AISM (2008) 60:205–224
DOI 10.1007/s10463-006-0083-3

Saddlepoint approximations for multivariate *M*-estimates with applications to bootstrap accuracy

Chris Field · John Robinson · Elvezio Ronchetti

Received: 4 March 2005 / Published online: 18 October 2006

Abstract We obtain marginal tail area approximations for the one-dimensional test statistic based on the appropriate component of the M-estimate for both standardized and Studentized versions which are needed for tests and confidence intervals. The result is proved under conditions which allow the application to finite sample situations such as the bootstrap and involves a careful discretization with saddlepoints being used for each neighbourhood. These results are used to obtain second-order relative error results on the accuracy of the Studentized and the tilted bootstrap. The tail area approximations are applied to a Poisson regression model and shown to have very good accuracy.

Keywords Empirical saddlepoint · Relative errors · Studentized M-estimates · Tail area approximation · Tilted bootstrap

1 Introduction

Let $\mathcal{F}$ be a class of distributions of X and let $\psi(x, \theta)$ be a score function which assumes values in $\Re^p$ for values of $\theta \in \Re^p$. Let $\theta(F)$ be the solution of

C. Field
Department of Mathematics and Statistics, Dalhousie University,
Halifax, NS, Canada B3H 3J5

J. Robinson (✉)
School of Mathematics and Statistics, University of Sydney, NSW, Sydney 2006, Australia
e-mail: johnr@maths.usyd.edu.au

E. Ronchetti
Department of Econometrics, University of Geneva, Geneva, Switzerland

$$E\psi(X,\theta) = 0. \tag{1}$$

Consider $\mathcal{F}_0 \subset \mathcal{F}$ such that the first element of $\theta(F_0)$, for $F_0 \in \mathcal{F}_0$ is equal to a specified value θ_{10}. Assume that we have independent, identically distributed observations $X_1,\dots,X_n$ from a distribution F_0. Denote the solution of the equations

$$\sum_{j=1}^{n} \psi(X_j,\theta) = 0$$

as the M-estimate T of $\theta(F)$. We consider an observed sample $x_1,\dots,x_n$, an observed statistic t, and we wish to test an hypothesis that the first component of $\theta(F)$, θ_1, equals θ_{10}. Throughout the paper P_0 will denote a probability based on some fixed distribution $F_0 \in \mathcal{F}_0$ and we will write $\theta_0 = \theta(F_0)$ for the corresponding parameter. We are interested in finding accurate approximations to a P-value for a test of the above hypothesis using T_1, the first component of T, as a test statistic.

If the distribution of T_1 were known we could find $P_0(T_1 \geq t_1)$, where t_1 is the first component of t. In general, this is not possible but we can consider an approximation of the Studentized statistic $p_s(a) = P_0((T_1 - \theta_{10})/S \geq a)$, where S is a consistent estimate of σ, the asymptotic standard deviation of $\sqrt{n}T_1$, and $a = (t_1 - \theta_{10})/s$ will be used throughout the paper. A first-order approximation gives the standard normal distribution. Higher order approximations can be obtained by means of Edgeworth or saddlepoint techniques, where we need to use empirical versions of these with estimated cumulants or an estimated cumulant-generating function, respectively. Finally, since the distribution F_0 is often not specified, it is natural to consider bootstrap approximations to the tail areas.

In this paper, we provide saddlepoint approximations to tail areas for the bootstrap case. From a theoretical point of view, they can be used to analyse the relative error properties of bootstrap approximations. We focus on saddlepoint techniques because they provide approximations where the *relative error* can be controlled. This allows us to go beyond typical results about absolute errors already available in the bootstrap literature. From a computational point of view, the saddlepoint approximation is an attractive alternative to the bootstrap especially when the number of bootstrap replicates has to be large to obtain a required level of accuracy. More specifically, our contributions to the literature are as follows.

In Sect. 2 we state the two main theorems which give a saddlepoint approximation to the tail area when the underlying distribution does not have a density. This opens up the application of the approximation in the bootstrap case (Sects. 3 and 4). In Theorem 1 we obtain a saddlepoint approximation to $P_0((T_1 - \theta_{10})/\sigma \geq y)$, where σ is the asymptotic standard deviation of $\sqrt{n}T_1$. This generalizes the result of Almudevar et al. (2000) which was obtained under the assumption of existence of densities. The proof is not given as it is a simpler version of the proof of Theorem 2, given in Sect. 6, in which we give a new saddlepoint approximation for $P_0((T_1 - \theta_{10})/S \geq y)$, where S is a consistent estimate of σ. The approximation in Theorem 2 is not fully relative but the absolute errors are kept exponentially small. These results are proved under the weak conditions which enable us to use them in bootstrap applications. The proof uses two essential ideas. The first is that the tilting necessary in the saddlepoint approach is performed on only some of the variables involved in the test statistic. This is similar to the approach in Jing et al. (2002). The second idea, which is an innovation in this paper, is that we need to relate the distribution of the test statistic to the behaviour of a set of equations in a small neighbourhood. Since we do not have densities, the saddlepoint is applied in neighbourhoods and then aggregated.

In Sect. 3 we consider the Studentized bootstrap and use Theorem 2 to show that its *relative error* is $\mathrm{O}_P(\sqrt{n}a^3 \vee n^{-1})$ for $a < n^{-1/3}$. This implies a relative error of order $\mathrm{O}_P(n^{-1})$ in the normal region and a relative error of order $\mathrm{O}_P(n^{-1/2})$ in a region beyond the normal region up to $a \sim \mathrm{O}(n^{-1/3})$. These results extend similar results for smooth functions of means obtained in Jing et al. (1994), Feuerverger et al. (1999), Robinson and Skovgaard (1998), and Jing et al. (2002).

An alternative bootstrap approach is to use a tilted bootstrap with P value $p_t^*(a) = \tilde{P}^*((\tilde{T}_1^* - t_1)/s \geq a)$, where s is a consistent estimate of σ computed in the original sample and the tilde indicates that we have used a bootstrap sample which has been tilted in order to satisfy the null hypothesis $\theta_1 = \theta_{10}$. In Sect. 4 we describe this approach and use Theorem 1 to show that its *relative error* is $\mathrm{O}(na^4 \vee n^{-1})$ for $a < n^{-1/3}$. This is similar to the result for the Studentized bootstrap in the normal region but not quite as good beyond that.

Finally, we illustrate the theoretical results with an example in Sect. 5 where we consider Poisson regression with three covariates. We compute the P-values using the Studentized and the tilted bootstrap and illustrate the accuracy of the tail area in Theorem 4 to the tilted bootstrap results. The computations are performed using Splus and avoid coding of complicated derivatives by using accurate numerical derivatives.

The proofs of the theorems are given in Sect. 6.

2 Two saddlepoint approximation theorems

In order to state the tail area results, we need to set up the notation. We write

$$\bar{L}_\theta = n^{-1} \sum_{j=1}^{n} \psi(X_j, \theta),$$
$$\bar{M}_\theta = n^{-1} \sum_{j=1}^{n} \psi'(X_j, \theta),$$

$$\bar{Q}_\theta = n^{-1}\sum_{j=1}^{n}\psi(X_j,\theta)\psi^T(X_j,\theta),$$

and

$$\bar{M}'_\theta = n^{-1}\partial\bar{M}_\theta/\partial\theta, \quad \bar{Q}'_\theta = n^{-1}\partial\bar{Q}_\theta/\partial\theta,$$

where $'$ denotes the derivative with respect to θ. Define

$$\hat{L}_\theta = \bar{M}_\theta^{-1}\bar{L}_\theta,$$

whenever $\det(\bar{M}_\theta) \neq 0$. For M-estimates, the asymptotic standard deviation of $\sqrt{n}T_1$ is σ, where

$$\sigma^2 = \left[(E_\theta\bar{M}_\theta)^{-1}E_\theta\bar{Q}_\theta(E\bar{M}_\theta)^{-1}\right]_{11}$$

with estimated standard deviation S, where

$$S^2 = \left(\bar{M}_T^{-1}\bar{Q}_T\bar{M}_T^{-1}\right)_{11}.$$

Denote the cumulant-generating function of $\psi(X_1,\theta)$ by

$$\kappa(\tau,\theta) = \log\int\exp(\tau^{\mathrm{T}}\psi(x,\theta))\mathrm{d}F_0(x) \tag{2}$$

and define $\tau(\theta)$ as the solution to

$$\frac{\partial\kappa(\tau,\theta)}{\partial\tau} = 0. \tag{3}$$

We will obtain results on the distribution of the standardized and the Studentized version of T_1. First we state a result on the standardized version under the conditions given in the Appendix (Sect. 6). For this standardized version we obtain an approximation with relative error $\mathrm{O}(1/n)$.

To state the result for Studentized T_1 we need some further notation. Let $U_{j\theta} = (L_{j\theta}, V_{j\theta}, W_{j\theta})$ be independent identically distributed random vectors with positive definite covariance matrix such that all elements of $\bar{M}_\theta$ and $\bar{Q}_\theta$ are linear forms of the sum of $(L_{j\theta}, V_{j\theta})$ and all elements of $\bar{M}'_\theta$ and $\bar{Q}'_\theta$ are linear forms of the sum of $U_{j\theta}$. Let the dimensions of the components of $U_{j\theta}$ be p, q, and r, respectively.

Let F_U be the distribution of $U_{j\theta}$ under F_0 and define the tilted variable $U^\tau = U_{\tau(\theta)\theta}$ to have distribution function

$$F^\tau(\ell,v,w) = \int_{(\ell',v',w')\leq(\ell,v,w)} \mathrm{e}^{\tau(\theta)^T\ell'-\kappa(\tau(\theta),\theta)}\mathrm{d}F_U(\ell',v',w').$$

Let $\Sigma_\tau = \mathrm{cov} U^\tau$, and let $\mu_\tau = EU^\tau$, where μ_τ has 0 in the first p components and $\theta = (\zeta, \eta)$ where ζ is scalar.

Theorem 1 *Suppose conditions* (*A1*)–(*A3*) *of Sect. 6 hold. Then*

$$P_0((T_1 - \theta_{10})/\sigma \geq y) = [1 - \Phi(\sqrt{n} w_1^\dagger(y))][1 + \mathrm{O}(1/n)], \tag{4}$$

where $w_1^\dagger(y) = w_1(y) - \log(w_1(y) G_1(y)/H_1'(y))/n w_1(y)$, $w_1(y) = \sqrt{2H_1(y)}$, *and for* $\zeta = \sigma y + \theta_{10}$,

$$H_1(y) = \inf_\eta \{-\kappa(\tau(\zeta, \eta), (\zeta, \eta))\} = -\kappa(\tau(\zeta, \tilde{\eta}), (\zeta, \tilde{\eta})),$$

and if tilde indicates that values are taken at $\tilde{\eta}$,

$$G_1(\zeta) = \frac{\sigma \tilde{J}}{\det \tilde{\Sigma}_{L\tau}^{1/2} \det \tilde{K}_{22}^{1/2}},$$

where $\Sigma_{L\tau} = \mathrm{cov}\, L_{\tau(\theta)\theta} = E_{\tau(\theta)} \bar{Q}_\theta$,

$$\begin{aligned} K_{22} &= \frac{\partial^2 \kappa(\tau(\zeta, \eta), (\zeta, \eta))}{\partial \eta^2} \\ &= \left(\left[\frac{\partial^2 \kappa(\tau, \theta)}{\partial \theta^2} - \frac{\partial^2 \kappa(\tau, \theta)}{\partial \theta \partial \tau} \left(\frac{\partial^2 \kappa(\tau, \theta)}{\partial \tau^2} \right)^{-1} \frac{\partial^2 \kappa(\tau, \theta)}{\partial \tau \partial \theta} \right]_{\tau = \tau(\theta)} \right)_{22}, \end{aligned}$$

where the subscript 22 *indicates the part of the matrix corresponding to* η, *and* J *is the expectation of the Jacobian of the transformation* $(\hat{\ell}, v) = g(\ell, v)$ *under the tilted distribution, namely* $E_{\tau(\theta)} \bar{M}_\theta$.

Note that we can show, after some computation, that

$$H_1'(y) = -\sigma \left[\frac{\partial \kappa(\tau, (\zeta, \eta))}{\partial \zeta} \right]_{\tau = \tau(\zeta, \tilde{\eta}), \eta = \tilde{\eta}}.$$

The proof of Theorem 1 is omitted as it follows in the same way as the proof of Theorem 2 given in Sect. 6.

The next theorem gives a result in the case of a Studentized statistic. Define the transformation $Z_\theta = (\hat{L}_\theta, \bar{V}_\theta + \hat{L}_\theta \frac{\partial \bar{V}_\theta}{\partial \theta}, \bar{W}_\theta) = g_1(\bar{L}_\theta, \bar{V}_\theta, \bar{W}_\theta)$ and let J_1 be the Jacobian of the transformation. Suppose $S = s(T_1, T_2, \bar{V}_T)$ and define the transformation $((T_1 - \theta_{10})/S, T_2, \bar{V}_T, \bar{W}_\theta) = g_2(T_1, T_2, \bar{V}_T, \bar{W}_\theta)$ and let J_2 be the Jacobian of this transformation. Let $(\xi, \eta, v, w) = g_2(\zeta, \eta, v, w)$ and define $\lambda(\xi, \eta, v, w) = -\kappa(\tau(\zeta, \eta), (\zeta, \eta))$ and $\Lambda(\xi, \eta, v, w) = \lambda(\xi, \eta, v, w) + u^{*T} u^*/2$, where $u^* = \Sigma_\tau^{-1/2}(u - \mu_\tau)$ and u^* has 0 in the first p components. Now define

$$H(\xi) = \inf_{\eta, v, w} \Lambda(\xi, \eta, v, w) = H(\xi, \tilde{\eta}, \tilde{v}, \tilde{w}) \tag{5}$$

and

$$h(\xi) = \inf_{\eta,v,w} \{\lambda(\xi, \eta, v, w)\}. \tag{6}$$

Theorem 2 *If Conditions (A1)–(A3) of Sect. 6 hold, then*

$$P_0((T_1 - \theta_{10})/S \geq y) = [1 - \Phi(\sqrt{n}w^{\dagger}(y))]\left[1 + \mathrm{O}\left(\frac{1}{n}\right)\right] + \mathrm{e}^{-nh(y)}\mathrm{O}\left(\frac{1}{n}\right), \tag{7}$$

where $w^{\dagger}(y) = w(y) - \log(w(y)G(y)/H'(y))/nw(y)$, $w(y) = \sqrt{2H(y)}$,

$$G(\xi) = \frac{\tilde{J}_1\tilde{J}_2}{\det\tilde{\Lambda}_{22}^{1/2}\det\tilde{\Sigma}_{\tau}^{1/2}\sqrt{2\pi/n}} \tag{8}$$

and Λ_{22} *denotes the submatrix of* $\partial^2\Lambda(z)/\partial z^2$ *for* $z = (\xi, \eta, v, w)$*, excluding the first row and column.*

Remark Because we tilt only on the variables $L_{j\theta}$ we are unable to obtain an approximation in the Studentized case where the errors are fully relative. However, we can get a substantial improvement over absolute errors as was possible in the case of smooth functions of means in Jing et al. (2002). It is worth noting that the improved result for Theorem 1 over that for Theorem 2, follows since in proving Theorem 1,

$$\inf_{\eta,v,w}\left[-\kappa(\tau(\zeta,\eta),(\zeta,\eta)) + u^{*T}u^*/2\right] = \inf_{\eta}\left[-\kappa(\tau(\zeta,\eta),(\zeta,\eta))\right] = -\kappa(\tau(\zeta,\tilde{\eta}),(\zeta,\tilde{\eta})),$$

whereas in Theorem 2, following the transformation to ξ which involves ζ, η, and v the minima used are given in (5) and (6).

3 Studentized bootstrap

In this section, we consider computing tail areas and P-values by using a studentized bootstrap. We are interested in $p_s(a) = P_0((T_1 - \theta_{10})/S > a)$, where the probability is computed under H_0. The bootstrap approximation proceeds as follows. Let $X_1^*, X_2^*, \ldots, X_n^*$ be a sample from the empirical distribution F_n and let T^* denote the solution of

$$\sum_{j=1}^{n} \psi(X_j^*, \theta) = 0. \tag{9}$$

In the studentized bootstrap we replace the tail area above by $p_s^*(a) = P^*((T_1^* - t_1)/S^* > a)$, where S^* is the bootstrap version of S. Our aim is to determine the accuracy of $p_s^*(a)$ relative to $p_s(a)$. The result is given in the next theorem.

Theorem 3 *If conditions (A1)–(A3) of Sect. 6 hold, then for $a < Cn^{-1/3}$, for some constant C,*

$$\frac{p_s^*(a)}{p_s(a)} = 1 + \mathrm{O}_P\left(\sqrt{n}a^3 \vee n^{-1}\right), \tag{10}$$

This ensures that the Studentized bootstrap has relative error $\mathrm{O}_P(1/n)$ for values of $a = \mathrm{O}(1/\sqrt{n})$ to relative error $\mathrm{O}_P(1/\sqrt{n})$ up to $a = \mathrm{O}(n^{-1/3})$. We note that in the case of the Studentized mean (which is a special case of the results considered here), under the assumption that $E\exp(tX^2) < \infty$ for t in an open neighbourhood of 0, the relative error can be kept of order $\mathrm{O}(\sqrt{n}a^3)$, that is of order $\mathrm{o}(1)$ for $a = \mathrm{o}(n^{-1/6})$. We are not able, under these conditions and with the methods used here, to extend Theorem 2 beyond $a = \mathrm{O}(n^{-1/3})$.

4 Bootstrap tilting

In the previous section, we showed that $p_s^*(a)$, the bootstrap approximation to the P value of the studentized statistic was accurate to relative order $\mathrm{O}_P(\sqrt{n}a^3 \vee n^{-1})$. Here we will look at a tilted bootstrap which will avoid the issue of studentizing and compare the accuracy of the P value for the tilted bootstrap to $p_s^*(a)$.

For the tilted bootstrap we will choose weights w_i which minimize the backward Kullback–Leibler distance between the weighted distribution and the distribution with weights $1/n$ subject to the constraints that, for each θ,

$$\sum_{i=1}^{n} w_i \psi(x_i, \theta) = 0 \tag{11}$$

and $\sum_{i=1}^{n} w_i = 1$. Thus, we minimize

$$\sum_{i=1}^{n} w_i \log(nw_i) - \beta^T \sum_{i=1}^{n} w_i \psi(x_i, \theta) + \gamma \left(\sum_{i=1}^{n} w_i - 1 \right) \tag{12}$$

with respect to w_i. This, together with the constraints, leads to

$$w_i = \mathrm{e}^{\beta(\theta)^T \psi(x_i,\theta) - \kappa^*(\beta(\theta),\theta)}, \tag{13}$$

where

$$\kappa^*(\beta, \theta) = \log \sum_{i=1}^{n} \mathrm{e}^{\beta^T \psi(x_i,\theta)}$$

and with $\beta(\theta)$ chosen so that (11) holds. The minimum of (12) is $-\kappa^*(\beta(\theta), \theta) + \log n$. Now if $\theta_1 = \theta_{10}$, we can choose θ_2 to minimize this by choice of θ_2 as the solution to

$$\sum_{i=1}^{n} w_i \beta(\theta)^T \partial\psi(x_i,\theta)/\partial\theta_2 = 0. \tag{14}$$

Denote the solution by $\tilde{\theta}$, where $\tilde{\theta}_1 = \theta_{10}$.

We now sample from the tilted empirical distribution with weights

$$\tilde{w}_i = \mathrm{e}^{\beta(\tilde{\theta})^T\psi(x_i,\tilde{\theta})-\kappa^*(\beta(\tilde{\theta}),\tilde{\theta})}. \tag{15}$$

We denote this empirical distribution by $\tilde{F}_n$ and the bootstrap sample as $\tilde{X}_i^*, i = 1,\ldots,n$. We now solve

$$\sum_{i=1}^{n} \psi(\tilde{X}_i^*, t) = 0$$

to get the estimate $\tilde{T}^*$. Our interest is in approximating the P value

$$p_t^*(a) = \tilde{P}^*((\tilde{T}_1^* - \theta_{10})/s > a),$$

where s is the standard deviation of $\sqrt{n}T_1$ computed with the original data and $a = (t_1 - \theta_{10})/s$, and comparing it with $p_s^*(a)$. To get this saddlepoint approximation we use Theorem 1 with $\kappa(\tau,\theta) = \log\sum_{i=1}^{n} \tilde{w}_i \exp(\tau^T\psi(x_i,\theta))$. The next theorem gives the result.

Theorem 4 *If conditions (A1)–(A3) of Sect. 6 hold, then for $a < Cn^{-1/4}$, for some constant C,*

$$\frac{p_s^*(a)}{p_t^*(a)} = 1 + \mathrm{O}(na^4 \vee n^{-1}). \tag{16}$$

This is not quite as good as the result for the Studentized bootstrap although it still gives relative error $\mathrm{O}(n^{-1})$ for $a < Cn^{-1/2}$ but we can only obtain relative error $\mathrm{o}(1)$ for $a = \mathrm{o}(n^{-1/4})$.

5 Numerical example

In this section, we illustrate the numerical accuracy of the tail areas approximations derived in the paper. Consider a Poisson regression model, $Y_i \sim \mathcal{P}(\mu_i)$, where

$$\log\mu_i = \theta_1 + \theta_2 x_{i2} + \theta_3 x_{i3} = x_i^T\theta \quad i = 1,\ldots,n$$

$x_i = (1, x_{i2}, x_{i3})^T, \theta = (\theta_1,\theta_2,\theta_3)^T$. We want to test the null hypothesis $H_0 : \theta_3 = \theta_{30} = 1$.

(x_{i2}, x_{i3}) are generated from a uniform distribution on $(0,1)$ for each sample and then Y_i are obtained as Poisson variables with mean μ_i. The parameter θ is set to $(1,1,1)^T$ and the sample size $n = 30$.

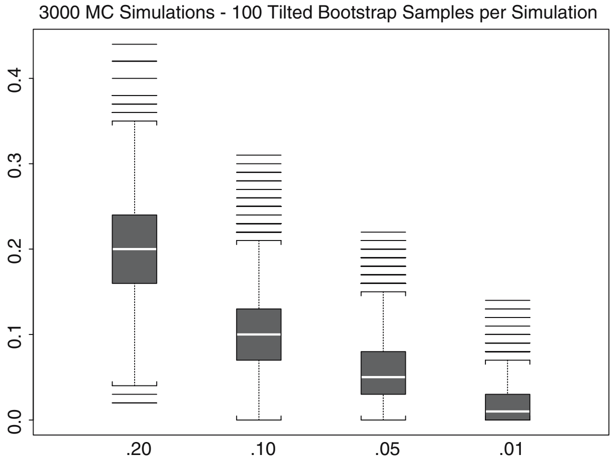

Fig. 1 The first panel gives tail probabilities for the saddlepoint approximation to the tilted bootstrap and approximate tail probabilities from 30,000 tilted bootstrap samples from one original sample. The second panel gives the relative errors of these two approximations

We consider the maximum likelihood estimator for the parameter θ, $\hat{\theta}$, the M-estimator defined by the equation

$$\sum_{i=1}^{n} \psi(Y_i, \theta) = 0,$$

where $\psi(Y_i, \theta) = (Y_i - \mu_i)x_i$ and $\mu_i = e^{x_i^T \theta}$.

We first consider the accuracy of the saddlepoint approximation of Theorem 1 by simulating a single sample of size 30 and obtaining the saddlepoint approximation to the tail probabilities, $p_t^*(a)$ for a sequence of values of a, for the tilted bootstrap for this sample. Then we obtain 30,000 tilted bootstrap samples from the original sample and get approximate tail area probabilities from these. These tail areas are plotted in the first panel of Fig. 1 together with the saddlepoint approximation from Sect. 4 (We note that derivatives used in the saddlepoint approximation are calculated numerically without loss of accuracy). In the second panel we plot the relative errors. It is clear that throughout the range an excellent approximation is obtained, illustrating the results of Theorem 1.

We also consider the accuracy of the tilted bootstrap to the true distribution. We take 10,000 Monte Carlo samples and for each sample compute $\hat{\theta}$, S, and $(\hat{\theta}_3 - \theta_{30})/S$. We approximate the tail areas corresponding to $p_s(a)$ of the Studentized statistic for $a = (0.5, 1.0, 1.5, 2.0)$ using these 10,000 Monte Carlo samples. Then we obtain 10 samples and from each we get 1,000 tilted bootstrap samples from which we get approximate tail probabilities corresponding to the four values of a. The mean (BSM) and standard deviation (BSSD) of these are given in Table 1. In addition, Table 1 gives the mean (SPM) and standard deviation (SPSD) of the 10 saddlepoint approximations corresponding to each

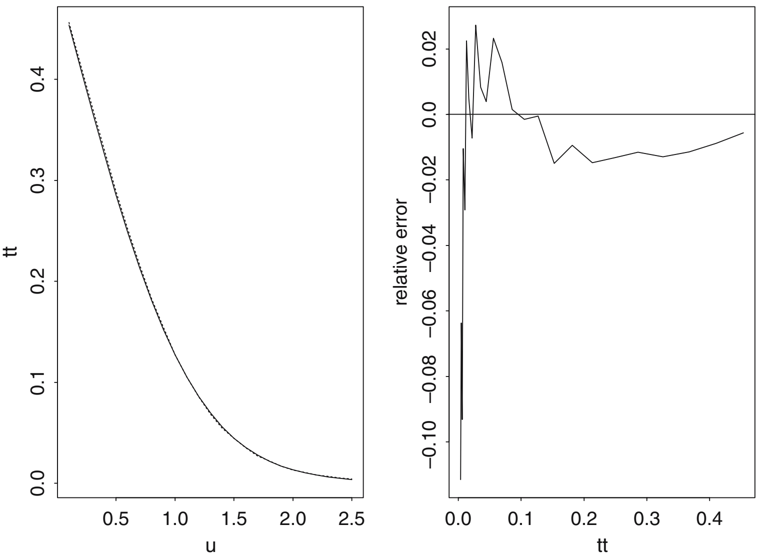

Fig. 2 Boxplots of 3,000 Studentized bootstrap p values corresponding to exact tail areas 0.2, 0.1, 0.05,0.01

Table 1 Values a, Monte Carlo approximation, mean and standard deviation of 10 saddlepoint approximations to tilted bootstrap, mean and standard deviation of 10 tilted bootstrap approximations using 1,000 bootstrap samples and standard deviation of the difference of the approximations

a	MC	SPM	SPSD	BSM	BSSD	SDDIF
0.5	0.314	0.298	0.046	0.296	0.050	0.015
1.0	0.164	0.152	0.051	0.153	0.054	0.012
1.5	0.068	0.069	0.041	0.074	0.042	0.008
2.0	0.020	0.030	0.028	0.032	0.028	0.008

of the four values of a. It also gives the standard error of the difference between the 10 pairs for each a value (SDDIF). The tilted bootstrap and the saddlepoint are seen to be very close from the last column (SDDIF) and much of the small variation can be explained by the fact that only 1,000 bootstrap samples were used. The approximation to the true distribution is not as good, as is to be expected from Theorems 3 and 4.

In addition, to examine the Studentized bootstrap, we take 3,000 Monte Carlo samples and obtain approximations to the quantiles of $(\hat{\theta}_3 - \theta_{30})/S$, corresponding to frequencies $0.2, 0.1, 0.05, 0.01$. For each of the 3,000 Monte Carlo samples we generate 100 nonparametric bootstrap samples, $(Y_i^*, x_{i1}^*, x_{i2}^*)$ for $i = 1, \ldots, n$ and for each of these we compute $(\theta_3^* - \hat{\theta}_3)/S^*$. Using the "a" given by the quantiles of the 3,000 Monte Carlo samples we compute the frequencies (out of 100) $p_s^*(a) = \sum \left((\theta_3^* - \hat{\theta}_3)/S^* \geq a\right)$. This provides 3,000 values for $p_s^*(a)$ which are represented in the boxplots of Fig. 2 and can be used to compare $p_s^*(a)$ with

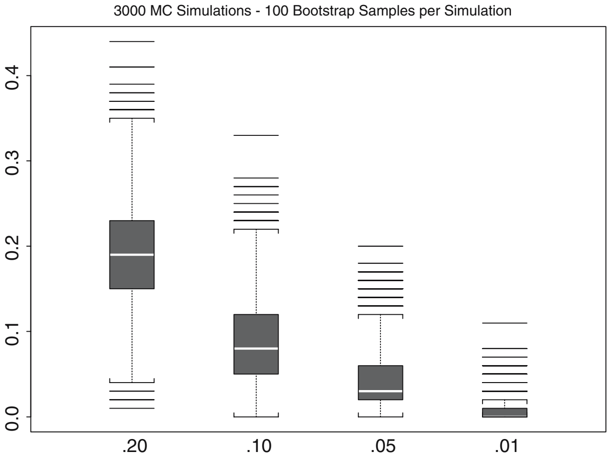

Fig. 3 Boxplots of 3,000 tilted bootstrap p values corresponding to exact tail areas 0.2, 0.1, 0.05,0.01

the exact tail areas $0.2, 0.1, 0.05, 0.01$. These give boxplots as expected for 3,000 random binomial $(100, p)$ variates with p taking the values $0.2, 0.1, 0.05, 0.01$. This is repeated with the tilted bootstrap to give Fig. 3.

Figures 2 and 3 show that the studentized bootstrap and the tilted bootstrap (which is not studentized) tail areas are equivalent. Both are centred around the exact values with the tilted bootstrap slightly less variable than the studentized bootstrap at least for the 0.2 and 0.1 tail areas.

Acknowledgements The authors thank Joanna Flemming for her help with the computation of the numerical example.

6 Appendix

6.1 Conditions

Consider independent, identically distributed observations $X_1, X_2, \ldots, X_n$ from a distribution F_0. We have a score function $\psi(X_1, \theta)$ which assumes values in $\Re^p$ such that $\int \psi(x, \theta)\mathrm{d}F_0(x) = 0$ has a solution θ_0. Suppose that $\psi(X_1, \theta)$ has a derivative $\psi'(X_1, \theta)$ with respect to θ, with probability 1, and assume

(A1) $\det\left(\int \psi'(x, \theta_0)\mathrm{d}F_0(x)\right) \neq 0$.

Then, if for some $\gamma > 0$, $\int \psi'(x, \theta)\mathrm{d}F_0(x)$ is continuous at all $\theta \in B^p_\gamma(\theta_0)$, the solution θ_0 is the unique solution in $B^p_\gamma(\theta_0)$, where by $B^p_\gamma(\theta_0)$ we mean a cube with side length 2γ of dimension p centred at θ_0.

Assume that

(A2) The elements of $\psi(X_1, \theta)$ and its first four derivatives with respect to θ exist and are bounded and continuous.

In order to apply an Edgeworth expansion we need a smoothness condition for the variables U^τ. Assume

(A3) $0 < c < \det\Sigma_\tau^{1/2} < C$ and if $\varphi_\tau(\xi) = E\mathrm{e}^{i\xi^T U^\tau}$, then $|\varphi_\tau(\xi)| < 1-\rho$, for $\rho > 0$ and for all $c < |\xi| < Cn^{d/2}$, where $d = p+q+r+1$.

Choose $0 < \epsilon < \frac{1}{4}|\det E_0\psi'(X_1,\theta_0)|$, $\gamma > 0$ and $B > 0$ and define the set E by

$$E = E(\epsilon,\gamma,B) \\ = \left\{|\det\bar{M}_\theta^{-1}\bar{Q}_\theta\bar{M}_\theta^{-1}| > \epsilon,\ \max|\bar{M}''_\theta| < B,\ |\hat{L}_\theta| < \frac{3}{4}\gamma,\ \text{for } \theta \in B_\gamma^p(\theta_0)\right\}. \quad (17)$$

Then the conditions (A1)–(A3) together with Cramér's large deviation theorem ensure that

$$P_0(E) > 1 - \mathrm{e}^{-cn}$$

for some $c > 0$ depending only on ϵ,γ,B. We can then restrict attention to $X \in E$ since for any event A

$$P_0(A) = P_0(A \cap E) + \mathrm{O}(\mathrm{e}^{-cn})$$

and we will be concerned only with approximations to probabilities of events with errors at least $\mathrm{O}(\mathrm{e}^{-cn})$. In the following, we will restrict attention to samples in E. Then

$$||\bar{M}_\theta^{-1}(\bar{M}_{\theta_0} - \bar{M}_\theta)|| < B\gamma/\epsilon < \frac{1}{4}$$

for $\theta \in B_\gamma^p(\theta_0)$ by choice of $\gamma < \epsilon/4B$. The inequality allows the application of Lemma 1 of Almudevar et al. (2000) with $\alpha = 1/4$ to show that there is a unique solution T of $\sum_{j=1}^n \psi(X_j,\theta) = 0$ in $B_\gamma^p(\theta_0)$.

6.2 Proof of Theorem 2

Since in the problem considered here no densities exist we find the probability of the tail event $\{(T_1-\theta_{10})/S \geq y\}$ by partitioning the space of (T, V_T) into small regions and approximating $P_0((T_1-\theta_{10})/S \geq y)$ by summing probabilities of the appropriate small regions. To do this we need to bound the probabilities of these small regions in the space of (T, V_T) by probabilities of regions in the space of $\bar{U}_\theta$. These bounds are derived in the technical lemma below. Next we use indirect Edgeworth approximations to these probabilities and an integral approximation to the sum of the indirect Edgeworth approximations. As part of this we find bounds for the errors of approximation.

Lemma 1 *Take* $\theta \in B^p_{\frac{3}{4}\gamma}(\theta_0)$*,* $v \in R^q$ *and* $0 < \delta < \frac{1}{4}\gamma$*. Then there is a* $C > 0$*, depending only on* B, ϵ *such that for* δ *chosen so that* $C\delta < \frac{1}{4}$*,*

$$\begin{aligned}
&\{\hat{L}_\theta \in B^p_{\delta(1-C\delta)}(0)\} \cap \left\{\bar{V}_\theta + \hat{L}_\theta\left[\frac{\partial \bar{V}_\theta}{\partial\theta}\right] \in B^q_{\delta(1-C\delta)}(v)\right\}\\
&\quad\subset \{T \in B^p_\delta(\theta)\} \cap \{\bar{V}_T \in B^q_\delta(v)\}\\
&\quad\subset \{\hat{L}_\theta \in B^p_{\delta(1+C\delta)}(0)\} \cap \left\{\bar{V}_\theta + \hat{L}_\theta\left[\frac{\partial \bar{V}_\theta}{\partial\theta}\right] \in B^q_{\delta(1+C\delta)}(v)\right\}. \qquad (18)
\end{aligned}$$

Proof Suppose $T \in B^p_\delta(\theta)$ and $V_T \in B^q_\delta(v)$. Expanding $\bar{L}_T = 0$ about θ and noting that in E, $|\bar{M}'_\theta|$ are bounded and that $|\det \bar{M}_\theta| > \epsilon$, we can choose

$$|\hat{L}_\theta - (\theta - T)| \leq C\delta^2$$

and then similarly

$$\left\|\bar{V}_T - V_\theta - \hat{L}_\theta\left[\frac{\partial \bar{V}_\theta}{\partial\theta}\right]\right\| \leq C\delta^2,$$

verifying the second inclusion of (18). Conversely, we can choose C such that

$$\sup_{\theta' \in B^p_\delta(\theta)} |\bar{M}^{-1}_{\theta'}\bar{M}_\theta - I_p| \leq \frac{1}{2}C\delta.$$

So from Lemma 1 of Almudevar et al. (2000), if $\hat{L}_\theta \in B^p_{\delta(1-C\delta)}(0)$ and δ is such that $C\delta < 1/4$, then there is a unique solution $T \in B^p_\delta(\theta)$. Also as before, if $\bar{V}_\theta + \hat{L}_\theta[\frac{\partial \bar{V}_\theta}{\partial\theta}] \in B^q_{\delta(1-C\delta)}(v)$, then $\bar{V}_T \in B^q_\delta(v)$. This concludes the proof of the lemma.

We want

$$\begin{aligned}
&P_0((T_1 - \theta_{10})/S \geq y)\\
&\quad= P_0(\{(T, \bar{V}_T) \in B^{p+q}_{\frac{3}{4}\gamma}(\theta_0, E_0\bar{V}_{\theta_0})\} \cap \{(T_1 - \theta_{10})/S \geq y\}) + \mathrm{O}(\mathrm{e}^{-cn}). \qquad (19)
\end{aligned}$$

Let (ζ_i, η_j, v_k), where i, j, k take values $\ldots, -2, -1, 0, 1, 2, \ldots$, be centres of cubes of side 2δ giving a partition of R^{p+q} with $(\zeta_0, \eta_0, v_0) = (\theta_{10}, \theta_{20}, E_0\bar{V}_{\theta_0})$. Denote by $\sum^\dagger$ the sum over $\{(i, j, k) : (\zeta_i, \eta_j, v_k) \in B^{p+q}_{\frac{3}{4}\gamma}(\theta_0, E_0\bar{V}_{\theta_0})$ and $\zeta_i/s(\zeta_i, \eta_j, v_k) \geq y\}$, where $s(\zeta, \eta, v)$ corresponds to S. Then

$$\begin{aligned}
&P_0((T_1 - \theta_{10})/S \geq y)\\
&\quad= \sum{}^\dagger P_0((T_1, T_2, \bar{V}_T) \in B^{p+q}_\delta(\zeta_i, \eta_j, v_k))(1 + \mathrm{O}(\delta)) + \mathrm{O}(\mathrm{e}^{-cn}) \qquad (20)
\end{aligned}$$

where the relative error $\mathrm{O}(\delta)$ is due to using the cubes touching the boundary of the region $\{(T_1 - \theta_{10})/S \geq y\}$ within $B^{p+q}_{\frac{3}{4}\gamma}(\theta_0, E_0\bar{V}_{\theta_0})$.

Now the lemma applied to the probability of this cube gives

$$P_0\left(\{\hat{L}_{(\zeta_i,\eta_j)} \in B^p_{\delta(1-C\delta)}(0)\} \cap \left\{\bar{V}_{(\zeta_i,\eta_j)} + \hat{L}_{(\zeta_i,\eta_j)}\left[\frac{\partial \bar{V}_\theta}{\partial\theta}\right]_{\theta=(\zeta_i,\eta_j)} \in B^q_{\delta(1-C\delta)}(v_k)\right\}\right)$$
$$< P_0((T_1, T_2, \bar{V}_T) \in B^{p+q}_\delta(\zeta_i, \eta_j, v_k))$$
$$< P_0\left(\{\hat{L}_{(\zeta_i,\eta_j)} \in B^p_{\delta(1+C\delta)}(0)\} \cap \left\{\bar{V}_{(\zeta_i,\eta_j)} + \hat{L}_{(\zeta_i,\eta_j)}\left[\frac{\partial \bar{V}_\theta}{\partial\theta}\right]_{\theta=(\zeta_i,\eta_j)} \in B^q_{\delta(1+C\delta)}(v_k)\right\}\right)$$

Take B_{km} to be a typical $B^p_{\delta(1-C\delta)}(0) \times B^q_{\delta(1-C\delta)}(v_k) \times B^r_\delta(w_m)$, or by a similar term with $1-C\delta$ replaced by $1+C\delta$, where m takes values $\ldots, -2, -1, 0, 1, 2 \ldots$. The w_m are centres of cubes of radius δ giving a partition of R^r with $w_0 = E_0(\bar{W}_{\theta_0})$. We can bound the sum in (20) by

$$\sum{}^{\dagger} \sum{}^{\ddagger} P_0(Z_{(\zeta_i,\eta_j)} \in B_{km}) + \mathrm{O}(\mathrm{e}^{-cn}),$$

where $\sum{}^{\ddagger}$ is a sum over m such that $|w_m| < \frac{3}{4}\gamma$ and where for the lower bound B_{km} has $1 - C\delta$ and $1 + C\delta$ for the upper bound.

Writing $u = (\ell, v, w)$, let

$$e_d(u, F^\tau) = \frac{\exp(-nu^{*T}u^*/2)}{(2\pi/n)^{(p+q+r)/2} \det \Sigma_\tau^{1/2}}\left(1 + \sum_{l=1}^{d} Q_{ln}(u^*\sqrt{n})\right).$$

Then using Theorem 1 of Robinson et al. (1990), we have

$$P_0(Z_\theta \in B_{km}) = P_0((\bar{L}_\theta, \bar{V}_\theta, \bar{W}_\theta) \in g_1^{-1}(B_{km}))$$
$$= \mathrm{e}^{n\kappa(\tau(\theta),\theta)}\left[\int_{g_1^{-1}(B_{km})} \mathrm{e}^{-n\ell^T\tau(\theta)} e_d((\ell, v, w), F^\tau)\mathrm{d}\ell\mathrm{d}v\mathrm{d}w + R\right], \tag{21}$$

where $R = R_1 + R_2 + R_3$ corresponding to the three residuals of Robinson et al. (1990). The first term in the last equation is equal to

$$\mathrm{e}^{n\kappa(\tau(\theta),\theta)} \int_{B_{km}} J_1(z)\mathrm{e}^{-n\ell^T\tau(\theta)} e_d(g_1^{-1}(z), F^\tau)\mathrm{d}z,$$

where $J(z)$ is the Jacobian of the transformation $z = g_1(\ell, v, w)$ and we write $g_1^{-1}(z) = (\ell(z), v(z), w(z))$. Noting that for this transformation $g_1^{-1}(0, v, w) = (0, v, w)$, we can approximate this integral by

$$P_0(Z_\theta \in B) = \mathrm{e}^{n\kappa(\tau(\theta),\theta)}[J_1(0, v_k, w_m)e_d((0, v_k, w_m), F^\tau)\delta^{p+q+r}(1 + \mathrm{O}(n\delta))+R].$$

Take $\delta = n^{-2}$ so that the term $\mathrm{O}(n\delta)$ is $\mathrm{O}(n^{-1})$. Then noting that $d = s - 3$ from Theorem 1 of Robinson et al. (1990) and that our result concerns means rather than sums,

$$R_1 < C\mathrm{vol}(g_1^{-1}(B))n^{(p+q+r)/2-(d+1)/2} < C\mathrm{vol}(g_1^{-1}(B))n^{-1} \tag{22}$$

if $d = p + q + r + 1$, where $\mathrm{vol}(A)$ is the volume of the set A. Also

$$R_3 < C\,\mathrm{sur}(g_1^{-1}(B))n^{(p+q+r)/2}\epsilon < C\mathrm{vol}(g_1^{-1}(B))n^{(p+q+r)/2}\epsilon/\delta, \tag{23}$$

where ϵ is the smoothing parameter in the theorem, where $\mathrm{sur}(A)$ is the surface area of the set A. Taking $\epsilon = n^{-(d+5)/2}$, we get

$$R_3 < C\mathrm{vol}(g_1^{-1}(B))n^{-1}.$$

To bound the other term we use (A4) from which we see

$$R_2 < C\mathrm{e}^{-cn}. \tag{24}$$

Approximating the sums by integrals, we have

$$\begin{aligned} P_0((T_1 - \theta_{10})/S \geq y) = &\int_{A_y} \mathrm{e}^{n\kappa(\tau(\zeta,\eta),(\zeta,\eta))}J_1(0, v, w)e_d((0, v, w), F^\tau)\mathrm{d}\zeta\,\mathrm{d}\eta\mathrm{d}v\mathrm{d}w \\ &\times (1 + \mathrm{O}(n^{-1})) + \int_{A_y} \mathrm{e}^{n\kappa(\tau(\zeta,\eta),(\zeta,\eta))}\mathrm{d}\zeta\,\mathrm{d}\eta\mathrm{d}v\mathrm{d}w\mathrm{O}(n^{-1}), \end{aligned} \tag{25}$$

where $A_y = \{(\zeta, \eta, v) : \{(\zeta - \theta_{01})/s(\zeta, \eta, v) \geq y\} \cap B^{p+q+r}_{\frac{3}{4}\gamma}(\theta_{10}, \theta_{20}, E_0\bar{V}_\theta, E_0\bar{W}_\theta)$ and where we may incorporate the exponential error term in the relative error term by bounding y by a sufficiently small constant.

Consider the transformation $(\xi, \eta, v, w) = g_2(\zeta, \eta, v, w)$, where $\xi = (\zeta - \theta_{10})/s(\zeta, \eta, v)$ with Jacobian $J_2(\xi, \eta, v, w)$. So we can write

$$\begin{aligned} P_0((T_1 - \theta_{10})/S \geq y) = &\int_y^v\!\!\int_D \frac{\mathrm{e}^{-n\Lambda(\xi,\eta,v,w)}J_1J_2}{(2\pi/n)^{(p+q+r)/2}\det\Sigma_\tau^{1/2}}\mathrm{d}\xi\,\mathrm{d}\eta\mathrm{d}v\mathrm{d}w(1 + \mathrm{O}(1/n)) \\ &+ \int_y^v\!\!\int_D \mathrm{e}^{-n\lambda(\xi,\eta,v,w)}\mathrm{d}\xi\,\mathrm{d}\eta\mathrm{d}v\mathrm{d}w\mathrm{O}(n^{-1}), \end{aligned} \tag{26}$$

where the first Edgeworth term in $e_d((0,v,w),F^\tau)$ integrates to zero by symmetry and the other terms are in the $O(1/n)$ relative error term, and ν and the sides of the rectangle D are chosen small enough so that $(y,\nu)\times D$ is in $B^{p+q+r-1}_{\frac{3}{4}\gamma}(0,E_0\bar{V}_{\theta_0},E_0\bar{W}_{\theta_0})$, and so the transformation is one to one and $\Lambda(\xi,\eta,v,w)$ and $\lambda(\xi,\eta,v,w)$ remain convex as functions of (ξ,η,v,w).

Now define H and h as in (5) and (6). Then using (A2)

$$P_0((T_1-\theta_{10})/S\geq y)=\int_y^\nu\left[\frac{\mathrm{e}^{-nH(\xi)}}{\sqrt{2\pi/n}}G(\xi)(1+O(1/n))+\mathrm{e}^{-nh(\xi))}O\left(\frac{1}{n}\right)\right]\mathrm{d}\xi. \tag{27}$$

Putting $w=w(\xi)=\sqrt{2H(\xi)}$ and $w^\dagger(\xi)=w(\xi)-\log(w(\xi)G(\xi)/H'(\xi))/nw(\xi)$ we can obtain (7) of Theorem 2 as in Jing and Robinson (1994).

6.3 Proof of Theorem 3

In order to prove the result, we need to have an approximation for the tail area which is valid both under sampling from F_0 and under bootstrap sampling. In particular the approximation must be valid for the situation when the quantity of interest does not have a density. Theorem 2 gives such a result covering both cases since Condition (A3) still holds for the bootstrap (see Weber and Kokic 1997). To apply Theorem 2 to the bootstrap, denote the cumulant-generating function of $\psi(X_1^*,\theta)$ by

$$\kappa^*(\tau,\theta)=\log\sum\exp(\tau^T\psi(x_i,\theta))/n. \tag{28}$$

Our interest is now in the approximation for the tail area

$$P^*((T_1^*-t_1)/S^*\geq a)=(1-\Phi(\sqrt{n}w^{\dagger *}(a)))(1+O(1/n)),$$

where P^* denotes the probability computed under F_n and $w^{\dagger *}(a)$ is defined in the same way as $w^\dagger$ in Theorem 2 with F_0 replaced by F_n.

Part of the argument follows closely that given in Sect. 2.1 of Feuerverger et al. (1999). To match their notation, write $\alpha(w(a))=w(a)G(a)/H'(a)$ and note that $w^\dagger$ here corresponds to w^* and α corresponds to ψ in their paper. As both $w(a)$ and $\alpha(w(a))$ are analytic functions of y in a neighbourhood of the origin, we obtain

$$w(a)=A_0+A_1a+A_2a^2+A_3a^3+A_4a^4+O(a^5)$$

and

$$\alpha(w(a))=B_0+B_1a+B_2a^2+B_3a^3+B_4a^4+O(a^5),$$

where the coefficients A_j and B_j depend on the cumulants of $\psi(X_i,\theta)$ and its derivatives under F^τ but not on n. We have a similar expression for w^* and $\alpha^*(w^*)$ from the bootstrap tail area.

By the same calculations as in the proof of Theorem 2 in Robinson et al. (2003) we obtain

$$H(0) = 0, \quad H'(0) = 0, \quad H''(0) = 1. \tag{29}$$

Therefore, from the expansion of w, we get $A_0 = 0$ and $A_1 = 1$. Moreover, by equating the integral in (27) taken over $\Re^1$ to 1, we obtain $\alpha(0) = 1$ and thus $B_0 = 1$. We want to consider the ratio

$$\frac{p_s^*(a)}{p_s(a)} = \frac{1-\Phi(\sqrt{n}w^{\dagger *})}{1-\Phi(\sqrt{n}w^{\dagger})}[1+\mathrm{O}(1/n)] + \frac{\mathrm{e}^{-nh(a))}}{1-\Phi(\sqrt{n}w^{\dagger})}\mathrm{O}(1/n). \tag{30}$$

The first term here is considered in the same way as in Sect. 2.1 of Feuerverger et al. (1999) and we can bound it by $1+\mathrm{O}_P(\sqrt{n}a^3)$ for $a < n^{-1/3}$. For the second term, we need to note that (29) holds and similarly that $h(0) = 0$, $h'(0) = 0$, $h''(0) = 1$, so $\exp(-nh(a))/(1-\Phi(\sqrt{n}w^{\dagger})) = \mathrm{O}(na^3)$. So if we restrict attention to $a = \mathrm{O}(n^{-1/3})$ we can see that the ratio in (30) is $1 + O_P(\sqrt{n}a^3)$, which, for $a = \mathrm{O}(n^{-1/3})$, is $1 + \mathrm{O}_P(1/\sqrt{n})$.

6.4 Proof of Theorem 4

In the case of the tilted bootstrap, the cumulant generating function of $\tilde{F}_n$ is given by

$$\tilde{\kappa}^*(\tau,\theta) = \log\left(\sum \tilde{w}_i \mathrm{e}^{\tau^T\psi(x_i,\theta)}\right). \tag{31}$$

This is used in Theorem 1 to obtain the saddlepoint approximation to the tilted bootstrap.

We first prove the second order accuracy of the Edgeworth in this case. This is needed to obtain the comparisons of the expansions of $w_1^\dagger(a)$ of Theorem 1 and $w^\dagger(a)$ of Theorem 2, used later in the proof. The first part of the following proof is related to that of DiCiccio and Romano (1990) but differs in significant ways. We could use the general results of Hall (1992) to give the Edgeworth results but it is more transparent to write them out directly. For simplicity we will neglect all terms of smaller order than $n^{-1/2}$ in the rest of this section. We have

$$
\begin{aligned}
p_s^*(a) &= P^*((T_1^* - t_1)/S^* \geq a) \\
&= P^*\left(\frac{T_1^* - t_1}{s} - \frac{S^{*2} - s^2}{2s^3} a \geq a\right) \\
&= 1 - \Phi\left(\sqrt{n}a(1 - \frac{a\mathrm{cov}(T_1^*, S^{*2})}{s^3}\right) + \frac{1}{\sqrt{n}} p(\sqrt{n}a)\phi(\sqrt{n}a) \\
&= 1 - \Phi(\sqrt{n}a) + \frac{\phi(\sqrt{n}a)}{\sqrt{n}}\left(p(\sqrt{n}a) - \sqrt{n}a^2 \frac{\mathrm{cov}(T_1^*, S^{*2})}{s^3}\right) \qquad (32)
\end{aligned}
$$

and, if $\tilde{s}$ is the variance of T_1 under the tilted distribution,

$$
\begin{aligned}
p_t^*(a) &= \tilde{P}^*((\tilde{T}_1^* - \theta_{10})/s \leq a) \\
&= P^*((\tilde{T}_1^* - \theta_{10})/\tilde{s} \leq a(1 - (\tilde{s}^2 - s^2)/2s^2)) \\
&= 1 - \Phi(\sqrt{n}a(1 - (\tilde{s}^2 - s^2)/2s^2)) + \frac{1}{\sqrt{n}} p(\sqrt{n}a)\phi(\sqrt{n}a) \\
&= 1 - \Phi(\sqrt{n}a) + \frac{\phi(\sqrt{n}a)}{\sqrt{n}}\left(p(\sqrt{n}a) - \sqrt{n}a \frac{\tilde{s}^2 - s^2}{2s^2}\right). \qquad (33)
\end{aligned}
$$

To show that

$$p_s^*(a) - p_t^*(a) = \mathrm{o}(1/\sqrt{n})$$

we need to show

$$\frac{\tilde{s}^2 - s^2}{s^2} = a \frac{\mathrm{cov}(T_1^*, S^{*2})}{s^3}. \qquad (34)$$

We can see that $t - \theta_0 = B_{\theta_0}^T \bar{L}_{\theta_0}$, where $B_{\theta_0}^T = \bar{M}_{\theta_0}^{-1}$ and $T^* - t = B_t^T \bar{L}_t$, so $T_1^* - t_1 = B_{t1}^T \bar{L}_t$. Let $\bar{Y}_\theta$ be the vector of all elements of $\bar{M}_\theta$ and $\bar{Q}_\theta$ and let $g(\bar{Y}_\theta) = s^2 = B_{t1}^T Q_t B_{t1}$. Then

$$S^{*2} = g(\bar{Y}_{T^*}) = g(\bar{Y}_t) + (\bar{Y}_t^* - \bar{Y}_t) g'(\bar{Y}_t) + (T^* - t)\frac{\partial \bar{Y}_t}{\partial t} g'(\bar{Y}_t).$$

So

$$\mathrm{cov}(T_1^*, S^{*2}) = B_{t1}^T C_{12} g'(\bar{Y}_t) + B_{t1}^T C_1 B_t \frac{\partial \bar{Y}_t}{\partial t} g'(\bar{Y}_t),$$

where $C_{12} = \mathrm{cov}(\bar{L}_t^*, \bar{Y}_t^*)$ and $C_1 = \mathrm{var}(\bar{L}_t^*)$.

Also $\tilde{T}_1^* - \theta_{10} = \tilde{B}_{t1}^T \sum_{i=1}^n \psi(\tilde{X}_i^*, \tilde{\theta})/n$ and $s^2 = g(\bar{Y}_t)$ and $\tilde{s}^2 = g(\tilde{Y}_{\tilde{\theta}})$, where by $\tilde{Y}_{\tilde{\theta}}$ we mean weighted means of $\psi(x_i, \tilde{\theta})$ and similar terms with weights $\tilde{w}_i$. Then

$$\tilde{s}^2 = s^2 + (\bar{Y}_{\tilde{\theta}} - \bar{Y}_t) g'(\bar{Y}_t) + (\tilde{Y}_{\tilde{\theta}} - \bar{Y}_{\tilde{\theta}}) g'(\bar{Y}_t). \qquad (35)$$

Now $\bar{Y}_{\tilde{\theta}} - \bar{Y}_t = (\tilde{\theta} - t)\partial \bar{Y}_t / \partial t$ and

$$\tilde{Y}_{\tilde{\theta}} = \sum_{i=1}^{n} \tilde{w}_i Y_{i\tilde{\theta}} = \bar{Y}_{\theta_0} + \lambda B_{t1}^T C_{12}$$

since $\tilde{w}_i = 1/n + \tilde{\lambda} B_{t1}^T \psi(x_i, t)$. Also $\tilde{\lambda} = -(t - \theta_{10})/s^2 = -a/s$ and expanding in (11) we have

$$\frac{1}{n}\sum_{i=1}^{n} \psi(x_i, \tilde{\theta}) + \tilde{\lambda} \sum_{i=1}^{n} \psi(x_i, t)\psi(x_i, t)^T = 0.$$

Now $\tilde{\theta} - t = \tilde{\lambda} B_t^T Q_t$, so using these in (35) gives (34).

In order to obtain a result on the relative error, we can use Theorems 1 and 2 to get

$$\frac{p_s^*(a)}{p_t^*(a)} = \frac{1 - \Phi(\sqrt{n} w_s^*)}{1 - \Phi(\sqrt{n} w_t^*)} \tag{36}$$

and then use Mill's ratio to get

$$\frac{1 - \Phi(\sqrt{n} w_s^*)}{1 - \Phi(\sqrt{n} w_t^*)} \leq n w_s^* |w_t^* - w_s^*| \mathrm{e}^{w_s^* |w_t^* - w_s^*|}. \tag{37}$$

We now expand the functions w_s^* and w_t^*. Since the expansion has the same form for each, we write an expansion for w^* noting that the coefficients will differ for the two functions. Now

$$\begin{aligned} w^*(a) &= a + A_2 a^2 + A_3 a^3 + \cdots - \frac{\log(1 + B_1 a + B_2 a^2 + \cdots)}{n(a + A_2 a^2 + A_3 a^3 + \cdots)} \\ &= a + A_2 a^2 + A_3 a^3 + \cdots - B_1/n - a B_2'/n + \cdots . \end{aligned}$$

As a result

$$n w_s^*(w_s^* - w_t^*) = na(A_{2s} - A_{2t})a^2 + na(A_{3s} - A_{3t})a^3 - na(B_{1s} - B_{1t})/n + \cdots . \tag{38}$$

Now $A_{2s} - A_{2t}$ and $B_{1s} - B_{1t}$ are both of order $\mathrm{O}(1/\sqrt{n})$ from the equivalence of the Edgeworth expansions up to order $\mathrm{O}(1/n)$ but $A_{3s} - A_{3t}$ can only be shown to be of order $\mathrm{O}(1)$. As a result we have that

$$\frac{1 - \Phi(\sqrt{n} w_s^*)}{1 - \Phi(\sqrt{n} w_t^*)} = 1 + \mathrm{O}(na^4 \vee n^{-1}) \tag{39}$$

if we restrict a to $\mathrm{O}(n^{-1/3})$.

References

Almudevar, A., Field, C., Robinson, J. (2000). The density of multivariate M-estimates. *Annals of Statistics*, *28*, 275–297.

DiCiccio, T., Romano, J. (1990). Nonparametric confidence limits by resampling methods and least favorable families. *International Statistical Revue*, *58*, 59–76.

Feuerverger, A., Robinson, J., Wong, A. (1999). On the relative accuracy of certain bootstrap procedures. *Canadian Journal of Statistics*, *27*, 225–236.

Hall, P. (1992). *The bootstrap and Edgeworth expansion*. New York: Springer.

Jing, B. Y., Robinson, J. (1994). Saddlepoint approximations for marginal and conditional probabilities of transformed variables. *Annals of Statistics*, *22*, 1115–1132.

Jing, B. Y., Feuerverger, A., Robinson, J. (1994). On the bootstrap saddlepoint approximations. *Biometrika*, *81*, 211–215.

Jing, B. Y., Kolassa, J., Robinson, J. (2002). Partial saddlepoint approximations for transformed means. *Scandinavian Journal of Statistics*, *29*, 721–731.

Robinson, J., Skovgaard, I. M. (1998). Bounds for probabilities of small errors for empirical saddlepoint and bootstrap approximations. *Annals of Statistics, 26*, 2369–2394.

Robinson, J., Höglund, T., Holst, L., Quine, M. P. (1990). On approximating probabilities for large and small deviations in R^d. *Annals of Probability*, *18*, 727–753.

Robinson, J., Ronchetti, E., Young, G. A. (2003). Saddlepoint approximations and tests based on multivariate M-estimates. *Annals of Statistics, 31*, 1154–1169.

Weber, N., Kokic, P. (1997). On Cramer's condition for Edgeworth expansions in the finite population case. *Theory of Stochastic Processes*, *3*, 468–474.