Topics in Confidence Distributions

Josephina Argyrou

Project submitted for the requirements of the
MSc in Statistics and Data Analysis

I would like to take this opportunity to thank my supervisor Prof. Nils Lid Hjort for this inspiring project and his constant support and guidance throughout its completion. Only rarely one finds a supervisor dedicate to help students achieve their goals and drive them to the fullest of their potential. The time and effort spent on this project was by far rewarding and it made me realise that research is fun and one cannot but falling absolutely in love with it. I also would like to thank all my family for their support and, particularly, my mother who has been there in difficult times providing support and courage to continue until my goals have been fulfilled.

Contents

List of Figures

List of Tables

1 Introduction

1.1 A few words

The concept of confidence distributions (CDs) and confidence curves (ccs) have gain interest as they offer a means to work out a $P(\psi \,|\, \text{data})$ in the frequentist framework. This seems to be of interest in the big data world. The concept has been conceptually challenging but notable efforts have been made towards clarification and extensions of definitions. Some of the literature sources include the book by Schweder and Hjort (2016) and the papers by Singh et al. (2005), Singh et al. (2007), Xie and Singh (2013), DeBlasi and Schweder (2016), Schweder (2017) and Hjort and Schweder (2017).

Refinements to CDs and ccs have rarely been considered. Only in Schweder and Hjort (2016, ch. 7) and DeBlasi and Schweder (2016), one can find methods to improve and refine the outcome of a CD or a cc. With this in mind, a suggestion is made towards this direction in section 3 and some results are proven. In section 2, a parametric modus of thinking is adopted, whereas in section 3, a nonparametric modus of thinking drives the discussion. Examples were provided in some occasions.

1.2 Definitions

Let P_θ denote a probability distribution with $\theta = (\psi, \chi)$ being a p-dimensional parameter belonging to some p-dimensional parameter space $\Theta = \Psi \times X$. It is assumed that ψ, the *focus* parameter of interest, is one-dimensional and χ, the *nuisance* parameter vector, is $(p-1)$-dimensional. Let $Y = \{Y_1, \ldots, Y_n\}$ denote a random sample from P_{θ_0}, with θ_0 being the true parameter vector. The sample realization, referred to as the (observed) data, is denoted as $y = y_{\text{obs}} = \{y_1, \ldots, y_n\}$ and the sample space is denoted as $\mathcal{Y}$.

Definition 1.1, which was formulated in Schweder and Hjort (2002) and utilised, for example, in Singh et al. (2005), Singh et al. (2007) and Xie and Singh (2013), formalises the two requirements that constitute the term *confidence distribution*. Schweder and Hjort (2016, ch. 4) extend the definition to accommodate 'less straightforward situations', such as, for example, cases where the range of the focus parameter is bounded or cases of multimodal likelihoods.

Definition 1.1 (Confidence Distribution)**.** A confidence distribution (CD) for the focus parameter ψ is a nondecreasing, right-continuous and data-dependent function $C(\psi, y) : \Psi \times \mathcal{Y} \to [0,1]$ with $(\psi, y) \mapsto \alpha$, where α denotes a confidence level, such that

(i) for any given data $y_{\text{obs}} \in \mathcal{Y}$, $C(\psi, y_{\text{obs}})$ is a cumulative distribution function on Ψ, and

(ii) at the true parameter value ψ_0, $C(\psi_0, Y)$ has the uniform $U(0,1)$ distribution.

In an explanatory effort, point (i) of Definition 1.1 views the CD as a function of the focus parameter ψ and requires the CD to be a distribution function on the parameter space Ψ. It shares many of the attractions of the posterior of the Bayesian paradigm, but it is not a posterior as such. Point (ii) of Definition 1.1 views the CD as a function of the random sample Y and requires the CD to follow the uniform distribution at the true value of the focus parameter, ψ_0. In essence, the requirement in point (ii) of Definition 1.1 renders a CD to have the correct coverage probability. Its importance in the construction of a CD is emphasised in Proposition 1.1 which offers a proof that, indeed, a CD has coverage accuracy.

Proposition 1.1. *Let C^{-1} denote the inverse function of a CD defined in Definition 1.1. Let, also, α_1 and α_2 denote confidence levels such that $\psi_{\alpha_1} = C^{-1}(\alpha_1)$ and $\psi_{\alpha_2} = C^{-1}(\alpha_2)$, where $C^{-1}(\alpha_i) = \inf\{\psi : C(\psi, Y) > \alpha_i\}$, for $i = 1, 2$. Then, a two-sided confidence interval $[\psi_{\alpha_1}, \psi_{\alpha_2}]$ has coverage probability $\alpha_2 - \alpha_1$.*

Proof. It is, first, recognised that a CD is a procedure for constructing random intervals. Then, by letting $\mathcal{I} = [\psi_{\alpha_1}, \psi_{\alpha_2}]$ denote such an interval, the probability that the random $\mathcal{I}$ will contain or cover the true parameter value, ψ_0, is

$$\begin{aligned} P_\theta\{\psi_0 \in \mathcal{I}\} &= P_\theta\{\psi_{\alpha_1} \leq \psi_0 \leq \psi_{\alpha_2}\} \\ &= P_\theta\{C(\psi_{\alpha_1}, Y) \leq C(\psi_0, Y) \leq C(\psi_{\alpha_2}, Y)\} \quad (1.1) \\ &= P_\theta\{\alpha_1 \leq C(\psi_0, Y) \leq \alpha_2\} \quad (1.2) \\ &= P_\theta\{C(\psi_0, Y) \leq \alpha_2\} - P_\theta\{C(\psi_0, Y) \leq \alpha_1\} \\ &= P\{\mathcal{U} \leq \alpha_2\} - P\{\mathcal{U} \leq \alpha_1\} \quad (1.3) \\ &= \alpha_2 - \alpha_1, \end{aligned}$$

where $\mathcal{U}$ denotes a random variable from $U(0,1)$ which is independent of θ. It is noticed that equations (1.1) and (1.2) follow from the properties described in Definition 1.1, i.e. that the CD is nondecreasing, right-continuous and data-dependent function that maps a tuple of the form (ψ, y) to a corresponding confidence level. Since $C(\psi_0, Y)$ is a random variable, equation (1.3) follows from point (ii) of Definition 1.1. It is clarified that $\theta = (\psi, \chi)$. The nuisance parameter vector χ is considered fixed but random. The proof is completed.

Corollaries 1.1 and 1.2 deal with one-sided intervals.

Corollary 1.1. *Let α denote a confidence levels such that $\psi_\alpha = C^{-1}(\alpha) = \inf\{\psi : C(\psi, Y) > \alpha\}$, where C^{-1} denote the inverse function of a CD defined in Definition 1.1. Then, a one-sided confidence interval of the form $(-\infty, \psi_\alpha]$ has coverage probability α.*

Proof. See Schweder and Hjort (2016, p. 59).

Corollary 1.2. *Let α denote a confidence levels such that $\psi_\alpha = C^{-1}(\alpha) = \inf\{\psi : C(\psi, Y) > \alpha\}$, where C^{-1} denote the inverse function of a CD defined in Definition 1.1. Then, a one-sided confidence interval of the form $[\psi_\alpha, \infty)$ has coverage probability $1 - \alpha$.*

Proof. The arguments are, essentially, the same as in the proof of Proposition 1.1 and summarised in

$$\begin{aligned} P_\theta\{\psi_0 \geq \psi_\alpha\} &= 1 - P_\theta\{\psi_0 < \psi_\alpha\} \quad (1.4) \\ &= 1 - P_\theta\{C(\psi_0, Y) < C(\psi_\alpha, Y)\} \\ &= 1 - P_\theta\{C(\psi_0, Y) < \alpha\} \\ &= 1 - P_\theta\{\mathcal{U} < \alpha\} \\ &= 1 - P\{\mathcal{U} \leq \alpha\} \\ &= 1 - \alpha, \end{aligned}$$

where $\mathcal{U}$ denotes a random variable from $U(0,1)$ which is independent of θ. It is noticed that equation (1.4) follows from basic and standard properties of probability distributions. The proof is completed.

Definition 1.2 handles the situation when point (ii) of Definition 1.1 applies only at the limit. Definition 1.2 formalises the discussion in Singh et al. (2005, p. 160) or Singh et al. (2007, p. 133) and extends Definition 3.1 of Schweder and Hjort (2016, p. 58).

Definition 1.2 (Asymptotic Confidence Distribution)**.** An asymptotic confidence distribution (aCD) for the focus parameter ψ is a nondecreasing and data-dependent function $C(\psi, y) : \Psi \times \mathcal{Y} \to [0, 1]$ with $(\psi, y) \mapsto \alpha$, where α denotes a confidence level, such that

(i) for any given data $y_{\text{obs}} \in \mathcal{Y}$, $C(\psi, y_{\text{obs}})$ is a cumulative distribution function on Ψ, and

(ii) at the true parameter value ψ_0, $C(\psi_0, Y_{1:n}) \longrightarrow_d U(0,1)$ as $n \to \infty$.

The dependence of $C(\psi_0, Y)$ on the sample size n is noticed by writing $Y_{1:n}$ for the sample Y.

Definition 1.3 is given in Schweder and Hjort (2016, p. 33) and repeated here for completeness and coherence of presentation.

Definition 1.3 (Pivot)**.** A function $\text{piv}(Y, \psi)$ of the data Y and the (unknown) focus parameter ψ is a pivot if its distribution function is independent of the full underlying parameter θ.

There are, though, cases where the distribution of a pivotal quantity is only independent of the underlying parameter at the limit. Definition 1.4 is relevant.

Definition 1.4 (Approximate Pivot)**.** A function $\text{piv}(Y, \psi)$ of the data Y and the (unknown) focus parameter ψ is an approximate pivot if its *limiting* distribution function is independent of the full underlying parameter θ. That is, $\text{piv}(Y, \psi)$ is independent of θ only asymptotically.

A more formal definition of what is called here an *approximate pivot* is given in Hall (1992b, p. 14) as follows: '*A function* $\text{piv}(Y, \psi)$ *is asymptotically pivotal if, for sequences of known constants* $\{a_n\}$ *and* $\{b_n\}$, $a_n \,\text{piv}(Y, \psi) + b_n$ *has a proper nondegenerate limiting distribution not depending on unknowns*'.

Another definition needed for subsequent discussion is the one referring to the term *confidence curve*. The definition is given in Schweder and Hjort (2016, p. 115) as Definition 4.3 and is presented, here, for completeness.

Definition 1.5 (Confidence Curve)**.** A confidence curve $\text{cc}(\psi, y) : \Psi \to [0,1]$ has as its level sets a nested family of confidence regions $R_\alpha(Y) = \{\psi : \text{cc}(\psi) \leq \alpha\}$ in Ψ, with $\alpha \in [0,1]$ being the confidence level. The confidence curve has these properties: (i) $\min_{\psi} \text{cc}(\psi, y) = \text{cc}(\widehat{\psi}(y), y) = 0$ for all outcomes of the data y, where $\widehat{\psi}(y)$ is a point estimate, (ii) $\text{cc}(\psi_0, Y)$ has a uniform distribution on the unit interval, when ψ_0 is the true value of the parameter.

A cc is always available, since it can be derived by using the probability integral transform. On the contrary, a CD is not always available. See, for example, the Fieller interval example (or Example 4.7) in Schweder and Hjort (2016, p. 118). But when a CD is available, a cc can be constructed as

$$cc(\psi, y) = |1 - 2\,C(\psi, y)| = \begin{cases} 1 - 2\,C(\psi, y) & \text{if } \psi \leq \widehat{\psi}_{.50} \\ 2\,C(\psi, y) - 1 & \text{if } \psi \geq \widehat{\psi}_{.50}, \end{cases} \tag{1.5}$$

where $\widehat{\psi}_{.50} = C^{-1}(\frac{1}{2})$ is the median of the confidence distribution. For a chosen confidence level, α, the two solutions of equation $cc(\psi, y_{\text{obs}}) = \alpha$, namely, $\psi_{\text{low}}(\alpha)$ and $\psi_{\text{up}}(\alpha)$, are such that

$$\text{confidence of } [\psi_{\text{low}}(\alpha)\,,\, \psi_{\text{up}}(\alpha)] = \alpha,$$

where $\psi_{\text{low}}(\alpha)$ is located to the left of $\widehat{\psi}_{.50}$ and $\psi_{\text{up}}(\alpha)$ is located to the right of $\widehat{\psi}_{.50}$. By (1.5),

$$C(\psi_{\text{low}}(\alpha), y_{\text{obs}}) = \tfrac{1}{2}(1 - \alpha) \qquad \text{and} \qquad C(\psi_{\text{up}}(\alpha), y_{\text{obs}}) = \tfrac{1}{2}(1 + \alpha).$$

2 Methods for good CD and cc approximation

This section reviews some of the techniques presented in Schweder and Hjort (2016) regarding (i) first-order large sample likelihood methods and (ii) improved approximations for confidence distributions. It draws, mainly, from chapters 2 and 7 of the book and the STK4180 course at UiO. Some examples are presented to clarify concepts and ideas and a few comments are given regarding each method, where appropriate. In the literature, no other, known, source deals with this aspect of confidence distributions.

2.1 Using distribution approximations for the maximum likelihood estimator

A confidence distribution and a confidence curve is sought to be constructed by using distribution approximations for the maximum likelihood estimators. But, first, a preliminary discussion is deemed necessary.

Let $Y = \{Y_1, \ldots, Y_n\}$ be a random sample from a parametric model with possible covariates $x_1, \ldots, x_n$ and unknown parameter vector $\theta = (\theta_1, \ldots, \theta_p)^t$, belonging to the parametric space $\Theta \in \mathbb{R}^p$. Then, in the case of an independently and identically distributed (i.i.d.) random sample, the simultaneous density for the full data set Y is given by $f_{\text{joint}}(y, \theta)$ and in the case, when covariate information is incorporated, the conditional density of $Y_i|x_i$ is given by $f_{\text{joint}}(y_i|x_i, \theta)$.

The *likelihood function*, $L(\theta)$, is defined to be the joint density when $Y = y_{\text{obs}}$. That is, for observed data, y_{obs}, the likelihood is viewed as a function of the parameter, θ. The *log-likelihood function* is defined as $\ell(\theta) = \log L(\theta)$. The maximum likelihood estimator, $\widehat{\theta}$, is the value of θ that maximises the likelihood function or, equivalently, the log-likelihood function. An important property of the maximum likelihood principle is the invariance with respect to, both, data transformation and parameter transformation. Under mild regularity conditions the surface of the log-likelihood is approximately quadratic.

Theorem 2.2 in Schweder and Hjort (2016, p. 27) gives distribution approximations for the maximum likelihood estimator, in the i.i.d. situation, and constitutes the basis for the simplest approximation to the distribution of a given focus parameter, ψ. The theorem is restated as Theorem 2.1, for coherence of presentation and the reader is referred to the book for further details.

Theorem 2.1 (Schweder and Hjort (2016))**.** *In the i.i.d. situation, let $\widehat{\theta} = \widehat{\theta}_n$ be the maximum likelihood estimator based on the first n observations. If the model holds, with θ_0 the true parameter, being an inner point of the parameter space, and with variance matrix at θ_0, $J(\theta_0)$, being of full rank, then under mild further regularity assumptions, as sample size n tends to infinity,*

$$\begin{aligned} \sqrt{n}(\widehat{\theta} - \theta_0) &\to_d J^{-1}U \sim N_p(0, J^{-1}), \\ D_n(\theta_0) = 2\{\ell_n(\widehat{\theta}) - \ell_n(\theta_0)\} &\to_d U^t J^{-1} U \sim \mathcal{X}^2_p, \end{aligned} \tag{2.1}$$

where $U \sim N_p(0, J)$.

The result at (2.1) still holds in a regression context, where, for each $i = 1, \ldots, n$, the random variable Y_i has density $f(y|x_i, \theta)$ for some covariate vector x_i. In practice, a consistent estimator of the limit distribution variance matrix is needed. Using Slutsky's theorem, the result at (2.1) remains true if $J(\theta_0)$ is replaced by a consistent estimator of choice. See, Schweder and Hjort (2016, p. 29), for two such estimators and discussion regarding the use of the observed rather that the expected Fisher information matrix.

For a sample of large or moderately large size, the following notes regarding the maximum likelihood estimator, $\widehat{\theta}$, are in order.

1. It is approximately unbiased.

2. Its distribution is approximately multinormal. Thus, by the properties of the normal distribution, the single components, $\widehat{\theta}_j$, and linear combinations of them are approximately normal. Also, any desired confidence procedures, such as confidence intervals and confidence distributions or curves, can, thus, be constructed easily with coverage and significance levels close to any values intended.

3. Its variance matrix achieves the Cramér-Rao lower bound for unbiased estimators, and is approximately equal to $J^{-1}(\theta_0)/n$. Thus, asymptotically, this is the best estimation strategy and no other alternatives perform better.

4. Its precision and associated confidence regions can, easily, be read off using the fact that, for any given or predetermined α, the set $\{\theta : D_n(\theta) \leq \Gamma_p^{-1}(\alpha)\}$, where the inverse Γ_p^{-1} denotes the quantile function of the $\mathcal{X}_p^2$ distribution, covers the true θ_0 with probability tending to α. Thus, it is not necessary to know or compute the variance matrix.

A note of warning is in order too. Firstly, in situations with many parameters, (i) the convergence towards the limit distribution may be slow and (ii) the implied approximation to the variance matrix of $\widehat{\theta}$, i.e. $J(\theta_0)/n$, may need modification and improvements. And, secondly, for a sample of small or moderately small size, there might be cases when the behaviour of $\widehat{\theta}$ is not to be trusted. Example **??** is such a case. Note, that the purpose of this example is to (i) demonstrate the construction of confidence distributions and confidence curves and (ii) compare the methods employed to achieve this aim. Section 3, offers a way to overcome the deficiency of $\widehat{\theta}$ in small or moderately sample size samples.

In 1.2, it is required that the focus parameter be one-dimensional. Thus, let $\psi = a(\theta) = a(\theta_1, \ldots, \theta_p)$ be the focus parameter. From the invariance property of maximum likelihood (ML) estimators,

$$\widehat{\psi}_{\text{ML}} = a(\widehat{\theta}_{\text{ML}}). \tag{2.2}$$

The so-called *delta method* gives, for $a(\theta_0)$ having smooth first-order derivatives in the p-parameters at θ_0,

$$\sqrt{n}\,(\,a(\widehat{\theta}) - a(\theta_0)\,) \to_d w^t Z = \sum_{j=1}^{p} w_j Z_j,$$

where $w = \partial a(\theta_0)/\partial\theta$, i.e. $w_j = \partial a(\theta_0)/\partial\theta_j$, for $j = 1, \ldots, p$ and $Z \sim \text{N}_p(0, J^{-1}(\theta_0))$. Concisely,

$$\sqrt{n}\,(\,\widehat{\psi} - \psi\,) \to_d w^t Z \sim \text{N}(0, \kappa^2) \qquad \text{where} \qquad \kappa^2 = w^t J^{-1}(\theta_0) w. \tag{2.3}$$

Thus, for any focus parameter $\psi = a(\theta)$, (2.3) constitutes a convenient and general large-sample recipe for constructing confidence procedures, such as confidence intervals and confidence curves and performing one- or two-sided hypothesis tests. By rearranging (2.3), write

$$V_n = \sqrt{n}\,(\,\widehat{\psi} - \psi\,)/\widehat{\kappa} \to_d \text{N}(0, 1) \tag{2.4}$$

and note that $\widehat{\kappa}$ is any consistent estimator of κ, such as $\widehat{\kappa} = \widehat{w}^t \widehat{J}^{-1} \widehat{w}$, with $\widehat{J}$ the Hessian matrix $-n^{-1}\partial^2 \ell_n(\widehat{\theta})/\partial\theta\partial\theta^t$ as output from the log-likelihood maximisation operation. Note that, $\widehat{w} = w(\widehat{\theta})$ which may be computed numerically if that is required. Asymptotically, V_n is a pivot, a property that makes this recipe even more attractive. That is, $V_n = V_n(Y, \psi)$ with a limiting distribution independent of ψ.

It is noticed that V_n at (2.4) is decreasing in ψ. By the theory developed in either Singh et al. (2007, p. 134) or Schweder and Hjort (2016, pp. 58–59), an asymptotic confidence distribution can be constructed as

$$\begin{aligned} C_n(\psi) &= 1 - \Phi(V_n(Y, \psi)) \\ &= 1 - \Phi(V_n), \end{aligned} \tag{2.5}$$

where Φ is the cumulative distribution function of a Standard Normal. By the probability integral transform, $\Phi(V_n)$ is uniformly distributed and, thus, $1 - \Phi(V_n)$ is uniformly distributed too. By Definition **??**, the $C_n(\psi)$ at (2.5) is the cumulative distribution function of a confidence distribution for ψ. Alternatively, the asymptotic confidence distribution at (2.5), is written as

$$C_{n,\text{alt}}(\psi) = \Phi\left(\frac{\sqrt{n}\,(\,\psi - \widehat{\psi}\,)}{\widehat{\kappa}}\right), \tag{2.6}$$

indicating that the asymptotic confidence distribution, $C_{n,\text{alt}}(\psi)$, is, simply, normally distributed with mean at the maximum likelihood estimator, $\widehat{\psi}$, and standard deviation $\widehat{\kappa}/\sqrt{n}$. Remark 3.3 in Schweder and Hjort (2016, p. 70), clarifies why it is preferable to work in terms of 'estimand minus estimator', thus, writing down confidence distributions in the form given at (2.6).

Theorem 2.1, enables the availability of a confidence distribution under the specified conditions. Thus, using (1.5), a confidence curve can also be constructed as

$$cc_n(\psi) = |1 - 2\,C_n(\psi)|.$$

2.2 Chi-squared approximation for the deviance

The $\widehat{\psi}$ estimator at (2.2) maximises, also, the *profile log-likelihood* given in

$$\ell_{n,\text{prof}}(\psi) = \max\{\ell_n(\theta) : a(\theta) = \psi\}.$$

Of course, $\widehat{\psi}$ maximises the *profile likelihood*, $L_{n,\text{prof}}(\psi) = \max\{L_n(\theta) : a(\theta) = \psi\}$, but it is preferred that the subsequent discussion be unfolded under the log-profile likelihood.

The *profile deviance* for a focus parameter ψ under consideration is defined to be

$$D_n(\psi) = 2\{\ell_{n,\text{prof}}(\widehat{\psi}) - \ell_{n,\text{prof}}(\psi)\}.$$

It is noticed that, $D_n(\psi)$ may be viewed as, both, (i) a curve computed from the observed data and (ii) a random variable for a given ψ value. Also, D_n is the twice log-likelihood-ratio statistic for testing $H_0 : \psi = \psi_0$ against $H_1 : \psi \neq \psi_0$, in that

$$D_n(\psi_0) = 2\,\log\frac{\max_{\text{all }\theta}\,L_n(\theta)}{\max_{\text{all }\theta\,:\,a(\theta)=\psi_0}\,L_n(\theta)} \tag{2.7}$$

Note that, under the null hypothesis the parameter dimension is $p-1$ and under the alternative hypothesis the parameter dimension is p. The deviance at (2.7) is a scale factor away from what is typically introduced in the literature about the deviance. See Schweder and Hjort (2016, p. 35), for further comments.

Theorem 2.4 in Schweder and Hjort (2016, p. 35) gives a chi-squared approximation for the profile deviance, in the i.i.d. situation, and provides another recipe to construct an asymptotic confidence distribution.

The theorem is restated as Theorem 2.2, for coherence of presentation and the reader is referred to the book for further details.

Theorem 2.2 (Schweder and Hjort (2016))**.** *Under conditions of the model and those described for Theorem 2.1, and under the true parameter θ_0 (so that the true value of the one-dimensional parameter ψ is $\psi_0 = a(\theta_0)$), assumed to be an inner point in the parameter space,*

$$D_n(\psi_0) = 2\{\ell_{n,\text{prof}}(\widehat{\psi}) - \ell_{n,\text{prof}}(\psi_0)\} \to_d \mathcal{X}^2_1. \tag{2.8}$$

Alternatively, the result at (2.8) is referred to as '*Wilks theorem*'.

Asymptotically, $D_n(\psi)$ is a pivot. That is, $D_n(\psi) = D_n(Y, \psi)$. Note that, it is a nonmonotone pivot. The statement at (2.8) depicts that, under the true parameter value θ_0, the distribution function of $D_n(\psi_0) = D_n(a(\theta_0))$, say K_n, converges to the distribution of a $\mathcal{X}^2_1$.

It is, now, shown how to derive an asymptotic confidence distribution. Essentially, the required confidence distribution is derived from the equitailed confidence intervals $\{\psi : D_n(\psi) \leq K_n^{-1}(\alpha)\}$, where α denotes a given or predefined probability level, by solving $D_n(\psi) = K_n^{-1}(\alpha)$. In the class of all well-behaved cases, the solution has two roots. Thus, define the signed deviance squared root by

$$r_n(\psi) = \text{sgn}(\psi - \widehat{\psi})\, D_n^{1/2}(\psi)$$

and note that it is monotone in ψ. Then, by letting H_n denote the distribution function of $r_n(\psi)$, construct the asymptotic confidence distribution as

$$C_n(\psi) = H_n(\text{sgn}(\psi - \widehat{\psi})\, D_n^{1/2}(\psi)).$$

It is emphasised that the signed deviance squared root is close to being a pivot and, as such, H_n does not depend on the parameter θ. Taking into account that the first-order asymptotic approximation regarding H_n corresponds to $H_n \to \Phi$, a simpler confidence distribution is derived as

$$C_{n,\text{smp}}(\psi) = \Phi(\text{sgn}(\psi - \widehat{\psi})\, D_n^{1/2}(\psi)).$$

A confidence curve is given by the $\mathcal{X}^2_1$ probability scaling as in

$$cc_n(\psi) = \Gamma_1(D_n(\psi)), \tag{2.9}$$

where Γ_1 denotes the $\mathcal{X}^2_1$ distribution function. The method at (2.9), often, has better precision that the one described in section 2.1. See, Schweder and Hjort (2016, p. 70) for further details.

2.3 Bartlett corrections for the deviance

Sections 2.1 and 2.2, reviewed two general likelihood based methods for constructing confidence distributions and confidence curves. Their common denominator is that both results rely on large-sample theory and that, in practice, the sample size need be at least moderately large. Here, an improvement to the finite sample behaviour of the deviance statistic is sought. The so-called '*Bartlett correction*' offers a way and its underlying idea is, thus, described and applied. A *Bartlett corrected deviance confidence curve* (Schweder and Hjort, 2016, p. 211) is, then, derived.

In Bartlett (1937, 1954) an improved likelihood ratio (LR) statistic with expected value closer to the expected value of a $\mathcal{X}^2_q$ distribution, where q is the difference between the dimensions of the parameter spaces under the alternative and null hypotheses, was suggested. Finding the exact distribution of the likelihood ratio statistic or at least a good approximation to it, when the null hypothesis is true, is challenging. The expected value of the improved likelihood ratio statistic, though, is closer to the expected value of the $\mathcal{X}^2_q$ distribution, whereas the likelihood ratio statistic's, itself, is not. Specifically, let the expected value of the likelihood ratio statistic, under the null hypothesis, be $E(LR) = q\{1 + c/n + O(n^{-2})\}$, where c is some constant that can be consistently estimated under the null hypothesis and n is, as usual, the sample size. Then, define the new likelihood statistics, LR^*, as $LR^* = LR/(1 + c/n)$ and notice that $E(LR^*) = q + O(n^{-2})$. This, 'diving by the right constant to get the mean right' (Schweder and Hjort, 2016, p. 211) type of modification is called the Bartlett correction. A generalisation of this idea is found in Lawley (1956). Also, Cordeiro and Cribari-Neto (2014) offer an introduction to the various aspects of Bartlett and Bartlett type corrections.

In the context of confidence curves, let a parametric model indexed by θ be considered. The focus parameter is $\psi = a(\theta)$. In section 2.2 a confidence curve for ψ was constructed via the $\mathcal{X}^2_1$ approximation to the distribution of the deviance. In Schweder and Hjort (2016, p. 210), it is noted that the accuracy of the $cc(\psi) = \Gamma_1(D_n(\psi))$ method, where Γ_1 denotes the $\mathcal{X}^2_1$ distribution function, depends on how effective the implied approximation $P_\theta\{cc(\psi) \le \alpha\} = P_\theta\{D_n(\psi) \le \Gamma^{-1}(\alpha)\} \doteq \alpha$, where α denotes a given or predefined probability level, is. The effectiveness of the implied approximation might depend on (i) the sample size n, (ii) the position of θ in the parameter space and the probability level α. Section 7.2 of Schweder and Hjort (2016) develops second-order theory for the deviance $D_n(\psi)$, in which, for a suitable $c(\theta)$,

$$E_\theta D_n(\psi) = 1 + c(\theta)/n + O(n^{-2}).$$

By letting $D^*_n(\psi) = D_n(\psi)/(1 + \varepsilon)$ be the new modified deviance and noting that $1 + \varepsilon$ is the mean of $D_n(\psi)$, a new confidence curve termed Bartlett corrected deviance confidence curve is defined as

$$cc^*(\psi) = \Gamma_1\left(\frac{D_n(\psi)}{E_\theta D_n(\psi)}\right) = \Gamma_1\left(\frac{D_n(\psi)}{1+\varepsilon}\right). \tag{2.10}$$

Note that the mean of $D_n(\psi)$, i.e. $1 + \varepsilon$, can be computed by simulation. Since the ε may depend on θ, the replicates of $D_n(\psi)$ should be computed at the required position θ in the parametric space. Implicitly, in (2.10) is that the distribution function of $D_n(\psi)$ is better approximated by $\Gamma_1(x/(1 + \varepsilon))$ than by $\Gamma_1(x)$, itself. For large n, theory secures that $\varepsilon \to 0$ and that $n\{E_\theta D_n(\psi) - 1\}$ converges to $c(\theta)$.

2.4 Examples

In this section some examples are provided to demonstrate the machinery and properties of confidence distributions and confidence curves. The examples are exercises retrieved from the STK4180 course website (http://www.uio.no/studier/emner/matnat/math/STK4180/). While this project was written, no solutions to these exercises were made available on the website. The R code accompanying the examples is given in the Appendix, see I.1 for Example 2.1, I.2 for Example 2.2, I.3 for Example 2.3 and I.4 for Examples 2.4 and 2.5.

Example 2.1 (A skewed distribution on the unit interval)**.** Let θ be a positive and unknown parameter. Consider a model with density $f(y, \theta) = \theta\, y^{\theta-1}$ for data on $[0, 1]$. The aim is to construct confidence distributions.

The log-likelihood function is given by

$$\ell_n(\theta) = \log L_n(\theta) = n \log \theta + (\theta - 1) \sum_{i=1}^{n} \log y_i$$

and, by differentiation with respect to θ, the maximum likelihood estimator is given by

$$\widehat{\theta}_n = -n \left(\sum_{i=1}^{n} \log Y_i \right)^{-1}.$$

A further differentiation of the log-likelihood function with respect to θ, gives the variance matrix as

$$J_n(\theta) = -E_\theta \frac{\partial^2 \ell(\theta)}{\partial \theta^2} = \frac{n}{\theta^2}.$$

By letting θ be the one-dimensional focus parameter, Theorem 2.1 depicts that $\sqrt{n}(\widehat{\theta} - \theta_0) \to_d \mathrm{N}(0, J^{-1})$, where θ_0 is the true parameter and J^{-1} needs to be consistently estimated. Note that, under the notation used presently, $J = J_n/n$. Let $\widehat{\kappa}^2 = J^{-1}(\widehat{\theta})$ be the required consistent estimator. Then, by Definition 1.1, an asymptotic confidence distribution is constructed as in

$$C_1(\theta) = 1 - \Phi \left(\frac{\widehat{\theta} - \theta}{\widehat{\kappa}/\sqrt{n}} \right).$$

Theorem 2.1, also, depicts that $D_n(\theta_0) = 2\{\ell_n(\widehat{\theta}) - \ell_n(\theta_0)\} \to_d \mathcal{X}^2_1$, where θ_0 is the true parameter. Hence, a second asymptotic confidence distribution is constructed as in

$$C_2(\theta) = 1 - \Phi(\mathrm{sgn}(\widehat{\theta} - \theta) D_n^{1/2}(\theta)).$$

For this model, the exact confidence distribution $C(\theta) = P_\theta\{\widehat{\theta} \geq \widehat{\theta}_{\text{obs}}\}$ can be computed using (i) exact probability calculus and (ii) simulations.

(i) Exact probability calculus

$$\begin{aligned} C_3(\theta) &= P_\theta\{\widehat{\theta} \geq \widehat{\theta}_{\text{obs}}\} \\ &= 1 - P_\theta \left\{ -n \left(\sum_{i=1}^{n} \log Y_i \right)^{-1} \leq \widehat{\theta}_{\text{obs}} \right\} \\ &= P_\theta \left\{ \sum_{i=1}^{n} (-\log Y_i) \leq \frac{n}{\widehat{\theta}_{\text{obs}}} \right\} \end{aligned}$$

Let $X = -\log Y$, generically. Then,

$$F(x) = P\{X \leq x\} = P\{-\log Y \leq x\} = P\{\log Y \geq -x\} = P\{Y \geq e^{-x}\} = 1 - P\{Y \leq e^{-x}\} = 1 - e^{-\theta x},$$

which is the cumulative distribution function of an exponential with parameter θ. That is, X follows an $\mathrm{Expo}(\theta) \equiv \mathrm{Gamma}(1, \theta)$.

Thus, $\sum_{i=1}^{n} (-\log Y_i)$ follows a $\mathrm{Gamma}(n, \theta)$.

(ii) Simulations

The probability integral transform is used. From the density of the model we have that $F(y,\theta) = y^{\theta}$. Let U be a random variable from the uniform $\mathcal{U}(0,1)$ distribution. Then, $F(Y) = U$ and $Y = U^{1/\theta}$. The confidence distribution is constructed as in

$$C_4(\theta) = \frac{1}{B}\sum_{j=1}^{B} I\{\widehat{\theta}_{\text{sim},j} \geq \widehat{\theta}_{\text{obs}}\},$$

where B is a suitably large number and $\widehat{\theta}_{\text{obs}} = -n\left(\sum_{i=1}^{n}\log y_i\right)^{-1}$.

All the above confidence distributions need to be evaluated for a grid of θ values. Given the data in appendix II.1, the four confidence distributions are plotted in Figure 2.1.

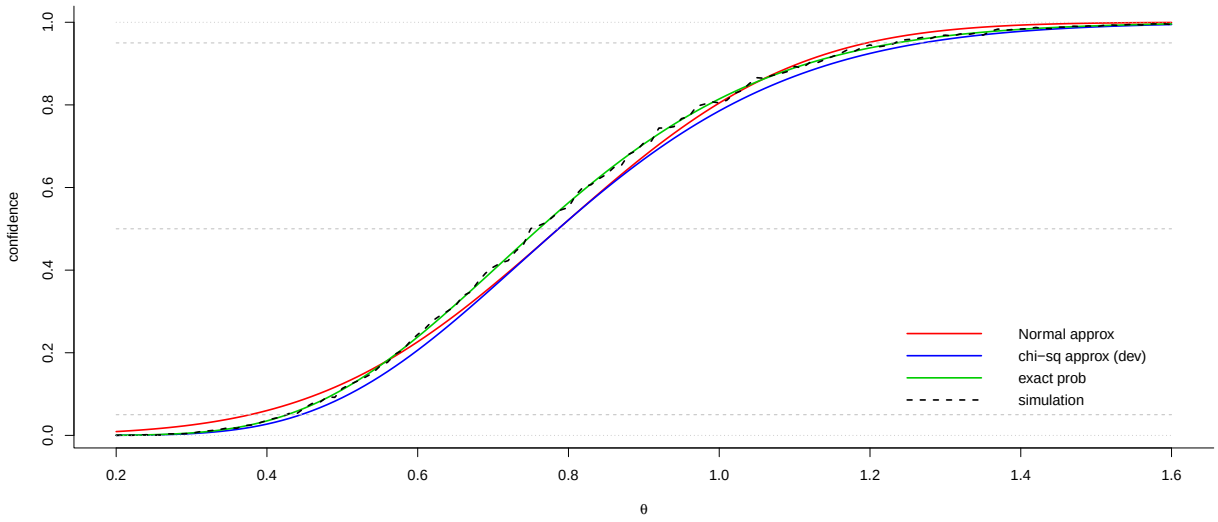

Figure 2.1: Confidence distributions for the data set given in II.1.

Example 2.2 (Light thinks it travels faster than anything)**.** Let

$$F_0(x) = \frac{\exp(x)}{1+\exp(x)} \qquad \text{and} \qquad f_0(x) = \frac{\exp(x)}{\{1+\exp(x)\}^2}.$$

Using maximum likelihood, the model with cumulative distribution function

$$F(y, \xi, \tau) = F_0\left(\frac{y-\xi}{\tau}\right)$$

is fitted to the data given in appendix II.2. The aim is to construct confidence curves for $p = P_{\xi,\tau}(Y \leq y_0)$, with $y_0 = 30.5$.

Fitting the data

The cumulative function $F_0(x)$ and the density $f_0(x)$ correspond to the cumulative function and density of the logistic distribution with parameters $(0, 1)$, which resembles the standard normal distribution but it has heavier tails.

The cumulative distribution function $F(y, \xi, \tau) = F_0\left(\frac{y-\xi}{\tau}\right)$ corresponds to the general logistic distribution with parameter (ξ, τ) and is written as

$$F(y, \xi, \tau) = F_0\left(\frac{y-\xi}{\tau}\right) = \frac{\exp(\frac{y-\xi}{\tau})}{1+\exp(\frac{y-\xi}{\tau})}$$

with density

$$\begin{aligned} f(y,\xi,\tau) &= \frac{\partial F(y,\xi,\tau)}{\partial y} \\ &= \left[\frac{1}{\tau}\exp\left(\frac{y-\xi}{\tau}\right)\left\{1+\exp\left(\frac{y-\xi}{\tau}\right)\right\} - \frac{1}{\tau}\exp\left(\frac{y-\xi}{\tau}\right)\exp\left(\frac{y-\xi}{\tau}\right)\right] \Big/ \left\{1+\exp\left(\frac{y-\xi}{\tau}\right)\right\}^2 \\ &= \frac{1}{\tau}\frac{\exp\left(\frac{y-\xi}{\tau}\right)}{\left\{1+\exp\left(\frac{y-\xi}{\tau}\right)\right\}^2} \\ &= \frac{1}{\tau} f_0\left(\frac{y-\xi}{\tau}\right). \end{aligned}$$

The likelihood function is

$$\begin{aligned} \mathcal{L}(\xi,\tau) &= \prod_{i=1}^{n} f(y_i;\xi,\tau) \\ &= \prod_{i=1}^{n} \frac{1}{\tau}\frac{\exp\left(\frac{y_i-\xi}{\tau}\right)}{\left\{1+\exp\left(\frac{y_i-\xi}{\tau}\right)\right\}^2} \\ &= \left(\frac{1}{\tau}\right)^n \prod_{i=1}^{n} \frac{\exp\left(\frac{y_i-\xi}{\tau}\right)}{\left\{1+\exp\left(\frac{y_i-\xi}{\tau}\right)\right\}^2}. \end{aligned}$$

and the log-likelihood function is

$$\begin{aligned}\ell(\xi,\tau) &= \log \mathcal{L}(\xi,\tau) \\ &= -n\log(\tau) + \frac{1}{\tau}\sum_{i=1}^{n}(y_i-\xi) - 2\sum_{i=1}^{n}\log\left\{1+\exp\left(\frac{y_i-\xi}{\tau}\right)\right\}.\end{aligned}$$

The model is fitted to the data using maximum likelihood. The negative log-likelihood function, $-\ell(\xi,\tau)$, is minimized using numerical optimization. The maximum likelihood estimates are found to be $(\widehat{\xi},\widehat{\tau}) = (27.61808, 2.838543)$. The associated standard deviation estimated for $(\widehat{\xi},\widehat{\tau})$ are found by inverting the Hessian matrix and applying the square root to the diagonal elements. These are 0.6157995 for $\widehat{\xi}$ and 0.2971452 for $\widehat{\tau}$.

Constructing confidence curves for $p = \Pr_{\xi,\tau}\{Y \le y_0\}$, with $y_0 = 30.5$

(i) The delta method is described in Schweder and Hjort (2016, p. 33). For a confidence distribution and a subsequent confidence curve based on the delta method, it is necessary to find

$$w_1 = \frac{\partial F(y,\xi,\tau)}{\partial \xi} \qquad \text{and} \qquad w_2 = \frac{\partial F(y,\xi,\tau)}{\partial \tau}.$$

These are

$$w_1 = \frac{\partial F(y,\xi,\tau)}{\partial \xi} = \frac{\partial}{\partial \xi}\frac{\exp\left(\frac{y-\xi}{\tau}\right)}{1+\exp\left(\frac{y-\xi}{\tau}\right)} = -\frac{1}{\tau}\frac{\exp\left(\frac{y-\xi}{\tau}\right)}{\left\{1+\exp\left(\frac{y-\xi}{\tau}\right)\right\}^2} = -\frac{1}{\tau}f_0\left(\frac{y-\xi}{\tau}\right) = -f(y,\xi,\tau)$$

and

$$\begin{aligned}w_1 &= \frac{\partial F(y,\xi,\tau)}{\partial \tau} = \frac{\partial}{\partial \tau}\frac{\exp\left(\frac{y-\xi}{\tau}\right)}{1+\exp\left(\frac{y-\xi}{\tau}\right)} = -\frac{y-\xi}{\tau^2}\frac{\exp\left(\frac{y-\xi}{\tau}\right)}{\left\{1+\exp\left(\frac{y-\xi}{\tau}\right)\right\}^2} = -\frac{y-\xi}{\tau^2}f_0\left(\frac{y-\xi}{\tau}\right) \\ &= -\frac{y-\xi}{\tau}\frac{1}{\tau}f_0(y,\xi,\tau) = -\frac{y-\xi}{\tau}f(y,\xi,\tau)\end{aligned}$$

Now, let $w = (w_1, w_2)^t$. Then, the confidence distribution is evaluated at a grid of p values as

$$C_{\text{nrm}}(p) = \Phi\left(\frac{p-\widehat{p}}{\widehat{k}}\right),$$

where $\widehat{p} = F_0\left(\frac{y_0-\widehat{\xi}}{\widehat{\tau}}\right)$, $\widehat{k} = \widehat{w}^t\widehat{J}^{-1}\widehat{w}$, with $\widehat{J}$ the Hessian matrix as given in R, and

$$\widehat{w} = \left(-f(y,\widehat{\xi},\widehat{\tau}),\, -\frac{y-\widehat{\xi}}{\widehat{\tau}}f(y,\widehat{\xi},\widehat{\tau})\right)^t.$$

The corresponding confidence curve is $cc_{\text{nrm}}(p) = |1-2C_{\text{nrm}}(p)|$, evaluated at the same grid of p values. Figure 2.2 displays the confidence curve.

(ii) The deviance method is described in Schweder and Hjort (2016, p. 35). For a confidence distribution and a subsequent confidence curve based on the deviance function $D(p)$ for p, it is necessary to consider the

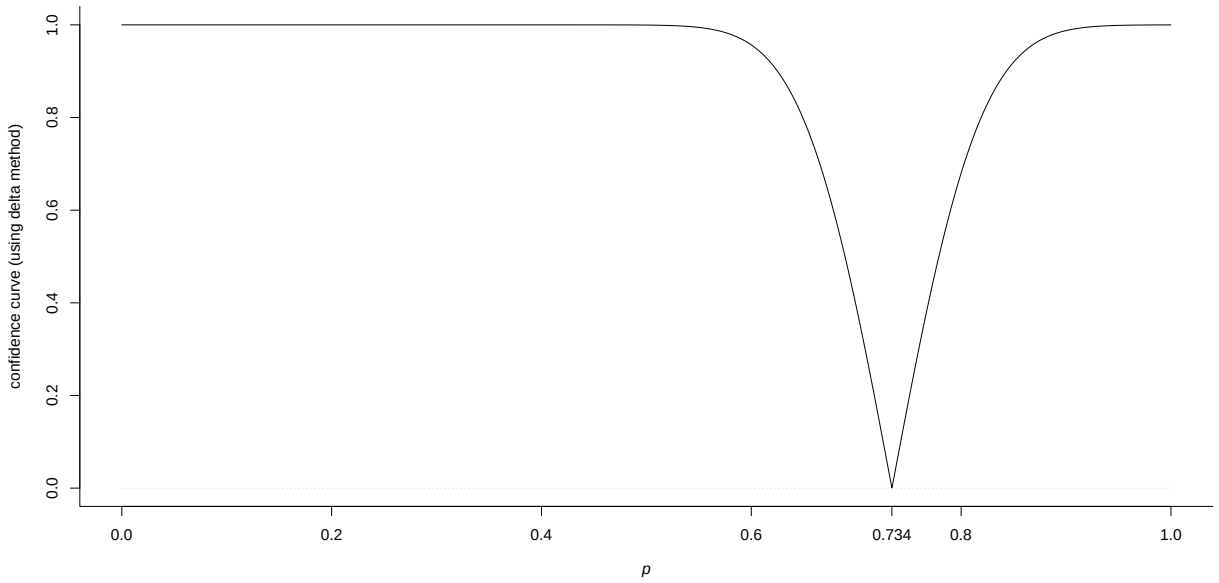

Figure 2.2: Confidence curve using the delta method.

profile log-likelihood modified in such a way so that it takes into account the constraint $F(y_0, \xi, \tau) = p$. This constrain is written as

$$\begin{aligned}
F(y_0, \xi, \tau) &= p \Longleftrightarrow F_0\left(\frac{y_0 - \xi}{\tau}\right) = p \\
\frac{\exp\left(\frac{y_0-\xi}{\tau}\right)}{1+\exp\left(\frac{y_0-\xi}{\tau}\right)} &= p \Longleftrightarrow (1-p)\exp\left(\frac{y_0 - \xi}{\tau}\right) = p \\
\exp\left(\frac{y_0 - \xi}{\tau}\right) &= \frac{p}{1-p} \Longleftrightarrow \frac{y_0 - \xi}{\tau} = \log\left(\frac{p}{1-p}\right) \\
y_0 &= \xi + \tau \log\left(\frac{p}{1-p}\right) \\
\xi &= y_0 - \tau \log\left(\frac{p}{1-p}\right).
\end{aligned}$$

Thus, the profile log-likelihood is given as

$$\begin{aligned}
\ell_{n,\text{prof}}(p) &= \max\left\{\ell_n(\xi, \tau) : F(y_0, \xi, \tau) = p\right\} \\
&= \max\left\{\ell_n(\xi, \tau) : \xi = y_0 - \tau \log\left(\frac{p}{1-p}\right)\right\},
\end{aligned}$$

which indicates that the ξ parameter is written as a linear combination of the τ parameter before optimization. Optimization is repeated for each p.

The deviance function $D(p)$ is given as

$$D(p) = 2\left\{\ell_{n,\text{prof}}(\widehat{p}) - \ell_{n,\text{prof}}(p)\right\}.$$

The confidence curve is evaluated at a grid of p values as $cc_{\mathrm{d}}(p) = \Gamma_1(D(p))$, where $\Gamma_1(\cdot)$ is the cumulative distribution function of the $\mathcal{X}_1^2$. Figure 2.3 displays the confidence curve. This curve is not symmetric as the one obtained by the delta method, but is tighter. In R, care in taken so that the grid of p values spans $(0.0001, 0.9999)$ to avoid infinity problems with log.

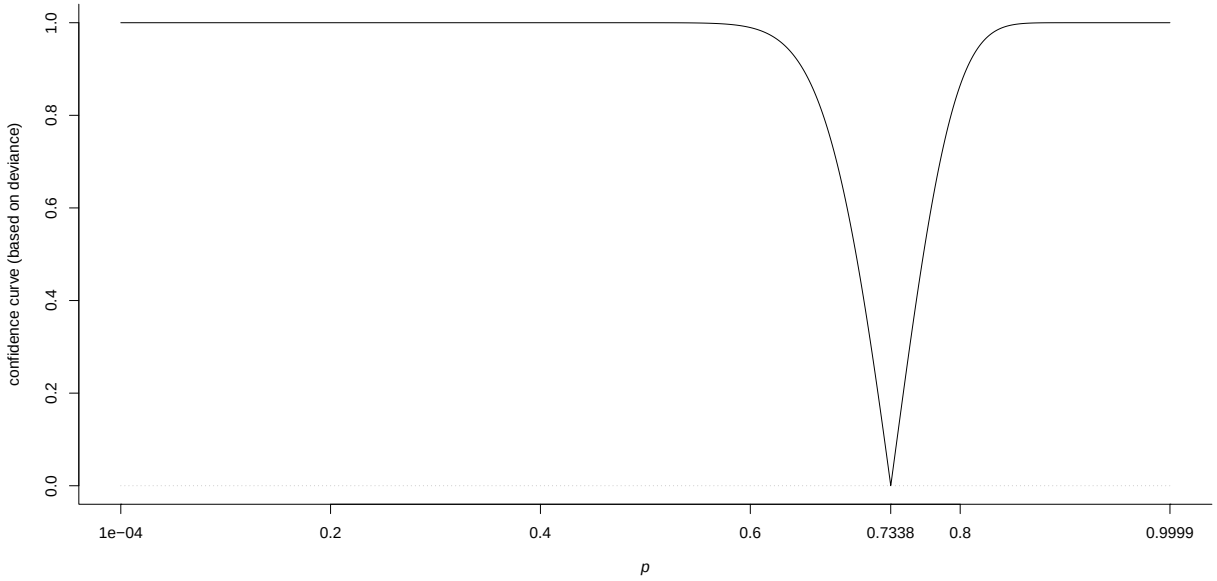

Figure 2.3: Confidence curve based on the deviance function.

(iii) Bartlett correction for the deviance is used. This method is described in Schweder and Hjort (2016, pp. 210–211). For a fine-tuned confidence curve based on the modified deviance function $\dfrac{D(p)}{1+\varepsilon}$, where $1+\varepsilon$ is the mean of $D(p)$, it is necessary to simulate a high number of replicates of $D(p)$ at the required position $(\widehat{\xi}, \widehat{\tau})$. For a large number of datapoints, n, ε will be close to zero. The deviance function is computed for a high number of simulated data sets of size 64, while p is kept fixed at $\widehat{p}$. The confidence curve is

$$cc_{\mathrm{md}}(p) = \Gamma_1\left(\frac{D(p)}{\mathrm{E}_{(\xi,\tau)}D(p)}\right),$$

where $\Gamma_1(\cdot)$ is the cumulative distribution function of the $\mathcal{X}_1^2$. Figure 2.4 displays the modified confidence curve and figure 2.5 compares $cc_{\mathrm{d}}(p)$ with $cc_{\mathrm{md}}(p)$.

It is noted that all methods estimate $\widehat{p}$ the same, as can be seen from the various plots.

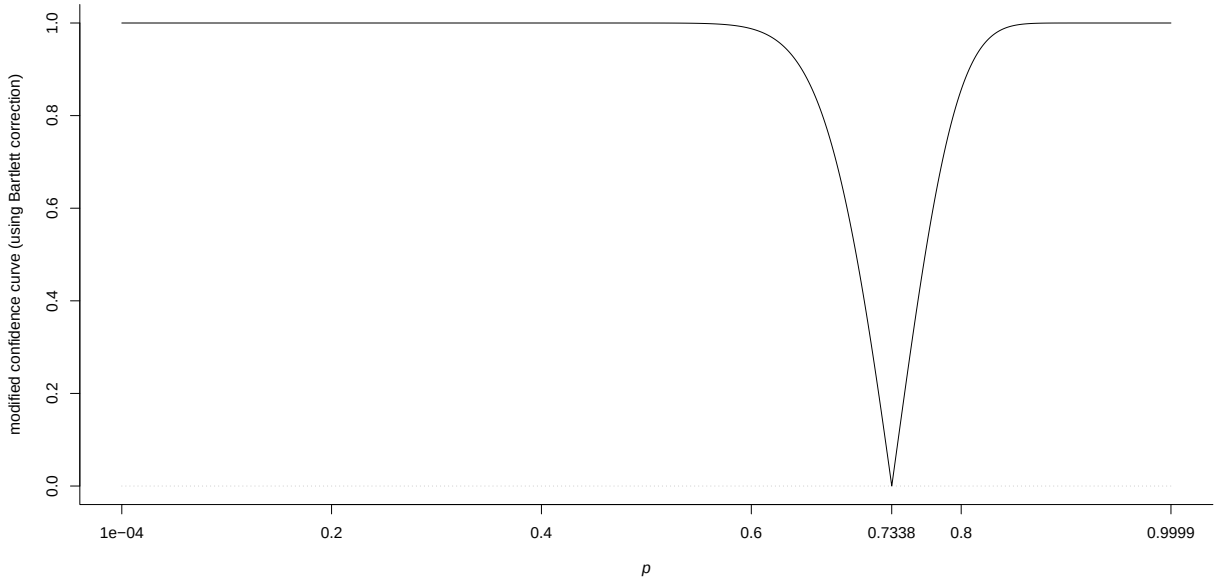

Figure 2.4: Confidence curve using Bartlett correction for the deviance.

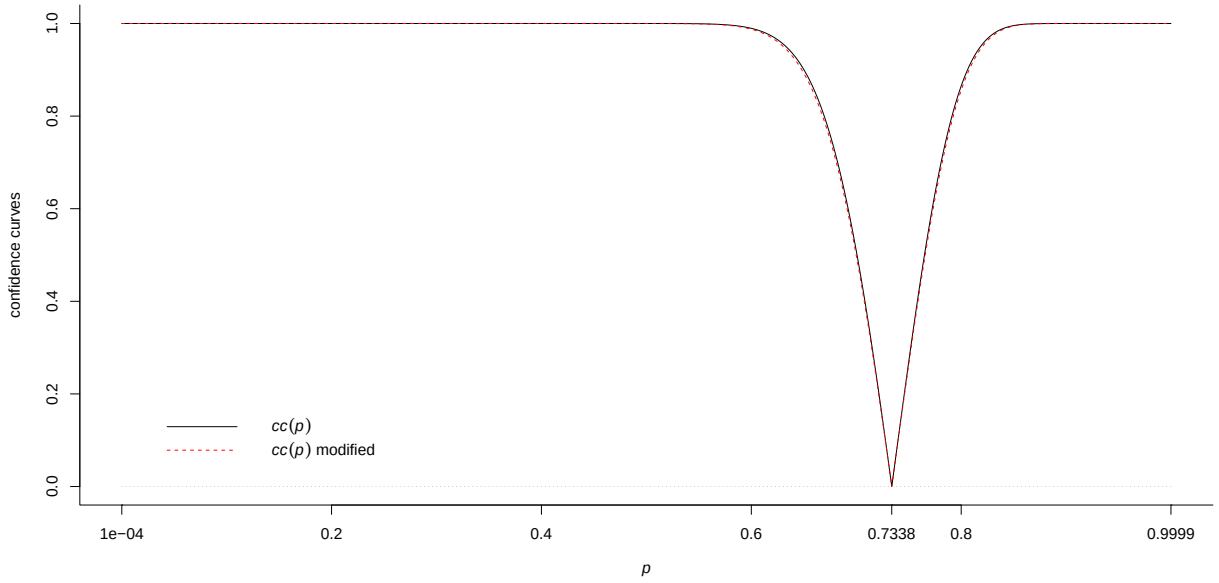

Figure 2.5: Comparison of $cc_{\mathrm{d}}(p)$ with $cc_{\mathrm{md}}(p)$. The difference is small as $\varepsilon = 0.032724$ and the mean of the deviance function at the required position is 1.032724. The sample size is 64.

Example 2.3 (Light thinks it travels faster than anything, continued)**.** Let

$$F_0(x) = \frac{\exp(x)}{1+\exp(x)} \qquad \text{and} \qquad f_0(x) = \frac{\exp(x)}{\{1+\exp(x)\}^2}.$$

Using maximum likelihood, the model with cumulative distribution function

$$F(y,\xi,\tau,\gamma) = \left\{F_0\left(\frac{y-\xi}{\tau}\right)\right\}^\gamma = \left[\frac{\exp\{(y-\xi)/\tau\}}{1+\exp\{(y-\xi)/\tau\}}\right]^\gamma$$

is fitted to the data given in appendix II.2. The aim is to construct a confidence curve for γ and check if the data support the model with $\gamma = 1$.

The cumulative distribution function is recognized to be the Type I: Skew-Logistic distribution. For details regarding this distribution, see, for example, Johnson et al. (1994) and Shao (2002). Its density is

$$\begin{aligned} f(y,\xi,\tau,\gamma) &= \frac{\partial F(y,\xi,\tau,\gamma)}{\partial y} \\ &= \frac{\gamma}{\tau}\frac{\exp\left(\gamma\,\frac{y-\xi}{\tau}\right)}{\left\{1+\exp\left(\frac{y-\xi}{\tau}\right)\right\}^{\gamma+1}}. \end{aligned}$$

For $\gamma = 1$, it is the density of the general logistic distribution, as in part (a).

The likelihood function is

$$\begin{aligned} \mathcal{L}(\xi,\tau,\gamma) &= \prod_{i=1}^n f(y_i;\xi,\tau,\gamma) \\ &= \prod_{i=1}^n \frac{\gamma}{\tau}\frac{\exp\left(\gamma\,\frac{y_i-\xi}{\tau}\right)}{\left\{1+\exp\left(\frac{y_i-\xi}{\tau}\right)\right\}^{\gamma+1}} \\ &= \left(\frac{\gamma}{\tau}\right)^n \prod_{i=1}^n \frac{\exp\left(\gamma\,\frac{y_i-\xi}{\tau}\right)}{\left\{1+\exp\left(\frac{y_i-\xi}{\tau}\right)\right\}^{\gamma+1}}. \end{aligned}$$

and the log-likelihood function is

$$\begin{aligned} \ell(\xi,\tau,\gamma) &= \log\mathcal{L}(\xi,\tau,\gamma) \\ &= n\log(\gamma) - n\log(\tau) + \frac{\gamma}{\tau}\sum_{i=1}^n (y_i-\xi) - (\gamma+1)\sum_{i=1}^n \log\left\{1+\exp\left(\frac{y_i-\xi}{\tau}\right)\right\}. \end{aligned}$$

The maximum likelihood estimates are $(\widehat{\xi},\widehat{\tau},\widehat{\gamma}) = (26.05397, 3.151787, 1.423737)$. The values reported are the ones output from `glogisfit`. The values output from the programmed function differ as in

```
c(xi.hat, tau.hat, gam.hat) - c(xihat, tauhat, gamhat)
     location         scale         shape
 6.630162e-04 -7.863028e-06 -4.151751e-05
```

The confidence curve is obtain through the profile log-likelihood for γ. That is,

$$\ell_{\text{prof}}(\gamma) = \max_{\text{all } (\xi,\tau)} \ell(\xi, \tau, \gamma) = \ell(\widehat{\xi}(\gamma), \widehat{\tau}(\gamma), \gamma),$$

where $(\widehat{\xi}(\gamma), \widehat{\tau}(\gamma))$ is the maximizer of $\ell(\xi, \tau, \gamma)$ for given γ and is displayed in figure 2.6.

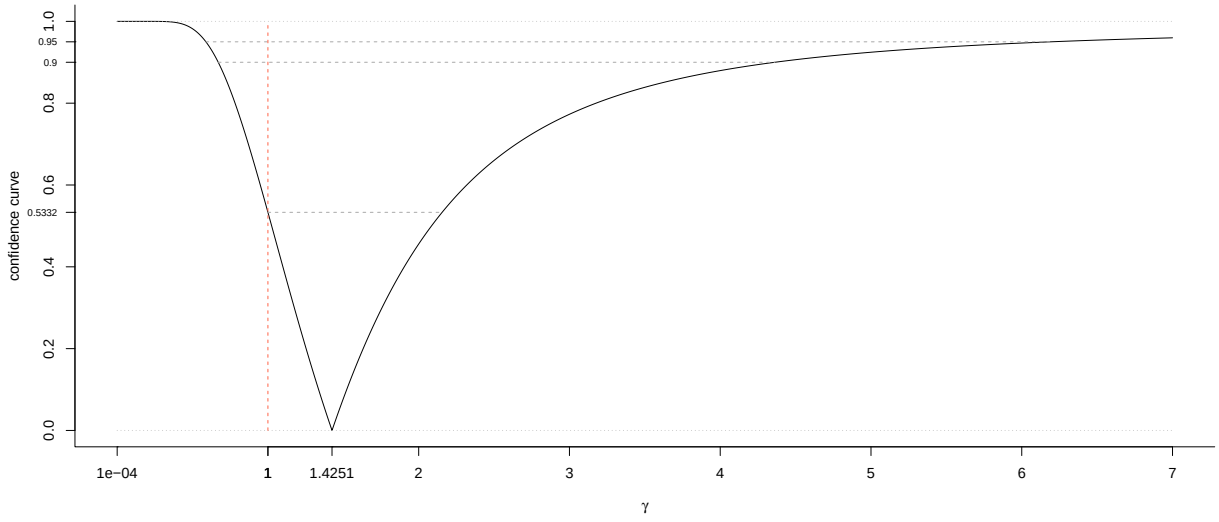

Figure 2.6: Confidence curve for γ. Confidence intervals above the level of 53.32% include the value of $\gamma = 1$.

A 90% confidence interval for $H_o = \gamma = 1$ is $[0.6711, 4.3621]$ and a 95% confidence interval is $[0.5891, 6.1911]$.

Example 2.4 (Parametric confidence for quantiles)**.** Let $Y_1, \ldots, Y_n$ be i.i.d. from $N(\mu, \sigma^2)$. Let, also, the p-quantile be $\psi_p = \mu + z_p\sigma$, with $z_p = \Phi^{-1}(p)$. The aim is to show that, as n increases,

$$\sqrt{n}(\widehat{\psi}_p - \psi_p) \to_d N(0, (1 + \tfrac{1}{2}z_p^2)\sigma).$$

Let Y be a random variable from the normal distribution with mean μ and variance σ^2, that is $Y \sim N(\mu, \sigma^2)$. Then, $Z = \dfrac{Y - \mu}{\sigma}$ is also normal, as a linear combination of a normally distributed random variable, with mean $\mathrm{E}\left(\dfrac{Y-\mu}{\sigma}\right) = \dfrac{1}{\sigma}\{\mathrm{E}(Y) - \mu\} = \dfrac{1}{\sigma}\{\mu - \mu\} = 0$ and variance $\mathrm{Var}\left(\dfrac{Y-\mu}{\sigma}\right) = \dfrac{1}{\sigma^2}\mathrm{Var}(Y) = \dfrac{1}{\sigma^2}\sigma^2 = 1$. That is, Z follows a standard normal distribution, $N(0, 1)$. Also, let F denote the cumulative distribution function of the general normal distribution, $N(\mu, \sigma^2)$, and, as usual, Φ be the cumulative distribution function of the standard normal distribution, $N(0, 1)$. Then,

$$F_Y(y) = \Pr\{Y \le y\} = \Pr\{Y - \mu \le y - \mu\} = \Pr\left\{\frac{Y-\mu}{\sigma} \le \frac{y-\mu}{\sigma}\right\} = \Pr\left\{Z \le \frac{y-\mu}{\sigma}\right\} = \Phi\left(\frac{y-\mu}{\sigma}\right).$$

Now, let ψ_p denote the p-quantile of a $N(\mu, \sigma^2)$. Then, by definition, $p = F(\psi_p)$. As $F(\psi_p) = \Phi\left(\dfrac{\psi_p - \mu}{\sigma}\right)$, $p = \Phi\left(\dfrac{\psi_p - \mu}{\sigma}\right)$. Φ is a cumulative distribution function, and as such is invertible at every point on the real line. Thus, $\Phi^{-1}(p) = \dfrac{\psi_p - \mu}{\sigma}$. Rearranging $\psi_p = \mu + \sigma\,\Phi^{-1}(p)$. By setting $z_p = \Phi^{-1}(p)$, ψ_p is, thus,

expressed as $\psi_p = \mu + z_p\,\sigma$.

Let $\widehat{\mu} = \dfrac{1}{n}\sum_{i=1}^{n} Y_i$. Then,

$$\widehat{\mu} \;=\; \frac{1}{n}\sum_{i=1}^{n} Y_i = \frac{1}{n}\sum_{i=1}^{n}(\mu + \sigma\, Z_i) = \frac{1}{n}\sum_{i=1}^{n}\mu + \frac{1}{n}\sigma\sum_{i=1}^{n} Z_i = \frac{1}{n}n\mu + \sigma\bar{Z} = \mu + \sigma\,\frac{1}{\sqrt{n}}N_1,$$

where Z_i are i.i.d. $N(0,1)$ and, thus, $\sum_{i=1}^{n} Z_i \sim N(0,n)$ or $\bar{Z} = \dfrac{1}{n}\sum_{i=1}^{n} Z_i \sim N\left(0, \dfrac{1}{n}\right) \equiv \dfrac{1}{\sqrt{n}}N(0,1)$. Equivalently, $\bar{Z} = \dfrac{1}{\sqrt{n}}N_1$ with $N_1 \sim N(0,1)$.

Also, $\dfrac{\widehat{\sigma}^2}{\sigma^2} \sim \dfrac{\mathcal{X}^2_\nu}{\nu}$, where $\nu = n-1$. Thus, $\widehat{\sigma}^2 \sim \sigma^2\,\dfrac{\mathcal{X}^2_\nu}{\nu}$ and $\widehat{\sigma} \sim \sigma\left(\dfrac{\mathcal{X}^2_\nu}{\nu}\right)^{1/2}$.

Thus,

$$\begin{aligned}\widehat{\psi}_p &= \widehat{\mu} + z_p\,\widehat{\sigma} \\ &= \mu + \left(\frac{\sigma}{\sqrt{n}}\right)N_1 + z_p\,\sigma\left(\frac{\mathcal{X}^2_\nu}{\nu}\right)^{1/2},\end{aligned}$$

and N_1 is independent of $\mathcal{X}^2_\nu$. Also, $\psi_p = \mu + z_p\,\sigma$.

Now,

$$\begin{aligned}\widehat{\psi}_p - \psi_p &= \mu + \left(\frac{\sigma}{\sqrt{n}}\right)N_1 + z_p\,\sigma\left(\frac{\mathcal{X}^2_\nu}{\nu}\right)^{1/2} - \mu - z_p\,\sigma \\ &= \left(\frac{\sigma}{\sqrt{n}}\right)N_1 + \left\{\left(\frac{\mathcal{X}^2_\nu}{\nu}\right)^{1/2} - 1\right\}z_p\,\sigma \\ &= \frac{\sigma}{\sqrt{n}}\left[N_1 + \sqrt{n}\left\{\left(\frac{\mathcal{X}^2_\nu}{\nu}\right)^{1/2} - 1\right\}\,z_p\right],\end{aligned}$$

or $\sqrt{n}(\widehat{\psi}_p - \psi_p) = \sigma\left[N_1 + \sqrt{n}\left\{\left(\dfrac{\mathcal{X}^2_\nu}{\nu}\right)^{1/2} - 1\right\}\,z_p\right]$.

An n increases, $n-1 \approx n$ and $\dfrac{\mathcal{X}^2_{n-1}}{n-1} \doteq \dfrac{\mathcal{X}^2_n}{n}$.

By the properties of $\mathcal{X}^2$ distribution, a $\mathcal{X}^2_n$ is the sum of n independent $\mathcal{X}^2_1$. That is, $\mathcal{X}^2_n = \sum_{i=1}^{n}\mathcal{X}^2_{1,i}$, where $\mathcal{X}^2_{1,1},\ldots,\mathcal{X}^2_{1,n}$ are independent of each other. Also, since $\forall\ i = 1,\ldots,n$, $\mathrm{E}\left(\mathcal{X}^2_{1,i}\right) = 1$ and $\mathrm{Var}\left(\mathcal{X}^2_{1,i}\right) = 2$,

$$\mathrm{E}\left(\frac{1}{n}\mathcal{X}^2_n\right) = \frac{1}{n}\mathrm{E}\left(\mathcal{X}^2_n\right) = \frac{1}{n}\mathrm{E}\left(\sum_{i=1}^{n}\mathcal{X}^2_{1,i}\right) = \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\left(\mathcal{X}^2_{1,i}\right) = \frac{1}{n}n = 1,$$

by the linearity of expectation.

By the Central Limit Theorem,

$$\begin{aligned}\sqrt{n}\left\{\left(\frac{\mathcal{X}_n^2}{n}\right)-1\right\} &= \sqrt{n}\left(\frac{1}{n}\mathcal{X}_n^2-1\right)\\ &= \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{X}_{1,i}^2-\mathrm{E}\left(\mathcal{X}_{1,i}^2\right)\right)\longrightarrow N(0,\mathrm{Var}\left(\mathcal{X}_{1,i}^2\right))\equiv N(0,2)\equiv\sqrt{2}\,N(0,1),\end{aligned}$$

By the delta method,

$$\sqrt{n}\left\{h\left(\frac{\mathcal{X}_n^2}{n}\right)-h(1)\right\}\longrightarrow_d h'(1)\,\sqrt{2}\,N(0,1),$$

where $h(x)=\sqrt{x}$ and $h'(x)=\dfrac{\partial h(x)}{\partial x}=\dfrac{1}{2\sqrt{x}}$. h is defined on $[0,\infty)$ for which $\dfrac{\mathcal{X}_n^2}{n}$ falls with probability 1 and has derivative in an interval around 1 that is continuous at that point (Schweder and Hjort, 2016, p. 451). Thus,

$$\sqrt{n}\left\{\left(\frac{\mathcal{X}_n^2}{n}\right)^{1/2}-1\right\}\longrightarrow_d\frac{1}{2}\sqrt{2}\,N(0,1)\equiv\frac{1}{\sqrt{2}}\,N(0,1)\equiv N\left(0,\frac{1}{2}\right)\equiv\frac{1}{\sqrt{2}}N_2,$$

with $N_2\sim N(0,1)$ and

$$\begin{aligned}\sqrt{n}\left(\widehat{\psi}_p-\psi_p\right) &\longrightarrow_d \sigma\left[N_1+\frac{1}{\sqrt{2}}\,N_2\,z_p\right]\equiv\sigma\left[N(0,1)+z_p\,N\left(0,\frac{1}{2}\right)\right]\\ &\equiv \sigma\,N\left(0,1+\frac{1}{2}\,z_p^2\right)\equiv N\left(0,\left(1+\frac{1}{2}\,z_p^2\right)\sigma^2\right).\end{aligned}$$

Equivalently,

$$\frac{\sqrt{n}\left(\widehat{\psi}_p-\psi_p\right)}{\left(1+\frac{1}{2}\,z_p^2\right)^{1/2}\sigma}\longrightarrow_d N(0,1),$$

or

$$\frac{\sqrt{n}\left(\psi_p-\widehat{\psi}_p\right)}{\left(1+\frac{1}{2}\,z_p^2\right)^{1/2}\sigma}\longrightarrow_d N(0,1),$$

as $A\sim N(0,1)$ has the same distribution as $-A\sim N(0,1)$.

Now, let, $W_n=\dfrac{\sqrt{n}\left(\psi_p-\widehat{\psi}_p\right)}{\left(1+\frac{1}{2}\,z_p^2\right)^{1/2}\sigma}$ and ψ_p be the focus parameter. By construction, ψ_p is a function of the two normal parameters, μ and σ^2. W_n is a 'large-sample pivot' as (1) is a function of the data, Y, (through $\widehat{\psi}_p$) and the focus parameter, ψ_p, and (2) has a limiting distribution independent of the parameters μ and σ^2. Because W_n is increasing in ψ_p, by the theory developed in Schweder and Hjort (2016), the approximate confidence distribution for ψ_p is

$$C_n(\psi_p)=\Phi\left(W_n\right)=\Phi\left(\frac{\sqrt{n}\left(\psi_p-\widehat{\psi}_p\right)}{\left(1+\frac{1}{2}\,z_p^2\right)^{1/2}\sigma}\right)$$

and the associated confidence curve is $cc_n(\psi_p)=|1-2C_n(\psi_p)|$.

The corresponding confidence curve is plotted in Figure 2.7.

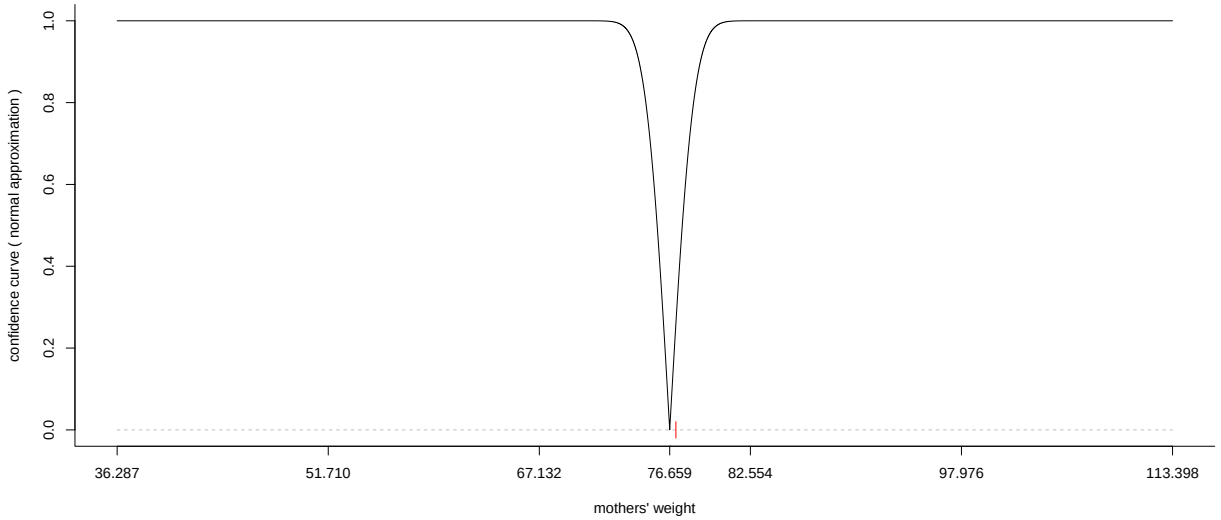

Figure 2.7: Approximate confidence curve for ψ_p. The tag indicates the sample quantile from `quantile` in R.

Example 2.5 (Nonparametric confidence for quantiles)**.** Let $Y_1, \ldots, Y_n$ be independent from a continuous and strictly increasing cumulative distribution function $F(y)$ on $[0, \infty]$ and let the p-quantile denoted by $\psi_p = F^{-1}(p)$. The aim is to construct nonparametric confidence distributions and demonstrate them using the data given in appendix II.3.

Note: This example is not directly related to the methods reviewed in section 2. It is added to demonstrate that, indeed, nonparametric confidence distributions can be constructed and to show how, following the instructions of the relevant exercise retrieved from the STK4180 website.

Let $Y_1, \ldots, Y_n$ be independent observations from a continuous and strictly increasing cumulative distribution function F on the half line $[0, \infty)$ and $Y_{(1)}, \ldots, Y_{(n)}$ be the corresponding ordered observations. Also, let $\psi_p = F^{-1}(p)$, for $p \in [0, 1]$. Then,

$$\begin{aligned} \Pr\left\{\psi_p \leq Y_{(j)}\right\} &= \Pr\left\{F^{-1}(p) \leq Y_{(j)}\right\} && (2.11)\\ &= \Pr\left\{F^{-1}(p) \leq F^{-1}\left(U_{(j)}\right)\right\} && (2.12)\\ &= \Pr\left\{p \leq F\left(F^{-1}\left(U_{(j)}\right)\right)\right\} && (2.13)\\ &= \Pr\left\{p \leq U_j\right\} && (2.14) \end{aligned}$$

Equation (2.11) is because $\psi_p = F^{-1}(p)$. Equation (2.12) is by the probability integral transform applied to the order statistics, see for example Reiss (1989, Theorem 1.2.5, pp. 17–18). Equations (2.13) and (2.14) is due to F being continuous and strictly increasing.

Let $Y_1, \ldots, Y_n$ be independent from a continuous and strictly increasing cumulative distribution function F. Then, from part (c), $\Pr\left\{\psi_p \leq Y_{(j)}\right\} = \Pr\left\{p \leq U_j\right\}$. Following example 4.4 of Schweder and Hjort (2016, p. 106), for p fixed, say p_{fix}, $\Pr\left\{\psi_{p_{\text{fix}}} \leq Y_{(j)}\right\} = \Pr\left\{p_{\text{fix}} \leq U_j\right\} = B_n(j-1, p_{\text{fix}})$, where $B_n(j-1, p_{\text{fix}})$ is the distribution for the binomial (n, p_{fix}) variable X that counts the number of uniform data points below p_{fix}.

This defines a nonparametric confidence distribution function $C(Y_{(j)})$ as

$$C(Y_{(j)}) = \Pr\{X \leq j-1\} = \Pr\{X < j\} = 1 - \Pr\{X \geq j\}.$$

But the event $\{X \geq j\}$ is equivalent to the event $\{U_j \leq p_{\text{fix}}\}$, as there are at least j of the ordered U_j in the interval $[0, p_{\text{fix}}]$ if and only if their jth smallest is in that interval. Thus,

$$C(Y_{(j)}) = 1 - \Pr\{X \geq j\} = 1 - \Pr\{U_j \leq p_{\text{fix}}\}$$

and the rest follow from the fact that the U_j has a beta distribution, with parameters $(j, n+1-j)$.

Figure 2.8, displays confidence curves for 0.10, 0.50, 0.90 quantiles in the same diagram. For the nonparametric quantiles the method described in Schweder and Hjort (2016, pp. 320–321) is implemented.

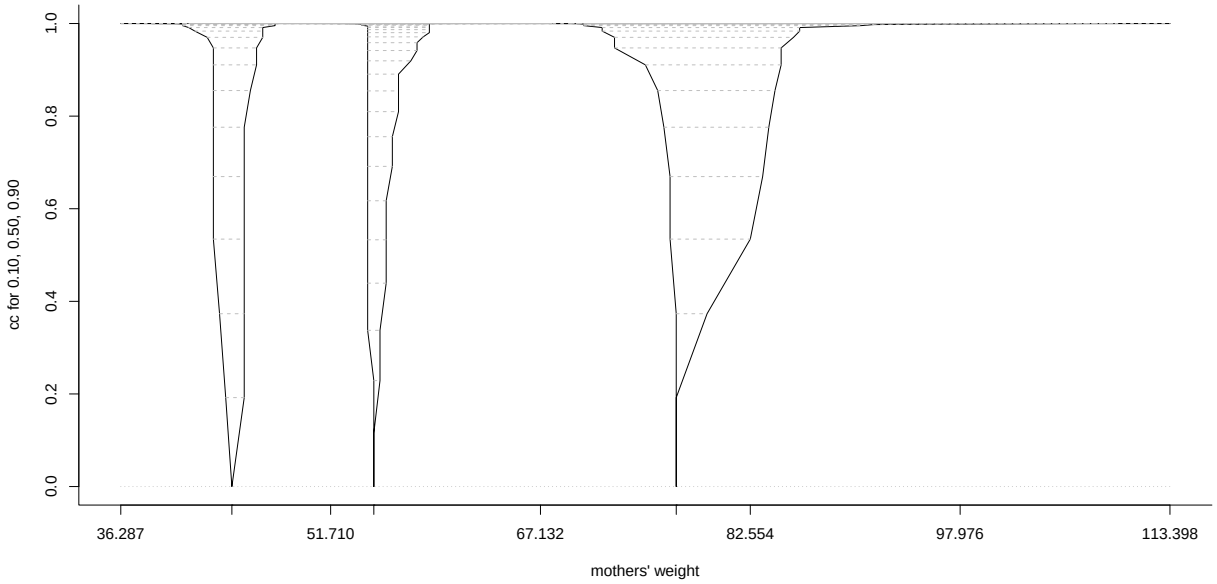

Figure 2.8: Confidence curves for 0.10, 0.50, 0.90 quantiles.

3 The t-bootstrap method

This section presents some contributions to the t-bootstrap method. It constitutes the main effort of this project and, thus, motivation, explanations and remarks are presented to help the reader understand the main points of the story and, perhaps, make suggestions for improvement. It is theoretical in nature, as results, had to be established in a rigorous manner, but a couple of examples are presented.

3.1 Derivation of the Confidence Distribution

Let the data, $y = \{y_1, \ldots, y_n\}$, be independent and identically distributed (i.i.d.) observations stemming from an unknown probability distribution P_θ of a univariate population. The focus is turned to a scalar parameter of interest, $\psi = a(\theta)$, where a is a function that maps the full, p-dimensional, parameter θ to some chosen predefined scalar parameter for which inference needs to be made. In particular, the focus is concentrated to a Studentised version of ψ, namely,

$$t = \frac{\psi - \widehat{\psi}_n}{\widehat{\tau}}, \tag{3.1}$$

where $\widehat{\psi}_n$ is an estimate of ψ and $\widehat{\tau}$ is a scale estimate, which forms the so called t-statistic. The term 'Studentised' hints at W. S. Gosset's approach, published under the pseudonym Student, of dividing a scalar centred quantity of the form $\gamma - \widehat{\gamma}_n$ by the estimated standard error of $\widehat{\gamma}_n$, $\widehat{se}(\widehat{\gamma}_n)$, i.e. the scale estimate. That is, the form given at (3.1) reflects the fact that a statistic is centred and scaled so that its mean and variance are 0 and 1, respectively. The choice of writing 'estimand minus estimator' and not 'estimator minus estimand', as has, traditionally, been the practise in the frequentist world, is intentional so as to have (i) the focus parameter in direct sight and (ii) a readily available increasing function in the focus parameter. See, Remark 3.3 in Schweder and Hjort (2016, p. 70) for further comments on this. For the present, what is interesting is to emphasise that pivotal character of t at 3.1, which renders it to have a distribution function independent of the underlying parameter. The statement that this distribution happens to tends to the Standard Normal distribution will be used at a later point.

Efron's bootstrap resampling scheme, originated in Efron (1979), is employed to provide an estimate to the true distribution of the t-statistic at (3.1), denoted as R. This scheme is particularly successful when the distribution of t at (3.1) is the same or asymptotically the same for each value of ψ, making t a pivot or an approximate pivot according to Definitions 1.3 and 1.4, respectively (Hall, 1992b, p. 83). Briefly, the resampling scheme requires, firstly, a suitably large number, say B, of bootstrap samples $y_b^* = \{y_{b1}^*, \ldots, y_{bn}^*\}$, where $b = 1, \ldots, B$, to be generated from $P_{\widehat{\theta}}$, an estimate of the unknown P_θ, and, secondly, the quantity $t^*(b) = \dfrac{\widehat{\psi}_n - \widehat{\psi}^*(b)}{\widehat{\tau}^*(b)}$, where $\widehat{\psi}^*(b)$ and $\widehat{\tau}^*(b)$ are the corresponding $\widehat{\psi}_n$ and $\widehat{\tau}$ values under the b bootstrap sample, to be computed for each of the $y_1^*, \ldots, y_B^*$ samples. The distribution of $t^*(b)$, denoted as $\widehat{R}$, is then used to estimate unbiasedly and consistently R. For a discussion on the number of bootstrap samples, B, required, see Hall (1986, 1989), although there is not a generally accepted recipe on how 'suitably large' the size of B should be, apart, perhaps, from the vague notion of 'the larger the better'.

Now, let the generated bootstrap samples be summarised as $y^* = \{y_1^*, \ldots, y_B^*\}$. Whether $\widehat{R}$ is independent of the underlying parameter ψ or not, a confidence distribution or an asymptotic confidence distribution is

easily constructed. Define, as in Schweder and Hjort (2016, p. 217),

$$C_{t\text{boot}}(\psi, y^*) = \widehat{R}\left(\frac{\psi - \widehat{\psi}_n}{\widehat{\tau}}\right) \tag{3.2}$$

and examine if either Definition 1.1 or 1.2 is satisfied. By construction, $\widehat{R}$ defines a cumulative distribution function (c.d.f.) on Ψ, the parameter space for ψ. At the true value of the parameter, ψ_0, though, the behavior of $\widehat{R}$ needs to be examined slightly more carefully since, by construction, the outcome of any bootstrap resampling scheme is a discrete distribution. The construction at (3.2) is seen as a function of y^*, a view that renders the construction a random character. When t at (3.1) is a perfect pivot, $\widehat{R}$ is exact, thus $\widehat{R} \equiv R$, and, for any $0 < \alpha < 1$, $P\{C_{t\text{boot}}(\psi_0, y^*) \leq \alpha\} = \alpha$, by the generalised form of the probability integral transform given in Bol'shev (1965). When t at (3.1) is an approximate pivot, $\widehat{R} \to_d R$, as $B \to \infty$, and, for any $0 < \alpha < 1$, $P\{C_{t\text{boot}}(\psi_0, y^*) \leq \alpha\} \doteq \alpha$, by applying the generalised form of the probability integral transform of Bol'shev (1965) to the limiting distribution. In the case of a perfect pivot, (3.2) is defined to be a CD as in Definition 1.1 and in the case of an approximate pivot, (3.2) is defined to be an aCD as in Definition 1.2. Operationally, $C_{t\text{boot}}(\psi, y^*)$ is obtained by the empirical distribution of $t^*(1), \ldots, t^*(B)$, that is,

$$C_{t\text{boot}}(\psi, y^*) = \frac{1}{B}\sum_{b=1}^{B} I\{t^*(b) \leq t_{\text{obs}}\}, \tag{3.3}$$

where I denotes the indicator function and $t_{\text{obs}} = \dfrac{\psi - \widehat{\psi}_n}{\widehat{\tau}}$. For each ψ on a grid of ψ values in Ψ, the steps involved in the operation described by (3.3) are summarised, algorithmically, in three steps: (i) Compute t_{obs}. (ii) Identify which values of the bootstrap sample, $\{t^*(b),\, b = 1, \ldots, B\}$, satisfy the relationship $t^*(b) \leq t_{\text{obs}}$. (iii) Compute the proportion of bootstrap samples less than or equal to t_{obs}. Note that, ideally, the entire Ψ should be covered, but for practical reasons, e.g. when plotting a CD or an aCD, Ψ is truncated to values close to $\widehat{\psi}_n$. Efron and Hastie (2016, ch. 11) demonstrate in a somewhat pedagogical manner the details of the operation at (3.3) by two examples (see, pp. 187, 191).

3.2 Introduction to the problem and summary of results

The t-bootstrap procedure is described, for example, in Efron and Tibshirani (1993, p. 160) and Davison and Hinkley (1997, p. 194) and is briefly discussed here as it provides a means of constructing confidence distributions. Depending on whether the t-statistic involved is a perfect pivot or an approximate one, the distribution of the t-statistic defines either a confidence distribution in the sense of Definition 1.1, or an asymptotic confidence distribution in the sense of Definition 1.2. The aim, here, is to suggest a transformed Studentised statistic that corrects for bias and skewness in the resulting confidence distributions. The suggestion is based on the discussion given in Kakizawa (1996) and the working framework is the '*smooth function model*' of Hall (1992b, p. 52), for which Edgeworth's expansions are valid as demonstrated in the seminal work of Bhattacharya and Ghosh (1978, 1980).

First, a description is given, in generic terms, on how and why the bootstrap gives a confidence distribution. Then, a brief discussion is given on the motivation behind the use of a monotone and invertible transformation such as the one discussed in Kakizawa (1996). Two Theorems, 3.1 and 3.2, are stated, in their most generality, to formalise the main results in Kakizawa (1996), as (i) such an effort has not been attempted before (at least, it is not know to have been) and (ii) the transformations form the basis for

further results. The first of these results concerns the accuracy of approximation of a confidence curve as this is described by Theorem 3.3. Then, a very brief description of the '*smooth function model*' is given, which constitutes the present working framework and restricts attention to pivotal statistics of certain form, say S_n. A brief discussion on the Edgeworth expansion is also given, as further results will be stated with its help. Another Theorem, 3.4, formalises the '*minimum moment conditions*', so that the Edgeworth expansion of a transformed S_n is still valid. Since the polynomials appearing in the Edgeworth expansion of S_n, depend on the characteristics of the underlying sampling distribution, these need to be estimated. A further Theorem, 3.5, formalises that under consistent estimators the resulting Edgeworth expansion of the transformed S_n still holds. Theorem 3.6 states that when the bootstrap procedure is used to approximate the distribution of either S_n or its transformed versions, cf. (3.6), then the order of the incurred error is either $o_P(n^{-1})$ or $o_P(n^{-(j+\nu)/2})$, for suitably specified values of j and ν. Theorem 3.7 establishes that the same order of incurred error as described by the previous theorem applies to the corresponding confidence distributions. It is noticed that one application of the considered monotone and invertible transformation applied to S_n results in a bias free confidence distribution that is unaffected from the main effect of skewness.

3.3 Monotone transformations

From a distributional point of view, a Studentised or pivotal statistic of the form at (3.1) follows the Standard Normal distribution, asymptotically. That is, when the size of the data is large inference can be made based on the Standard Normal distribution, which is symmetric, as a consequence of the Central Limit Theorem. The confidence intervals constructed then, have coverage error $O(n^{-1})$. The bootstrapped confidence intervals, though, which are based on the resampled t^* values, have coverage error $O(n^{-3/2})$ and this is considered to be a major advantage of the bootstrap procedure. See, e.g., Hall (1992b, pp. 13–16, 83), Efron and Tibshirani (1993, p. 174) and Barndorff-Nielsen and Cox (1994, pp. 298–301) on how pivotal quantities improve the performance of the bootstrap. But, it is not always the case that the size of the data is large as their might be lots of practical reasons behind the unavailability of data. When the size of the data is small, the bootstrap distribution of a Studentised or pivotal statistic is asymmetric, i.e. skewed, and ways for correcting the asymmetry in the distribution are sought. A consequence of the asymmetry in the distribution of a statistic is that any constructed confidence intervals would have higher coverage error that constructed confidence intervals based on a statistic with a symmetric distribution. The well studied Edgeworth expansions, which have proven to a be a powerful tool, give the theoretical insight into this as the first term, of size $n^{-1/2}$, in an expansion gives a description of the error in the usual Standard Normal approximation. This error is entirely due to skewness and, in particular, its main effect. An effective way to eliminate or remove the main effect of skewness from the distribution of a pivotal statistic is to apply a transformation to the current statistic so that the resulting, new, statistic has a more symmetric distribution. This will be reflected to the bootstrap distribution when a bootstrap scheme is employed. In this way, the high coverage accuracy obtained by applying confidence procedures to a symmetric distribution is retained. See, Hall (1992a, p. 221) who suggested to use monotone and invertible transformations to assist with this task. It is kept in mind that confidence distributions provide confidence intervals for any chosen level of confidence and methods for constructing them should be safeguarded against the undesirable effects of asymmetry.

Kakizawa (1996) provides a method for finding a transformation that is monotone and invertible over the whole real line. The method is general enough to be applicable to a variety of statistics and, thus, not constrained to the case of Studentised or pivotal statistics, which, by the way, is the case under consideration here. The idea that drives the suggested method is described, briefly and in very general terms, as follows. Let $S = S_n$ denote a statistic of order $O_p(1)$. Also, assume that S admits an Edgeworth expansion of the

form $P(S \leq x) = \Phi(x) + n^{-k/2}\, h(x) + O(n^{-(k+1)/2})$, where $k \geq 1$ is integer and $h(\cdot)$ is some polynomial dependent on ϕ, the density of the Standard Normal. A function, $f_{S,n}$, that depends on the S statistic and the sample size n, is sought such that the transformed statistic $T(S) = S + f_{S,n}$ is monotone with respect to S and $f_{S,n} = O_P(n^{-k/2})$. The O_P-notation is defined in Hall (1992b, pp. xii–xiii) and, simply, denotes a random variable that is, in probability, of the specified order. By setting $f_{S,n} = n^{-k/2}\, H(S)$, where $H(\cdot)$ is some polynomial to be specified, this perturbation leads to the distribution of the transformed statistic $T(S)$ having

$$\begin{aligned} P(T(S) \leq x) &= P(S \leq T^{-1}(x)) \\ &\doteq P(S \leq x - n^{-k/2}\, H(x)) \\ &\doteq \Phi(x) - n^{-k/2}\, H(x)\, \phi(x) + n^{-k/2}\, h(x), \end{aligned} \tag{3.4}$$

where $H(x) = h(x)/\phi(x)$ reduces the error of approximation. Hence, the necessity of requiring that $f_S = O_P(n^{-k/2})$ and the transformed statistic be monotone. It would have been extremely interesting to see the proofs regarding these results in Kakizawa (1996), but they are not offered. The paper gives a discussion on the existence of such transformations using a chi-squared limiting distribution and provides some numerical studies. Here, the task of going over the details of a proof is undertaken in Theorems 3.1 and 3.2. First, the monotonicity of such a transformation is shown and then the idea depicted at (3.4) is presented in a more formal way. Both results are given in generality. Remarks follow the statement of the results for clarification.

Theorem 3.1. *Let ν be some fixed integer greater that or equal to 1, $\pi_j(x)$ be, for every $j = 1, \ldots, \nu$, a polynomial on the real line and c be a constant such that $c \geq \frac{1}{4}$. Define the indefinite integral*

$$\int \left\{ \frac{\mathrm{d}}{\mathrm{d}y} \pi_k(y) \right\}^2 \mathrm{d}y = \Pi_k(y) + \textit{const.},$$

where const. is some arbitrary constant. Then, the transformation $T_{j\circ 0}(x) = (T_j \circ \ldots \circ T_0)(x)$ with

$$T_{k\circ 0}(x) = T_{(k-1)\circ 0}(x) + n^{-k/2}\, \pi_k(T_{(k-1)\circ 0}(x)) + c\, n^{-(k+1)/2}\, \Pi_k(T_{(k-1)\circ 0}(x)),$$

for some $k \in \{1, \ldots, j\}$ and $T_0(x) = x$, is strictly monotone increasing for every $j = 1, \ldots, \nu$.

Proof. By mathematical induction on j. The derivative argument is used in Hall (1992b, pp. 123–124). Without loss of generality, let $c = \frac{1}{4}$. If, though, $c > \frac{1}{4}$, it sufficient to complete the square as in $ca^2 + a + 1 = \left\{ \sqrt{c}a + \frac{1}{2\sqrt{c}} \right\}^2 - \frac{1}{4c} + 1$, for a suitably chosen a.

Let $j = 1$. Then, $T_1(x) = x + n^{-1/2}\, \pi_1(x) + \frac{1}{4}\, n^{-1}\, \Pi_1(x)$. Differentiation of $T_1(x)$ with respect to x, obtains

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}x} T_1(x) &= 1 + n^{-1/2}\, \frac{\mathrm{d}}{\mathrm{d}x} \pi_1(x) + \tfrac{1}{4}\, n^{-1}\, \frac{\mathrm{d}}{\mathrm{d}x} \Pi_1(x) \\ &= 1 + n^{-1/2}\, \frac{\mathrm{d}}{\mathrm{d}x} \pi_1(x) + \tfrac{1}{4}\, n^{-1} \left\{ \frac{\mathrm{d}}{\mathrm{d}x} \pi_1(x) \right\}^2, \end{aligned}$$

by elementary calculations and an employment of the fundamental theorem of calculus. Observing that

$$1 + n^{-1/2}\, \frac{\mathrm{d}}{\mathrm{d}x} \pi_1(x) + \tfrac{1}{4}\, n^{-1} \left\{ \frac{\mathrm{d}}{\mathrm{d}x} \pi_1(x) \right\}^2 = \left\{ \tfrac{1}{2}\, n^{-1/2}\, \frac{\mathrm{d}}{\mathrm{d}x} \pi_1(x) + 1 \right\}^2,$$

which is always positive, it follows immediately that $\dfrac{\mathrm{d}}{\mathrm{d}x} T_1(x) > 0$, for every x on the real line and thus, $T_1(x)$ is strictly monotone increasing for $j = 1$.

Assuming now that, for some $j > 1$, $T_{j\circ 0}(x)$ is strictly monotone increasing for every x on the real line, the

aim is to show that for $j+1$, $T_{(j+1)\circ 0}(x)$ is strictly monotone increasing for every x on the real line. The transformation $T_{(j+1)\circ 0}(x)$ is written as

$$T_{(j+1)\circ 0}(x) = (T_{j+1} \circ \ldots \circ T_0)(x) = T_{j+1}(T_{j\circ 0}(x)).$$

Thus,

$$T_{j+1}(T_{j\circ 0}(x)) = T_{j\circ 0}(x) + n^{-(j+1)/2}\, \pi_{j+1}(T_{j\circ 0}(x)) + \tfrac{1}{4}\, n^{-(j+2)/2}\, \Pi_{j+1}(T_{j\circ 0}(x)).$$

Differentiation of $T_{(j+1)\circ 0}(x)$ with respect to x obtains

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}x} T_{j+1}(T_{j\circ 0}(x)) &= \frac{\mathrm{d}}{\mathrm{d}x} T_{j\circ 0}(x) + n^{-(j+1)/2}\, \frac{\mathrm{d}}{\mathrm{d}x} \pi_{j+1}(T_{j\circ 0}(x))\, \frac{\mathrm{d}}{\mathrm{d}x} T_{j\circ 0}(x) \\
&+ \tfrac{1}{4}\, n^{-(j+2)/2}\, \frac{\mathrm{d}}{\mathrm{d}x} \Pi_{j+1}(T_{j\circ 0}(x))\, \frac{\mathrm{d}}{\mathrm{d}x} T_{j\circ 0}(x) \\
&= \left[1 + n^{-(j+1)/2}\, \frac{\mathrm{d}}{\mathrm{d}x} \pi_{j+1}(T_{j\circ 0}(x)) + \tfrac{1}{4}\, n^{-(j+2)/2} \left\{ \frac{\mathrm{d}}{\mathrm{d}x} \pi_{j+1}(T_{j\circ 0}(x)) \right\}^2 \right] \frac{\mathrm{d}}{\mathrm{d}x} T_{j\circ 0}(x),
\end{aligned}$$

by elementary calculations, the chain rule and the fundamental theorem of calculus. Observing that

$$\begin{aligned}
&1 + n^{-(j+1)/2}\, \frac{\mathrm{d}}{\mathrm{d}x} \pi_{j+1}(T_{j\circ 0}(x)) + \tfrac{1}{4}\, n^{-(j+2)/2} \left\{ \frac{\mathrm{d}}{\mathrm{d}x} \pi_{j+1}(T_{j\circ 0}(x)) \right\}^2 \quad > \\
&1 + n^{-(j+1)/2}\, \frac{\mathrm{d}}{\mathrm{d}x} \pi_{j+1}(T_{j\circ 0}(x)) + \tfrac{1}{4}\, n^{-(j+1)} \quad \left\{ \frac{\mathrm{d}}{\mathrm{d}x} \pi_{j+1}(T_{j\circ 0}(x)) \right\}^2,
\end{aligned}$$

since $-\dfrac{j+2}{2} = -(j+1) + \dfrac{1}{2} > -(j+1)$, and that

$$1 + n^{-(j+1)/2}\, \frac{\mathrm{d}}{\mathrm{d}x} \pi_{j+1}(T_{j\circ 0}(x)) + \tfrac{1}{4}\, n^{-(j+1)} \left\{ \frac{\mathrm{d}}{\mathrm{d}x} \pi_{j+1}(T_{j\circ 0}(x)) \right\}^2 = \left\{ \tfrac{1}{2}\, n^{-(j+1)/2}\, \frac{\mathrm{d}}{\mathrm{d}x} \pi_{j+1}(T_{j\circ 0}(x)) + 1 \right\}^2,$$

which is always positive, by the induction hypothesis that $\dfrac{\mathrm{d}}{\mathrm{d}x} T_{j\circ 0}(x) > 0$, it follows immediately that $\dfrac{\mathrm{d}}{\mathrm{d}x} T_{(j+1)\circ 0}(x) > 0$, for every x on the real line and thus, $T_{(j+1)\circ 0}(x)$ is strictly monotone increasing.
The proof is completed as the requirements of mathematical induction are fulfilled.

Corollary 3.1. *Let ν be some fixed integer greater that or equal to 1, $\pi_j(x)$ be, for every $j = 1, \ldots, \nu$, a polynomial on the real line and c be a constant such that $c \geq \frac{1}{4}$. Define the indefinite integral*

$$\int \left\{ \frac{\mathrm{d}}{\mathrm{d}y} \pi_k(y) \right\}^2 \mathrm{d}y = \Pi_k(y) + \textit{const.},$$

where const. is some arbitrary constant. Then, the transformation $T_{j\circ 0}(x) = (T_j \circ \ldots \circ T_0)(x)$ with

$$T_{k\circ 0}(x) = T_{(k-1)\circ 0}(x) + n^{-k/2}\, \pi_k(T_{(k-1)\circ 0}(x)) + c\, n^{-(k+1)/2}\, \Pi_k(T_{(k-1)\circ 0}(x)),$$

for some $k \in \{1, \ldots, j\}$ and $T_0(x) = x$, is one-to-one for every $j = 1, \ldots, \nu$.

Proof. By definition as follows. The transformation $T_{j\circ 0}(x)$ is strictly monotone increasing on the real line for every $j = 1, \ldots, \nu$, by Theorem 3.1. Take arbitrary j and any two distinct $x_1, x_2 \in \mathbb{R}$ such that $x_1 < x_2$. Then, $T_{j\circ 0}(x_1) < T_{j\circ 0}(x_2)$ by the monotonicity of the transformation. Thus, $T_{j\circ 0}(x_1) \neq T_{j\circ 0}(x_2)$ and, by definition, $T_{j\circ 0}(x)$ is one-to-one. Since j is arbitrary, $T_{j\circ 0}(x)$ is one-to-one on the real line for every

$j = 1, \ldots, \nu$. The proof is completed.

Theorem 3.2. *Let ν be some fixed integer greater that or equal to 1. Also, let the cumulative distribution function of a statistic S_n admit, uniformly in x on the whole real line, an asymptotic expansion of the form*

$$\begin{aligned} F(x) &\equiv P(S_n \leq x) \\ &= \Xi(x) + \sum_{\ell=1}^{\nu} n^{-\ell/2}\, \pi_{0,\ell}(x)\, \xi(x) + O(n^{-(\nu+1)/2}) \end{aligned} \tag{3.5}$$

as $n \to \infty$, where $\pi_{0,\ell}(x)$ is, for every $\ell = 1, \ldots, \nu$, a polynomial on the real line depended on the characteristics of the underlying distribution of the sample values but not depended on n, Ξ is the limiting cumulative distribution function and ξ is the corresponding probability density function. Define the indefinite integral

$$\int \left\{ \frac{\mathrm{d}}{\mathrm{d}y} \pi_{(k-1),k}(y) \right\}^2 \mathrm{d}y = \Pi_{(k-1),k}(y) + \textit{const.},$$

where const. is some arbitrary constant. Then, the transformed statistic $T_{j\circ 0}(S_n) = (T_j \circ \ldots \circ T_0)(S_n)$ with

$$T_{k\circ 0}(S_n) = T_{(k-1)\circ 0}(S_n) + n^{-k/2}\, \pi_{(k-1),k}(T_{(k-1)\circ 0}(S_n)) + c\, n^{-(k+1)/2}\, \Pi_{(k-1),k}(T_{(k-1)\circ 0}(S_n)), \tag{3.6}$$

for some $k \in \{1, \ldots, j\}$, $T_0(S_n) = S_n$ and $c \geq \frac{1}{4}$, obtains

$$\begin{aligned} G_j(x) &\equiv P(T_{j\circ 0}(S_n) \leq x) \\ &= \Xi(x) + O(n^{-(j+1)/2}), \end{aligned} \tag{3.7}$$

for every $j = 1, \ldots, \nu$.

Remark 3.1. The following notes are in order.

1. Subscript of $\pi_{(k-1),k}(\cdot)$: The first integer, i.e. $k-1$, indicates that the polynomial associated with the expansion of the $T_{(k-1)\circ 0}(\cdot)$ statistic is used for the current statistic $T_{k\circ 0}(\cdot)$ and the second integer, i.e. k, indicates the order, see $n^{-k/2}$, which it corresponds to.

2. Limiting distribution at (3.5) and (3.7): Theorem 3.2 is stated in general terms in an attempt to show that the transformation applies, for any fixed $\epsilon > 0$, to '***any*** *statistic $S = S_\epsilon$ of order $O_P(1)$ whose distribution is* $\mathrm{pr}(S \leq x) = \Xi(x) + \epsilon^k\, \xi(x) + O(\epsilon^{k+1})$ *for some $k > 0$, where Ξ is the limiting distribution of $S = S_\epsilon$*' (Kakizawa, 1996, pp. 923–924). This include statistics whose limiting distribution is the Standard Normal and statistics whose limiting distribution is the chi-square. Note that, here, $\psi(\cdot)$ is such that $\xi(x) = \pi_k(x) \dfrac{\mathrm{d}}{\mathrm{d}x} \Xi(x)$, where $\pi_k(\cdot)$ is some polynomial.

Proof. By mathematical induction on j and a repeated Taylor's series expansion argument. For brevity, the arguments are demonstrated first and then the induction's steps are described with reference to the arguments.
Arguments: Let k be some positive integer. The transformed statistic $T_{k\circ 0}(S_n)$ is written, conveniently, as $T_{k\circ 0}(S_n) = T_k(T_{(k-1)\circ 0}(S_n))$, using properties of composite functions. Thus,

$$T_k(T_{(k-1)\circ 0}(S_n)) = T_{(k-1)\circ 0}(S_n) + n^{-k/2}\, \pi_{(k-1),k}(T_{(k-1)\circ 0}(S_n)) + c\, n^{-(k+1)/2}\, \Pi_{(k-1),k}(T_{(k-1)\circ 0}(S_n)) \tag{3.8}$$

and

$$\begin{aligned} G_k(x) &\equiv P(T_{k\circ 0}(S_n) \leq x) \\ &= P(T_k(T_{(k-1)\circ 0}(S_n)) \leq x) \\ &= P(T_{(k-1)\circ 0}(S_n) \leq T_k^{-1}(x)), \end{aligned} \tag{3.9}$$

where x is some known value on the real line. As such, x is considered fixed.
Firstly, a Taylor's series expansion of $T_k^{-1}(x)$ is computed and written in the convenient form of $T_k^{-1}(x) = x - n^{-k/2}\,\pi_{(k-1),k}(x) + O(n^{-(k+1)/2})$ following the motivation driving (3.4). The same line of arguments can be found, also, in Yanagihara and Yuan (2005, pp. 234–235), though, here, they are treated in a more general context. Secondly, the expansion for $G_k(x)$ is derived, assuming that the expansion for $G_{k-1}(x)$ is available.
By equation (3.8), write

$$T_k(x) = x + n^{-k/2}\,\pi_{(k-1),k}(x) + c\,n^{-(k+1)/2}\,\Pi_{(k-1),k}(x). \tag{3.10}$$

By Theorem 3.1, T_k is strictly monotone increasing and, as a polynomial, T_k is continuous over the whole real line. Thus, T_k is invertible and let T_k^{-1} denote its inverse. It follows, then, from equation (3.10), that

$$x = T_k^{-1}(x + n^{-k/2}\,\pi_{(k-1),k}(x) + c\,n^{-(k+1)/2}\,\Pi_{(k-1),k}(x)),$$

where the $n^{-k/2}\,\pi_{(k-1),k}(x) + c\,n^{-(k+1)/2}\,\Pi_{(k-1),k}(x)$ part, which is fixed as x is fixed, is regarded as some departure away from x. A Taylor's series expansion (see, e.g., formula (8.1) in Bhattacharya and Rao (2010, p. 57) gives

$$x = T_k^{-1}(x) + \left\{n^{-k/2}\,\pi_{(k-1),k}(x) + c\,n^{-(k+1)/2}\,\Pi_{(k-1),k}(x)\right\}\,\frac{\mathrm{d}}{\mathrm{d}x}T_k^{-1}(x) + O(n^{-(k+1)/2})$$

and by gathering terms of the same order,

$$x = T_k^{-1}(x) + n^{-k/2}\,\pi_{(k-1),k}(x)\,\frac{\mathrm{d}}{\mathrm{d}x}T_k^{-1}(x) + O(n^{-(k+1)/2}). \tag{3.11}$$

The derivative of $T_k^{-1}(x)$ with respect to x needs to be computed. The inverse function theorem gives,

$$\frac{\mathrm{d}}{\mathrm{d}x}T_k^{-1}(x) = \left\{\frac{\mathrm{d}}{\mathrm{d}w}T_k(w)\Big|_{w=T_k^{-1}(x)}\right\}^{-1},$$

where an evaluation of the derivative of $T_k(w)$ with respect to w at $T_k^{-1}(x)$ results in

$$\frac{\mathrm{d}}{\mathrm{d}x}T_k^{-1}(x) = \left\{1 + n^{-k/2}\,\frac{\mathrm{d}}{\mathrm{d}w}\pi_{(k-1),k}(w)\Big|_{w=T_k^{-1}(x)} + O(n^{-(k+1)/2})\right\}^{-1}. \tag{3.12}$$

The derivative of $\pi_{(k-1),k}(w)$ with respect to w evaluated at $T_k^{-1}(x)$ is computed next. A Taylor's series expansion is used again. It is convenient to write, for the moment, that $T_k^{-1}(x) = x + O(n^{-k/2})$, after an elementary rearrangement of equation (3.11), so that $w = T_k^{-1}(x) = x + O(n^{-k/2})$. Recall that $\pi_{(k-1),k}$ is a

polynomial. Hence,

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}w}\pi_{(k-1),k}(w)\bigg|_{w=T_k^{-1}(x)} &= \frac{\mathrm{d}}{\mathrm{d}w}\pi_{(k-1),k}(w)\bigg|_{w=x+O(n^{-k/2})} \\ &= \frac{\mathrm{d}}{\mathrm{d}w}\pi_{(k-1),k}(w)\bigg|_{w=x} + O(n^{-k/2}),\end{aligned}$$

Therefore, equation (3.12) reduces to

$$\frac{\mathrm{d}}{\mathrm{d}x}T_k^{-1}(x) = \left\{1 + n^{-k/2}\,\frac{\mathrm{d}}{\mathrm{d}w}\pi_{(k-1),k}(w)\bigg|_{w=x} + O(n^{-(k+1)/2})\right\}^{-1}. \tag{3.13}$$

By letting $z_x = n^{-k/2}\,\dfrac{\mathrm{d}}{\mathrm{d}w}\pi_{(k-1),k}(w)\bigg|_{w=x} + O(n^{-(k+1)/2})$ notice that z_x depends on x but it is fixed as x is fixed. Using the well known expansion formula for $(1+z_x)^{-1}$, equation (3.13) is written as

$$\frac{\mathrm{d}}{\mathrm{d}x}T_k^{-1}(x) = 1 - n^{-k/2}\,\frac{\mathrm{d}}{\mathrm{d}w}\pi_{(k-1),k}(w)\bigg|_{w=x} + O(n^{-(k+1)/2}). \tag{3.14}$$

Equations (3.11) and (3.14), combined together, give

$$x = T_k^{-1}(x) + n^{-k/2}\,\pi_{(k-1),k}(x) + O(n^{-(k+1)/2}),$$

from which, by an elementary rearrangement,

$$T_k^{-1}(x) = x - n^{-k/2}\,\pi_{(k-1),k}(x) + O(n^{-(k+1)/2}). \tag{3.15}$$

Now, for some ω on the real line, let

$$\begin{aligned}G_{k-1}(\omega) &\equiv P(T_{(k-1)\circ 0}(S_n) \le \omega) \\ &= \Xi(\omega) + \sum_{\ell=k}^{\nu} n^{-\ell/2}\,\pi_{(k-1),\ell}(\omega)\,\xi(\omega) + O(n^{-(\nu+1)/2}),\end{aligned} \tag{3.16}$$

with $G_0(\omega) \equiv F(\omega)$, for $k=1$, and $T_0(S_n) = S$. Then, equation (3.9) is expanded as in

$$\begin{aligned}G_k(x) &\equiv P(T_{(k-1)\circ 0}(S_n) \le T_k^{-1}(x)) \\ &= \Xi(T_k^{-1}(x)) + \sum_{\ell=k}^{\nu} n^{-\ell/2}\,\pi_{(k-1),\ell}(T_k^{-1}(x))\,\xi(T_k^{-1}(x)) + O(n^{-(\nu+1)/2}).\end{aligned} \tag{3.17}$$

Regarding the $-n^{-k/2}\,\pi_{(k-1),k}(x) + O(n^{-(k+1)/2})$ part of equation (3.15) as some departure away from x, which is considered fixed, a Taylor's series expansion argument is employed to $\Xi(T_k^{-1}(x))$, $\pi_{(k-1),\ell}(T_k^{-1}(x))$ and $\xi(T_k^{-1}(x))$, respectively, in

$$\begin{aligned}\Xi(T_k^{-1}(x)) &= \Xi(x) + \left\{-n^{-k/2}\,\pi_{(k-1),k}(x) + O(n^{-(k+1)/2})\right\}\,\frac{\mathrm{d}}{\mathrm{d}x}\Xi(x) + O(n^{-(k+1)/2}) \\ &= \Xi(x) - n^{-k/2}\,\pi_{(k-1),k}(x)\,\xi(x) + O(n^{-(k+1)/2}),\end{aligned} \tag{3.18}$$

$$\begin{aligned}\pi_{(k-1),\ell}(T_k^{-1}(x)) &= \pi_{(k-1),\ell}(x) + \left\{-n^{-k/2}\,\pi_{(k-1),k}(x) + O(n^{-(k+1)/2})\right\}\,\frac{\mathrm{d}}{\mathrm{d}x}\pi_{(k-1),\ell}(x) + O(n^{-(k+1)/2}), \\ &= \pi_{(k-1),\ell}(x) + O(n^{-k/2}), \quad \text{for all } \ell = k,\ldots,\nu\end{aligned} \tag{3.19}$$

and

$$\begin{aligned}\xi(T_k^{-1}(x)) &= \xi(x) + \left\{-n^{-k/2}\,\pi_{(k-1),k}(x) + O(n^{-(k+1)/2})\right\}\,\frac{\mathrm{d}}{\mathrm{d}x}\xi(x) + O(n^{-(k+1)/2}) \\ &= \xi(x) + O(n^{-k/2}). \end{aligned} \tag{3.20}$$

Attention is drawn, firstly, on the middle term of equation (3.17), which by substitution of equations (3.19) and (3.20) into equation (3.17) and use of properties of the O-notation, is simplified as in

$$\begin{aligned}\sum_{\ell=k}^{\nu} n^{-\ell/2}\,\pi_{(k-1),\ell}(T_k^{-1}(x))\,\xi(T_k^{-1}(x)) &= \sum_{\ell=k}^{\nu} n^{-\ell/2}\,\left\{\pi_{(k-1),\ell}(x) + O(n^{-k/2})\right\}\,\left\{\xi(x) + O(n^{-k/2})\right\} \\ &= \sum_{\ell=k}^{\nu} n^{-\ell/2}\,\pi_{(k-1),\ell}(x)\,\xi(x) + O(n^{-k}) \\ &= n^{-k/2}\,\pi_{(k-1),k}(x)\,\xi(x) + O(n^{-(k+1)/2}). \end{aligned} \tag{3.21}$$

Then, equations (3.18) and (3.21) are substituted into equation (3.17) to give

$$\begin{aligned}G_k(x) &\equiv P(T_{k\circ 0}(S_n) \le x) \\ &= P(T_{(k-1)\circ 0}(S_n) \le T_k^{-1}(x)) \\ &= \Xi(x) + O(n^{-(k+1)/2}), \end{aligned} \tag{3.22}$$

after some elementary algebraic calculations, where same terms are cancelled out. As shown next, induction concludes the proof.

Induction on j: Let k be j in the *Arguments*. Then, for $j = 1$, equation (3.16) is, for $\omega = x$, identical to equation (3.5) and equation (3.22) gives

$$\begin{aligned}G_1(x) &\equiv P(T_1(S_n) \le x) \\ &= \Xi(x) + O(n^{-1}). \end{aligned} \tag{3.23}$$

Thus, equation (3.23) verifies the claim in equation (3.7) for $j = 1$. Assume now that, for some positive integer $j > 1$,

$$\begin{aligned}G_j(x) &\equiv P(T_{j\circ 0}(S_n) \le x) \\ &= \Xi(x) + O(n^{-(j+1)/2}). \end{aligned} \tag{3.24}$$

The aim is to show that for the next positive integer, i.e. $j + 1$,

$$\begin{aligned}G_{j+1}(x) &\equiv P(T_{(j+1)\circ 0}(S_n) \le x) \\ &= \Xi(x) + O(n^{-(j+2)/2}). \end{aligned} \tag{3.25}$$

Let k be $j + 1$ in the *Arguments*. Then, for $k = j + 1$, equation (3.16) becomes, for $\omega = x$, the induction hypothesis, which is written concisely in equation (3.24). Equation (3.17), for $k = j + 1$, follows, then, and finally equation (3.22) gives, for $k = j + 1$,

$$\begin{aligned}G_{j+1}(x) &\equiv P(T_{(j+1)\circ 0}(S_n) \le x) \\ &= P(T_{j\circ 0}(S_n) \le T_{j+1}^{-1}(x)) \\ &= \Xi(x) + O(n^{-(j+2)/2}). \end{aligned} \tag{3.26}$$

Equation (3.26) verifies the claim in equation (3.25). Thus, equation (3.7) applies for every $j = 1, \ldots, \nu$. The proof is completed as the requirements of mathematical induction are fulfilled.

3.4 A corrected confidence curve

The following result gives the order of correction to a confidence curve when transformations are taken into account.

Theorem 3.3. *Let ν be some fixed integer greater than or equal to 1 and $S_n = S_n(\psi)$ be a pivotal statistic whose asymptotic expansion is given at (3.5) with limiting cumulative distribution function Ξ. Define the asymptotic confidence curve, $cc(\psi) = \Xi(S_n(\psi))$. Also, for each $j = 1, \ldots, \nu$, define the asymptotic confidence curve, $cc_{tr,j}(\psi) = \Xi(T_{j\circ 0}(S_n(\psi)))$. Then,*

$$cc_{tr,j}(\psi) = cc(\psi) + O_P(n^{-(j+1)/2}).$$

Remark 3.2. For notational convenience, the dependence on y has been dropped from, both, the statistic and the defined confidence curves.

Proof. By Definition 1.5 and Theorem (3.2), the proof is completed.

3.5 Smooth function model

The following discussion will be unfolded under the '*smooth function model*' of Hall (1992b, pp. 52, 238), which encompasses, for example, problems where the parameter to be estimated, ψ at (3.1), is either a mean, or a variance, or a ratio of means or variances, or a difference of means or variances, or a correlation coefficient and view things in the nonparametric modus of thinking. The seminal paper by Bhattacharya and Ghosh (1978), with correction in Bhattacharya and Ghosh (1980), establishes the validity of the formal Edgeworth expansion under this model. In addition, the model allows for '*approximate moments*' of the asymptotic distribution of a pivotal type statistic, which is written as a function of sample averages (see below), to be computed by the use of the so-called *delta method* discussed in, e.g., Bhattacharya and Ghosh (1978, pp. 435, 438), Hall (1992b, pp. 72, 76) and Small (2010, pp. 99–111), thus, overcoming possible existence of moments problems and permitting considerations of, both, theoretical and practical interest to be benefited. Formulae regarding cumulants can be expanded as a power series of the form

$$\kappa_{j,n} = n^{-(j-2)/2}(k_{j,1} + n^{-1}\, k_{j,2} + n^{-2}\, k_{j,3} + \ldots), \tag{3.27}$$

where $j \geq 1$ and the constants $k_{j,\cdot}$ depend, among other things, on the moments of the underlying sampling distribution. It is noted that, since for a statistic that is centred and scaled its limiting mean, $\kappa_{1,n}$, is 0 and limiting variance, $\kappa_{2,n}$, is 1, it follows that $k_{1,1} = 0$ and $k_{2,1} = 1$. It is noted, also, that thorough discussions on cumulant expansions may be found in James (1955, 1958) and James and Mayne (1962), for example. Further, Hall (1992b, p. 83) explains properties of bootstrap methods for estimating distributions, among other things, by giving an Edgeworth view of the bootstrap under the framework of this model. Hence, a brief description of the '*smooth function model*' is given. For the current presentation, attention is restricted to data stemming from a univariate population.

Let $\{Y_1, Y_2, \ldots, Y_n\}$ be independent and identically distributed random variables from a sufficiently smooth probability distribution, P_θ. Each Y_i, $i = 1, \ldots, n$ is one-dimensional. The '*smooth function model*' consists of expressing the parameter of interest as a smooth function of population means. To this end, let the d-dimensional, independent and identically distributed, random vectors $X_i = (Y_i, Y_i^2, \ldots, Y_i^d)^t$, be

defined for each $i = 1, \ldots, n$. The value of the dimension d, depends on which is the targeting or focus parameter, ψ, to be estimated. Table 3.1 reports the dimension, d, for four cases of frequent interest. See, discussion in Hall (1992b, p. 52–53).

ψ	d
mean	2
variance	4
variance ratio	4
correlation coefficient	5

Table 3.1: Value of dimension d for some cases of focus parameter ψ.

Let h be a smooth function such that $h : \mathbb{R}^d \to \mathbb{R}$. Also, let μ denote, for each $i = 1, \ldots, n$, the mean of X_i such that $\mu = E(X_i) = \big(E(Y_i)\,,\, E(Y_i^2)\,,\, \ldots\,,\, E(Y_i^d)\big)^t$. The focus parameter, ψ, is, then, expressed as a function of μ by writing $\psi = h(\mu)$ and $\widehat{\psi}_n$ is expressed as a function of sample averages by writing $\widehat{\psi}_n = h(\bar{X})$, where

$$\bar{X} = \left(n^{-1} \sum_{i=1}^{n} Y_i\,,\, n^{-1} \sum_{i=1}^{n} Y_i^2\,,\, \ldots\,,\, n^{-1} \sum_{i=1}^{n} Y_i^d \right)^t .$$

It is assumed that $\widehat{\psi}_n$ has asymptotic variance $n^{-1} g^2(\mu)$, where $g : \mathbb{R}^d \to \mathbb{R}$ is a known smooth function. That is, $g^2(\mu) = \lim_{n\to\infty} \mathrm{var}(n^{1/2}\widehat{\psi}_n)$. An estimate of $g(\mu)$ is given by $g(\bar{X})$. In this context, the pivotal t-statistic at (3.1) takes the form

$$S_n = \frac{n^{1/2}\,(\,h(\mu) - h(\bar{X})\,)}{g(\bar{X})}, \tag{3.28}$$

where the dependence on the sample size, n, is emphasised and notational convenience permits the change from t to S_n. Although, it would have been easier to consider cases where the population variance is known and, thus, $g(\mu)$ is known, it is hardly ever the case, in practice, to have a known variance and the t-statistic is presented in the form at (3.28). Essentially, the statistic is seen as a function, say, A, such that $A : \mathbb{R}^d \to \mathbb{R}$ with $z \mapsto n^{1/2}\,(\,h(\mu) - h(\bar{z})\,)/g(\bar{z})$ and $A(\mu) = 0$, although, this might obscure, a bit, its pivotal character, cf. Definitions 1.3 and 1.4, but still it is recognised as a pivotal quantity of some form.

3.6 Edgeworth Expansions

Now that the working framework has been, briefly, described, a short introduction to the main aspects of the well known Edgeworth expansion is given, as further results are based on the assumption of such expansions. Let S_n denote a statistic whose limiting distribution is Standard Normal. Then, an Edgeworth expansion of the statistic's distribution function, $P(S_n \leq x)$, is an expression in powers of $n^{-1/2}$ of the form

$$\begin{aligned} P(S_n \leq x) &= \Phi(x) + n^{-1/2}\, p_1(x)\, \phi(x) + n^{-1}\, p_2(x)\, \phi(x) + \ldots + n^{-\nu/2}\, p_\nu(x)\, \phi(x) + \ldots \\ &= \Phi(x) + \sum_{\ell=1}^{\infty} n^{-\ell/2}\, p_\ell(x)\, \phi(x), \end{aligned} \tag{3.29}$$

where the p_ν's denote polynomials of degree $3\nu - 1$, for integer $\nu \geq 1$. It is easily seen that, odd indexed polynomials have even degree, constituting even functions and, correspondingly, even indexed polynomials have odd degree, constituting odd functions. It is important to note that, the polynomial functions need to obey this property, since otherwise, the coverage error of a confidence procedure for a focus parameter

at hand, ψ, based on the Normal approximation to the distribution of a Studentised statistic of the form at (3.1) would be of a larger size and there is no remedy guarantee when bootstrap is called to action. See, e.g., Hall (1992b, pp. 33, 49, 320–321) for further comments and an example on this. Here, confidence procedures include, of course, confidence distributions or confidence curves. The p_ν polynomial functions have coefficients that depends on the cumulants. As in Hall (1992b, p. 47), the first polynomial, i.e. p_1, shall be called a '*skewness correction*' and the second polynomial, i.e. p_2, shall be called a '*correction for kurtosis and for the secondary effect of skewness*'. Note that, the product $p_\nu(x)\,\phi(x)$ tends to 0 as x tends to either $-\infty$ or ∞, for integer $\nu \geq 1$.

In the majority of cases, though, expansions of the form at (3.29) are divergent as infinite series and only available as asymptotic series or asymptotic expansions. Therefore, the term *Edgeworth expansion* shall, henceforth, be used to mean an expression of the form

$$\begin{aligned} P(S_n \leq x) &= \Phi(x) + n^{-1/2}\,p_1(x)\,\phi(x) + n^{-1}\,p_2(x)\,\phi(x) + \ldots + n^{-\nu/2}\,p_\nu(x)\,\phi(x) + o(n^{-\nu/2}) \\ &= \Phi(x) + \sum_{\ell=1}^{\nu} n^{-\ell/2}\,p_\ell(x)\,\phi(x) + o(n^{-\nu/2}). \end{aligned} \tag{3.30}$$

To be more precise, the expression at (3.30) is taken to mean that if the expansion at (3.29) is truncated after a predetermined or given number of terms, i.e. a fixed ν, then the remainder is $o(n^{-\nu/2})$ and has a smaller order than the order of the last included term. It does not matter if the order of the last included term is not stated or specified exactly. Nevertheless, the expansion at (3.30) is valid for fixed ν, as $n \to \infty$.

A detailed account on the derivation of expansions such as those described by (3.29) may be found, for example, in Petrov (1975, ch. VI), Barndorff-Nielsen and Cox (1989, sec. 4.2), Field and Ronchetti (1990, sec. 2.3.b), Hall (1992b, ch. 2) and Small (2010, sec. 7.10). By skimming the surface of derivation details, some formulae are, intentionally, presented, here, to enable a brief discussion on a following example. The notation will mostly follow Hall (1992b), with necessary changes, of course, but the essence of the arguments can be found in all of the cited references just listed. Their common denominator is that, they all use the characteristic function to say something about errors in Normal approximations or, expressed differently, to fine-tune the approximation provided by the Central Limit Theorem with assistance from cumulants. Note that the characteristic function of a Standard Normal is given by $e^{-t^2/2}$, for real t.

Let λ_n denote the characteristic function of a statistic with limiting Standard Normal distribution, say, S_n. Then,

$$\begin{aligned} \lambda_n(t) &= E\{\exp(itS_n)\} \\ &= \exp\left\{\kappa_{1,n}\,it + \tfrac{1}{2}\,\kappa_{2,n}\,(it)^2 + \ldots + \tfrac{1}{j!}\,\kappa_{j,n}\,(it)^j + \ldots\right\}. \end{aligned} \tag{3.31}$$

It can be seen that, by taking $\log \lambda_n(t)$, the jth cumulant, $\kappa_{j,n}$, is defined as the coefficient of $\frac{1}{j!}\,(it)^j$ in the resulting power series expansion. Theorem 2.1 in Hall (1992b, p. 53) establishes the validity of the power series at (3.27) under the '*smooth function model*' and, thus, it can be substituted, for each integer $j \geq 1$, into the expression at (3.31). After some lengthly algebraic calculations, the characteristic function is written

as in

$$\begin{aligned}\lambda_n(t) &= \exp\Big\{-\tfrac{1}{2}t^2 + n^{-1/2}\,\big[\,k_{1,2}\,it + \tfrac{1}{3!}\,k_{3,1}\,(it)^3\,\big] + n^{-1}\,\big[\,\tfrac{1}{2!}\,k_{2,2}\,(it)^2 + \tfrac{1}{4!}\,k_{4,1}\,(it)^4\,\big] + \ldots\Big\} \\ &= \exp\big\{-\tfrac{1}{2}t^2\big\}\cdot \\ &\quad \exp\Big\{n^{-1/2}\,\big[\,k_{1,2}\,it + \tfrac{1}{3!}\,k_{3,1}\,(it)^3\,\big] + n^{-1}\,\big[\,\tfrac{1}{2!}\,k_{2,2}\,(it)^2 + \tfrac{1}{4!}\,k_{4,1}\,(it)^4\,\big] + \ldots\Big\}. \qquad (3.32)\end{aligned}$$

Using the form of the well known expansion formula for e^z, the second $\exp\{\ldots\}$ at (3.32) is expanded and, after some algebraic efforts, the characteristic function is, further, written as in

$$\begin{aligned}\lambda_n(t) &= \exp\big\{-\tfrac{1}{2}t^2\big\}\,\Bigg(1 + n^{-1/2}\,\big[\,k_{1,2}\,it + \tfrac{1}{6}\,k_{3,1}\,(it)^3\,\big] + \\ &\quad n^{-1}\,\Big[\,\tfrac{1}{2!}\,k_{2,2}\,(it)^2 + \tfrac{1}{24}\,k_{4,1}\,(it)^4 + \tfrac{1}{2}\,\big[\,k_{1,2}\,it + \tfrac{1}{6}\,k_{3,1}\,(it)^3\,\big]^2\Big] + O(n^{-3/2})\Bigg). \qquad (3.33)\end{aligned}$$

More compactly,

$$\lambda_n(t) = e^{-t^2/2} + n^{-1/2}\,r_1(it)\,e^{-t^2/2} + n^{-1}\,r_2(it)\,e^{-t^2/2} + \ldots + n^{-j/2}\,r_j(it)\,e^{-t^2/2} + \ldots, \qquad (3.34)$$

where the r_j's denote polynomials of degree no more than $3j$, for integer $j \geq 1$. It is easily seen that, odd indexed polynomials are odd and even indexed polynomials are even. By comparing (3.33) and (3.34), the first two r_j polynomials are given in

$$r_1(it) = k_{1,2}\,it + \tfrac{1}{6}\,k_{3,1}\,(it)^3$$

and

$$\begin{aligned}r_2(it) &= \tfrac{1}{2!}\,k_{2,2}\,(it)^2 + \tfrac{1}{24}\,k_{4,1}\,(it)^4 + \tfrac{1}{2}\,\big[\,k_{1,2}\,it + \tfrac{1}{6}\,k_{3,1}\,(it)^3\,\big]^2 \\ &= \tfrac{1}{2!}\,[\,k_{2,2} + k_{1,2}^2\,]\,(it)^2 + \tfrac{1}{24}\,[\,k_{4,1} + 4\,k_{1,2}\,k_{3,1}\,]\,(it)^4 + \tfrac{1}{72}\,k_{3,1}^2\,(it)^6.\end{aligned}$$

To derive an expression of the form at (3.29), the expansion at (3.34) needs to be inverted. The arguments for the inversion are initiated by noticing that

$$\lambda_n(t) = \int_{-\infty}^{\infty} e^{itx} dP(S_n \leq x) \qquad \text{and} \qquad e^{-t^2/2} = \int_{-\infty}^{\infty} e^{itx} d\Phi(x).$$

Then, by using a Fourier-Stieltjes transform of the form

$$\int_{-\infty}^{\infty} e^{itx} d\{-\mathrm{HER}_{j-1}(x)\phi(x)\} = (it)^j\,e^{-t^2/2},$$

where $j \geq 1$, integer, the first two p_j polynomials in the expression at (3.29) are given by

$$\begin{aligned}p_1(x) &= -k_{1,2}\,\mathrm{HER}_0(x) - \tfrac{1}{6}\,k_{3,1}\,\mathrm{HER}_2(x) \\ &= -k_{1,2} - \tfrac{1}{6}\,k_{3,1}\,(x^2 - 1) \qquad (3.35)\end{aligned}$$

and

$$\begin{aligned}p_2(x) &= \tfrac{1}{2!}\,[\,k_{2,2} + k_{1,2}^2\,]\,\mathrm{HER}_1(x) + \tfrac{1}{24}\,[\,k_{4,1} + 4\,k_{1,2}\,k_{3,1}\,]\,\mathrm{HER}_3(x) + \tfrac{1}{72}\,k_{3,1}^2\,\mathrm{HER}_5(x) \\ &= -x\,\big\{\tfrac{1}{2!}\,[\,k_{2,2} + k_{1,2}^2\,] + \tfrac{1}{24}\,[\,k_{4,1} + 4\,k_{1,2}\,k_{3,1}\,]\,(x^2 - 3) + \tfrac{1}{72}\,k_{3,1}^2\,(x^4 - 10\,x^2 + 15)\big\}. \qquad (3.36)\end{aligned}$$

Note that, for some nonnegative integer, ℓ, the function $\mathrm{HER}_\ell(\cdot)$ denotes the so-called *ℓth Chebyshev-Hermite polynomial*. Chebyshev-Hermite polynomials are treated in Petrov (1975, p. 137). Details regarding Hermite functions and their derivation can be found, for example, in Arfken et al. (2013, sec. 18.1). Table 3.2 reports the first six Hermite polynomials.

ℓ	$\ell-1$	$\mathrm{HER}_{\ell-1}(x)$
1	0	1
2	1	x
3	2	x^2-1
4	3	$x\,(x^2-3)$
5	4	$x^4-6\,x^2+3$
6	5	$x\,(x^4-10\,x^2+15)$

Table 3.2: The first six Chebyshev-Hermite polynomials.

Without stating, yet, any explicit regularity conditions, although it has been assumed that these hold, let an Edgeworth expansion such as the one in

$$P(S_n \leq x) \quad = \quad \Phi(x) + n^{-1/2}\,p_1(x)\,\phi(x) + n^{-1}\,p_2(x)\,\phi(x) + \ldots + n^{-\nu/2}\,p_\nu(x)\,\phi(x) + o(n^{-\nu/2})$$

be available under the '*smooth function model*'. For completeness, it remains to provide a basic insight on how the four constants, $k_{1,2}$, $k_{2,2}$, $k_{3,1}$ and $k_{4,1}$, in the expression of the first two p_j polynomials at (3.35) and (3.36) are computed. This aims at further unfolding the machinery of the '*smooth function model*'. The example 2.1 in Hall (1992b, pp. 71–73) lends some of its details due to (i) its familiarity and (ii) its somewhat tractability regarding the algebraic computations.

Hence, let the focus parameter ψ be the mean, m, of a univariate population with unknown variance $\beta^2 > 0$, from which $\{Y_1, \ldots, Y_n\}$ are i.i.d. random variables. Table 3.1 reports that in the case when the focus parameter under study is the mean, $d = 2$, and, thus, for each $i = 1, \ldots, n$, $X_i = (Y_i, Y_i^2)^t$ and $\mu = E(X_i) = (E(Y_i), E(Y_i^2))^t$, where $E(Y_i) = m$ and $E(Y_i^2) = E^2(Y_i) + \mathrm{var}(Y_i) = m^2 + \beta^2$. Also,

$$\bar{X} = \left(n^{-1} \sum_{i=1}^{n} Y_i \,,\, n^{-1} \sum_{i=1}^{n} Y_i^2 \right)^t = \left(\bar{Y} \,,\, n^{-1} \sum_{i=1}^{n} Y_i^2 \right)^t .$$

Define, conveniently, the functions $h : \mathbb{R}^2 \to \mathbb{R}$ with $(z_1, z_2)^t \mapsto z_1$ and $g : \mathbb{R}^2 \to \mathbb{R}$ with $(z_1, z_2)^t \mapsto \left\{ z_2 - z_1^2 \right\}^{1/2}$. Then, $h(\mu) = m$, $h(\bar{X}) = \bar{Y}$ and $g(\bar{X}) = \left\{ n^{-1} \sum_{i=1}^{n} Y_i^2 - \bar{Y}^2 \right\}^{1/2}$.

Since $S_n = A(\bar{X}) = \dfrac{n^{1/2}\,(\,h(\mu) - h(\bar{X})\,)}{g(\bar{X})}$ and $\widehat{\beta}^2 = n^{-1} \sum_{i=1}^{n} (Y_i - \bar{Y})^2 = n^{-1} \sum_{i=1}^{n} Y_i^2 - \bar{Y}^2$, the statistic at (3.28) becomes $S_n = \dfrac{n^{1/2}\,(\,m - \bar{Y}\,)}{\widehat{\beta}}$. Notice that the biased estimator for β has been used. For the unbiased estimator, $S_n^{\mathrm{unb}} = \left(1 - \frac{1}{n}\right) S_n$. Since, the statistic S_n is invariant under the transformation from Y_i to Y_i/β, it is assumed, without loss of generality, that, for each $i = 1, \ldots, n$, $E(Y_i) = 0$ and $E^2(Y_i) = 1$. Thus, $\mathrm{var}(Y_i) = 1$ and $S_n = -\dfrac{n^{1/2}\,\bar{Y}}{\widehat{\beta}}$.

Following Remark 3.3 in Schweder and Hjort (2016, p. 70), the distribution of $-S_n$ is considered further on

since $-S_n = \dfrac{n^{1/2}\,\bar{Y}}{\widehat{\beta}}$. An algebraic trick and the well known expansion formula for $(1+z_Y)^{-1/2}$, assist in breaking down $-S_n$ as in

$$\begin{aligned}
-S_n \;=\; \frac{n^{1/2}\,\bar{Y}}{\widehat{\beta}} \;=\; n^{1/2}\,\bar{Y}\,\widehat{\beta}^{-1} \;&=\; n^{1/2}\,\bar{Y}\,\left\{n^{-1}\,\sum_{i=1}^{n} Y_i^2 - \bar{Y}^2\right\}^{-1/2} \\
&=\; n^{1/2}\,\bar{Y}\,\left\{1 + n^{-1}\,\sum_{i=1}^{n} Y_i^2 - n^{-1}\,n - \bar{Y}^2\right\}^{-1/2} \\
&=\; n^{1/2}\,\bar{Y}\,\left\{1 + n^{-1}\,\sum_{i=1}^{n} (Y_i^2 - 1) - \bar{Y}^2\right\}^{-1/2} \\
&=\; n^{1/2}\,\bar{Y}\,\left[1 - \tfrac{1}{2}\,n^{-1}\,\sum_{i=1}^{n} (Y_i^2 - 1) + \tfrac{1}{2}\,\bar{Y}^2 + \right. \\
&\qquad \left. \tfrac{3}{8}\,\left\{n^{-1}\,\sum_{i=1}^{n} (Y_i^2 - 1)\right\}^2 + O_P(n^{-3/2})\right]. \qquad (3.37)
\end{aligned}$$

It should be noted that, Theorem 2.1 in Hall (1992b, p. 53) requires to take cumulants of a Taylor series approximant of $-S_n$ rather than of $-S_n$ itself, but to lessen algebraic burden it suffice to work with the cumulants of $-S_n$ itself for the present case. See, also, footnotes in Hall (1992b, pp. 72, 99). Expressions for $E(-S_n)$, $E(\{-S_n\}^2)$, $E(\{-S_n\}^3)$ and $E(\{-S_n\}^4)$ are given using the corresponding expressions for S_n in Hall (1992b, pp.72–73). The latter ones are derived using lengthly algebraic computations that involve the linearity of expectation, the random variables being i.i.d. and the definitions of skewness, γ, and kurtosis κ. Thus,

$$E(-S_n) = -E(S_n) = \tfrac{1}{2}\,n^{-1/2}\,\gamma + O(n^{-3/2}),$$

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

$$E(\{-S_n\}^3) = -E(S_n^3) = \tfrac{7}{2}\,n^{-1/2}\,\gamma + O(n^{-3/2}),$$

$$E(\{-S_n\}^4) = E(S_n^4) = 3 + n^{-1}\,(28\gamma^2 - 2\kappa + 24) + O(n^{-2}),$$

where the linearity of expectation was used. The cumulants are, then, given by

$$\kappa_{1,n} = E(-S_n) = \tfrac{1}{2}\,n^{-1/2}\,\gamma + O(n^{-3/2}),$$

$$\kappa_{2,n} = E(\{-S_n\}^2) - [E(-S_n)]^2 = 1 + \tfrac{1}{4}\,n^{-1}\,(7\gamma^2 + 12) + O(n^{-2}),$$

$$\kappa_{3,n} = E(\{-S_n\}^3) - 3\,E(\{-S_n\}^2)\,E(-S_n) + 2\,[E(-S_n)]^3 = 2\,n^{-1/2}\,\gamma + O(n^{-3/2}),$$

$$\begin{aligned}
\kappa_{4,n} &= E(\{-S_n\}^4) - 4\,E(\{-S_n\}^3)\,E(-S_n) - 3\,[E(\{-S_n\}^2)]^2 + 12\,E(\{-S_n\}^2)\,[E(-S_n)]^2 - 6\,[E(-S_n)]^4 \\
&= n^{-1}\,(12\gamma^2 - 2\kappa + 6) + O(n^{-2}),
\end{aligned}$$

from which, using the expression at (3.27),

$$k_{1,2} = \tfrac{1}{2}\,\gamma, \qquad k_{2,2} = \tfrac{1}{4}\,(7\gamma^2 + 12), \qquad k_{3,1} = 2\,\gamma, \qquad k_{4,1} = 12\gamma^2 - 2\kappa + 6.$$

The constants $k_{1,2}$, $k_{2,2}$, $k_{3,1}$ and $k_{4,1}$, are substituted into (3.35) and (3.36) to give after elementary algebraic calculations

$$p_1(x) = -\tfrac{1}{6}\,\gamma\,(2x^2 + 1)$$

and

$$p_2(x) = x\,\left\{\tfrac{1}{12}\,\kappa\,(x^2-3) - \tfrac{1}{18}\,\gamma^2\,(x^4+2x^2-3) - \tfrac{1}{4}\,(x^2+3)\right\}.$$

Notice that the sign of $k_{1,2}$, $k_{3,1}$ is positive and the sign of $p_1(x)$ is negative, whereas, the corresponding ones in Hall (1992b, pp. 72–73) have the opposite sings. This is a consequence of considering $-S_n$ in place of S_n. The signs of $k_{2,2}$, $k_{4,1}$ and $p_2(x)$ remain, as checked, the same.

3.7 Minimum moment conditions

Recall that in the '*smooth function model*', the d-dimensional i.i.d. random vectors $X_i = (Y_i, Y_i^2, \ldots, Y_i^d)^t$, were defined for each $i = 1, \ldots, n$, where $\{Y_1, Y_2, \ldots, Y_n\}$ denoted i.i.d. random variables from a sufficiently smooth probability distribution of a univariate population. Theorem 2.2 in Hall (1992b, p. 56) states sufficient regularity conditions under which an Edgeworth expansion is available uniformly in x on the whole real line. By letting ν be some fixed integer greater that or equal to 1, S_n be a pivotal statistic of the form given at (3.28) and the generic X be $(Y, Y^2, \ldots, Y^d)^t$, write the Edgeworth expansion as in

$$P(S_n \le x) = \Phi(x) + \sum_{\ell=1}^{\nu} n^{-\ell/2}\,\pi_{0,\ell}(x)\,\phi(x) + o(n^{-\nu/2}), \tag{3.38}$$

where, for $\ell = 1, \ldots, \nu$, the polynomials $\pi_{0,\ell}(\cdot)$ are of degree at most $3\ell - 1$, even for odd ℓ and odd for even ℓ, with coefficients depending on moments of X up to the $(\ell+2)$th order. Then, these conditions may be summarised in (i) $E(\|X\|^{\nu+2}) < \infty$ and (ii) the distribution of X is nonsingular. Equivalently, (i) requires the existence of moments up to the $(\nu+2)$th order moment and (ii) requires that X has a nondegenerate absolutely continuous component, i.e. X has a proper density function. As discussed in Hall (1987), the polynomial terms before the remainder $o(n^{-\nu/2})$ are depended only on the moments up to the $(\nu+2)$th order one, hence point (i). Regarding point (ii), Cramér's continuity condition, i.e. $\limsup_{\|t\|\to\infty} \|\lambda(t)\| < 1$, where $\lambda(\cdot)$ denotes the characteristic function, is stated instead, which holds if the distribution of X is nonsingular.

When a monotone transformation of the form given at (3.6) is applied to a pivotal statistic S_n of the form given at (3.28), stronger conditions need to hold so that the Edgeworth expansion corresponding to the *transformed* statistic is available uniformly in x on the whole real line. This is because the transformation, automatically, eliminates the terms of the ℓth order, thus, the associated moment effects up to the $(\ell+2)$th order moment. The resulting Edgeworth expansion remains valid only if moments up to the $((\ell+1)+2)$th order, i.e. $(\ell+3)$th moment, exist i.e. are finite. Theorem 3.4 gives a statement of this and is regarded to give '*minimal moment conditions*' when monotone and invertible transformations of the form at (3.6) are considered.

Theorem 3.4. *Let ν be some fixed integer greater that or equal to 1. Also, let the cumulative distribution function of a pivotal statistic S_n of the form given at (3.28) admit, uniformly in x on the whole real line, the Edgeworth expansion*

$$\begin{aligned} F(x) &\equiv P(S_n \le x) \\ &= \Phi(x) + \sum_{\ell=1}^{\nu} n^{-\ell/2}\,\pi_{0,\ell}(x)\,\phi(x) + o(n^{-\nu/2}) \end{aligned}$$

as $n \to \infty$, where $\pi_{0,\ell}(x)$ is, for every $\ell = 1, \ldots, \nu$, a polynomial on the real line of degree no more that $3\ell - 1$, even for odd ℓ and odd for even ℓ, and depended on the characteristics of the underlying sampling distribution but not on n. Also, the cumulative distribution function Φ and probability density function ϕ

correspond to the Standard Normal ones. Define the indefinite integral

$$\int \left\{ \frac{\mathrm{d}}{\mathrm{d}y} \pi_{(k-1),k}(y) \right\}^2 \mathrm{d}y = \Pi_{(k-1),k}(y) + const.,$$

where const. is some arbitrary constant. Then, the transformed statistic $T_{j\circ 0}(S_n) = (T_j \circ \ldots \circ T_0)(S_n)$ *with*

$$T_{k\circ 0}(S_n) = T_{(k-1)\circ 0}(S_n) + n^{-k/2}\, \pi_{(k-1),k}(T_{(k-1)\circ 0}(S_n)) + c\, n^{-(k+1)/2}\, \Pi_{(k-1),k}(T_{(k-1)\circ 0}(S_n)),$$

for some $k \in \{1, \ldots, j\}$, $T_0(S_n) = S_n$ *and* $c \geq \frac{1}{4}$, *obtains an Edgeworth's type expansion of the form*

$$\begin{aligned} G_j(x) &\equiv P(T_{j\circ 0}(S_n) \leq x) \\ &= \Phi(x) + o(n^{-(j+1)/2}), \end{aligned} \tag{3.39}$$

uniformly in x *over the whole real line, for every* $j = 1, \ldots, \nu$, *provided that the underlying sampling distribution (i) has finite absolute moments up to the* $(\nu+3)$*th order moment and (ii) is nonsingular.*

Remark 3.3. The following notes are in order.

1. The expansion at (3.39) is called Edgeworth's *type* expansion, since not all powers of $n^{-1/2}$ are present. The term Edgeworth expansion is used for an expression such as the one at (3.30).

2. The conditions may be restated using the generic $X = (Y, Y^2, \ldots, Y^d)^t$, from the description of the '*smooth function model*', as (i) $E(\|X\|^{\nu+3}) < \infty$ and (ii) the distribution of X is nonsingular.

Proof. Omitted. The techniques presented in Hall (1987) may be extended to provide a rigorous proof.

3.8 Estimation of unknown polynomials

In practice, the polynomials appearing in the Edgeworth expansion of the pivotal statistic S_n at (3.38) need to be estimated. Theorem 3.5, states that when consistent estimators are used the results of Theorem 3.4 still hold. When ν is taken to be equal to 1, the stated result, improves upon the result stated as Theorem 1 in Abramovitch and Singh (1985, p. 117) since correction for monotonicity is taken into account. From a practical point of view, though, i.e. when actual applications are considered, one should pay attention to the phenomenon of overcorrection which may occur in cases when the size of the data is not sufficiently large. If the transformation is applied too many times, then too many corrections, in the sense of the idea behind (3.4) and Theorem 3.2, are incorporated with the undesirable result of having a worse normal approximation than that initially intended without the correction. See Hall (1983, pp. 569–570), for a further elaboration. Niki and Konishi (1986) warn, too, against using many terms in an Edgeworth type expansion to safeguard against spurious oscillations at the tails of the distribution of a transformed statistic that may result in poor fit to the exact distribution. Thus, the possibility of overcorrection should not be overlooked when the sample size is not sufficiently large. A simple suggestion would be to use either $\nu = 1$ or $\nu = 2$, depending, of course, on the application. A further suggestion, which perhaps could shed a more precise light, would be to produce a plot of values for the νth approximation against the ν values or, even, tabulate them, see, e.g. Table 1 in Kakizawa (1996, p. 926).

Theorem 3.5. *Let* ν *be some fixed integer greater that or equal to* 1. *Also, let* S_n *be either a pivotal statistic of the form given at (3.28) or a transformed version of it according to the monotone and invertible transformations described by (3.8). Assume that the cumulative distribution function of a statistic* S_n *admits,*

uniformly in x on the whole real line, the Edgeworth expansion

$$\begin{aligned} F(x) &\equiv P(S_n \leq x) \\ &= \Phi(x) + n^{-\nu/2}\, q_{0,\nu}(x)\, \phi(x) + o(n^{-\nu/2}), \end{aligned} \tag{3.40}$$

where the degree $3\nu-1$ polynomial $q_{0,\nu}$ depends on the characteristics of the underlying sampling distribution but does not depend on n. Also, let $\widehat{q}_{0,\nu,n}$ be an estimator of $q_{0,\nu}$ that satisfies

$$P(|\widehat{q}_{0,\nu,n} - q_{0,\nu}(S_n)| > \varepsilon) = o(n^{-1/2}), \tag{3.41}$$

for every $\varepsilon > 0$ and given ν. Define the indefinite integral

$$\int \left\{ \frac{\mathrm{d}}{\mathrm{d}y} q_{0,\nu}(y) \right\}^2 \mathrm{d}y = Q_{0,\nu}(y) + \textit{const.},$$

where const. is some arbitrary constant. Let $\widehat{Q}_{0,\nu}(S_n)$ be the corresponding estimator of $Q_{0,\nu}(y)$. Then, the transformed statistic $\widetilde{T}_\nu(S_n)$ given as $\widetilde{T}_\nu(S_n) = S_n + n^{-\nu/2}\, \widehat{q}_{0,\nu}(S_n) + c\, n^{-(\nu+1)/2}\, \widehat{Q}_{0,\nu}(S_n)$, with $c \geq \frac{1}{4}$, obtains

$$\begin{aligned} \widetilde{G}_\nu(x) &\equiv P(\widetilde{T}_\nu(S_n) \leq x) \\ &= \Phi(x) + o(n^{-(\nu+1)/2}), \end{aligned} \tag{3.42}$$

uniformly in x on the whole real line.

Remark 3.4. Implicitly, it is assumed that all the necessary moments to write down the expansions at (3.41) and (3.42) exist.

Proof. Let $\overline{T}_\nu(S_n) = S_n + n^{-\nu/2}\, q_{0,\nu}(S_n) + c\, n^{-(\nu+1)/2}\, Q_{0,\nu}(S_n)$. Then, by Theorem 3.2, $\overline{T}_\nu(S_n)$ obtains $P(\overline{T}_\nu(S_n) \leq x) = \Phi(x) + o(n^{-\nu/2})$. By Theorem 3.5, the $(\nu+3)$th absolute moment of the underlying sampling distribution exists, i.e. is finite, and, thus, by Theorem 2 in Bhattacharya and Ghosh (1978, p. 436), $n^{(\nu+1)/2} \sup_x |P(\overline{T}_\nu(S_n) \leq x) - \Phi(x)| \to 0$. Since, condition at (3.41) holds by assumption and since $\widehat{Q}_{0,\nu}(S_n)$ is, also, a consistent estimator, the claim at (3.42) follows by Slutsky's Theorem. The proof is completed.

3.9 Accuracy of approximation and confidence distributions

The work done so far aimed at establishing a rigorous framework around the use of the monotone and invertible transformations given by (3.6). Theorem 3.6, establishes the accuracy order of a bootstrap approximation to the distribution of either a statistic or its transformed versions, cf. (3.6), leading one closer to the construction of an asymptotic confidence distribution. Essentially, the theorem states that when the bootstrap procedure is used to approximate the distribution of either a statistic or its transformed versions, cf. (3.6), then the order of error incurred is either $o_P(n^{-1})$ or $o_P(n^{-(j+\nu)/2})$, for suitably specified j and values. Theorem 3.7 establishes that the same order of incurred error as described by the previous theorem applies to the corresponding constructed confidence distributions. It is noticed that one application of the considered monotone and invertible transformation applied to S_n results in a bias free confidence distribution that is unaffected from the main effect of skewness.

Theorem 3.6. *Let S_n be the statistic given at (3.28) and its transformed version $T_{j\circ 0}(S_n)$ given by (3.6), for each $j = 1, \ldots, \nu$. Also, let ν is some fixed integer greater than or equal to 2 and denote by $\mathcal{F}$ the*

underlying sampling distribution. Assume that the S_n statistic admits, uniformly in x over the whole real line, the Edgeworth expansion

$$P(S_n \leq x) = \Phi(x) + \phi(x) \sum_{\ell=1}^{\nu} n^{-\ell/2}\, p_{0,\ell}(x, \mathcal{F}) + o(n^{-\nu/2}), \tag{3.43}$$

and that, for each $j = 1, \ldots, \nu$, its transformed version $T_{j\circ 0}(S_n)$ admits, uniformly in x over the whole real line, the Edgeworth's type expansion

$$P(T_{j\circ 0}(S_n) \leq x) = \Phi(x) + \phi(x) \sum_{\ell=j+1}^{\nu} n^{-\ell/2}\, p_{{}_{tr},\ell}(x, \mathcal{F}) + o(n^{-\nu/2}). \tag{3.44}$$

For, either, S_n or its transformed version, $T_{j\circ 0}(S_n)$, where $j = 1, \ldots, \nu$, denote the corresponding cumulative distribution function by $R_n(\cdot, \widehat{\psi}_n)$ and the corresponding bootstrapped cumulative distribution function by $R_n^*(\cdot, \widehat{\psi}_n)$. Then,

$$R_n^*(S_n, \widehat{\psi}_n) = R_n(S_n, \psi_n) + o_P(n^{-1}) \tag{3.45}$$

and, for each $j = 1, \ldots, \nu$,

$$R_n^*(T_{j\circ 0}(S_n), \widehat{\psi}_n) = R_n(T_{j\circ 0}(S_n), \psi_n) + o_P(n^{-(j+\nu)/2}). \tag{3.46}$$

Remark 3.5. For $j = 1$ and $\nu = 2$, it is noticed that $R_n^*(T_1(S_n), \widehat{\psi}_n) = R_n(T_1(S_n), \psi_n) + o_P(n^{-3/2})$, which is the same order of accuracy as achieved by the prepivoting method of Beran (1987, 1988). Prepivoting consists of transforming a pivotal statistic by its estimated bootstrapped cumulative distribution function. Note that, the transformations regarded here are of algebraic nature, whereas in Beran (1987, 1988), the transformation is of computational nature. Both of these methods, though, have the same order of accuracy and the choice between them is just a matter of taste and preference.

Proof. Let ν be some fixed integer greater than or equal to 1. Then, by the results presented in Beran (1987, pp. 467–468) and Beran (1988, pp. 690–692), it is known that the bootstrapped cumulative distribution function $R_n^*(\cdot, \widehat{\psi}_n)$ can be expanded uniformly in its first argument and locally uniform in its second argument. Thus, for (3.45),

$$\begin{aligned}
R_n^*(S_n, \widehat{\psi}_n) &= \Phi(S_n) + \phi(S_n) \sum_{\ell=1}^{\nu} n^{-\ell/2}\, p_{0,\ell}(S_n, \widehat{\mathcal{F}}_n) + o_P(n^{-\nu/2}) \\
&= \Phi(S_n) + \\
&\quad \phi(S_n) \sum_{\ell=1}^{\nu} n^{-\ell/2} \left[\, p_{0,\ell}(S_n, \widehat{\mathcal{F}}_n) - p_{0,\ell}(S_n, \mathcal{F}) + p_{0,\ell}(S_n, \mathcal{F}) \,\right] + \\
&\quad o_P(n^{-\nu/2}) \\
&= \Phi(S_n) + \phi(S_n) \sum_{\ell=1}^{\nu} n^{-\ell/2}\, p_{0,\ell}(S_n, \mathcal{F}) + \\
&\quad \phi(S_n) \sum_{\ell=1}^{\nu} n^{-\ell/2} \left[\, p_{0,\ell}(S_n, \widehat{\mathcal{F}}_n) - p_{0,\ell}(S_n, \mathcal{F}) \,\right] + o_P(n^{-\nu/2}) \\
&= R_n(S_n, \psi) + o_P(n^{-1}),
\end{aligned}$$

since $p_{0,\ell}(S_n, \widehat{\mathcal{F}}_n) - p_{0,\ell}(S_n, \mathcal{F}) = o_P(n^{-1/2})$ and by the properties of the o_P-notation. The same arguments

are repeated for (3.46). Hence,

$$\begin{aligned}
R_n^*(T_{j\circ 0}(S_n), \widehat{\psi}_n) &= \Phi(T_{j\circ 0}(S_n)) + \phi(T_{j\circ 0}(S_n)) \sum_{\ell=j+1}^{\nu} n^{-\ell/2}\, p_{\mathrm{tr},\ell}(T_{j\circ 0}(S_n), \widehat{\mathcal{F}}_n) + o_P(n^{-\nu/2}) \\
&= \Phi(T_{j\circ 0}(S_n)) + \\
&\quad \phi(T_{j\circ 0}(S_n)) \sum_{\ell=j+1}^{\nu} n^{-\ell/2}\, [\, p_{\mathrm{tr},\ell}(T_{j\circ 0}(S_n), \widehat{\mathcal{F}}_n) - p_\ell(T_{j\circ 0}(S_n), \mathcal{F}) + p_{\mathrm{tr},\ell}(T_{j\circ 0}(S_n), \mathcal{F})\,] + \\
&\quad o_P(n^{-\nu/2}) \\
&= \Phi(T_{j\circ 0}(S_n)) + \phi(T_{j\circ 0}(S_n)) \sum_{\ell=j+1}^{\nu} n^{-\ell/2}\, p_{\mathrm{tr},\ell}(T_{j\circ 0}(S_n), \mathcal{F}) + \\
&\quad \phi(T_{j\circ 0}(S_n)) \sum_{\ell=j+1}^{\nu} n^{-\ell/2}\, [\, p_{\mathrm{tr},\ell}(T_{j\circ 0}(S_n), \widehat{\mathcal{F}}_n) - p_\ell(T_{j\circ 0}(S_n), \mathcal{F})\,] + o_P(n^{-\nu/2}) \\
&= R_n(T_{j\circ 0}(S_n), \psi) + o_P(n^{-\max\{j+2,\nu\}/2}),
\end{aligned}$$

since $p_{\mathrm{tr},\ell}(T_{j\circ 0}(S_n), \widehat{\mathcal{F}}_n) - p_{\mathrm{tr},\ell}(T_{j\circ 0}(S_n), \mathcal{F}) = o_P(n^{-1/2})$ and by the properties of the o_P-notation. The proof is completed.

The results regarding more accurate confidence distributions as produced by the use of the suggested transformations, cf. (3.6), are stated as Theorem 3.7 and Corollary 3.2.

Theorem 3.7. *Let ν be some fixed integer greater than or equal to 1. Suppose that $S_n = S_n(\psi)$ is a pivotal statistic which admits, uniformly in x over the whole real line, the asymptotic expansion given at (3.43) and that its transformed version $T_{j\circ 0}(S_n)$, given by (3.6) for each $j = 1, \ldots, \nu$, admits, uniformly in x over the whole real line, the asymptotic expansion given at (3.44). For, either, S_n or its transformed version, $T_{j\circ 0}(S_n)$, where $j = 1, \ldots, \nu$, denote the corresponding cumulative distribution function by $R_n(\cdot, \widehat{\psi}_n)$ and the corresponding bootstrapped cumulative distribution function by $R_n^*(\cdot, \widehat{\psi}_n)$. Define the confidence distributions $C(\psi) = R_n(S_n, \psi_n)$ and $C_{tr,j}(\psi) = R_n(T_{j\circ 0}(S_n), \psi_n)$, for each $j = 1, \ldots, \nu$. Also, define the asymptotic confidence distributions $C_{boot}(\psi) = R_n^*(S_n, \psi_n)$ and $C_{boot,tr,j}(\psi) = R_n^*(T_{j\circ 0}(S_n), \psi_n)$, for each $j = 1, \ldots, \nu$. Then,*

$$C_{boot}(\psi) = C(\psi) + o_P(n^{-1})$$

and

$$C_{boot,tr,j}(\psi) = C_{tr,j}(\psi) + o_P(n^{-\max\{j+2,\nu\}/2}).$$

Remark 3.6. For notational convenience, the dependence on y has been dropped from, both, the statistics and the defined confidence distributions.

Proof. By substitution and Theorem (3.6), the proof is completed.

Corollary 3.2. *For $j = 1$ and $\nu = 2$,*

$$C_{boot,tr,1}(\psi) = C_{tr,1}(\psi) + o_P(n^{-3/2}).$$

Remark 3.7. The derived aCD has, thus, been corrected for the main effect of skewness.

Proof. Immediate.

4 Limitations and Extensions

In this section a brief discussion on the limitations and extensions regarding this project are discussed. Of course, some of the topics discussed here could easily be regarded as future research, see section 5. Future research is regarded in its broader sense, not necessarily limited or confined to methods related to the ones discussed in this project.

In section 2, a review of three methods was given and a few examples explained and demonstrated various aspect of the discussed methods in some detail. The list could have been added the abc-bootstrap method, see Schweder and Hjort (2016, p. 217), the saddlepoint approximations and the magic formula, see Schweder and Hjort (2016, p. 218) and the median-bias correction, see Schweder and Hjort (2016, p. 214) and DeBlasi and Schweder (2016), along with further examples to demonstrate the underlying techniques and machinery.

In section 3, the discussion evolved around the i.i.d. case without mentioning any possible extensions to the regression context. From the perspective of applications, perhaps, it would have been more interesting. From a theoretical perspective, the same effort needs to be undertaken to show how the suggested transformations fit in the regression framework. Section 3 lacks of examples that show the machinery in action. The concentration was, mainly, to establish the theoretical results presented. Some simple examples to demonstrate the techniques have been prepared but due to time constraints they were not included. The two-sample problem should and must be in the list of examples to apply the suggested methods. As a further step, comparisons need to be made with currently available methods. In particular, in the class of exponential models, where it is known that an optimality theorem exists, cf. Theorem 5.11 in Schweder and Hjort (2016, p. 173), it is desirable to see how well the currently presented methods and the optimality theorem compete with each other. Nuisance parameter have not been included in the discussion presented in this project, but the paper by Zheng et al. (2017) may provide some guidelines for further work.

5 Further Research

In this section an attempt is made to suggest possible routes of further exploration regarding the world of confidence distributions (CDs) and confidence curves (ccs). Theoretical results and interesting applications can be suggested once things are turned around. Before continuing, though, it is worth mentioning some notable sources as a starting point or an introduction to the concepts and ideas behind CDs and ccs. These include, the book by Schweder and Hjort (2016), which supplements concepts and ideas with a plethora of examples and the papers by Singh et al. (2005), Singh et al. (2007), Xie and Singh (2013), DeBlasi and Schweder (2016), Schweder (2017) and Hjort and Schweder (2017). A further note would be that CDs and ccs should be regarded as another tool in the statisticians' bag and, as such, should be utilised to complement the inference task and provide insightful solutions when applied. Possible paths or directions for future research may include the following:

1. CDs and ccs under model misspecification:
   Viraswami and Reid (1996, 1998) develop higher-order asymptotic results under model misspecification. In Viraswami and Reid (1996), robust versions of the score statistic and Wald statistic are treated, whereas in Viraswami and Reid (1998), the likelihood-ratio statistic and the adjusted likelihood-ratio statistic are treated. From the perspective of CDs and ccs, perhaps, more interesting is their work regarding the likelihood-ratio statistic. They notice that under model misspecification, Bartlett correction cannot be utilised to improve the $\mathcal{X}^2_1$ approximation to the distribution of the statistic. They solve the problem, though, by utilising the polynomial approach of Cordeiro and Ferrari (1991) and they obtain an improved version of the uncorrected likelihood ratio having a $\mathcal{X}^2_1$ to third order. They work similarly for the adjusted likelihood-ratio statistic. In the literature, an extension to the polynomial approach of Cordeiro and Ferrari (1991) is given by Kakizawa (1997). A suggestion, therefore, would be to implement this extension and, then, form the necessary theoretical background to construct an asymptotic cc.

2. Optimal confidence for exponential families:
   In Schweder and Hjort (2016, sec. 5.5) a theorem that constructs uniformly most powerful CDs in the exponential class of models is presented. The examples given in sections 5.5 and 8.2 of the book suggest that the computation of the conditional probability needed for the CD is challenging in the not so easy class of models. Saddlepoint approximations have been suggested in the literature to approximate conditional distributions, see e.g., Kolassa (1996), Skovgaard (1987), Butler (2007) and Butler et al. (2008), for theoretical and practical considerations. The paper by Theodosopoulos (2007) is concerned with general random variables and constructs a lower bound for the tails of their distribution using, only, knowledge of their moment generating function. The procedure shares numerous methodological similarities with the development of saddlepoint approximation. For example, it uses titling to centre the power series expansion at the desired tail of the distribution. The difference, though, is that the titling procedure is seen as a nonlinear optimisation problem and the added degrees of freedom enables it to produce tighter and more broadly applicable lower bounds than existing tail approximations. It is suggested that the methodology described in Theodosopoulos (2007) be adopted in the optimal CDs framework to provide a method, perhaps in the form of an algorithm, to approximate the conditional distribution in the non so easy cases.

3. An alternative to the bootstrap:
   Efron's bootstrap has been one of the most valuable tools in a practitioner's tool of bag and is so famous that needs not any introductions. Recently, a paper was publishes that in the small sample cases offers an alternative to the bootstrap. In particular, the methodology developed in Jayadeva and Soumy

(In Press) is concentrated in augmenting small data sets so that the original underlying distribution is least distorted. It is a non-iterative technique that uses principal components and least-squares to add extra samples in the eigenspace of the original small size data set. It is suggested that in the CDs framework, their technique be used to over improved approximations as bootstrap is used and the two methods be compared.

4. Empirical likelihood:
   In Schweder and Hjort (2016, sec. 11.5) a brief discussion is given on the empirical likelihood and some theoretical results in the i.i.d. and regression cases are delivered. The paper by Jing et al. (2017) offers a simple transformation of the empirical likelihood with the aim to alleviate the under-coverage of confidence regions. In the spirit of producing more accurate CDs and ccs it is suggested that the methodology in Jing et al. (2017) in be adopted. Their transformation is capable of providing '*substantially more accurate condence regions without adding theoretical or computational complexity*', as they mention. The necessary criteria concerning accuracy and consistency for good transformations are laid down and a transformed empirical likelihood meeting these criteria is arrived at. Their effort produces an attractive and accurate empirical likelihood that is easy to use and '*surprisingly accurate even in small sample and in multidimensional situations*', as the mention. It is suggested that their criteria be adopted in the CDs and ccs framework to produced more accurate CDs and ccs.

Appendix

I R codes

I.1 Example 2.1

```
y.obs <-
c(0.013, 0.054, 0.234, 0.286, 0.332, 0.507, 0.703, 0.763, 0.772, 0.920)
y.obs

n.y.obs <- length(y.obs) ; n.y.obs

theta.hat <- - n.y.obs / sum(log(y.obs)) ; theta.hat

theta.vals <- seq(0.2, 1.6, by=0.01) ; theta.vals ; length(theta.vals)

conf.lev <- 0.90 ; conf.lev

n.sims <- 5*10^3 ; n.sims

C.sims <- 0*(1:n.sims) ; C.sims
C.sim <- 0*theta.vals ; C.sim
for (j in 1:length(theta.vals)) {
theta <- theta.vals[j]
for (ss in 1:n.sims) {
ysim <- runif(n.y.obs)^(1/theta)
C.sims[ss] <- 1/mean(-log(ysim)) }
C.sim[j] <- mean(1*(C.sims>=theta.hat)) }

q.N <- ( theta.hat - theta.vals ) / ( theta.hat / sqrt(n.y.obs) ) ; q.N

deviance.f <-
-2 * ( ( n.y.obs * log(theta.vals) + theta.vals * sum(log(y.obs)) ) -
( n.y.obs * log(theta.hat) + theta.hat * sum(log(y.obs)) ) ) ; deviance.f

plot(theta.vals, pnorm(q.N, 0, 1, lower.tail=F), type="l",
bty="l", main="", xlab=expression(theta), ylab="confidence", col=2, lwd=1.8)
lines(theta.vals, pnorm(sqrt(deviance.f)*sign(theta.vals-theta.hat), 0, 1),
col=4, lwd=1.8)
lines(theta.vals, pgamma(n.y.obs / theta.hat, n.y.obs, theta.vals),
col=3, lwd=1.8)
lines(theta.vals, C.sim, lty=2, col=1, lwd=1.8)

#cc.theta <- pchisq(deviance.f,1) ; cc.theta
#med.theta <- theta.vals[which(cc.theta==min(cc.theta))] ; med.theta
#C.theta <- 0.5 * c(1-cc.theta[theta.vals<med.theta],
#1+cc.theta[theta.vals>=med.theta])
```

```
#lines(theta.vals, C.theta, col=6)

segments(min(theta.vals), (1-conf.lev)/2, max(theta.vals),
(1-conf.lev)/2, lty=2, col='grey')
segments(min(theta.vals), 1/2, max(theta.vals), 1/2, lty=2, col='grey')
segments(min(theta.vals), (1+conf.lev)/2, max(theta.vals),
(1+conf.lev)/2, lty=2, col='grey')
segments(min(theta.vals), c(0,1), max(theta.vals), c(0,1),
lty=3, col='grey')
legend(1.2, 0.3,
legend=c("Normal approx", "chi-sq approx (dev)", "exact prob", "simulation"),
lty=c(1,1,1,2), col=c(2,4,3,1), bty="n", lwd=1.8)
```

I.2 Example 2.2

```
# read data
yy.dat <- scan("speed_of_light.txt", skip=1) ; yy.dat

# get number of datapoints
nn <- length(yy.dat) ; nn

# require stats package
# for logistic density
library(stats)

# define log-likelihood function
loglik.f <- function(prms) {
sum(dlogis(yy.dat, location=prms[1], scale=prms[2], log=T))
}

# define minus log-likelihood
# the function to be minimized
mloglik.f <- function(prms) {
-loglik.f(prms)
}

# use optim to perform minimization
# starting values by trial and error
optim.results <- optim(c(3,2), mloglik.f, hessian=T)

# view results
optim.results

# check convergence
optim.results$conv == 0   # TRUE #

# get maximum likelihood estimates
```

```
xi.hat <- optim.results$par[1] ; xi.hat     # 27.61808  #
tau.hat <- optim.results$par[2] ; tau.hat   #  2.838543 #

# get Hessian matrix
J.tot <- optim.results$hessian ; J.tot

# get standard deviation estimates
sd.xi <- sqrt(diag(solve(J.tot)))[1] ; sd.xi     # 0.6157995 #
sd.tau <- sqrt(diag(solve(J.tot)))[2] ; sd.tau   # 0.2971452 #

##############################################################################

# estimate p
y0 <- 30.5 ; y0
p.hat <- plogis(y0, location=xi.hat, scale=tau.hat) ; p.hat   # 0.7340526 #

# check p estimate
qq.pnt <- (y0-xi.hat)/tau.hat ; qq.pnt
p.hat.chk <- exp(qq.pnt) / (1+exp(qq.pnt)) ; p.hat.chk
p.hat.chk == p.hat

# estimate vector of derivatives, ww
w1.hat <- -dlogis(y0, location=xi.hat, scale=tau.hat) ; w1.hat
w2.hat <- -(y0-xi.hat)/tau.hat *
dlogis(y0, location=xi.hat, scale=tau.hat) ; w2.hat
ww.hat <- c(w1.hat, w2.hat) ; ww.hat

# check ww components estimates
chk1 <- -exp(qq.pnt) / ( tau.hat * (1+exp(qq.pnt))^2 ) ; chk1
chk1 == w1.hat
chk2 <- qq.pnt * w1 ; chk2
chk2 == w2.hat

# define k hat
k.hat <- t(ww.hat) %*% J.tot %*% ww.hat ; k.hat

# delta method #

# define a grid of p values
p.seq <- seq(0, 1, by=0.0001) ; p.seq

# get confidence distribution
qq.C <- (p.seq - p.hat) / k.hat ; qq.C
C.p <- pnorm(qq.C) ; C.p

# get confidence curve
```

```
cc.p <- abs(1-2*C.p) ; cc.p

# plot confidence curve (using delta method)
plot(p.seq, cc.p, bty="l", type="l", lty=1,
xlab=expression(italic(p)), ylab="confidence curve (using delta method)")
axis(1, at=round(p.hat,3), label=T, las=1)
matlines(p.seq, 0+0*p.seq, lty=3, col="grey")
#matlines(p.seq, 1+0*p.seq, lty=3, col="grey")

# deviance #

# define a grid of p values
# take care as p/(1-p) \in (0,1)
p.val <- seq(0.0001, 0.9999, by=0.0001) ; p.val

# compute profile log-likelihood
loglik.prof.val <- 0*p.val ; loglik.prof.val
for (pp in 1:length(p.val)) {
mloglik.f.cc <- function(prm2) {
prm1 <- y0 - prm2 * log( p.val[pp]/(1-p.val[pp]) )
-sum(dlogis(yy.dat, location=prm1, scale=prm2, log=T))
}
loglik.prof.val[pp] <- -optimize(mloglik.f.cc, c(0.000001,10^10))$objective
}

# check
loglik.f(c(xi.hat, tau.hat))
max(loglik.prof.val)
p.val[which.max(loglik.prof.val)]

# compute deviance
dev.f <- 2*(max(loglik.prof.val) - loglik.prof.val) ; dev.f

# check
p.val[which(dev.f == min(dev.f))]
# or #
p.val[which.max(-dev.f)]

# get confidence curve
cc.d.p <- pchisq(dev.f, 1) ; cc.d.p

# check
p.val[which(cc.d.p == 0)]
# or #
p.val[which.max(-cc.d.p)]
```

```
# plot confidence curve (based on deviance)
plot(p.val, cc.d.p, bty="l", type="l", lty=1, xaxt="n",
xlab=expression(italic(p)), ylab="confidence curve (based on deviance)")
axis(1, at=seq(0.2,0.8,by=0.2), label=T)
axis(1, at=0.0001, label=T)
axis(1, at=0.9999, label=T)
axis(1, at=p.val[which.max(-cc.d.p)], label=T, las=1)
matlines(p.val, 0+0*p.val, lty=3, col="grey")
#matlines(p.val, 1+0*p.val, lty=3, col="grey")
#abline(v=p.val[which.max(-cc.d.p)], lty=3, col="tomato")

# Bartlett correction for the deviance #
# via simulation #

# define deviance function
dev.val <- function(yydat, yy0) {
m.loglik <- function(prm.2) {
prm.1 <- yy0 - prm.2 * log( p.hat/(1-p.hat) )
-sum(dlogis(yydat, location=prm.1, scale=prm.2, log=T))
}
tau.aux <- optimize(m.loglik, c(0.000001,10^10))$min
xi.aux <- yy0 - tau.aux * log( p.hat/(1-p.hat) )
2*(loglik.f(c(xi.hat, tau.hat)) - loglik.f(c(xi.aux, tau.aux)))
} # end dev.val

# check
loglik.f(c(xi.hat, tau.hat))
dev.val(yy.dat, y0)

# find mean of deviance statistic
n.sim <- 10^4 ; n.sim
Dn.p <- 0*(1:n.sim)
for (ss in 1:n.sim) {
yy.sim <- rlogis(nn, location=xi.hat, scale=tau.hat)
Dn.p[ss] <- dev.val(yy.sim, y0)
}
mean.dev.stat <- mean(Dn.p) ; mean.dev.stat   # 1.032724 #

# get modified confidence curve
mod.dev <- dev.f/mean.dev.stat ; mod.dev
cc.d.p.m <- pchisq(mod.dev, 1) ; cc.d.p.m

# check
p.val[which(cc.d.p.m == 0)]
p.val[which(cc.d.p == 0)]
```

```
# plot modified confidence curve
plot(p.val, cc.d.p.m, bty="l", type="l", lty=1, xaxt="n",
xlab=expression(italic(p)),
ylab="modified confidence curve (using Bartlett correction)")
axis(1, at=seq(0.2,0.8,by=0.2), label=T)
axis(1, at=0.0001, label=T)
axis(1, at=0.9999, label=T)
axis(1, at=p.val[which.max(-cc.d.p.m)], label=T, las=1)
matlines(p.val, 0+0*p.val, lty=3, col="grey")
#matlines(p.val, 1+0*p.val, lty=3, col="grey")
#abline(v=p.val[which.max(-cc.d.p.m)], lty=3, col="tomato")

# plot confidence curves together
# cc(p) and modified cc(p)
matplot(p.val, cbind(cc.d.p, cc.d.p.m), bty="l",
type="l", lty=1:2, col=1:2, xaxt="n",
xlab=expression(italic(p)), ylab="confidence curves")
axis(1, at=seq(0.2,0.8,by=0.2), label=T)
axis(1, at=0.0001, label=T)
axis(1, at=0.9999, label=T)
axis(1, at=p.val[which.max(-cc.d.p.m)], label=T, las=1)
matlines(p.val, 0+0*p.val, lty=3, col="grey")
legend(0.01, 0.18,
legend=c(expression(italic(cc(p))),
expression(paste(italic(cc(p))," modified"))),
lty=1:2, col=1:2, bty="n")
```

I.3 Example 2.3

```
# read data
yy.d <- scan("speed_of_light.txt", skip=1) ; yy.d

# get number of datapoints
nn <- length(yy.d) ; nn

# require glogis package
# for generalized logistic density
# Type I: Skew-Logitic
library(glogis)

# define log-likelihood function
loglik.f <- function(prms) {
sum(dglogis(yy.d, location=prms[1], scale=prms[2], shape=prms[3], log=T))
}

# define minus log-likelihood
# the function to be minimized
```

```
mloglik.f <- function(prms) {
-loglik.f(prms)
}

# use optim to perform minimization
# starting values by trial and error
optim.results <- optim(c(2,2.5,3), mloglik.f, hessian=T)

# view results
optim.results

# check convergence
optim.results$conv == 0   # TRUE #

# get maximum likelihood estimates
xi.hat <- optim.results$par[1] ; xi.hat     # 26.04327  #
tau.hat <- optim.results$par[2] ; tau.hat   #  3.153894 #
gam.hat <- optim.results$par[3] ; gam.hat   #  1.426836 #

# compare with
glogisfit(yy.d)

# continue with
xihat <- glogisfit(yy.d)$param[1] ; xihat     # 26.05397  #
tauhat <- glogisfit(yy.d)$param[2] ; tauhat   #  3.151787 #
gamhat <- glogisfit(yy.d)$param[3] ; gamhat   #  1.423737 #
# as output from glogisfit
# though
# difference is small
# as seen in
c(xi.hat, tau.hat, gam.hat) - c(xihat, tauhat, gamhat)

# define profile log-likelihood for gamma
prf.ll.g.aux <- function(gam.v) {
aux.f <- function(prms) {
-sum(dglogis(yy.d, location=prms[1], scale=prms[2], shape=gam.v, log=T))
}
-optim(c(xihat,tauhat), aux.f)$val
} # end prf.ll.g

# define a grid of gamma values
gam.val <- seq(0.0001, 7.001, by=0.001) ; gam.val

# compute profile log-likelihood at gam.val
prf.ll.g <- 0*gam.val ; prf.ll.g
for (gg in 1:length(gam.val)) {
```

```
prf.ll.g[gg] <- prf.ll.g.aux(gam.val[gg])
}

# compute deviance
prf.Dn.g <- 2*(max(prf.ll.g) - prf.ll.g) ; prf.Dn.g

# check
max(prf.ll.g)
loglik.f(c(xihat, tauhat, gamhat))

# check
gam.val[which.max(-prf.Dn.g)]
gamhat

# get confidence curve
cc.d.g <- pchisq(prf.Dn.g,1) ; cc.d.g

# compute level, g1.lev
# for which confidence intervals above g1.lev
# include the value of \gamma = 1
aux.lev <- 2*(max(prf.ll.g) - prf.ll.g.aux(1)) ; aux.lev
g1.lev <- pchisq(aux.lev, 1) ; g1.lev
g1.lev*100
cc.aux.lev <- gam.val[cc.d.g<=g1.lev] ; cc.aux.lev

# find confidence intervals
ci.l <- 0.95 ; ci.l
ci.aux <- gam.val[cc.d.g <= ci.l] ; ci.aux
ci.lv <- c(min(ci.aux), max(ci.aux)) ; ci.lv

# plot confidence curve
plot(gam.val, cc.d.g, bty="l", type="l", lty=1, xaxt="n",
xlab=expression(italic(gamma)), ylab="confidence curve")
axis(1, at=seq(1,7,by=1), label=T)
axis(1, at=min(gam.val), label=T)
axis(1, at=gam.val[which.max(-cc.d.g)], label=T, las=1)
axis(1, at=1, label=T)
axis(2, at=round(g1.lev,4), label=T, cex.axis=0.7, las=2)
matlines(gam.val, 0+0*gam.val, lty=3, col="grey")
matlines(gam.val, 1+0*gam.val, lty=3, col="grey")
matlines(cc.aux.lev, g1.lev+0*cc.aux.lev, lty=2, col="darkgrey")
#matlines(ci.aux, ci.l+0*ci.aux, lty=2, col="darkgrey")
matlines(1+0*seq(0,1,by=0.001), seq(0,1,by=0.001), lty=2, col="tomato")
```

I.4 Examples 2.4 and 2.5

```
# read data
```

```
sb.d <- scan("smallbabies_data.txt", skip=3) ; sb.d

# arrange in matrix form
dat.mat <- matrix(sb.d, ncol=11, byrow=T) ; dat.mat

# get hold of mothers' weight
wgt <- dat.mat[,4] ; wgt

# convert mothers' weight from pounds to kilograms
# 1 pound = 0.45359237 kg
wgt.kg <- wgt / 2.204623 ; wgt.kg

# define function to compute cc
# normal approximation of (a)

cc.norm.appr <- function(yy.dat, p.qntl, psi.val) {

# get number of observations
n.yy <- length(yy.dat)

# compute p-quantile of N(0,1)
z.p <- qnorm(p.qntl)

# estimate mean and sd for yy.dat
mu.hat <- mean(yy.dat)
sd.hat <- sd(yy.dat)

# estimate p-quantile
psi.hat <- mu.hat + z.p * sd.hat

# define pivot
sd.piv <- sqrt((1 + 0.5 * z.p^2)) * sd.hat
t.piv <- sqrt(n.yy) * (psi.hat - psi.val) / sd.piv

# compute CD
CC.norm.appr <- pnorm(t.piv)

# compute cc
abs(1-2*CC.norm.appr)

} # end cc.norm.appr

# define auxiliary function to compute cc
# eq.(11.3) of CLP book, p.321

levels.f <- function(n.obs, p.qntl, aa, bb) {
```

```
# set number of observations
nn <- n.obs

# define integrand
integrand <- function(u) {
( 1 - pbeta((p.qntl-u)/(1-u), bb-aa, nn-bb+1) ) * dbeta(u, aa, nn-aa+1)
}

# integrate
integrate(integrand, lower=0, upper=p.qntl)$value

} # end levels.f

# define function to compute cc
# nonparametrically
# eq.(11.3) of CLP book, p.321

cc.nprm <- function(yy.dat, p.qntl) {

# order observations
yy.sort <- sort(yy.dat)

# get number of observations
nn <- length(yy.dat)

# define [nn*p.qntl]
upp <- min(ceiling(nn*p.qntl)-1, nn-ceiling(nn*p.qntl))

# keep cc values
cc.mat <- 0*(1:(2*upp+1))%*%t(1:2)

# compute cc
for (jj in 0:upp) {
aa <- ceiling(nn*p.qntl) - jj
bb <- ceiling(nn*p.qntl) + jj
cc.mat[c(upp+1-jj, upp+1+jj), 1] <- yy.sort[c(aa,bb)]
cc.mat[c(upp+1-jj, upp+1+jj), 2] <- levels.f(nn, p.qntl, aa, bb)
}

# output cc values and upp
list(cc.mat=cc.mat, upp=upp)

} # end cc.nprm

# define a grid of \psi values
```

```
psi.vals <- seq(min(wgt.kg), max(wgt.kg), by=0.0001)

# compute cc based on
# normal approximation of (a)

# apply cc.norm.appr function
cc.1 <- cc.norm.appr(wgt.kg, 0.90, psi.vals)

# check
mean(wgt.kg) + qnorm(0.90) * sd(wgt.kg)
psi.vals[which.max(-cc.1)]

# compare with 90% sample quantile
quantile(wgt.kg, 0.90)
quantile(wgt.kg, 0.90) - psi.vals[which.max(-cc.1)]

# and plot cc based on
# normal approximation of (a)
plot(psi.vals, cc.1, bty="l", type="l", lty=1, xaxt="n",
ylab="confidence curve ( normal approximation of (a) )",
xlab="mothers' weight")
axis(1, at=round(seq(min(wgt.kg), max(wgt.kg), len=6),3), labels=T)
axis(1, at=round(psi.vals[which.max(-cc.1)],3), labels=T)
matlines(psi.vals, 0+0*psi.vals, lty=2, col="grey")
aux1 <- seq(-1, 1, length=10)
matlines(quantile(wgt.kg, 0.90)+0*aux1, 0.02*aux1, lty=1, col=2)

# compute cc
# nonparametrically

# apply cc.nprm function
cc.3 <- cc.nprm(wgt.kg, 0.90)$cc.mat

# check
min(cc.3[,2])

# check
quantile(wgt.kg, 0.90)
cc.3[,1][which.min(cc.3[,2])]

# and plot cc
# nonparametrically
matplot(cc.3[,1], cc.3[,2], bty="l", type="l", lty=1,
ylab="confidence curve ( nonparametrically )", xlab="mothers' weight")
axis(1, at=round(cc.3[,1][which.min(cc.3[,2])],3), labels=T)
matlines(cc.3[,1], 0+0*cc.3[,1], lty=3, col="grey")
```

```
# display confidence intervals
cc.3.upp <- cc.nprm(wgt.kg, 0.90)$upp
for (ll in 0:cc.3.upp) {
matlines(cc.3[(cc.3.upp+1-ll):(cc.3.upp+1+ll),1],
0*cc.3[(cc.3.upp+1-ll):(cc.3.upp+1+ll),1] + cc.3[(cc.3.upp+1-ll),2],
lty=2, col="grey")
}

# require MASS package
# for truehist
library(MASS)

# explore data
truehist(wgt.kg, bty="l", xlab="mothers' weight", col="lightgrey")

# display ccs for 0.10, 0.50, 0.90 quantiles
# in the same diagram

cc.010 <- cc.nprm(wgt.kg, 0.10)$cc.mat
cc.050 <- cc.nprm(wgt.kg, 0.50)$cc.mat
cc.090 <- cc.3

# check
quantile(wgt.kg, 0.10)                # 45.17779 #
cc.010[,1][which.min(cc.010[,2])]     # 44.45204 #
quantile(wgt.kg, 0.50)                # 54.88467 #
cc.050[,1][which.min(cc.050[,2])]     # 54.88467 #
quantile(wgt.kg, 0.90)                # 77.11069 #
cc.090[,1][which.min(cc.090[,2])]     # 77.11069 #

cc010.upp <- cc.nprm(wgt.kg, 0.10)$upp
cc050.upp <- cc.nprm(wgt.kg, 0.50)$upp
cc090.upp <- cc.3.upp

plt.xlim <- c(min(cc.010[,1], cc.050[,1], cc.090[,1]),
max(cc.010[,1], cc.050[,1], cc.090[,1])) ; plt.xlim

plot(sort(wgt.kg), xlim=plt.xlim, ylim=c(0,1), bty="l", type="n",
xaxt="n", ylab="cc for 0.10, 0.50, 0.90", xlab="mothers' weight")
axis(1, at=round(seq(min(wgt.kg), max(wgt.kg), len=6),3), label=T)
axis(1, at=round(cc.010[,1][which.min(cc.010[,2])],3), labels=F)
axis(1, at=round(cc.050[,1][which.min(cc.050[,2])],3), labels=F)
axis(1, at=round(cc.090[,1][which.min(cc.090[,2])],3), labels=F)
matlines(sort(wgt.kg), 0+0*sort(wgt.kg), lty=3, col="grey")
```

```
matlines(cc.010[,1], cc.010[,2], type="l", lty=1)
for (ll in 0:cc010.upp) {
matlines(cc.010[(cc010.upp+1-ll):(cc010.upp+1+ll),1],
0*cc.010[(cc010.upp+1-ll):(cc010.upp+1+ll),1] + cc.010[(cc010.upp+1-ll),2],
lty=2, col="grey")
}

matlines(cc.050[,1], cc.050[,2], type="l", lty=1)
for (ll in 0:cc050.upp) {
matlines(cc.050[(cc050.upp+1-ll):(cc050.upp+1+ll),1],
0*cc.050[(cc050.upp+1-ll):(cc050.upp+1+ll),1] + cc.050[(cc050.upp+1-ll),2],
lty=2, col="grey")
}

matlines(cc.090[,1], cc.090[,2], type="l", lty=1)
for (ll in 0:cc090.upp) {
matlines(cc.090[(cc090.upp+1-ll):(cc090.upp+1+ll),1],
0*cc.090[(cc090.upp+1-ll):(cc090.upp+1+ll),1] + cc.090[(cc090.upp+1-ll),2],
lty=2, col="grey")
}
```

II Data sets

II.1 Example 2.1

```
0.013, 0.054, 0.234, 0.286, 0.332, 0.507, 0.703, 0.763, 0.772, 0.920
```

II.2 Example 2.2

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

II.3 Examples 2.4 and 2.5

```
low birthweights data from
https://feb.kuleuven.be/public/u0043181/modelselection/index.html
```

References

Abramovitch, L. and Singh, K. (1985). Edgeworth corrected pivotal statistics and the bootstrap, *The Annals of Statistics*, **13 (1)**: 116–132.

Arfken, G., Weber, H. and Harris, F. E. (2013). *Mathematical Methods for Physicists: A Comprehensive Guide*, Oxford: Academic Press, 7th edn.

Barndorff-Nielsen, O. E. and Cox, D. R. (1989). *Asymptotic Techniques for Use in Statistics*, no. 31 in Monographs on Statistics and Applied Probability, London: Chapman and Hall.

Barndorff-Nielsen, O. E. and Cox, D. R. (1994). *Inference and Asymptotics*, no. 52 in Monographs on Statistics and Applied Probability, New Delhi: Chapman and Hall/CRC.

Bartlett, M. S. (1937). Properties of sufficiency and statistical tests, *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, **160 (901)**: 268–282.

Bartlett, M. S. (1954). A note on the multiplying factors for various χ^2 approximations, *Journal of the Royal Statistical Society. Series B (Methodological)*, **16 (2)**: 296–298.

Beran, R. (1987). Prepivoting to reduce level error of confidence sets, *Biometrika*, **74 (3)**: 457–468.

Beran, R. (1988). Prepivoting test statistics: A bootstrap view of asymptotic refinements, *Journal of the American Statistical Association*, **83 (403)**: 687–697.

Bhattacharya, R. N. and Ghosh, J. K. (1978). On the validity of the formal Edgeworth expansion, *The Annals of Statistics*, **6 (2)**: 434–451.

Bhattacharya, R. N. and Ghosh, J. K. (1980). Correction to 'On the validity of the formal Edgeworth expansion', *The Annals of Statistics*, **8 (6)**: 1399.

Bhattacharya, R. N. and Rao, R. R. (2010). *Normal Approximation and Asymptotic Expansions*, no. 64 in Classics in Applied Mathematics, Philadelphia: Society for Industrial and Applied Mathematics.

Bol'shev, L. N. (1965). On the construction of confidence limits, *Theory of Probability and Its Applications*, **10 (1)**: 173–177.

Butler, R., Sutton, R., Booth, J. and Strickland, P. O. (2008). Simulation-assisted saddlepoint approximation, *Journal of Statistical Computation and Simulation*, **78 (8)**: 731–745.

Butler, R. W. (2007). *Saddlepoint Approximations with Applications*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge: Cambridge University Press.

Cordeiro, G. M. and Cribari-Neto, F. (2014). *An Introduction to Bartlett Correction and Bias Reduction*, Heidelberg: Springer.

Cordeiro, G. M. and Ferrari, S. L. (1991). A modified score test statistic having chi-squared distribution to order n^{-1}, *Biometrika*, **78 (3)**: 573–582.

Davison, A. C. and Hinkley, D. V. (1997). *Bootstrap Methods and their Application*, Cambridge Series in Statistical and Probabilistic Mathematics, New York: Cambridge University Press.

DeBlasi, P. and Schweder, T. (2016). Confidence distributions from likelihoods by median bias correction, *arXiv:1607.05136 [math.ST]*, (pp.1–19).

Efron, B. (1979). Bootstrap methods: another look at the jackknife, *The Annals of Statistics*, **7 (1)**: 1–26.

Efron, B. and Hastie, T. (2016). *Computer Age Statistical Inference: Algorithms, Evidence, and Data Science*, Institute of Mathematical Statistics Monographs, New York: Cambridge University Press.

Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*, no. 57 in Monographs on Statistics and Applied Probability, New York: Chapman & Hall.

Field, C. and Ronchetti, E. (1990). *Small Sample Asymptotics*, vol. 13 of *Lecture Notes - Monograph Series*, California: Institute of Mathematical Statistics.

Hall, P. (1983). Inverting an Edgeworth expansion, *The Annals of Statistics*, **11 (2)**: 569–576.

Hall, P. (1986). On the number of bootstrap simulations required to construct a confidence interval, *The Annals of Statistics*, **14 (4)**: 1453–1462.

Hall, P. (1987). Edgeworth expansion for Student's t statistic under minimal moment conditions, *The Annals of Probability*, **15 (3)**: 920–931.

Hall, P. (1989). On efficient bootstrap simulation, *Biometrika*, **76 (3)**: 613–617.

Hall, P. (1992a). On the removal of skewness by transformation, *Journal of the Royal Statistical Society. Series B (Methodological)*, **54 (1)**: 221–228.

Hall, P. (1992b). *The Bootstrap and Edgeworth Expansion*, Springer Series in Statistics, New York: Springer-Verlag.

Hjort, N. L. and Schweder, T. (2017). Confidence distributions and related themes, *Journal of Statistical Planning and Inference*, in press.

James, G. S. (1955). Cumulants of a transformed variate, *Biometrika*, **42 (3/4)**: 529–531.

James, G. S. (1958). On moments and cumulants of systems of statistics, *Sankhyā: The Indian Journal of Statistics*, **20 (1/2)**: 1–30.

James, G. S. and Mayne, A. J. (1962). Cumulants of functions of random variables, *Sankhyā: The Indian Journal of Statistics, Series A*, **24 (1)**: 47–54.

Jayadeva, S. S. and Soumy, S. (In Press). EigenSample: A non-iterative technique for adding samples to small datasets, *Applied Soft Computing*.

Jing, B.-Y., Tsao, M. and Zhou, W. (2017). Transforming the empirical likelihood towards better accuracy, *The Canadian Journal of Statistics*, **45 (3)**: 340–352.

Johnson, N. L., Kotz, S. and Balakrishnan, N. (1994). *Continuous Univariate Distributions*, vol. II, John Wiley & Sons, Inc.

Kakizawa, Y. (1996). Higher order monotone Bartlett-type adjustment for some multivariate test statistics, *Biometrika*, **83 (4)**: 923–927.

Kakizawa, Y. (1997). Higher-order Bartlett-type adjustment, *Journal of Statistical Planning and Inference*, **65 (2)**: 269–280.

Kolassa, J. E. (1996). Higher-Order Approximations to Conditional Distribution Functions, *The Annals of Statistics*, **24 (1)**: 353–364.

Lawley, D. N. (1956). A general method for approximating to the distribution of likelihood ratio criteria, *Biometrika*, **43 (3/4)**: 295–303.

Niki, N. and Konishi, S. (1986). Effects of transformations in higher order asymptotic expansions, *Annals of the Institute of Statistical Mathematics*, **38 (1)**: 371–383.

Petrov, V. V. (1975). *Sums of Independent Random Variables*, Berlin: Springer-Verlag.

Reiss, R.-D. (1989). *Approximate Distributions of Order Statistics: With Applications to Nonparametric Statistics*, Springer-Verlag.

Schweder, T. (2017). Confidence is epistemic probability for empirical science, *Journal of Statistical Planning and Inference*, in press.

Schweder, T. and Hjort, N. L. (2002). Confidence and likelihood, *Scandinavian Journal of Statistics*, **29 (2)**: 309–332.

Schweder, T. and Hjort, N. L. (2016). *Confidence, Likelihood, Probability: Statistical Inference with Confidence Distributions*, Cambridge Series in Statistical and Probabilistic Mathematics, New York: Cambridge University Press.

Shao, Q. (2002). Maximum Likelihood Estimation for Generalised Logistic Distributions, *Communications in Statistics - Theory and Methods*, **31 (10)**: 1687–1700.

Singh, K., Xie, M. and Strawderman, W. E. (2005). Combining information from independent sources through confidence distributions, *The Annals of Statistics*, **33 (1)**: 159–183.

Singh, K., Xie, M. and Strawderman, W. E. (2007). *Complex Datasets and Inverse Problems: Tomography, Networks and Beyond*, vol. 54 of *Lecture Notes–Monograph Series*, chap. Confidence distribution (CD) – distribution estimator of a parameter, pp. 132–150, Beachwood, Ohio, USA: Institute of Mathematical Statistics.

Skovgaard, I. M. (1987). Saddlepoint Expansions for Conditional Distributions, *Journal of Applied Probability*, **24 (4)**: 875–887.

Small, C. G. (2010). *Expansions and Asymptotics for Statistics*, no. 115 in Monographs on Statistics and Applied Probability, Boca Raton: Chapman and Hall/CRC.

Theodosopoulos, T. (2007). A reversion of the Chernoff bound, *Statistics & Probability Letters*, **77 (5)**: 558–565.

Viraswami, K. and Reid, N. (1996). Higher-order asymptotics under model misspecification, *The Canadian Journal of Statistics*, **24 (2)**: 263–278.

Viraswami, K. and Reid, N. (1998). A note on the likelihood-ratio statistic under model misspecification, *The Canadian Journal of Statistics*, **26 (1)**: 161–168.

Xie, M. and Singh, K. (2013). Confidence distribution, the frequentist distribution estimator of a parameter: A review, *International Statistical Review*, **81 (1)**: 3–39.

Yanagihara, H. and Yuan, K.-H. (2005). Four improved statistics for contrasting means by correcting skewness and kurtosis, *British Journal of Mathematical and Statistical Psychology*, **58 (2)**: 209–237.

Zheng, G., Xiong, J., Li, Q., Xu, J. and A. Yuan, J. L. G. (2017). Evaluating the accuracy of small *p*values in genetic association studies using Edgeworth expansions, *Scandinavian Journal of Statistics*, in press.