice of C_3 .

LEMMA 5.7.

$$A_0^2 = [p(1-p) + n^{-\frac{1}{2}} \{2\pi^{-1} p(1-p)\}^{\frac{1}{2}}] f(\xi_p)^{-2} + O(n^{-1}),$$

and

$$A_1 = 2\{pf'(\xi_p)f(\xi_p)^{-2} - 1\} + O(n^{-\frac{1}{2}}).$$

Proof of Lemma 5.7. Elementary calculations show that $\mu_j = \log\{(r-1)/(j-1)\} = (r-j)r^{-1} + \frac{1}{2}(r-j)^2 r^{-2} + O(n^{-3/2})$, and $\sum_{m_{0j} \leq k \leq m_{1j}} k^{-2} = |r-j|r^{-2} + O(n^{-2})$ uniformly in $|j-r| \leq Cn^{\frac{1}{2}}$, any $C > 0$. The first identity follows from identity 0.131 of Gradshteyn and Ryzhik (1965), p.2. Using expansion (5.23) for the w_j 's, and noting that $\gamma = p + n^{-1}(j-r) + O(n^{-1})$ and $r = np + O(1)$, we have

$$\begin{aligned} A_0^2 &= n \sum_{j=1}^n \{\mu_j^2 H'(\mu)^2 + \mu_j^3 H'(\mu) H''(\mu) + \left(\sum_{k=m_{0j}}^{m_{1j}} k^{-2} \right) H'(\mu)^2\} w_j + O(n^{-1}) \\ &= n \left[H'(\mu)^2 (np)^{-2} \sum_j'' (r-j)^2 w_j + \{H'(\mu)^2 + H'(\mu) H''(\mu)\} (np)^{-3} \right. \\ &\quad \times \sum_j'' (r-j)^3 w_j + H'(\mu)^2 (np)^{-2} \sum_j'' |j-r| w_j \left. \right] + O(n^{-1}) \\ &= n \left(H'(\mu)^2 (np)^{-2} \sum_j'' (r-j)^2 (\{np(1-p)\}^{-\frac{1}{2}} \phi[(j-r)\{np(1-p)\}^{-\frac{1}{2}}] \right. \\ &\quad + C_{1p} n^{-3/2} (j-r) \phi[(j-r)\{np(1-p)\}^{-\frac{1}{2}}] \\ &\quad \left. + C_{2p} n^{-5/2} (j-r)^3 \phi[(j-r)\{np(1-p)\}^{-\frac{1}{2}}] \right) \end{aligned}$$

$$\begin{aligned}
& + \{H'(\mu)^2 + H'(\mu)H''(\mu)\}(np)^{-3} \\
& \times \sum_j'' (r-j)^3 \{np(1-p)\}^{-\frac{1}{2}} \phi[(j-r)\{np(1-p)\}^{-\frac{1}{2}}] \\
& + H'(\mu)^2 (np)^{-2} \sum_j'' |r-j| \{np(1-p)\}^{-\frac{1}{2}} \phi[(j-r)\{np(1-p)\}^{-\frac{1}{2}}] \Big) \\
& + O(n^{-1}),
\end{aligned}$$

where C_{1p} and C_{2p} are constants depending on p arising from an expansion of the weights w_j about $\{np(1-p)\}^{-\frac{1}{2}} \phi[(j-r)\{np(1-p)\}^{-\frac{1}{2}}]$. Approximating sums by integrals in the above expression, we get

$$\begin{aligned}
A_0^2 &= n^{-1} p^{-2} H'(\mu)^2 \{np(1-p)\}^{-\frac{1}{2}} \\
& \times \int_{r-C_4 n^{\frac{1}{2}}}^{r+C_4 n^{\frac{1}{2}}} \{(x-r)^2 + |x-r|\} \phi[(x-r)\{np(1-p)\}^{-\frac{1}{2}}] dx \\
& + n^{-5/2} \left(p^{-2} H'(\mu)^2 C_{1p} \int_{r-C_4 n^{\frac{1}{2}}}^{r+C_4 n^{\frac{1}{2}}} (x-r) \phi[(x-r)\{np(1-p)\}^{-\frac{1}{2}}] dx \right. \\
& - p^{-3} \{p(1-p)\}^{-\frac{1}{2}} \{H'(\mu)^2 + H'(\mu)H''(\mu)\} \\
& \times \int_{r-C_4 n^{\frac{1}{2}}}^{r+C_4 n^{\frac{1}{2}}} (x-r)^3 \phi[(x-r)\{np(1-p)\}^{-\frac{1}{2}}] dx \Big) \\
& + n^{-7/2} p^{-2} H'(\mu)^2 C_{2p} \int_{r-C_4 n^{\frac{1}{2}}}^{r+C_4 n^{\frac{1}{2}}} (x-r)^5 \phi[(x-r)\{np(1-p)\}^{-\frac{1}{2}}] dx \\
& + O(n^{-1}) \\
& = n^{-1} p^{-2} H'(\mu)^2 \{np(1-p)\}^{-\frac{1}{2}} \int_{-\infty}^{\infty} (x^2 + |x|) \phi[x\{np(1-p)\}^{-\frac{1}{2}}] dx \\
& + n^{-5/2} \left(C_1(\mu, p) \int_{-\infty}^{\infty} x \phi[x\{np(1-p)\}^{-\frac{1}{2}}] dx \right. \\
& + C_2(\mu, p) \int_{-\infty}^{\infty} x^3 \phi[x\{np(1-p)\}^{-\frac{1}{2}}] dx \Big) \\
& + n^{-7/2} C_3(\mu, p) \int_{-\infty}^{\infty} x^5 \phi[x\{np(1-p)\}^{-\frac{1}{2}}] dx + O(n^{-1}) \\
& = p^{-2} H'(\mu)^2 [p(1-p) + \{2n^{-1} \pi p(1-p)\}^{\frac{1}{2}}] + O(n^{-1}),
\end{aligned}$$

where $C_1(\mu, p)$, $C_2(\mu, p)$, and $C_3(\mu, p)$ are coefficients depending on μ and p . Similarly, it follows that

$$A_0^2 A_1 = 2n H'(\mu) H''(\mu) \sum_{j=1}^n \mu_j^2 w_j + O(n^{-\frac{1}{2}})$$

$$\begin{aligned}
&= 2nH'(\mu)H''(\mu)(np)^{-2} \\
&\quad \times \sum_j'' (j-r)^2 \{np(1-p)\}^{-\frac{1}{2}} \phi[(j-r)\{np(1-p)\}^{-\frac{1}{2}}] + O(n^{-\frac{1}{2}}) \\
&= 2n^{-1}p^{-2}H'(\mu)H''(\mu)\{np(1-p)\}^{-\frac{1}{2}} \int_{-\infty}^{\infty} x^2 \phi[x\{np(1-p)\}^{-\frac{1}{2}}] dx \\
&\quad + O(n^{-\frac{1}{2}}) \\
&= 2p^{-1}(1-p)H'(\mu)H''(\mu) + O(n^{-\frac{1}{2}}).
\end{aligned}$$

Using (5.33) to write $H^{(i)}(\mu)$ in terms of p , ξ_p and $f(\xi_p)$ for $i = 0, 1, 2$ completes the proof.

Employing the fact that $E(\delta^2) = O(n^{-1})$, we have $S \equiv n^{\frac{1}{2}}\delta = O_p(1)$. Also, since $E(\delta_j^2) = O(n^{-2}|j-r|) = O(n^{-3/2})$ uniformly in $|j-r| \leq Cn^{\frac{1}{2}}$, any $C > 0$, we have

$$\begin{aligned}
S_1 &\equiv n^{1/4}U_1 = 2A_0^{-2}n^{5/4} \sum_{j=1}^n \delta_j \mu_j H'(\mu)^2 w_j + O_p(n^{-\frac{1}{2}}) \\
&= 2A_0^{-2}n^{5/4} \sum_j'' n^{-3/4} n^{-\frac{1}{2}} H'(\mu)^2 w_j + O_p(n^{-\frac{1}{2}}) \\
&= O_p(n^{5/4-3/4-\frac{1}{2}}) = O_p(1),
\end{aligned}$$

and

$$\begin{aligned}
S_2 &\equiv n^{\frac{1}{2}}U_2 = A_0^{-2}n^{3/2} \sum_{j=1}^n (\delta_j^2 - E\delta_j^2) H'(\mu + \mu_j)^2 w_j \\
&= A_0^{-2}n^{3/2} H'(\mu)^2 \sum_j'' (\delta_j^2 - E\delta_j^2) w_j + O_p(n^{-\frac{1}{2}}) \\
&= O_p\left(\sum_j'' w_j\right) = O_p(1),
\end{aligned}$$

the second-last identity following since $E\{(\delta_j^2 - E\delta_j^2)^2\} = E(\delta_j^4) + O(n^{-4}|j-r|^2) = O(\sum_{m_{0j} \leq k, l \leq m_{1j}} k^{-2}l^{-2} + n^{-4}|j-r|^2) = O(n^{-4}|j-r|^2) = O(n^{-3})$ uniformly in $|j-r| \leq Cn^{\frac{1}{2}}$.

Next we treat the problem of centring the sample quantile X_{nr} . Since $[np]$ does not increase smoothly as n increases, but rather jumps in steps of one unit, we study $X_{nr} - \eta_p$, where $\eta_p \equiv H(\mu)$, instead of $X_{nr} - \xi_p$. To see what effect this recentring has on our conclusions, note that $\mu = \sum_{r \leq k \leq n} k^{-1} = \log(n/r) - (n-r)(2nr)^{-1} + O(n^{-2}) = -\log p - (np)^{-1}\{\epsilon_n + \frac{1}{2}(1-p)\} + O(n^{-2})$, whence

$$\eta_p = \xi_p + n^{-1}\{\epsilon_n + \frac{1}{2}(1-p)\}f(\xi_p)^{-1} + O(n^{-2}),$$

where $\epsilon_n \equiv [np] + 1 - np$. Furthermore, $\hat{\sigma} = \{n^{-1}p(1-p)\}^{\frac{1}{2}} f(\xi_p)^{-1} + O_p(n^{-3/4})$, and so

$$(X_{nr} - \xi_p)/\hat{\sigma} = (X_{nr} - \eta_p)/\hat{\sigma} + \{np(1-p)\}^{-\frac{1}{2}} \{\epsilon_n + \frac{1}{2}(1-p)\} + O_p(n^{-3/4}). \quad (5.36)$$

Observe that

$$X_{nr} - \eta_p = H(\mu + \delta) - H(\mu) = \delta H'(\mu) + \frac{1}{2} \delta^2 H''(\mu) + O_p(n^{-3/2}). \quad (5.37)$$

This expansion, together with (5.35) gives us

$$\begin{aligned} (X_{nr} - \eta_p)/\hat{\sigma} &= n^{\frac{1}{2}} \{ \delta H'(\mu) + \frac{1}{2} \delta^2 H''(\mu) \} A_0^{-1} \{ 1 - \frac{1}{2}(A_1 \delta + U_1 + U_2) + \frac{3}{8} U_1^2 \} \\ &\quad + O_p(n^{-3/4}) \\ &= B_0 S [1 - n^{-1/4} \frac{1}{2} S_1 + n^{-\frac{1}{2}} \{ \frac{1}{2}(B_1 - A_1)S - \frac{1}{2} S_2 + \frac{3}{8} S_1^2 \}] + O_p(n^{-3/4}), \end{aligned}$$

where $B_0 \equiv H'(\mu)/A_0$ and $B_1 \equiv H''(\mu)/H'(\mu)$. By Taylor expansion, the moment generating function $\psi(\theta)$ of $Z \equiv (X_{nr} - \eta_p)/(B_0 \hat{\sigma})$ satisfies

$$\begin{aligned} \psi(\theta) &\equiv E\{\exp(\theta Z)\} \\ &= E\left(\exp(\theta S) \exp\left[\theta S \left\{ n^{-\frac{1}{2}} \frac{1}{2}(B_1 - A_1)S - \frac{1}{2}(n^{-1/4} S_1 + n^{-\frac{1}{2}} S_2) \right. \right. \right. \\ &\quad \left. \left. \left. + \frac{3}{8} n^{-\frac{1}{2}} S_1^2 \right\} \right] \right) + O(n^{-3/4}) \\ &= E\left(\exp(\theta S) \left[1 + \theta S \left\{ n^{-\frac{1}{2}} \frac{1}{2}(B_1 - A_1)S - \frac{1}{2}(n^{-1/4} S_1 + n^{-\frac{1}{2}} S_2) \right. \right. \right. \right. \\ &\quad \left. \left. \left. + \frac{3}{8} n^{-\frac{1}{2}} S_1^2 \right\} + \frac{1}{8} (\theta S S_1)^2 \right] \right) + O(n^{-3/4}) \\ &= E\left(e^{\theta S} \left[1 - n^{-1/4} \frac{1}{2} \theta S S_1 + n^{-\frac{1}{2}} \left\{ \frac{1}{2}(B_1 - A_1) \theta S^2 - \frac{1}{2} \theta S S_2 \right. \right. \right. \right. \\ &\quad \left. \left. \left. + \frac{3}{8} \theta S S_1^2 + \frac{1}{8} \theta^2 S^2 S_1^2 \right\} \right] \right) + O(n^{-3/4}). \quad (5.38) \end{aligned}$$

Next we treat in succession the terms of orders 1, $n^{-1/4}$ and $n^{-\frac{1}{2}}$ on the right-hand side of (5.38). Write σ_S^2 and $\sigma_{S_i}^2$ for variances of S and S_i respectively, and ρ_{SS_i} for the coefficient of correlation between S and S_i . Since $E\{\exp(\beta_k t)\} = (1-t)^{-1} e^{-t} = \exp\{\frac{1}{2} t^2 + \frac{1}{3} t^3 + O(t^4)\}$ as $t \rightarrow 0$,

$$\begin{aligned} E(e^{\theta S}) &= \exp \left[\frac{1}{2} \sum_{k=r}^n (n^{\frac{1}{2}} k^{-1} \theta)^2 + \frac{1}{3} \sum_{k=r}^n (n^{\frac{1}{2}} k^{-1} \theta)^3 + O \left\{ \sum_{k=r}^n (n^{\frac{1}{2}} k^{-1})^4 \right\} \right] \\ &= \exp(\frac{1}{2} \sigma_S^2 \theta^2) \{ 1 + n^{-\frac{1}{2}} \frac{1}{6} p^{-2} (1-p^2) \theta^3 + O(n^{-1}) \}, \quad (5.39) \end{aligned}$$

the last line following since $\sum_{r \leq k \leq n} k^{-3} = \frac{1}{2}p^{-2}(1 - p^2)$.

To compute the term of order $n^{-1/4}$, put

$$d_j \equiv 2n^{5/4} A_0^{-2} \{H(\mu + \mu_j) - H(\mu)\} H'(\mu + \mu_j) w_j,$$

$D_n = 0$, and

$$D_k \equiv \begin{cases} \sum_{j=1}^k d_j, & \text{if } 1 \leq k \leq r-1, \\ \sum_{j=k+1}^n d_j, & \text{if } r \leq k \leq n-1. \end{cases}$$

Then,

$$S_1 = \sum_{j=1}^n d_j \delta_j = \sum_{k=1}^{r-1} k^{-1} D_k \beta_k - \sum_{k=r}^{n-1} k^{-1} D_k \beta_k = S_4 - S_3,$$

say. By independence of the β_k 's,

$$-E(SS_1 e^{\theta S}) = E(SS_3 e^{\theta S}) = (\partial/\partial\theta)(\partial/\partial\omega)E\{\exp(\theta S + \omega S_3)\} \Big|_{\omega=0}. \quad (5.40)$$

Now,

$$\begin{aligned} & E\{\exp(\theta S + \omega S_3)\} \\ &= \prod_{k=r}^n E \exp\{k^{-1}(n^{\frac{1}{2}}\theta + D_k\omega)\beta_k\} \\ &= \exp\left(\frac{1}{2} \sum_{k=r}^n \{k^{-1}(n^{\frac{1}{2}}\theta + D_k\omega)\}^2 + \frac{1}{3} \sum_{k=r}^n \{k^{-1}(n^{\frac{1}{2}}\theta + D_k\omega)\}^3 \right. \\ &\quad \left. + O\left[\sum_{k=r}^n \{k^{-1}(n^{\frac{1}{2}}\theta + D_k\omega)\}^4\right]\right). \end{aligned} \quad (5.41)$$

Using expansion (5.23) for w_j , it follows for $k \geq r$,

$$\begin{aligned} |D_k| &\leq C_1 n^{5/4} \sum_{j=k+1}^n n^{-1} |j-r| n^{-\frac{1}{2}} \exp\{-C_2 n^{-1}(j-r)^2\} \\ &\leq C_3 n^{3/4} \exp\{-C_2 n^{-1}(k-r)^2\}, \end{aligned}$$

whence it follows, for bounded θ and ω , that

$$\sum_{k=r}^n \{k^{-1}(n^{\frac{1}{2}}\theta + D_k\omega)\}^4 \leq C_4 \sum_{k=r}^n [n^{-2} + n^{-1} \exp\{-C_2 n^{-1}(k-r)^2\}] \leq C_5 n^{-\frac{1}{2}},$$

and

$$\begin{aligned}
 & \sum_{k=r}^n |\{k^{-1}(n^{\frac{1}{2}}\theta + D_k\omega)\}^3 - (k^{-1}D_k\omega)^3| \\
 & \leq C_6 \sum_{k=r}^n n^{-3} \{|n^{\frac{1}{2}}\theta|^3 + (n^{\frac{1}{2}}\theta)^2 |D_k\omega| + |n^{\frac{1}{2}}\theta| (D_k\omega)^2\} \\
 & \leq C_7 n^{-\frac{1}{2}}.
 \end{aligned}$$

This indicates that

$$\begin{aligned}
 & E\{\exp(\theta S + \omega S_3)\} \\
 & = \exp(\tfrac{1}{2}\sigma_S^2\theta^2 + \rho_{SS_3}\sigma_S\sigma_{S_3}\theta\omega + \tfrac{1}{2}\sigma_{S_3}^2\omega^2 + C_8\omega^3) + O(n^{-\frac{1}{2}}),
 \end{aligned}$$

which in view of (5.40) suggests that

$$\begin{aligned}
 -E(SS_1e^{\theta S}) & = (\partial/\partial\theta)\rho_{SS_3}\sigma_S\sigma_{S_3}\theta\exp(\tfrac{1}{2}\sigma_S^2\theta^2) + O(n^{-\frac{1}{2}}) \\
 & = \rho_{SS_3}\sigma_S\sigma_{S_3}(1 + \sigma_S^2\theta^2)\exp(\tfrac{1}{2}\sigma_S^2\theta^2) + O(n^{-\frac{1}{2}}). \quad (5.42)
 \end{aligned}$$

The complete veracity of this result is easily deduced by working with the exact moment generating function at (5.41).

To treat the term of order $n^{-\frac{1}{2}}$ in (5.38), observe that if (V, W) has a joint Normal distribution with covariance vector $(\sigma_V^2, \sigma_W^2, \rho\sigma_V\sigma_W)$, and if we define $h(a, b) \equiv E\{V^a W^b \exp(\theta V)\}$, for $a \geq 0$, $b \geq 0$, then

$$\begin{aligned}
 h(2, 0) & = \sigma_V^2(1 + \sigma_V^2\theta^2)\exp(\tfrac{1}{2}\sigma_V^2\theta^2), \\
 h(1, 1) & = \rho\sigma_V\sigma_W(1 + \sigma_V^2\theta^2)\exp(\tfrac{1}{2}\sigma_V^2\theta^2), \\
 h(1, 2) & = (\sigma_V\sigma_W)^2\theta\{1 + \rho^2(2 + \sigma_V^2\theta^2)\}\exp(\tfrac{1}{2}\sigma_V^2\theta^2),
 \end{aligned}$$

and

$$h(2, 2) = (\sigma_V\sigma_W)^2\{(1 + 2\rho^2) + \sigma_V^2\theta^2(1 + 5\rho^2) + (\sigma_V^2\theta^2)^2\rho^2\}\exp(\tfrac{1}{2}\sigma_V^2\theta^2).$$

Therefore, noting that (S, S_1, S_2) has a limiting joint Normal distribution, we conclude that up to terms of order $n^{-1/4}$, and for $T = S, S_1$, or S_2 ,

$$E(STe^{\theta S}) = \rho_{ST}\sigma_S\sigma_T(1 + \sigma_S^2\theta^2)\exp(\tfrac{1}{2}\sigma_S^2\theta^2),$$

$$E(SS_1^2e^{\theta S}) = \sigma_S^2\sigma_{S_1}^2\theta\{1 + \rho_{SS_1}^2(2 + \sigma_S^2\theta^2)\}\exp(\tfrac{1}{2}\sigma_S^2\theta^2),$$

and

$$\begin{aligned}
 E(S^2S_1^2e^{\theta S}) & = \sigma_S^2\sigma_{S_1}^2\{1 + 2\rho_{SS_1}^2 + \sigma_S^2\theta^2(1 + 5\rho_{SS_1}^2) \\
 & \quad + (\sigma_S^2\theta^2)^2\rho_{SS_1}^2\}\exp(\tfrac{1}{2}\sigma_S^2\theta^2).
 \end{aligned}$$

Substituting into (5.38) the above expressions, (5.39), and (5.40), and collecting terms, we get

$$\begin{aligned} \psi(\theta) = & \left(1 + n^{-1/4} \frac{1}{2} \rho_{SS_3} \sigma_S \sigma_{S_3} \theta (1 + \sigma_S^2 \theta^2) \right. \\ & + n^{-1/2} \left[\frac{1}{2} (B_1 - A_1) \sigma_S^2 \theta (1 + \sigma_S^2 \theta^2) - \frac{1}{2} \rho_{SS_2} \sigma_S \sigma_{S_2} \theta (1 + \sigma_S^2 \theta^2) \right. \\ & + \frac{1}{6} p^{-2} (1 - p^2) \theta^3 + (\sigma_S \sigma_{S_1} \theta)^2 \left\{ \left(\frac{1}{2} + \rho_{SS_1}^2 \right) + \sigma_S^2 \theta^2 \left(\frac{1}{8} + \rho_{SS_1}^2 \right) \right. \\ & \left. \left. + \frac{1}{8} (\sigma_S^2 \theta^2)^2 \rho_{SS_1}^2 \right\} \right] \left. \right) \exp\left(\frac{1}{2} \sigma_S^2 \theta^2\right) + O(n^{-3/4}). \end{aligned}$$

Replacing θ by θ/σ_S and inverting the Laplace transform, we see that

$$\begin{aligned} & P\{(X_{nr} - \eta_p)/(B_0 \sigma_S \hat{\sigma}) \leq z\} \\ &= \Phi(z) - n^{-1/4} \frac{1}{2} \rho_{SS_3} \sigma_{S_3} z^2 \phi(z) \\ & \quad - n^{-1/2} \left[\left\{ \frac{1}{2} (B_1 - A_1) \sigma_S - \frac{1}{2} \rho_{SS_2} \sigma_{S_2} \right\} z^2 + \frac{1}{6} p^{-2} (1 - p^2) \sigma_S^{-3} (z^2 - 1) \right. \\ & \quad + \sigma_{S_1}^2 z \left\{ \left(\frac{1}{2} + \rho_{SS_1}^2 \right) + \left(\frac{1}{8} + \rho_{SS_1}^2 \right) (z^2 - 3) \right. \\ & \quad \left. \left. + \frac{1}{8} \rho_{SS_1}^2 (z^4 - 10z^2 + 15) \right\} \right] \phi(z) + O(n^{-3/4}). \end{aligned} \quad (5.43)$$

The final step is to evaluate the various constants appearing in (5.43), making use of (5.23), (5.33) and Lemma 5.7. Keeping in mind that $B_0 = H'(\mu)/A_0$ and $B_1 = H''(\mu)/H'(\mu)$, we deduce that

$$\sigma_S^2 = nE(\delta^2) = n \sum_{k=r}^n k^{-2} = p^{-1} (1 - p) + O(n^{-1}),$$

$$B_0 \sigma_S = -1 + n^{-1/2} \{2\pi p(1 - p)\}^{-1/2} + O(n^{-1}),$$

and

$$B_1 - A_1 = 1 - pf'(\xi_p) f(\xi_p)^{-2} + O(n^{-1/2}).$$

Now,

$$\begin{aligned} D_k & \sim 2n^{5/4} A_0^{-2} H'(\mu) \sum_{j=k+1}^n \mu_j w_j \\ & \sim -2n^{3/4} \{p/(1 - p)\}^{1/2} \phi[(k - r)\{np(1 - p)\}^{-1/2}], \end{aligned}$$

for $k \geq r$, whence

$$\begin{aligned} \sigma_{S_1}^2 &= \sigma_{S_3}^2 + \sigma_{S_4}^2 = 2\sigma_{S_3}^2 + O(n^{-1/2}) \\ &= 2 \sum_{k=r}^{n-1} k^{-2} D_k^2 + O(n^{-1/2}) \end{aligned}$$

$$\begin{aligned}
&= 8n^{3/2}p(1-p)^{-1}(2\pi)^{-1} \sum_{k=r}^{n-1} k^{-2} \exp[-(k-r)^2 \{np(1-p)\}^{-1}] + O(n^{-\frac{1}{2}}) \\
&= 8n^{3/2}p(1-p)^{-1}(2\pi)^{-1}(np)^{-2} \sum_{k=r}^{n-1} \exp[-(k-r)^2 \{np(1-p)\}^{-1}] \\
&\quad + O\left(n^{-\frac{1}{2}} + n^{-\frac{1}{2}} \sum_{k=r}^{n-1} r^{-1}(k-r) \exp[-(k-r)^2 \{np(1-p)\}^{-1}]\right) \\
&= 4\{p(1-p)\}^{-\frac{1}{2}}\pi^{-1} \int_0^\infty e^{-x^2} dx + O\left(n^{-\frac{1}{2}} + n^{-1} \int_0^\infty xe^{-x^2} dx\right) \\
&= 2\{\pi p(1-p)\}^{-\frac{1}{2}} + O(n^{-\frac{1}{2}}),
\end{aligned}$$

and

$$\begin{aligned}
-\rho_{SS_1}\sigma_S\sigma_{S_1} &= \rho_{SS_3}\sigma_S\sigma_{S_3} = n^{\frac{1}{2}} \sum_{k=r}^{n-1} k^{-2} D_k \\
&= -2n^{5/4}\{p/(1-p)\}^{\frac{1}{2}}(np)^{-2} \sum_{k=r}^{n-1} \phi[(k-r)\{np(1-p)\}^{-\frac{1}{2}}] \\
&\quad + O(n^{-\frac{1}{2}}) \\
&= -2n^{-3/4}\{p/(1-p)\}^{\frac{1}{2}}p^{-2}\{np(1-p)\}^{\frac{1}{2}} \int_0^\infty \phi(x) dx + O(n^{-\frac{1}{2}}) \\
&= -n^{-1/4}p^{-1} + O(n^{-\frac{1}{2}}),
\end{aligned}$$

where the error may be shown to be of the stated order in the same way as for $\sigma_{S_1}^2$. Finally,

$$\begin{aligned}
\rho_{SS_2}\sigma_S\sigma_{S_2} &= E(SS_2) \\
&= n^2 A_0^{-2} E\left[\left(\sum_{k=r}^n k^{-1} \beta_k\right) \left\{\sum_{j=1}^n (\delta_j^2 - E\delta_j^2) H'(\mu + \mu_j)^2 w_j\right\}\right] \\
&= n^2 A_0^{-2} E(\beta_1^3) \sum_{j=r+1}^n H'(\mu + \mu_j)^2 w_j \sum_{k=r}^{j-1} k^{-3} = O(n^{-\frac{1}{2}}).
\end{aligned}$$

Hence, by (5.43), on collecting terms and rearranging, we obtain

$$\begin{aligned}
&P\{(X_{nr} - \eta_p)/\hat{\sigma} \leq z\} \\
&= \Phi(z) + n^{-\frac{1}{2}} \left(\frac{1}{4} \{\pi p(1-p)\}^{-\frac{1}{2}} z(z^2 + 1 + 2\sqrt{2}) \right. \\
&\quad + \frac{1}{6} \{p(1-p)\}^{-\frac{1}{2}} (1+p)(z^2 - 1) - \frac{1}{2} z^2 [\{p(1-p)\}^{-\frac{1}{2}} p \\
&\quad \left. + \{p(1-p)\}^{\frac{1}{2}} f'(\xi_p) f(\xi_p)^{-2}] \right) \phi(z) + O(n^{-3/4}). \quad (5.44)
\end{aligned}$$

It remains only to replace η_p by ξ_p on the left-hand side of (5.44). Noting (5.36), if we replace η_p by ξ_p there, we must subtract $\{np(1-p)\}^{-\frac{1}{2}}\{\frac{1}{2}(1-p) + \epsilon_n\}\phi(z)$ from the right-hand side of (5.44).

6.4. Proof of Theorem 5.4. Equation (5.11a) follows directly from Theorem 5.1. To derive (5.11b), define $H(x) \equiv F^{-1}(e^{-x})$ as in the proof of Theorem 5.3, and by Rényi's representation, let $Z_1, \dots, Z_n$ denote independent, unit-mean, negative exponential variables such that $X_{nj} = H(\sum_{j \leq k \leq n} k^{-1} Z_k)$, for $1 \leq j \leq n$. Put $A_s \equiv \sum_{s \leq k \leq n} k^{-1} Z_k$, and $a_s \equiv \sum_{s \leq k \leq n} k^{-1}$. By Taylor expansion, $X_{n,s+1} - X_{n,s} = D + R$, where $D \equiv -s^{-1}H'(a_s)Z_s$ and

$$R \equiv -s^{-1}Z_s \int_0^1 \{H'(A_s - ts^{-1}Z_s) - H'(a_s)\} dt.$$

Now, given $\zeta > 0$, with probability tending to one as $n \rightarrow \infty$, $X_{n,s+1} = H(\sum_{s+1 \leq k \leq n} k^{-1} Z_k) \in (\xi_p - \frac{1}{2}\zeta, \xi_p + \frac{1}{2}\zeta)$. Therefore, since f is continuous at ξ_p , and $H'(x) = -e^{-x}/f\{H(x)\}$, we see that $R = o_p(n^{-1})$. Thus,

$$\begin{aligned} \hat{\sigma}_J^2 &= n^{-2}s^{-1}(n-1)(n-s)H'(a_s)^2 Z_s^2 + o_p(n^{-1}) \\ &= n^{-1}p(1-p)f(\xi_p)^{-2} Z_s^2 + o_p(n^{-1}), \end{aligned}$$

whence

$$\hat{\sigma}_J^2 \sigma^{-2} = Z_s^2 + o_p(1).$$

The result follows on noting that Z_s^2 has a Weibull distribution with parameters 1 and $\frac{1}{2}$.

6.5. Proof of Theorem 5.5. We use the same notation as for the proof of Theorem 5.3. From (5.33) and (5.37), we have

$$\begin{aligned} X_{nr} - \xi_p &= X_{nr} - \eta_p + O(n^{-1}) = \delta H'(\mu) + o_p(n^{-\frac{1}{2}}) \\ &= -pf(\xi_p)^{-1}n^{-\frac{1}{2}}S + o_p(n^{-\frac{1}{2}}), \end{aligned} \quad (5.45)$$

and from Theorem 5.4,

$$\hat{\sigma}_J = n^{-\frac{1}{2}}\{p(1-p)\}^{\frac{1}{2}}f(\xi_p)^{-1}Z_s + o_p(n^{-\frac{1}{2}}). \quad (5.46)$$

Combining (5.45) and (5.46), we obtain

$$(X_{nr} - \xi_p)/\hat{\sigma}_J = -\{p/(1-p)\}^{\frac{1}{2}}Z_s^{-1}S + o_p(1). \quad (5.47)$$

Trivially,

$$Z_s^{-1}S = Z_s^{-1}n^{\frac{1}{2}} \sum_{k=r}^n k^{-1}\beta_k = Z_s^{-1}n^{\frac{1}{2}} \sum_{k=s+1}^n k^{-1}\beta_k + o_p(1) = Z_s^{-1}T + o_p(1),$$

where $T \equiv n^{\frac{1}{2}} \sum_{s+1 \leq k \leq n} k^{-1}\beta_k$. By independence of the β_k 's, Z_s and T are independent and the characteristic function of T ,

$$E(e^{i\theta T}) = \exp(-i\theta n^{\frac{1}{2}} \sum_{k=s+1}^n k^{-1}) \prod_{k=s+1}^n (1 - i\theta n^{\frac{1}{2}} k^{-1})^{-1},$$

converges to $\exp[-t^2 p / \{2(1-p)\}]$, the characteristic function of a Normal random variable with zero mean and variance $p^{-1}(1-p)$. Hence, it follows from (5.47) that the limiting distribution of $(X_{nr} - \xi_p) / \hat{\sigma}_J$ is that of the ratio of a Standard Normal random variable to an independent, unit-mean, negative exponential random variable. The distribution function of this statistic is

$$G(z) \equiv \frac{1}{2} + \text{sgn}(z)\Phi(-|z|^{-1})\exp\{(2z^2)^{-1}\}$$

for $z \neq 0$, with $G(0) = \frac{1}{2}$.

APPENDIX A5.1. Proof of (5.3).

Using an Edgeworth expansion for the binomial distribution function (see Gnedenko and Kolmogorov (1968), p.213), writing $y = (s - np)\{np(1 - p)\}^{-\frac{1}{2}}$, and noting that $p = \frac{1}{2}$, we see that when $n = n_k$,

$$\begin{aligned} P\{\xi_p \leq \hat{F}^{-1}(n^{-1}s)\} &= P\{Bi(n, p) \leq s\} \\ &= \Phi(y) + 2n^{-\frac{1}{2}}([ky] - ky + \frac{1}{2})\phi(y) + o(n^{-\frac{1}{2}}), \end{aligned}$$

where $[\cdot]$ denotes the integer part function. Change variable from s to $t = s - 2k^2 - 2k$, an integer. Then $y = sk^{-1} - 2k = 2 + tk^{-1}$, ky is an integer, and

$$\begin{aligned} A(y) &\equiv |P\{\xi_p \leq \hat{F}^{-1}(n^{-1}s)\} - (1 - \alpha)| \\ &= |\Phi(y) + n^{-\frac{1}{2}}\phi(y) - (1 - \alpha)| + o(n^{-\frac{1}{2}}). \end{aligned}$$

If t is chosen to minimize this quantity then, since $1 - \alpha = \phi(2)$, we must have $tk^{-1} \rightarrow 0$ and

$$A(y) = n^{-\frac{1}{2}}\phi(2)|2t + 1| + o(n^{-\frac{1}{2}}),$$

whose asymptotic minimum value, achieved at $t = 0$ or $t = -1$, is $n^{-\frac{1}{2}}\phi(2) + o(n^{-\frac{1}{2}})$.

Bibliography

- Abramovitch, L. and Singh, K. (1985) Edgeworth corrected pivotal statistics and the bootstrap. *Ann. Statist.* **13**, 116–132.
- Abramowitz, M. and Stegun, I. (1972) *Handbook of Mathematical Functions*, 9'th Edn. Dover, New York.
- Aspin, A.A. (1948) An examination and further development of a formula arising in the problem of comparing two mean values. *Biometrika* **35**, 88–96.
- Babu, G.J. (1986) A note on bootstrapping the variance of sample quantile. *Ann. Inst. Statist. Math.* **38**, 439–443.
- Babu, G.J. and Singh, K. (1983) Inference on means using the bootstrap. *Ann. Statist.* **11**, 999–1003.
- Barnard, G.A. (1963) Discussion of Bartlett, M.S. (1963), *J. Roy. Statist. Soc. Ser. B* **25**, 294.
- Barndorff-Nielsen, O.E. and Cox, D.R. (1984) Bartlett adjustments to the likelihood ratio statistic and the distribution of the maximum likelihood estimator. *J. Roy. Statist. Soc. Ser. B* **46**, 483–495.
- Barndorff-Nielsen, O.E. and Hall, P. (1988) On the level error after Bartlett adjustment of the likelihood ratio statistic. *Biometrika* **75**, 374–378.
- Bartlett, M.S. (1937) Properties of sufficiency and statistical tests. *Proc. Roy. Soc. A* **160**, 268–282.
- Behrens, B.-U. (1929) Ein Beitrag zur Fehlen-Berechnung bei wenigen Beobachtungen. *Landwirtsch. Jb.* **68**, 807–837.
- Beran, R. (1982) Estimated sampling distributions: the bootstrap and competitors. *Ann. Statist.* **10**, 212–225.
- Beran, R. (1984a) Jackknife approximations to bootstrap estimates. *Ann. Statist.* **12**, 101–118.
- Beran, R. (1984b) Bootstrap methods in statistics. *Jber. d. Dt. Math. Verein.* **86**, 14–30.
- Beran, R. (1987a) Prepivoting to reduce level error of confidence sets. *Biometrika* **74**, 457–468.

- Beran, R. (1987b) Discussion of Wu, C.F.J. (1987), *Ann. Statist.* **14**, 1295–1298.
- Beran, R. (1988) Prepivoting test statistics: A bootstrap view of the Behrens-Fisher problem, the Bartlett adjustment, and nonparametric analogs. *J. Amer. Statist. Assoc.*, in press.
- Bhattacharya, R.N. and Ghosh, J.K. (1978) On the validity of the formal Edgeworth expansion. *Ann. Statist.* **6**, 435–451.
- Bhattacharya, R.N. and Ranga Rao, R. (1976) *Normal Approximation and Asymptotic Expansions*. Wiley, New York.
- Bickel, P.J. (1974) Edgeworth expansions in nonparametric statistics. *Ann. Statist.* **2**, 1–20.
- Bickel, P.J. and Freedman, D. (1980) On Edgeworth expansions and the bootstrap. Unpublished manuscript.
- Bickel, P.J. and Freedman, D. (1981) Some asymptotic theory for the bootstrap. *Ann. Statist.* **9**, 1196–1217.
- Bloch, D.A. and Gastwirth, J.L. (1968) On a simple estimate of the reciprocal of the density function. *Ann. Math. Statist.* **39**, 1083–1085.
- Burkholder, D.L. (1973) Distribution function inequalities for martingales. *Ann. Probab.* **1**, 19–42.
- Chambers, J.M. (1967) On methods of asymptotic approximation for multivariate distributions. *Biometrika* **54**, 367–383.
- Chebyshev, P.L. (1890) Sur deux théorèmes relatifs aux probabilités. *Acta Math.* **14**, 305–315.
- Chibishov, D.M. (1972) An asymptotic expansion for the distribution of a statistic admitting an asymptotic expansion. *Theor. Probab. Appl.* **17**, 620–630.
- Chibishov, D.M. (1973a) An asymptotic expansion for a class of estimators containing maximum likelihood estimators. *Theor. Probab. Appl.* **18**, 295–303.
- Chibishov, D.M. (1973b) An asymptotic expansion for the distribution of sums of a special form with application to minimum contrast estimates. *Theor. Probab. Appl.* **18**, 649–661.
- Chung, K.L. (1974) *A Course in Probability Theory*. Academic Press, New York.

- Cornish, E.A. and Fisher, R.A. (1937) Moments and cumulants in the specification of distributions. *Rev. Inst. Int. Statist.* **5**, 307–322.
- Cramér, H. (1928) On the composition of elementary errors. *Skand. Akt.* **11**, 13–74 and 141–180.
- Cramér, H. (1946) *Mathematical Methods of Statistics*. Princeton University Press, Princeton.
- Csörgő, M. (1983) *Quantile Processes with Statistical Applications*. SIAM, Philadelphia.
- David, H.A. (1981) *Order Statistics*, 2nd Edn. Wiley, New York.
- David, F.N. and Johnson, N.L. (1954) Statistical treatment of censored data, I. *Biometrika* **41**, 228–240.
- Davis, A.W. (1976) Statistical distributions in univariate and multivariate Edgeworth populations. *Biometrika* **63**, 661–670.
- Diaconis, P. and Efron, B. (1983) Computer-intensive methods in statistics. *Scientific American* **248**, 96–108.
- DiCiccio, T.J. and Romano, J.P. (1988) A review of bootstrap confidence intervals. (With discussion.) *J. Roy. Statist. Soc. Ser. B* **50**, 338–370.
- Duncan, G.T. and Layard, M.W.J. (1973) A Monte-Carlo study of asymptotically robust tests for correlation coefficients. *Biometrika* **60**, 551–558.
- Edgeworth, F.Y. (1904) The law of error. *Trans. Camb. Phil. Soc.* **20**, 36–65.
- Efron, B. (1979a) Bootstrap methods: another look at the jackknife. *Ann. Statist.* **7**, 1–26.
- Efron, B. (1979b) Computers and the theory of statistics: thinking the unthinkable. *SIAM Review* **21**, 460–480.
- Efron, B. (1981a) Nonparametric standard errors and confidence intervals. (With discussion.) *Canadian J. Statist.* **9**, 139–172.
- Efron, B. (1981b) Nonparametric estimates of standard error: the jackknife, the bootstrap, and other methods. *Biometrika* **58**, 589–599.
- Efron, B. (1982) *The Jackknife, the Bootstrap and Other Resampling Plans*. SIAM, Philadelphia.
- Efron, B. (1987) Better bootstrap confidence intervals. (With discussion.) *J.*

- Amer. Statist. Assoc.* **82**, 171–200.
- Efron, B. and Gong, G. (1983) A leisurely look at the bootstrap, the jackknife and cross-validation. *Amer. Statistician* **37**, 36–48.
- Efron, B. and Tibshirani, R. (1986) Bootstrap measures for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science* **1**, 54–77.
- Esseen, C.G. (1945) Fourier Analysis of Distribution Functions: A Mathematical Study of the Laplace-Gaussian Law. *Acta Math.* **77**, 1–125.
- Falk, M. (1986) On the estimation of the quantile density function. *Statist. Probab. Letters* **4**, 69–73.
- Fisher, R.A. (1921) On the probable error of a coefficient of correlation deduced from a small sample. *Metron* **1**, 3–32.
- Fisher, R.A. (1935) The fiducial argument in statistical inference. *Ann. Eugen.* **6**, 391–398.
- Gayen, A.K. (1949) The distribution of “Student’s t ” in random samples of any size drawn from nonnormal universes. *Biometrika* **36**, 353–369.
- Gayen, A.K. (1950) The variance ratio in random samples of any size drawn from nonnormal universes. *Biometrika* **37**, 236–255.
- Geary, R.C. (1936) The distribution of “Student’s” ratio for nonnormal samples. *Supp. J. Roy. Statist. Soc.* **3**, 178–184.
- Geary, R.C. (1947) Testing for normality. *Biometrika* **34**, 209–242.
- Ghosh, M., Parr, W.C., Singh, K., and Babu, G.J. (1984) A note on bootstrapping the sample median. *Ann. Statist.* **12**, 1130–1135.
- Gnedenko, B.V. and Kolmogorov, A.N. (1968) *Limit Distributions for Sums of Independent Random Variables*. Addison-Wesley, London.
- Gradshteyn, I.S. and Ryzhik, I.M. (1965) *Tables of Integrals, Series and Products*. Academic Press, New York.
- Guenther, W.C. (1969) Shortest confidence intervals. *Amer. Statistician* **23**, 22–25.
- Hall, P. (1983) Inverting an Edgeworth expansion. *Ann. Statist.* **11**, 569–576.
- Hall, P. (1986a) On the bootstrap and confidence intervals. *Ann. Statist.* **14**, 1431–1452.

- Hall, P. (1986b) On the number of bootstrap simulations required to construct a confidence interval. *Ann. Statist.* **14**, 1453–1462.
- Hall, P. (1987a) On the bootstrap and likelihood-based confidence intervals. *Biometrika* **74**, 481–493.
- Hall, P. (1987b) Discussion of Wu, C.F.J. (1987), *Ann. Statist.* **14**, 1311–1312.
- Hall, P. (1987c) Edgeworth expansions for Student's t statistic under minimal moment conditions. *Ann. Probab.* **15**, 920–931.
- Hall, P. (1988a) Theoretical comparison of bootstrap confidence intervals. (With discussion.) *Ann. Statist.* **16**, 927–985.
- Hall, P. (1988b) On symmetric bootstrap confidence intervals. *J. Roy. Statist. Soc. Ser. B* **50**, 35–45.
- Hall, P. (1989) *An Edgeworth View of the Bootstrap*. Unpublished manuscript.
- Hall, P., DiCiccio, T.J., and Romano, J.P. (1988) On smoothing and the bootstrap. *Ann. Statist.*, in press.
- Hall, P. and Heyde, C.C. (1980) *Martingale Limit Theory and its Application*. Academic Press, New York.
- Hall, P. and Martin, M.A. (1988a) On the bootstrap and two-sample problems. *Austral. J. Statist.* **30A**, 179–192.
- Hall, P. and Martin, M.A. (1988b) On bootstrap resampling and iteration. *Biometrika*, in press.
- Hall, P. and Martin, M.A. (1988c) Exact convergence rate of bootstrap quantile variance estimator. *Probab. Theor. and Rel. Fields*, in press.
- Hall, P. and Martin, M.A. (1988d) A note on the accuracy of bootstrap percentile method confidence intervals for a quantile. *Statist. Probab. Letters*, in press.
- Hall, P. and Martin, M.A. (1988e) On the error incurred using the bootstrap variance estimate when constructing confidence intervals for quantiles. *J. Roy. Statist. Soc. Ser. B*, submitted.
- Hall, P., Martin, M.A., and Schucany, W.R. (1988) Better nonparametric bootstrap confidence intervals for the correlation coefficient. *J. Statist. Computa. and Simul.*, submitted.

- Hall, P. and Sheather, S.J. (1988) On the distribution of a Studentized quantile. *J. Roy. Statist. Soc. Ser. B* **50**, 381–391.
- Harter, H.L. (1964) Criteria for best substitute interval estimator with an application to the normal distribution. *J. Amer. Statist. Assoc.* **59**, 1133–1140.
- Hartigan, J.A. (1969) Using subsample values as typical values. *J. Amer. Statist. Assoc.* **64**, 1303–1317.
- Hartigan, J.A. (1971) Error analysis by replaced samples. *J. Roy. Statist. Soc. Ser. B* **33**, 98–110.
- Hartigan, J.A. (1975) Necessary and sufficient conditions for asymptotic joint normality of a statistic and its subsample values. *Ann. Statist.* **3**, 573–580.
- Hartigan, J.A. (1986) Discussion of Efron and Tibshirani (1986), *Statistical Science* **1**, 75–77.
- Hayakawa, T. (1976) Asymptotic expansion of the distribution of the likelihood ratio criteria for homogeneity of parameters. In *Essays in Probability and Statistics*, Ed. S. Ikeda et al. Shinko Tsusho, Tokyo, pp. 265–285.
- Hinkley, D.V. and Wei, B.-C. (1984) Improvements of jackknife confidence limit methods. *Biometrika* **71**, 331–339.
- Hoeffding, W. (1963) Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **52**, 13–30.
- Hsu, P.L. (1945) The approximate distributions of the mean and variance of a sample of independent variables. *Ann. Math. Statist.* **16**, 1–29.
- Hope, A.C.A. (1968) A simple Monte Carlo test procedure. *J. Roy. Statist. Soc. Ser. B* **30**, 582–598.
- James, G.S. (1955) Cumulants of a transformed variable. *Biometrika* **42**, 529–531.
- James, G.S. (1958) On moments and cumulants of systems of statistics. *Sankhyā* **20**, 1–30.
- James, G.S. and Mayne, A.J. (1962) Cumulants of functions of random variables. *Sankhyā Ser. A* **24**, 47–54.
- Kaigh, W.D. (1983) Quantile interval estimation. *Commun. Statist.-Theor. Meth.* **12**, 2427–2443.

- Kendall, M.G. and Stuart, A. (1977) *The Advanced Theory of Statistics*, Vols. 1 and 2, 4'th Edn. Griffin, London.
- Lawless, J.F. (1981) Mean squared error properties of generalized ridge estimators. *J. Amer. Statist. Assoc.* **76**, 462-466.
- Lawley, D.N. (1956) A general method for approximating the distribution of the likelihood ratio criteria. *Biometrika* **43**, 295-303.
- Lehmann, E.L. (1986) *Testing Statistical Hypotheses*, 2'nd Edn. Wiley, New York.
- Lemmer, H.H. (1982) Shrinkage estimators. In *Encyclopedia of Statistical Science*, Vol. 8. Ed. S. Kotz and N.L. Johnson. Wiley, New York. pp. 452-456.
- Liu, R.Y. and Singh, K. (1987) On a partial correction by the bootstrap. *Ann. Statist.* **15**, 1713-1718.
- Loh, W.-Y. (1987) Calibrating confidence coefficients. *J. Amer. Statist. Assoc.* **82**, 155-162.
- Lunneborg, C.E. (1985) Estimating the correlation coefficient - the bootstrap approach. *Psych. Bull.* **98**, 209-215.
- Maritz, J.S. and Jarrett, R.G. (1978) A note on estimating the variance of the sample median. *J. Amer. Statist. Assoc.* **73**, 194-196.
- Marriot, F.H.C. (1979) Barnard's Monte Carlo test: how many simulations? *Appl. Statist.* **28**, 75-77.
- McKean, J.W. and Schrader, R.M. (1984) A comparison of methods for Studentizing the sample quantile. *Commun. Statist. Ser. B - Simula. Computa.* **13**, 751-773.
- Mehta, J.S. and Srinivasan, R. (1971) Estimation of the mean by shrinkage to a point. *J. Amer. Statist. Assoc.* **66**, 86-90.
- Miller, R.G. (1964) A trustworthy jackknife. *Ann. Math. Statist.* **39**, 1594-1605.
- Miller, R.G. (1974) The jackknife - a review. *Biometrika* **61**, 1-17.
- Parr, W.C. (1983) A note on the jackknife, the bootstrap and delta method estimators of bias and variance. *Biometrika* **70**, 719-722.
- Parr, W.C. (1985) Jackknifing differentiable statistical functionals. *J. Roy. Statist. Soc. Ser. B* **47**, 56-66.
- Parzen, E. (1979) Nonparametric statistical data modeling. (With discussion.) *J.*

- Amer. Statist. Assoc.* **74**, 105–131.
- Pearson, K. and Pearson, M.V. (1931) On the mean character and variance of a ranked individual, and the mean and variance of the intervals between ranked individuals, I: symmetrical distributions (normal and rectangular). *Biometrika* **23**, 364–397.
- Pearson, K. and Pearson, M.V. (1932) On the mean character and variance of a ranked individual, and the mean and variance of the intervals between ranked individuals, II: case of certain skew curves. *Biometrika* **24**, 204–279.
- Petrov, V.V. (1975) *Sums of Independent Random Variables*. Springer, Berlin.
- Pfanzagl, J. (1973a) Asymptotic expansions related to minimum contrast estimators. *Ann. Statist.* **1**, 993–1026.
- Pfanzagl, J. (1973b) The accuracy of the normal approximation for estimates of vector parameters. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **25**, 171–198.
- Pfanzagl, J. (1979) Nonparametric minimum contrast estimators. In *Selecta Statistica Canadiana, V.*, pp. 105–140.
- Phillips, P.C.B. (1977) A general theorem in the theory of asymptotic expansions as approximations to the finite sample distributions of econometric estimators. *Econometrica* **45**, 1517–1534.
- Pratt, J.W. (1961) Length of confidence intervals. *J. Amer. Statist. Assoc.* **56**, 549–567.
- Pratt, J.W. (1963) Shorter confidence intervals for the mean of a normal distribution with known variance. *Ann. Math. Statist.* **34**, 574–586.
- Quenouille, M.H. (1949) Approximate tests of correlation in time series. *J. Roy. Statist. Soc. Ser. B* **11**, 68–84.
- Quenouille, M.H. (1956) Notes on bias in estimation. *Biometrika* **43**, 353–360.
- Rasmussen, J. (1987) Estimating correlation coefficients: bootstrap and parametric approaches. *Psych. Bull.* **101**, 136–139.
- Reiss, R.-D. (1974) On the accuracy of the normal approximation for quantiles. *Ann. Probab.* **2**, 741–744.

- Reiss, R.-D. (1976) Asymptotic expansions for sample quantiles. *Ann. Probab.* **4**, 249–258.
- Rényi, A. (1953) On the theory of order statistics. *Acta Math. Acad. Sci. Hung.* **4**, 191–232.
- Robinson, G.K. (1982) Behrens-Fisher Problem. In *Encyclopedia of Statistical Science*, Vol. 1. Ed. S. Kotz and N.L. Johnson. Wiley, New York. pp. 205–209.
- Rosenblatt, M. (1971) Curve estimates. *Ann. Math. Statist.* **42**, 1815–1842.
- Sargan, J.D. (1976) Econometric estimators and the Edgeworth expansion. *Econometrica* **44**, 421–448.
- Schenker, N. (1985) Qualms about bootstrap confidence intervals. *J. Amer. Statist. Assoc.* **80**, 360–361.
- Sen, P.K. (1959) On the moments of sample quantiles. *Calcutta Statist. Assoc. Bull.* **9**, 1–19.
- Sen, P.K. (1988) Functional jackknifing: rationality and general asymptotics. *Ann. Statist.* **16**, 450–469.
- Sheather, S.J. (1986a) A finite sample estimate of the variance of the sample median. *Statist. Probab. Letters* **4**, 337–342.
- Sheather, S.J. (1986b) An improved data-based algorithm for choosing the window width when estimating a density at a point. *Comput. Statist. Data Anal.* **4**, 61–65.
- Sheather, S.J. (1987) Assessing the accuracy of the sample median: estimated standard errors versus interpolated confidence intervals. In *Statistical Data Analysis Based on the L_1 -Norm*. Ed. Y. Dodge, pp. 203–215. North Holland, Amsterdam.
- Sheather, S.J. and Maritz, J.S. (1983) An estimate of the asymptotic standard error of the sample median. *Austral. J. Statist.* **25**, 109–122.
- Sheather, S.J. and McKean, J.W. (1987) A comparison of testing and confidence interval methods for the median. *Statist. Probab. Letters* **6**, 31–36.
- Siddiqui, M.M. (1960) Distribution of quantiles in samples from a bivariate population. *Journ. Res. Nat. Bur. Stand.* **64B**, 145–150.

- Silverman, B.W. and Young, G.A. (1987) The bootstrap: to smooth or not to smooth? *Biometrika* **74**, 469–479.
- Singh, K. (1981) On the asymptotic accuracy of Efron's bootstrap. *Ann. Statist.* **9**, 1187–1195.
- Singh, K. (1987) Discussion of Wu, C.F.J. (1987), *Ann. Statist.* **14**, 1328–1330.
- Tukey, J.W. (1958) Bias and confidence in not quite large samples. (Abstract.) *Ann. Math. Statist.* **29**, 614.
- Thompson, J.R. (1968) Some shrinkage techniques for estimating the mean. *J. Amer. Statist. Assoc.* **63**, 113–122.
- Van Zwet, W.R. (1964) *Convex Transformations of Random Variables*. Mathematical Centre Tracts No. 7. Mathematisch Centrum, Amsterdam.
- Wallace, D.L. (1958) Asymptotic approximations to distributions. *Ann. Math. Statist.* **29**, 635–654.
- Welch, B.L. (1947) The generalization of "Student's" problem when several different population variances are involved. *Biometrika* **34**, 28–35.
- Welsh, A.H. (1987) Kernel estimates of the sparsity function. In *Statistical Data Analysis Based on the L_1 -Norm*. Ed. Y. Dodge, pp. 369–377. North Holland, Amsterdam.
- Welsh, A.H. (1988) Asymptotically efficient estimation of the sparsity function at a point. *Statist. Probab. Letters*, in press.
- Wilson, P.D. and Tonascia, J. (1971) Tables for shortest confidence intervals on the standard deviation and variance ratio from normal distributions. *J. Amer. Statist. Assoc.* **66**, 909–912.
- Withers, C.S. (1983) Expansions for the distribution and quantiles of a regular functional of the empirical distribution with applications to nonparametric confidence intervals. *Ann. Statist.* **11**, 577–587.
- Withers, C.S. (1984) Asymptotic expansions for distributions and quantiles with power series cumulants. *J. Roy. Statist. Soc. Ser. B* **46**, 389–396.
- Wu, C.F.J. (1987) Jackknife, bootstrap and other resampling plans in regression analysis. (With discussion.) *Ann. Statist.* **14**, 1261–1350.
- Young, G.A. (1988) A note on bootstrapping the correlation coefficient. *Biometrika* **75**, 370–373.

"... and you will know the truth, and the truth will make you free"

John 8:32