Computational Statistics (2005) 20:231-244

Computational Statistics

Saddlepoint approximations to studentized bootstrap distributions based on M-estimates

Diego Kuonen [1]

[1] Statoo Consulting, PO Box 107, 1015 Lausanne, Switzerland.

Summary

Saddlepoint methods can provide extremely accurate approximations to resampling distributions. This article applies them to distributions of studentized bootstrap statistics based on robust M-estimates. As examples we consider the studentized versions of Huber's M-estimate of location, of its initially MAD scaled version, and of Huber's proposal 2. The studentized version of Huber's proposal 2 seems to be a preferable measure of location. Remarks on implementation and related problems are given.

Keywords: Bootstrap, Huber's proposal 2, Measure of location, M-estimation, Robust estimation, Saddlepoint, Studentized bootstrap.

1 Introduction

Saddlepoint approximations (Daniels, 1954) are highly accurate tools for finding approximations to the density or the distribution of a statistic. Detailed references on their appplication can be found in Jensen (1995), Kuonen (2001) and Davison and Hinkley (1997) for bootstrap analysis. There are several ways of using bootstrap results in approximating the density or distribution functions of pivots. A general device, described briefly in Section 2, is the ***studentized bootstrap*** or ***bootstrap-t method.*** Bootstrap analysis presents an important application of saddlepoint approximation. Daniels and Young (1991) generalised Davison and Hinkley's (1988) saddlepoint approximations to the studentized bootstrap mean. Their approach was extended by Davison *et al.* (1995); see Davison and Hinkley (1997, Section 9.5.3). This ***integration approach*** will be discussed in more detail in Section 3. In Section 4 we apply this to studentized versions of robust M-estimates of location. Remarks on implementation are given in Section 5.

2 Studentized bootstrap

Pivotal quantities play a crucial role for inference. This raises the question: what is a good pivot in nonparametric settings? A reasonable answer is

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

where T is the estimator of a parameter θ and V is a consistent estimator of $\mathrm{var}(T)$. In many cases the studentized statistic Z is called an ***approximate pivot***; its distribution is approximately the same for each value of θ. One simple approximation is to take Z to be $N(0,1)$. This is often valid as $n \to \infty$ but is only an approximation for finite samples. A slightly better approximation is given by Student's t distribution, but although one can expect that Z will behave like a t-statistic, there is no guarantee of having $n-1$ degrees of freedom. A more accurate procedure is to estimate the distribution of Z from replicates of the studentized bootstrap statistic, $Z^\star = (T^\star - t)/V^{\star 1/2}$, where t denotes the observed value of the statistic T, and $T^\star$ and $V^\star$ are based on a simulated sample, $X_1^\star, \ldots, X_n^\star$. Hence we can obtain accurate approximations without normal or Student theory assumptions. This procedure estimates the distribution of Z directly from the data.

3 Saddlepoint approximations

Consider a sample $x_1, \ldots, x_n$, thought of as the outcome of n independent and identically distributed random variables $X_1, \ldots, X_n$ whose probability density function (PDF) and cumulative distribution function (CDF) are denoted by f and F. Suppose that the statistic of interest, T, and some nuisance statistics, $S = (S_1, \ldots, S_{d-1})^T$, are the solution to the d estimating equations

$$U(t, s) = \sum_{i=1}^{n} \psi(x_i, t, s) = 0, \tag{2}$$

where $\psi(x_i, t, s)$ is a $d \times 1$ vector. The bootstrap quantities $T^\star$ and $S^\star$ are the solutions of the equation $U^\star(t, s) = \sum \psi(x_i, t, s) f_i^\star = 0$, where the bootstrap frequencies $(f_1^\star, \ldots, f_n^\star)$ have a multinomial distribution with denominator n and resampling probability vector $(p_1, \ldots, p_n)$; typically $p_i \equiv n^{-1}$. The cumulant generating function of $U^\star(t, s)$ for fixed t and s is

$$K(\zeta, t, s) = n \log \left[\sum_{i=1}^{n} p_i \exp\{\zeta^T \psi(x_i, t, s)\} \right] \tag{3}$$

and the saddlepoint approximation to the PDF of $U^\star(t, s)$ at u is

$$f_{U^\star}(u) = (2\pi)^{-d/2} |K''(\hat{\zeta}, t, s)|^{-1/2} \exp\left\{ K(\hat{\zeta}, t, s) - \hat{\zeta}^T u \right\}, \tag{4}$$

where $\hat{\zeta} = \hat{\zeta}(u)$, the saddlepoint, satisfies the $d \times 1$ saddlepoint equation $\partial K(\zeta, t, s)/\partial \zeta = u$ and $K''(\hat{\zeta}, t, s)$ is the $d \times d$ matrix of second derivatives with respect to ζ, and $|\cdot|$ denotes determinant; see also Davison and Hinkley (1997, Section 9.5.2). To get the saddlepoint approximations to the marginal PDF and CDF of $T^\star$ we need the Jacobian for the transformation from $U^\star$ to $T^\star$ and $S^\star$. A good approximation is given by

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

where $w(x_i, \hat{\zeta}, t, s) = p_i \exp\{\hat{\zeta}^T \psi(x_i, t, s)\} / \sum p_k \exp\{\hat{\zeta}^T \psi(x_k, t, s)\}$. Hence using the fact that $\Pr(T^\star = t, S^\star = s) = J(\hat{\zeta}, t, s) \Pr\{U^\star(t, s) = 0\}$ and using (4), the saddlepoint approximation to the joint density of $T^\star$ and $S^\star$ is

$$f_{T^\star, S^\star}(t, s) = J(\hat{\zeta}, t, s) (2\pi)^{-d/2} |K''(\hat{\zeta}, t, s)|^{-1/2} \exp\left\{ K(\hat{\zeta}, t, s) \right\}, \tag{5}$$

where $\hat{\zeta} = \hat{\zeta}(t, s)$ is the solution of the d equations $\partial K(\zeta, t, s)/\partial \zeta = 0$, and $K''(\hat{\zeta}, t, s)$ is the $d \times d$ matrix with elements $\partial^2 K(\hat{\zeta}, t, s)/\partial \zeta \partial \zeta^T$. We now apply Laplace's method to the integral of (5). Provided that the matrix $\partial^2 K(\hat{\zeta}, t, s)/\partial s \partial s^T$ is negative definite, the Laplace approximation to the integral of (5) with respect to $s = (s_1, \ldots, s_{d-1})$ at t can be obtained as shown in Davison and Hinkley (1997, pages 479–480). The resulting approximate marginal density of $T^\star$ at t is

$$f_{T^\star}(t) = J(\tilde{\zeta}, t, \tilde{s}) \left\{ 2\pi |K''(\tilde{\zeta}, t, \tilde{s})| \left| \frac{\partial^2 K(\hat{\zeta}, t, \tilde{s})}{\partial s \partial s^T} \right| \right\}^{-1/2} \exp\left\{ K(\tilde{\zeta}, t, \tilde{s}) \right\}, \tag{6}$$

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

$$\frac{\partial K(\zeta, t, s)}{\partial \zeta} = 0, \qquad \frac{\partial K(\zeta, t, s)}{\partial s} = 0. \tag{7}$$

These $2d-1$ equations can be solved using packaged routines within S-PLUS or R. To do so one can minimise the sum of squares of the derivatives in (7), ensuring that the minimum value is zero. For starting values, note that when t equals its sample value, say t_0, we have approximately that $\tilde{\zeta} = 0$ and that $\tilde{s}$ takes its values from the original data. Furthermore, $\partial^2 K(\hat{\zeta}, t, s)/\partial s \partial s^T$ in (6) denotes the $(d-1) \times (d-1)$ matrix whose detailed form is given in Davison and Hinkley (1997, page 480) and Kuonen (2001, pages 83–84). However, it only depends on the derivatives of $\psi(x_i, t, s)$, which can be easily obtained. Finally, the saddlepoint approximation to the CDF of $T^\star$ at t is

$$F_{T^\star}(t) = \Phi\left\{w + \frac{1}{w}\log\left(\frac{v}{w}\right)\right\}, \tag{8}$$

where $\Phi(\cdot)$ denotes the standard normal CDF, and

$$w = \operatorname{sign}(t - t_0)\left\{2K(\tilde{\zeta}, t, \tilde{s})\right\}^{1/2}, \tag{9}$$

$$v = -\frac{\partial K(\tilde{\zeta}, t, \tilde{s})}{\partial t} |K''(\tilde{\zeta}, t, \tilde{s})|^{1/2} \left|\frac{\partial^2 K(\hat{\zeta}, t, \tilde{s})}{\partial s \partial s^T}\right|^{1/2} \{J(\tilde{\zeta}, t, \tilde{s})\}^{-1}. \tag{10}$$

The only new quantity needed is $\partial K(\tilde{\zeta}, t, \tilde{s})/\partial t$, which is easily obtained. The problems in this integration approach are mainly computational. The only analytical work required is knowledge of the estimating equation function $\psi(x_i, t, s)$ and its derivatives, which define $T^\star$ and $S^\star$.

4 Studentized M-estimates

We now consider studentized versions of Huber's M-estimate of location, of its initially MAD scaled version and of Huber's proposal 2, which is especially appropriate when both location and scale parameters are of interest. The nonparametric delta method can be used to obtain an approximate variance V for these M-estimates. Let $\hat{\sigma}$ be a suitable scale estimate. The nonparametric linear delta estimate of the variability of the estimator $\hat{\theta}$ is (Kuonen, 2001, Remark 5.3)

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

where $\psi_k(\cdot)$ denotes Huber's function $\psi_k(x) = \min\{k, \max(x, -k)\}$, and prime denotes differentiation with respect to x_i; we take $k = 1.345$ throughout. Using $T = \hat{\theta}$, $V = v_L$ and $\theta = \theta_0$, the value of θ on the original data $x_1, \ldots, x_n$, the studentized statistic (1) can be rewritten as

$$Z = \frac{\hat{\theta} - \theta_0}{v_L^{1/2}} = \frac{\hat{\theta} - \theta_0}{\hat{\sigma}} \left\{ \sum_{i=1}^{n} \psi_k^2 \left(\frac{x_i - \hat{\theta}}{\hat{\sigma}} \right) \right\}^{-1/2} \sum_{i=1}^{n} \psi_k' \left(\frac{x_i - \hat{\theta}}{\hat{\sigma}} \right). \tag{11}$$

4.1 Huber's M-estimate of location

Let $\hat{\theta}$ be Huber's M-estimate of location defined as the solution to $\sum \psi_k(x_i - \theta) = 0$. Setting $\hat{\sigma} = 1$ in (11) yields the studentized version of Huber's M-estimate of location. Thus its bootstrap version is

$$Z^\star = (\hat{\theta}^\star - \theta_0) \left\{ \sum_{i=1}^{n} \psi_k^2 (x_i^\star - \hat{\theta}^\star) \right\}^{-1/2} \sum_{i=1}^{n} \psi_k' (x_i^\star - \hat{\theta}^\star), \tag{12}$$

where $\hat{\theta}^\star$ is Huber's M-estimate issued from the bootstrap sample. For our purposes we now simply use (2) with the $d = 2$ estimating equations

$$\psi(x_i, z, s) = \begin{pmatrix} \psi_k(x_i - s) \\ zn^{-1} \left\{ \sum_{i=1}^{n} \psi_k^2 (x_i - s) \right\}^{1/2} - (s - \theta_0) \psi_k'(x_i - s) \end{pmatrix}, \tag{13}$$

where $s = \hat{\theta}$ is the nuisance statistic and z is the statistic of interest, *i.e.* the studentized version of Huber's M-estimate of location. To calculate the marginal saddlepoint approximations of $Z^\star$, (6) and (8) using (9) and (10), we can get easily from (13) the needed matrices of derivatives.

As an example we consider data on the estimated percentage of shrimp in $n = 18$ shrimp cocktails (Staudte and Sheather, 1990, page 134). The data contain two outliers — one in each tail. Apart from a small underestimation of the bootstrap density in the lower tail the saddlepoint approximation to the PDF given in the right panel of Figure 1 is accurate. For the bootstrap distribution all approximations, including the saddlepoint approximation to the CDF, work reasonably well (right panel of Figure 1). Applications to other data (Kuonen, 2001, Section 6.4.1) suggest that bootstrapping the studentized version of Huber's M-estimate of location may be questionable. First, for data containing several outliers the resulting saddlepoint approximation breaks down by not capturing multi-modality of the histogram of bootstrap replicates. Second, the bootstrap may break down by delivering extreme bootstrap values in the tails of the distribution. Highly skewed data estimates of both location and scale overcome these problems.

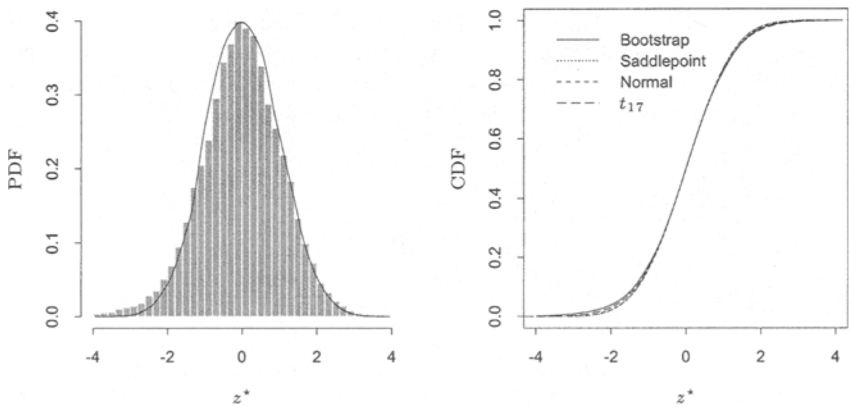

Figure 1: *Saddlepoint approximations for the shrimp data. Left panel: histogram of* $R = 49'999$ *replicates of the studentized version of Huber's* M*-estimate of location. The solid line represents the saddlepoint approximation to its bootstrap density. Right panel: bootstrap, saddlepoint, normal and Student-t approximation estimates of the CDF.*

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

An example of an estimator which merges location and scale estimates is the standardised M-estimator of location $\hat{\theta}$ with initial MAD scaling defined by $\sum \psi_k\{(x_i - \theta)/\hat{\sigma}\} = 0$ with MAD $\hat{\sigma} = \beta \text{med}\{|x_1 - \text{med}(x_1, \ldots, x_n)|, \ldots, |x_n - \text{med}(x_1, \ldots, x_n)|\}$, where med$(\cdot)$ denotes the median and β equals $1/\Phi^{-1}(3/4)$ to obtain Fisher consistency for a Gaussian standard deviation. Hampel *et al.* (1986, page 105) recommend the use of MAD scaling for M-estimators. Equation (11) delivers the studentized version of Huber's M-estimate of location. Its bootstrap version is

$$Z^\star = \frac{\hat{\theta}^\star - \theta_0}{\sigma_0} \left\{ \sum_{i=1}^{n} \psi_k^2 \left(\frac{x_i^\star - \hat{\theta}^\star}{\sigma_0} \right) \right\}^{-1/2} \sum_{i=1}^{n} \psi_k' \left(\frac{x_i^\star - \hat{\theta}^\star}{\sigma_0} \right), \tag{14}$$

where $\hat{\theta}^\star$ is Huber's M-estimate of location with initial MAD scaling issued from the bootstrap sample, θ_0 is its observed value and σ_0 the observed MAD. Using $\sigma_0 = 1$ reduces (14) to (12). This connection can be exploited below.

It can easily be verified that the $d = 2$ estimating equations are

$$\psi(x_i, z, s) = \left(\begin{array}{c} \psi_k\{(x_i - s)/\sigma_0\} \\ zn^{-1}\left\{\sum_{i=1}^{n} \psi_k^2\left(\frac{x_i - s}{\sigma_0}\right)\right\}^{1/2} - \frac{s - \theta_0}{\sigma_0}\psi_k'\left(\frac{x_i - s}{\sigma_0}\right) \end{array} \right), \tag{15}$$

where the nuisance statistic is $s = \hat{\theta}$ and z is the studentized version of Huber's M-estimate of location with initial MAD scaling. Second, the derivatives needed to get the saddlepoint approximations of $Z^\star$ can then be derived from those in the previous section. The saddlepoint approximation for the PDF of $T^\star = Z^\star$ and $S^\star = S_1^\star = \hat{\theta}^\star$ can then be obtained by means of (6), and for the CDF approximation one needs to apply (8) with (9) and (10).

Previously some limitations of the studentized version of Huber's M-estimate of location were mentioned. Kuonen (2001, Section 6.4.2) remarked that multi-modality of the the bootstrap density does not occur when initial MAD scaling is applied. Initial MAD scaling also provided a remedy for other data sets where the non-scaled versions did not work properly. As an example consider data on the capacities of $n = 17$ male Moriori skulls (Barnett and Lewis, 1994, page 40). Due to one single outlier the data are positively skewed. The saddlepoint approximation to the PDF overestimates the upper tail of the bootstrap density; see the left panel of Figure 2. The CDF approximations, given in the right panel of Figure 2, underestimate the upper tail of the bootstrap distribution. Addditional examples (Kuonen, 2001, Section 6.4.2) establish that the saddlepoint approximation to the bootstrap PDF or CDF over- or underestimates the tails. The scale estimate $\hat{\sigma}$ was fixed to the one from the original data set. This corresponds to assuming that the true variance is known. One may prefer to take a bootstrap scale estimate, *i.e.* take a separate scale estimate for each bootstrap sample, not a single overall scale estimate. Location M-estimators are usually not scale-invariant. This problem can be solved by using an estimator like Huber's proposal 2.

4.3 Huber's proposal 2

We suppose that $\hat{\theta}$ and $\hat{\sigma}$ are M-estimates found by simultaneous solution of

$$\sum_{i=1}^{n} \psi_k\left(\frac{x_i - \theta}{\sigma}\right) = 0, \qquad \sum_{i=1}^{n} \psi_k^2\left(\frac{x_i - \theta}{\sigma}\right) = n\gamma, \tag{16}$$

where $\psi_k(\cdot)$ denotes Huber's function. Taking $k = \infty$ and $\gamma = 1$ gives the classical estimates $\hat{\theta} = \bar{x}$ and $\hat{\sigma}^2 = n^{-1}\sum(x_i - \bar{x})^2$. Huber (1964, 1981) treats questions of existence and uniqueness of solutions to (16). Moreover, he set $\gamma = 2\Phi(k) - 1 - 2k\phi(k) + 2k^2[1 - \Phi(k)]$ in his *proposal 2* in order to get Fisher consistency, and he remarked that his proposal 2 is not sensitive

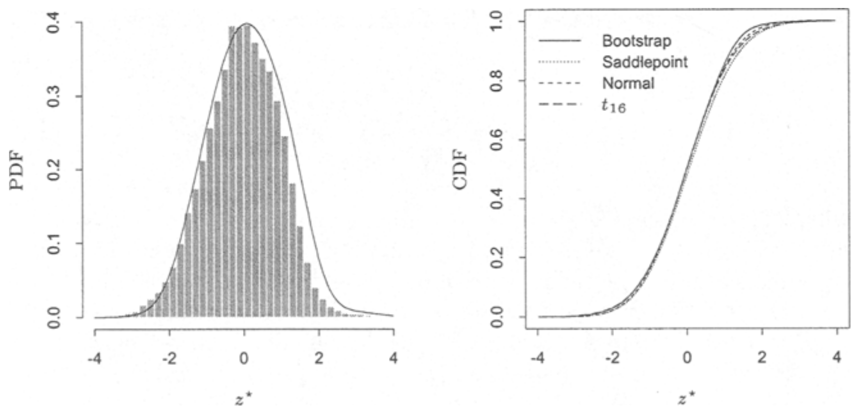

Figure 2: *Saddlepoint approximations for the Moriori data. Left panel: histogram of $R = 49'999$ replicates of the studentized version of Huber's M-estimate of location with initial MAD scaling. The solid line represents the saddlepoint approximation to its bootstrap density. Right panel: bootstrap, saddlepoint, normal and Student-t approximation estimates of the CDF.*

to a 'wrong' choice of k. We considered $k = 1.345$, thus $\gamma = 0.71$. Using $\hat{\theta}$ and $\hat{\sigma}$, the simultaneous solutions of (16), the studentized version of Huber's proposal 2 is given by (11). Its bootstrap version is

$$\begin{aligned} Z^\star &= \frac{\hat{\theta}^\star - \theta_0}{\hat{\sigma}^\star} \left\{ \sum_{i=1}^{n} \psi_k^2 \left(\frac{x_i^\star - \hat{\theta}^\star}{\hat{\sigma}^\star} \right) \right\}^{-1/2} \sum_{i=1}^{n} \psi_k' \left(\frac{x_i^\star - \hat{\theta}^\star}{\hat{\sigma}^\star} \right) \\ &= \frac{\hat{\theta}^\star - \theta_0}{\hat{\sigma}^\star} (n\gamma)^{-1/2} \sum_{i=1}^{n} \psi_k' \left(\frac{x_i^\star - \hat{\theta}^\star}{\hat{\sigma}^\star} \right), \end{aligned} \tag{17}$$

where $\hat{\theta}^\star$ and $\hat{\sigma}^\star$ are based on the bootstrap sample $x_1^\star, \ldots, x_n^\star$, and θ_0 is the location estimate of the original data set. Davison and Hinkley (1997, Example 9.20) noted that (17) can also be written as

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

In denoting $n^{-1} \sum \psi_k' \{(x_i^\star - \hat{\theta}^\star)/\hat{\sigma}^\star\}$ by $s_2^\star$ we have at $Z^\star = z$ that $\hat{\sigma}^\star z = (\hat{\theta}^\star - \theta_0)(\gamma/n)^{-1/2} s_2^\star$. It yields that $\hat{\theta}^\star$ equals $\theta_0 + z(\gamma/n)^{1/2}\hat{\sigma}^\star/s_2^\star$ and therefore $\tilde{e}_i = (x_i^\star - \hat{\theta}^\star)/\hat{\sigma}^\star = \sigma_0 e_i^\star / s_1^\star - z(\gamma/n)^{1/2}/s_2^\star$, where $s_1^\star$ denotes $\hat{\sigma}^\star$, $e_i^\star = (x_i^\star - \theta_0)/\sigma_0$ are the studentized bootstrap values and θ_0, σ_0 are the observed values of the location and of the scale estimate, respectively. Hence we can apply the marginal saddlepoint approximations (6) and (8), using (9) and (10), with $T = Z$, $S = (S_1, S_2) = (\hat{\sigma},\, n^{-1}\sum \psi_k'\{(x_i - \hat{\theta})/\hat{\sigma}\})$ and

$$\psi(x_i, z, s_1, s_2) = \begin{pmatrix} \psi_k(\tilde{e}_i) \\ \psi_k^2(\tilde{e}_i) - \gamma \\ \psi_k'(\tilde{e}_i) - s_2 \end{pmatrix}. \tag{18}$$

Note that $ns_2^\star$ is the number of observations contained in the interval $(\hat{\theta}^\star - k\hat{\sigma}^\star,\, \hat{\theta}^\star + k\hat{\sigma}^\star)$, and thus it would be unwise to treat $s_2^\star$ as continuous. Hence the sampling version of s_2 is fixed to its observed value, *i.e.* $s_2^\star = s_2 = n^{-1}\sum \psi_k'\{(x_i - \theta_0)/\sigma_0\}$, and (18) can be modified by dropping off its third component. Therefore, $T = Z$ and $S = S_1 = \hat{\sigma}$ are the solutions to the remaining $d = 2$ estimating equations. The derivatives needed to calculate the saddlepoint approximation are again easily obtained; see Davison and Hinkley (1997, Example 9.20) and Kuonen (2001, page 99).

Examples of the application of saddlepoint approximations can be found in Kuonen (2001, Section 6.4.3). They illustrate that the resulting saddlepoint approximations are very accurate, implying that it is preferable to use the studentized version of Huber's proposal 2 instead of the studentized version of Huber's M-estimate of location with initial MAD scaling.

5 Miscellany

In this section we will illustrate the problems that may occur when applying the integration approach presented in Section 3 and applied in Section 4.

5.1 Diagnostics

Daniels and Young (1991, Section 5) noticed that care is needed when applying a Laplace approximation to (5) in order to obtain the marginal approximation to the density of $T^\star$ given in (6). This problem occurs because to apply a Laplace approximation we assume the matrix $\partial^2 K(\hat{\zeta}, t, s)/\partial s \partial s^T$ to be negative definite, *i.e.* the cumulant generating function $K(\hat{\zeta}, t, s)$ to be concave. But, there is no guarantee that $K(\hat{\zeta}, t, s)$ is a concave function of t and s. Daniels and Young (1991, page 171) and Davison and Hinkley (1997, page 482) noticed that fortunately this difficulty is much rarer in large samples. This is something the author can confirm as well. The only times we encountered non-concavity of $K(\hat{\zeta}, t, s)$ was with highly-skewed data sets with moderate values of n. Using the studentized version of Huber's proposal 2 this is for instance the case with Darwin's data, which contain $n = 15$ observations issued from an experiment to examine the superiority of cross-fertilised plants over self-fertilised plants (Davison and Hinkley, 1997, page

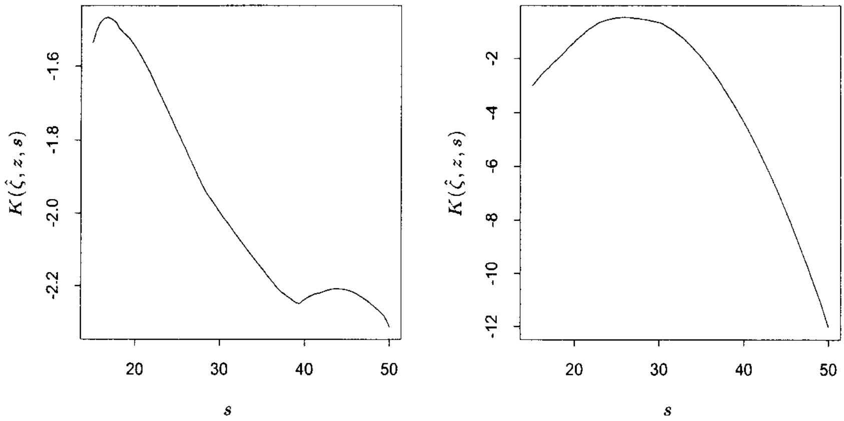

Figure 3: *The cumulant generating function (3) as a function of s, for $z = -2$ (left panel) and $z = 1$ (right panel), using the studentized version of Huber's proposal 2 applied to Darwin's data.*

186). For fixed values of $t = z$, namely $z = -2$ and 1, Figure 3 shows $K(\hat{\zeta}, z, s)$ as a function of s. The left panel illustrates that there does not exist an unique maximum of $K(\hat{\zeta}, -2, s)$ or similarly an unique minimum of $h(s) = -K(\hat{\zeta}, -2, s)$. The unique minimum of $h(s)$, $\tilde{s}$ say, should exist within the interior of the domain of integration for s. Uniqueness of $\tilde{s}$ is needed in order to apply the Laplace approximation and to ensure that the major contribution of the integral of (5) with respect to s is interior to the region of integration. The wireframe plots in Figure 4 show what happens when we apply the studentized version of Huber's proposal 2 to Darwin's data. The multi-modality of the joint density (5) is shown. The peaks shown in this figure would result in an inaccurate marginal approximation as the Laplace approximation only accounts for the dominant centre peak. Computation using a 50×50 grid took about 42 minutes in average CPU time on a Sun SPARC Ultra 60 workstation with 1Gb RAM using S-PLUS. It would be wise to use diagnostic plots like the one given in Figure 4. The drawback is that the CPU time needed is large, slowing down the integration approach. Another approach to diagnose a possible failure of Laplace's method is to check the signs of the eigenvalues of the real symmetric $(d-1) \times (d-1)$ matrix $\partial^2 K(\hat{\zeta}, z, s)/\partial s \partial s^T$ at a range of values of s: if all eigenvalues are negative Laplace approximation can be used, as this fact ensures the convexity of $h(s)$, the matrix $-\partial^2 K(\hat{\zeta}, z, s)/\partial s \partial s^T$ being then positive definite, and hence the uniqueness of $\tilde{s}$.

In the examples considered by the author this behaviour did not occur, so

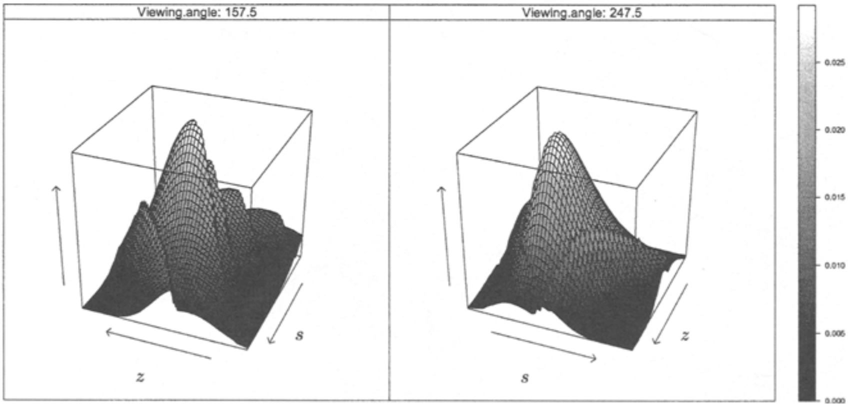

Figure 4: *Wireframe plots for the joint density (5) using the studentized version of Huber's proposal 2 applied to Darwin's data. Left panel: using a viewing angle of 157.5. Right panel: using a viewing angle of 247.5.*

we always considered a Laplace approximation. Moreover, if the difficulty is thought to have arisen one could simply jitter the values of interest and recalculate the quantities needed for the integration approach. Nevertheless, an approximation to the marginal distribution of $T^\star = Z^\star$ can always be found by numerical integration of (5) with respect to s; especially if there exists at least one positive eigenvalue of $\partial^2 K(\hat{\zeta}, z, s)/\partial s \partial s^T$.

5.2 Numerical integration

The safest procedure to apply to the joint density (5) would be to integrate it numerically with respect to $s = (s_1, \ldots, s_{d-1})$. Computationally such numerical integration is most conveniently performed using direct function evaluation of (5) on a regular grid. Numerical integration methods for use in S-PLUS or R are discussed in detail in Kuonen (2001, Appendix A) and summarised in Kuonen (2003). Integrals over infinite domains should be transformed to a finite region in view of the accuracy and convergence of the quadrature method in use. The transformations which performed best are listed in Kuonen (2001, Section 6.5.2.2), and Kuonen (2001, Section 6.5.2.3) presented a method which can be used to get an idea of the range of the studentized statistics, namely by using the range of their empirical influence values. All these ideas were then used in performing the numerical integration. However, the examples considered clearly illustrated the drawbacks of numerical integration: it becomes useless in practice as the running time is outperformed by direct simulation of the bootstrap replicates. Indeed, for the studentized version of Huber's proposal 2 applied to Darwin's data we computed, besides

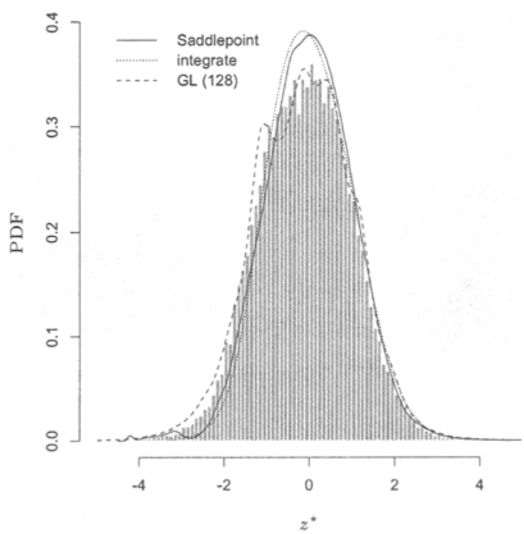

Figure 5: *Histogram of $R = 49'999$ replicates of the studentized version of Huber's proposal 2 for Darwin's data with saddlepoint,* `integrate` *and 128-point Gauss–Legendre (GL) approximations of the PDF.*

the $R = 49'999$ bootstrap replicates, the following PDF approximations: the saddlepoint approximation based on Laplace's method (see Section 3), and approximations obtained through numerical integration. Based on the comments in Kuonen (2001, Appendix A) we used as quadrature methods the S-PLUS function `integrate`, which implements uni-dimensional adaptive 15-point Gauss–Kronrod quadrature, and a 128-point Gauss–Legendre (GL) rule. For Darwin's data the results are shown in Figure 5. The saddlepoint approximation and `integrate` underestimate the lower tail and overestimates the centre. The 128-point GL rule seems to perform best, being the only approximation that captured the multi-modality in the centre and the behaviour in the lower tail. The approximate average CPU times for these computations were 712 minutes for the brute-force bootstrap computation, 39 minutes for the saddlepoint approximation, 184 minutes using `integrate` and 37 minutes using a 128-point GL approximation. The `integrate` function is very slow compared to the other approximations, but it still is significantly faster than brute-force bootstrap computation. Additional examples show that with `integrate` the computation can become very time-intensive and may give inaccurate PDF approximations. But, a GL rule with 128 points seems to be a good choice in practice. In order to approximate the CDF of $Z^\star$ the values of $F_{T^\star}(t)$ were calculated for 50 values of t equally spaced between the empirical influence values and a spline was used to interpolate between these values. In our context, it is important to note that the use of a 128-point GL rule implies 16'384 function evaluations solving each time the equations (7), which is very computer-intensive. For instance, at $t = z^\star = 1$, using Darwin's data and the studentized version of Huber's proposal 2, the approximation to $F_{T^\star}(1)$ using a 128-point GL rule took about 1'238 minutes in average CPU time. The use of a 64-point GL rule decreases this to about 36 minutes but decreases the accuracy. Hence the time needed to get the approximation for the entire range of the 50 values is clearly outperformed by the bootstrap running time of 712 minutes.

In summary, this and the examples in Kuonen (2001, Section 6.5.2) clearly illustrate the drawbacks of numerical integration for the computation of the CDF of $T^\star$. They become useless in practice as their running time is outperformed by direct simulation. One may think that this is due to the use of interpreted languages like S-PLUS or R, but we do not think that this is the case as numerical integration in multi-dimensional problems still raises many open questions.

6 Conclusion

In this article we considered saddlepoint approximations to studentized bootstrap versions of Huber's M-estimate of location, of its initially MAD scaled version and of Huber's proposal 2. The saddlepoint approximations are based on a Laplace approximation and hence care is needed with their application. Fortunately, experience shows that the problems described do not occur often. Therefore the need to perform a very computer-intensive numerical integration as mentioned disappears. However, it may always be wise to use diagnostic plots like the ones presented here. Hampel *et al.* (1986, page 105) recommended the use of MAD scaling for M-estimators instead of Huber's proposal 2, and noted that the latter is less reliable. However, we observed in the bootstrap context that Huber's proposal 2 seems to be a preferable measure of location which is not unduly affected by outliers or asymmetry in the data. Finally, note that the presented saddlepoint approximations form the basis for the construction of confidence intervals; see Kuonen (2001, Chapter 7) for a detailed discussion.

Acknowledgement

The author expresses his deep gratitude to his PhD thesis supervisor A. C. Davison for his suggestions relating to this work, which was supported by a grant from the Swiss National Science Foundation.

References

Barnett, V. and Lewis, T., 1994. *Outliers in Statistical Data* (3rd ed.). New York: Wiley.

Daniels, H. E., 1954. Saddlepoint approximations in statistics. *Annals of Mathematical Statistics*, 25, 631–650.

Daniels, H. E. and Young, G. A., 1991. Saddlepoint approximation for the studentized mean, with an application to the bootstrap. *Biometrika*, 78, 169–179.

Davison, A. C. and Hinkley, D. V., 1988. Saddlepoint approximations in resampling methods. *Biometrika*, 75, 417–431.

Davison, A. C. and Hinkley, D. V., 1997. *Bootstrap Methods and Their Application.* Cambridge: Cambridge University Press.

Davison, A. C., Hinkley, D. V. and Worton, B. J., 1995. Accurate and efficient construction of bootstrap likelihoods. *Statistics and Computing*, 5, 257–264.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. and Stahel, W. J., 1986. *Robust Statistics, the Approach Based on Influence Functions.* New York: Wiley.

Huber, P. J., 1964. Robust estimation of a location parameter. *Annals of Mathematical Statistics*, 35, 73–101.

Huber, P. J., 1981. *Robust Statistics.* New York: Wiley.

Jensen, J. L., 1995. *Saddlepoint Approximations.* Oxford: Oxford University Press.

Kuonen, D., 2001. *Computer-intensive statistical methods: with applications in bootstrap and robust inference.* PhD Thesis N. 2449, Department of Mathematics, Swiss Federal Institute of Technology, CH–1015 Lausanne. (stat.kuonen.com/thesis/)

Kuonen, D., 2003. Numerical integration in S-PLUS or R: A survey. *Journal of Statistical Software*, 8, 1–14.

Staudte, R. G. and Sheather, S. J., 1990. *Robust Estimation and Testing.* New York: Wiley.