GENERAL SADDLEPOINT APPROXIMATION METHODS FOR SMOOTH FUNCTIONS OF M-ESTIMATES WITH BOOTSTRAP APPLICATIONS

D.F. Andrews and A. Feuerverger
University of Toronto

Procedures are developed and implemented for computing general saddlepoint approximations for statistics defined by estimating equations and functions of these. Our approach is based on the fact that such statistics can be approximated as finite linear combinations of products of centered, normalized averages, and that cumulants of such approximants may be evaluated to any desired accuracy. The resulting approximations are useful in a wide variety of applications and may be computed using computer algebra routines. The application of these procedures to replace bootstrap sampling (in the case when the underlying distribution is an empirical cdf) is discussed. Mathematica code implementing these procedures is available at: `http://www.utstat.utoronto.ca/david/expand.dist.nb`

1. Introduction

This paper is concerned with the development and implementation of general saddlepoint approximations to the distributions of statistics belonging to a broad class, namely those which may be represented as smooth functions of M-estimators of identically and independently distributed variables leading to empirical versions related to bootstrap distributions. This work follows the earlier developments of Young and Daniels (1991) and DiCiccio, Martin and Young (1992a,b). Symbolic and numerical implementation of the procedures is carried out using Mathematica (Wolfram, 1988) and is based on some extensions of the general saddlepoint approximation method of Easton and Ronchetti (1986). See also Barndorff-Nielsen and Cox (1989), Wang (1992), Jing, Feuerverger and Robinson (1994) and Hu and Kalbfleisch (2000). Our approach relies fundamentally on the fact that, in general, such statistics can be approximated arbitrarily well as finite linear combinations (of a certain form) of products of centered, normalized averages, and that in turn, the cumulants of such approximants are straightforward to evaluate to any desired accuracy.

In Sections 2–4, the main methodological and computational details are presented. Specifically, in Section 2 it is shown firstly that a single M-estimator may be approximated in the form (2.9)–(2.10) and consequently that any smooth function of M-estimators may be approximated as in (2.13)–(2.15). We then consider a typical approximation (2.14) of this type and in Section 3 discuss how its cumulant structure may be estimated to any

Keywords and phrases: bootstrap approximation; Edgeworth expansion; M-estimator; saddlepoint approximation; symbolic computation.

accuracy. These approximate cumulants are then used, in Section 4, as input to the general saddlepoint approximation (4.1)–(4.2) and its extensions; it turns out that only the first four cumulants are needed to achieve absolute error of the required accuracies.

In Section 5 we show how these methods can be adapted to handle the bootstrap distribution context, when the underlying distribution is not known. This involves replacing the unknown distribution function of the data by the empirical distribution when evaluating the cumulant expansions, and making an appropriate pivotal adjustment for the statistic of interest. Finally in Section 6 some typical applications of our algorithms are illustrated; specifically, we obtain the distributions associated with the t-like statistics formed from location and scale M-estimates based on MLE's of the Gaussian and Cauchy families under various distributional assumptions.

It is a key point of this paper that all of the methods described here are easily implemented for machine computation, and in fact depend primarily on only two key symbolic routines—one for computing expansions of general cumulants for smooth statistics, and one for computing expansions of roots of smooth functions. This has been carried out in Mathematica. In particular, it turns out that the computations for any particular problem require specification only of the functions defining the M-estimates and of the function defining the statistic of interest which is formed from these. Consequently our routines are extremely simple to apply to a wide range of statistics and underlying distributions. The Mathematica code is available at: `http://www.utstat.utoronto.ca/david/expand.dist.nb`

2. An approximation for functions of M-estimators

Suppose that $X_1, X_2, \ldots, X_n$ are iid from some distribution indexed by θ, with θ_0 being the assumed true value, and consider first a single M-estimator $\hat{\theta}$ assumed to be the solution of the equation

$$\tilde{\psi}_n(\theta) \equiv \frac{1}{n}\sum_{i=1}^{n} \psi(X_i, \theta) = 0. \tag{2.1}$$

Denoting expectation with respect to θ_0 by E_0, we shall require that

$$E_0 \psi(X, \theta_0) = 0 \tag{2.2}$$

and

$$E_0 \psi'(X, \theta_0) \neq 0$$

in order that $\hat{\theta}$ be $\sqrt{n}$-consistent for θ_0. (See, for example, Huber, 1977.) We wish to develop a particular sequence of increasingly accurate approximations to the root $\hat{\theta}$ of the equation (2.1). To this end, we carry out a

Taylor expansion of (2.1) up to kth order:

$$\tilde{\psi}_n(\theta) = \sum_{j=0}^{k} \frac{(\theta-\theta_0)^j}{j!} \tilde{\psi}_n^{(j)}(\theta_0) + R_{k+1} \tag{2.3}$$

which may also be written as

$$\tilde{\psi}_n(\theta) = \sum_{j=1}^{k} \frac{(\theta-\theta_0)^j}{j!} E_0\psi^{(j)}(X,\theta_0) + \sum_{j=0}^{k} \frac{(\theta-\theta_0)^j}{j!} \frac{Z_n^{(j)}(\theta_0)}{\sqrt{n}} + R_{k+1} \tag{2.4}$$

where

$$Z_n(\theta) \equiv \sqrt{n} \cdot \left[\frac{1}{n}\sum_{1}^{n}\{\psi(X_i,\theta) - E_\theta\psi(X,\theta)\}\right] \tag{2.5}$$

and

$$Z_n^{(j)}(\theta) \equiv \sqrt{n} \cdot \left[\frac{1}{n}\sum_{1}^{n}\{\psi^{(j)}(X_i,\theta) - E_\theta\psi^{(j)}(X,\theta)\}\right]. \tag{2.6}$$

Here the superscripts (j) on ψ represent differentiation with respect to θ, and we have omitted the $j = 0$ term in the first sum of (2.4) in view of (2.2). The remainder term appearing in (2.3) and (2.4) may be given as

$$R_{k+1} \equiv \frac{(\theta^* - \theta_0)^{k+1}}{(k+1)!} \tilde{\psi}_n^{(k+1)}(\theta^*) = O_P(n^{-(k+1)/2})$$

for some θ^* between θ and θ_0, and it is of the order indicated whenever $\theta^* = \theta_0 + O_P(n^{-1/2})$ provided only that $\psi^{(k+1)}(X,\theta)$ is uniformly integrable in a neighbourhood of θ_0. For the development to follow, it is important to note that the terms $Z_n^{(j)}(\theta_0)$ appearing in (2.4) are normalized, centered averages and hence are of *precise* order $O_P(1)$ provided $E|\psi^{(j)}|^2 < \infty$; i.e., they are of order $O_P(1)$ and are not of any lower order. (For a more exact definition, see Hall, 1992, pp. xii–xiii.)

We now describe the particular sequence of approximate roots to (2.1) which we require. Firstly, setting (2.4) equal to 0 and using $k = 1$ leads to

$$(\theta-\theta_0) \cdot E_0\psi'(X,\theta_0) + \frac{Z_n(\theta_0)}{\sqrt{n}} + (\theta-\theta_0)\frac{Z_n'(\theta_0)}{\sqrt{n}} + R_2 = 0 \tag{2.7}$$

and then (upon ignoring the third and fourth terms here) to our first approximate root $\hat{\theta}_1$ which we shall write as

$$\hat{\theta}_1 - \theta_0 = \frac{\delta_0}{\sqrt{n}},$$

where

$$\delta_0 = \frac{-Z_n(\theta_0)}{E_0\psi'(X,\theta_0)}. \tag{2.8}$$

Note that δ_0 is an average (and therefore depends upon n) and is $O_P(1)$ exactly. Further, since the M-estimator $\hat{\theta}$ satisfies (2.7)— with the third and fourth terms of (2.7) then being $O_P(n^{-1})$—and because $(\hat{\theta}_1 - \theta_0) \cdot E_0\psi'(X,\theta_0) + Z_n(\theta_0)/\sqrt{n} = 0$ we have (upon equating) that $(\hat{\theta} - \theta_0) \cdot E_0\psi' + O_P(n^{-1}) = (\hat{\theta}_1 - \theta_0) \cdot E_0\psi'$ and therefore that $\hat{\theta}_1 - \hat{\theta} = O_P(n^{-1})$.

Now our sequence of approximate roots $\hat{\theta}_l$ will be constructed inductively to have the form

$$\sqrt{n} \cdot (\hat{\theta}_{l+1} - \theta_0) = \delta_0 + \frac{\delta_1}{n^{1/2}} + \frac{\delta_2}{n} + \cdots + \frac{\delta_l}{n^{l/2}} \tag{2.9}$$

for $l = 1, 2, \ldots$, where the δ_i's are $O_P(1)$, depend upon n but not upon l, and $\hat{\theta}_l - \hat{\theta} = O_P(n^{-(l+1)/2})$. Further each δ_i will have the special form

$$\delta_i = \sum_{j=1}^{J} c_j \cdot Z_{j_1} \cdots Z_{j_{i+1}}, \tag{2.10}$$

i.e., a linear combination of products of exactly $i+1$ normalized, centered averages Z_j of the type (2.5) and (2.6). Thus suppose that $\hat{\theta}_l$ has been determined and that we seek $\hat{\theta}_{l+1}$. Then again setting (2.4) equal to 0, but this time with $k = l+1$ and $\theta = \hat{\theta}_{l+1} \equiv \hat{\theta}_l + \delta_l/n^{(l+1)/2}$ we are led to

$$\begin{aligned} &\frac{-\delta_l}{n^{l/2}} \cdot E_0\psi'(X,\theta_0) \\ &= \left\{ \sum_{j=1}^{l+1} \frac{(\hat{\theta}_l - \theta_0)^j}{j!} E_0\psi^{(j)}(X,\theta_0) \right. \\ &\qquad \left. + \sum_{j=0}^{l+1} \frac{(\hat{\theta}_l - \theta_0)^j}{j!} \frac{Z_n^{(j)}(\theta_0)}{\sqrt{n}} + O_P(n^{-(l+2)/2}) \right\} + R_{l+2} \end{aligned} \tag{2.11}$$

where the $O_P(n^{-(l+2)/2})$ term shown within the braces arose upon replacing $\hat{\theta}_{l+1}$'s with $\hat{\theta}_l$'s in the $j > 1$ terms of both sums there. Note also that the two sums within the braces then total to $O_P(n^{-(l+1)/2})$ by virtue of the previous iteration. Therefore, if we define

$$\delta_l = \frac{-n^{(l+1)/2}}{E_0\psi'(X,\theta_0)} \cdot \left\{ \sum_{j=1}^{l+1} \frac{(\hat{\theta}_l)^j}{j!} E_0\psi^{(j)}(X,\theta_0) + \sum_{j=0}^{l+1} \frac{(\hat{\theta}_l)^j}{j!} \frac{Z_n^{(j)}(\theta_0)}{\sqrt{n}} \right\}, \tag{2.12}$$

we will have $\delta_l = O_P(1)$ and also $\hat{\theta}_{l+1} \equiv \hat{\theta}_l + \delta_l/n^{l+1} = \hat{\theta} + O_P(n^{-(l+2)/2})$. Note also from (2.8) and (2.12) that the δ_i's, as defined here, will be linear combinations of products of $i+1$ centered, normalized averages as in (2.10). The above procedure is a symbolic analogue of a quasi Newton–Raphson iteration; see Andrews and Stafford (1993, 2000).

Next suppose that we have a multivariate M-estimator whose components $\hat{\theta}^1, \ldots, \hat{\theta}^k$, correspond to parameters $\theta^1, \ldots, \theta^k$ with true values $\theta_0^1, \cdots, \theta_0^k$, and that this multivariate M-estimator is based on the ψ-functions $\psi_1 \equiv \psi_1(X_i, \theta^1, \cdots, \theta^k), \cdots, \psi_k \equiv \psi_k(X_i, \theta^1, \ldots, \theta^k)$. The methods of the previous paragraph extend directly to allow us to write each $\hat{\theta}^j$ in the form (2.9), (2.10) where any of the terms Z_j's appearing in (2.10) can now be any of the terms (2.5), (2.6) corresponding to any of the ψ-functions $\psi_1, \ldots, \psi_k$. Now suppose further that we are interested in some smooth function, say g, of these M-estimators. (By 'smooth' we mean that g has the number of derivatives required for the expansion in equation (2.13) below to hold.) We shall view $\hat{g} \equiv g(\hat{\theta}^1, \ldots, \hat{\theta}^k)$ as being the statistic of interest, and $g \equiv g(\theta_0^1, \ldots, \theta_0^k)$ as the quantity of inferential interest. Then by substituting approximations which are the vector analogues of (2.9) (for each component of the M-estimator) into a straightforward Taylor expansion for g we are led to the fact that

$$\sqrt{n} \cdot (\hat{g} - g) = \nu_l + O_P(n^{-(l+1)/2}) \tag{2.13}$$

where

$$\nu_l = \gamma_0 + \frac{\gamma_1}{n^{1/2}} + \frac{\gamma_2}{n} + \cdots + \frac{\gamma_l}{n^{l/2}} \tag{2.14}$$

and the γ_i's are each $O_P(1)$ linear combinations of products of $i+1$ centered, normalized averages as in (2.10):

$$\gamma_i = \sum_{j=1}^{J} c_j \cdot Z_{j_1} \cdots Z_{j_{i+1}}. \tag{2.15}$$

In this way we are led to consider the members of the sequence of truncated expansions ν_l as approximants to the quantity $\sqrt{n} \cdot (\hat{g} - g)$. Further, the distributions of the approximants ν_l will themselves be approximated (in Section 3) by means of general saddlepoint approximations.

Of course the $O_P(n^{-(l+1)/2})$ error term introduced when using (2.14) to approximate $\sqrt{n} \cdot (\hat{g} - g)$ will affect the order of approximation to the distribution functions and their densities. To consider the relevant technical issues, suppose $U_n = V_n + \epsilon_n$. If $\epsilon_n \to 0$ in probability to some order, then the cdf's of V_n will approximate those of U_n absolutely and uniformly to the same order provided the cdf's of V_n converge to a continuous cdf. This simple result does not automatically carry over to density functions. However, if for some $\alpha > 0$ the $(V_n, n^\alpha \cdot \epsilon_n)$'s have densities which converge uniformly to a continuous bivariate density, then the densities of V_n may be used as uniform approximants of order $O(n^{-\alpha})$ to those of U_n. Such conditions (and their variants) are not very convenient to deal with, although clearly they may often be expected to hold under the smoothness typical in situations of practical interest. In any case, it should be noted that it is tail areas—and therefore distribution functions—that are generally of greatest interest in most applications.

3. Cumulant structure of the approximations.

To apply the general saddlepoint approximation methods of Section 4 to statistics ν_l of the form (2.14), it will be necessary to consider how their cumulants may be evaluated, and how the number l of terms appearing in (2.14) should be selected. Now in evaluating the kth order cumulant $\text{cum}(\nu_l, \cdots, \nu_l)$, we first make the substitutions (2.14) and (2.15) and apply linearity; this leads to an expression of the form

$$\sum_{m=k}^{k(l+1)} n^{-m/2} \sum_j d_j \cdot \text{cum}(Y_{j,1}, \ldots, Y_{j,k}) \tag{3.1}$$

where each of the Y's here is a product of Z's arising from (2.15), and the overall number of such Z's within the cumulant shown in (3.1) totals to m. Following McCullagh (1987, Section 3.1), we refer to m as the *degree* and to k as the *order* of this cumulant. In (3.1) note that the index m of the first sum ranges over the degree, i.e., over the total number of Z's appearing within the component cumulant shown there, and the second sum results upon collecting those terms which are of degree m. Next observe that if the kth order cumulant of degree m shown in (3.1) is decomposed into sums of products of *ordinary* cumulants of Z's in the usual way (as in McCullagh, Section 3.2) then the sum of the orders of the ordinary cumulants in any such product is always m. Further, by standard arguments, any such ordinary cumulant of Z's having order r can be seen to be $O(n^{-(r-2)/2})$, except of course the order $r = 1$ cumulants, which will be zero. This is based on the fact that our Z's are normalized averages of variables and assumes that these variables possess the required number of cumulants, a condition which is easily checked in any particular application. (Recall that our Z's derive ultimately from (2.5) and (2.6).) Using such arguments, it may be seen that to obtain the first cumulant of the approximating statistics (2.14) of interest correct to some order $O_P(n^{-c/2})$ say, only requires that we maintain the terms up to $l = c$ in the approximation (2.14). Similarly, to obtain like accuracy of the second cumulant also requires $l = c$, while for like accuracy of the third and fourth cumulants will require $l = c - 1$ and $l = c - 2$ respectively. In particular, using ν_2, i.e., $l = 2$, allows us to obtain the first four cumulants correct to orders $O_P(n^{-1})$, and at this level of accuracy, these evaluations would not change if we were to use ν_l with $l > 2$.

Some further results on the nature of the cumulants for statistics of the form (2.14) will in fact be required. In this regard, we refer to Theorem 2.1 of Hall (1992) and its proof. Our statistics ν_l are analogous to Hall's U_{nr}'s, except that the U_{nr}'s are built up from a finite number of averages, while the ν_l's are based not only on the ψ_j's but also on their derivatives, with the lth term in ν_l including terms up to the lth derivative of the ψ_j's. Nevertheless, the method of Hall's proof remains applicable and, letting $k_{j,n}/n^{(j-2)/2}$ here denote the jth cumulant of ν_l, we are led to obtain the expansions

$$\frac{k_{j,n}}{n^{(j-2)/2}} = n^{-(j-2)/2} \cdot \left(c_{j,1} + \frac{c_{j,2}}{n} + \frac{c_{j,3}}{n^2} + \cdots \right) \tag{3.2}$$

provided $j > 1$, where the $c_{j,i}$ are constants which are sums of products of ordinary cumulants. For $j = 1$ the result is

$$\frac{k_{1,n}}{n^{1/2}} = n^{-1/2} \cdot \left(c_{1,1} + \frac{c_{1,2}}{n} + \frac{c_{1,3}}{n^2} + \cdots \right). \tag{3.3}$$

Our notation here was selected to emphasize the typical orders of magnitude of the cumulants of the ν_l, and the defined quantities $k_{1,n}$, $k_{2,n}$, $k_{3,n}$, $k_{4,n}$, etc., will all be $O(1)$ in n when the required cumulants exist. Note that the values of the $c_{j,i}$ depend upon l, but become constant for $l \geq j$. Note also that for *standardized* statistics, we will have $c_{1,i} = 0$ for $i \geq 1$, and $c_{2,1} = 1$, $c_{2,i} = 0$ for $i > 1$. Consider then the quantity ν_l where l is considered now to be fixed. Then as a consequence of the foregoing, we may write

$$\operatorname{cum}(\nu_l) = \frac{k_{1,n}}{\sqrt{n}} = \frac{c_{1,1}}{\sqrt{n}} + O(n^{-3/2}) \tag{3.4}$$

$$\operatorname{cum}(\nu_l, \nu_l) = k_{2,n} = c_{2,1} + \frac{c_{2,2}}{n} + O(n^{-2}) \tag{3.5}$$

$$\operatorname{cum}(\nu_l, \nu_l, \nu_l) = \frac{k_{3,n}}{\sqrt{n}} = \frac{c_{3,1}}{\sqrt{n}} + O(n^{-3/2}) \tag{3.6}$$

$$\operatorname{cum}(\nu_l, \nu_l, \nu_l, \nu_l) = \frac{k_{4,n}}{n} = \frac{c_{4,1}}{n} + O(n^{-2}). \tag{3.7}$$

Note that $k_{1,n}/\sqrt{n}$ is the bias in $\hat{g}$ viewed as an estimator of g. The relations (3.4)–(3.7) will prove useful below.

4. General saddlepoint approximations

Now the general saddlepoint approximations that we make use of extend the one of Easton and Ronchetti (1986) whose approximation may be deduced formally from the usual saddlepoint approximation for $\bar{X}$ in a simple way.

In fact by an obvious substitution of $R(t/n)/n$ for the cumulant generating function (and a variable change $nt \to t$) we obtain the following form of a saddlepoint approximation for the density function of a general statistic $Y \equiv Y_n \equiv Y_n(X_1, \ldots, X_n)$ based on an iid sample $X_1, \ldots, X_n$ from some density f:

$$f_Y(y) \sim \left(\frac{1}{2\pi R''(t)}\right)^{1/2} \cdot \exp[R(t) - ty] \tag{4.1}$$

where the saddlepoint t is defined by

$$R'(t) = y \tag{4.2}$$

and $R(t)$ is the cumulant generating function of Y. For some general background, see for example Reid (1988) or Section 3.3 of Field and Ronchetti (1990). Easton and Ronchetti (1986) give general conditions which ensure that (4.1) will have uniform error of order $O(n^{-1})$ when $R(t)$ is suitably approximated; these conditions only require that Y possess a valid Edgeworth expansion of the required order. Although this uniform approximation error is *absolute* and not *relative*, we have found that in many situations these general saddlepoint approximations are more accurate than the corresponding Edgeworth expansions, likely because these saddlepoint approximations are density-like objects; a similar finding was reported by Easton and Ronchetti. See also Wang (1992). For some details concerning the nature of the error in Edgeworth approximation, see, for example, Theorem 2.2 of Field and Ronchetti (1990) and Theorem 2.2 of Hall (1992) and references cited therein.

For statistics of the form (2.14), it should be noted that the method of Easton and Ronchetti may be readily extended to include the usual $1 + O(n^{-1})$ correction factor to the saddlepoint approximation thus giving an (absolute) approximation which is in fact *correct* to $O(n^{-1})$. The technical argument for this relies fundamentally on the nature of the cumulant expansion (3.2) for the statistics ν_l and the results that follow from this concerning validity of the higher order Edgeworth expansions, as in Hall (1992, Sections 2.3 and 2.4). With valid Edgeworth expansions thus established, one may then argue as in Easton and Ronchetti, to conclude that if the saddlepoint approximation (4.1) is multiplied by the first correction factor

$$\eta_1 = 1 + \frac{1}{n} \cdot \left[\frac{1}{8}\left\{\frac{R^{(4)}(t)}{[R''(t)]^2}\right\} - \frac{5}{24}\left\{\frac{R'''(t)}{[R''(t)]^{3/2}}\right\}^2\right] \tag{4.3}$$

then the resulting approximation is improved so that the true density of Y now equals

$$f_Y(y) \cdot \eta_1 + O(n^{-3/2}). \tag{4.4}$$

(This correction factor is cited, for example, in Field and Ronchetti (1990).) Further, when $R(t)$ is not available (or does not exist) it may be replaced by a polynomial involving the first four cumulants only—and this is sufficient even in (4.3) for (4.4) to hold, provided y is restricted to lie within a normal range $|y - EY| \leq c/\sqrt{n}$. In fact these four cumulants may themselves be estimated, and need only be accurate up to and including the $O(n^{-1})$ terms in order for (4.4) to hold in that range. (Analogous extensions of these results which incorporate higher order correction factors, can also be established in this way.) Alternatively, note that such results may be established directly through term by term comparison of expansions of the Edgeworth and saddlepoint approximations; this stems ultimately from the fact that the saddlepoint approximation can be related to the Edgeworth via tilting. These results will then hold for statistics of the form ν_l provided only that all moments (of the ψ-functions and their derivatives) required for the Edgeworth expansion to be valid exist.

Now using the notation of (3.2)—where the cumulants of Y are given by $k_{1,n}/\sqrt{n}$, $k_{2,n}$, $k_{3,n}/\sqrt{n}$, $k_{4,n}/n, \dots$, we have

$$R(t) = \frac{k_{1,n}}{\sqrt{n}} \cdot t + k_{2,n} \cdot \frac{t^2}{2!} + \frac{k_{3,n}}{\sqrt{n}} \cdot \frac{t^3}{3!} + \frac{k_{4,n}}{n} \cdot \frac{t^4}{4!} + \cdots . \tag{4.5}$$

so that the saddlepoint equation (4.2) then becomes

$$k_{2,n} \cdot t + k_{3,n} \cdot \frac{t^2}{2} + k_{4,n} \cdot \frac{t^3}{6} + \cdots = y - \frac{k_{1,n}}{\sqrt{n}} \tag{4.6}$$

and this may be solved iteratively for t in the manner of (2.12) to yield an expression of the form (2.9) for t correct to any desired order. From (4.6) and the fact that y is restricted to the normal range we see that t will have order $O(n^{-1/2})$ and therefore the first correction factor to the saddlepoint approximation may be taken as

$$\eta_1 = 1 + \frac{1}{n} \cdot \left[\frac{3c_{4,1}}{c_{2,1}^2} - \frac{5c_{3,1}^2}{c_{2,1}^3} \right] \tag{4.7}$$

which has full $O(n^{-1})$ absolute accuracy. (Here the $c_{j,i}$ are as in (3.2).) With this correction factor, the general saddlepoint approximation discussed here will typically be accurate to within $O(n^{-3/2})$ uniformly on finite intervals of the normal range. (For unrestricted uniform and absolute convergence, the polynomial approximation to $R(t)$ must be adjusted; see, for example, Wang (1992).) Note that to take advantage of this level of accuracy requires that the $O(n^{-1})$ term in $\sqrt{n} \cdot t$ be estimated correctly, and that ν_l of equation (2.14) be used with $l \geq 2$.

5. The bootstrap application

Suppose now that the underlying distribution of the X's is not known, so that the cumulants involving the $Z_n^{(j)}(\theta_0)$'s of (2.5), (2.6) cannot be evaluated either analytically or numerically—a situation in which bootstrap Monte Carlo would be entertained. In that case the "normalized" cumulant quantities $k_{1,n}, k_{2,n}, k_{3,n}, k_{4,n}$ of (4.5) and (3.2) may be estimated empirically from our sample to the usual $O_P(n^{-1/2})$ statistical accuracy by means of replacing theoretical (i.e., population) moments in expansions of the cumulants by their corresponding sample averages. Substitution of these empirical cumulants directly into (4.1) is then seen to lead to an approximation having error $O_P(n^{-1/2})$. This, of course, is not of sufficient accuracy to take advantage of the saddlepoint approximation method.

In fact, in view of (3.4)–(3.7), an examination of (4.1), (4.2), (4.5) and (4.7) reveals that for any fixed y, the second order saddlepoint approximation $\tilde{f}_Y(y) \equiv f_Y(y) \cdot \eta_1$ to a statistic of the type $Y = \nu_l$ has the form

$$\tilde{f}_Y(y) = f_Y(y) \cdot \eta_1 = h\left(\frac{c_{1,1}}{\sqrt{n}}, c_{2,1}, \frac{c_{2,2}}{n}, \frac{c_{3,1}}{\sqrt{n}}, \frac{c_{4,1}}{n}\right) \tag{5.1}$$

where h is a differentiable function. This *computable* function has an empirical version:

$$\widehat{\tilde{f}}_Y(y) = f_Y(y) \cdot \eta_1 = h\left(\frac{\hat{c}_{1,1}}{\sqrt{n}}, \hat{c}_{2,1}, \frac{\hat{c}_{2,2}}{n}, \frac{\hat{c}_{3,1}}{\sqrt{n}}, \frac{\hat{c}_{4,1}}{n}\right) \tag{5.2}$$

in which the quantities $\hat{c}_{1,1}$, $\hat{c}_{2,1}$, $\hat{c}_{2,2}$, $\hat{c}_{3,1}$, $\hat{c}_{4,1}$ are obtained from the same expansions as those for $c_{1,1}$, $c_{2,1}$, $c_{2,2}$, $c_{3,1}$, $c_{4,1}$, but by replacing the ordinary cumulants appearing in these expansions by their empirical versions, each of which have the usual $O_P(n^{-1/2})$ statistical error of estimation. It therefore follows that

$$\widehat{\tilde{f}}_Y(y) - \tilde{f}_Y(y) = O_P(n^{-1/2}) \tag{5.3}$$

as was asserted.

On the other hand, suppose it were *known* that $c_{1,1} = 0$ and $c_{2,1} = 1$, i.e., that

$$k_{n,1} = O(n^{-1}) \tag{5.4}$$

$$k_{n,2} = 1 + O(n^{-1}) \tag{5.5}$$

and suppose that these *known* values of 0 and 1 were used in place of $\hat{c}_{1,1}$ and $\hat{c}_{2,1}$ in (5.2). In that case we would clearly have

$$\widehat{\tilde{f}}_Y(y) - \tilde{f}_Y(y) = O_P(n^{-1}). \tag{5.6}$$

Further, note that if the distribution of $\hat{g}$ was approximately symmetric so that $c_{3,1} = 0$, i.e., $k_3 = O(n^{-1})$ then the resulting error would be only $O_P(n^{-3/2})$. The conditions (5.4) and (5.5) however may be attained, and to achieve this we need only apply the saddlepoint approximations to our statistic after it has been studentized to an adequate degree of approximation:

$$\frac{\nu_2 - \hat{c}_{1,1}/\sqrt{n}}{\sqrt{\hat{c}_{2,1}}}. \tag{5.7}$$

This result should be compared to the usual bootstrap result which states that the sampling distribution of a quantity such as $(\hat{\theta}-\theta)/s_\theta$ differs from its bootstrap distribution by $O_P(n^{-1})$. Note that in our case, the denominator term $\hat{c}_{2,1}$ has an explicitly specified form, while the centering term $\hat{c}_{1,1}/\sqrt{n}$ in our numerator automatically incorporates a bias correction; this may be expected to result in a statistic which can be approximated more precisely by its bootstrap analogue (i.e., by the plug-in estimate) than would be the case in the ordinary bootstrap context.

6. A numerical example

As an example of the application of the methods that have been discussed in this paper, in this section we show how they may be used to approximate the distributions of the t-like statistics associated with two different location-scale M-estimators—namely those arising from the MLE's in the Gaussian and in the Cauchy location-scale families. Thus let μ and σ respectively represent the location and scale parameters associated with a random variable X and $\hat{\mu}$ and $\hat{\sigma}$ be the corresponding M-estimators. The parameters are considered as being defined via the functions $\psi_1(Y)$ and $\psi_2(Y)$, where $Y = (X - \mu)/\sigma$, by the equations $E[\psi_i(Y)] = 0$, while the estimates based on a sample $\{x_j\}$ are defined by the equations $\text{Avg}[\psi_i(y_j)] = 0$, $i = 1, 2$, where $y_j = (x_j - \mu)/\sigma$. Specifically, in the Gaussian case we use:

$$\psi_1(y) = y$$
$$\psi_2(y) = y^2 - 1$$

while in the Cauchy case we use:

$$\psi_1(y) = y/(1 + y^2)$$
$$\psi_2(y) = 1 - 2/(1 + y^2).$$

For use in inference concerning the hypothesis $\mu = 0$, the t-like statistic associated with $\hat{\mu}$ and $\hat{\sigma}$ is defined as $T = n^{1/2}\hat{\mu}/\hat{\sigma}$.

The expansions for T involve expected values of the $\psi_i(Y)$ and their derivatives. If X is a random variable with a given density function, these expectations may be evaluated directly by one-dimensional, numerical, integration. If X arises from a discrete empirical cdf, the expectations are obtained as averages involving the $\psi_i(\hat{y}_j)$ and their derivatives where $\hat{y}_j = (x_j - \hat{\mu})/\hat{\sigma}$.

Now let T^* denote the version of T that has been studentized in the manner of (5.7). The approximate cumulants of T^* were computed simply using the approximate relation

$$K_{T^*}(u) \doteq \hat{K_T}\left(\frac{u}{\widehat{\text{Var}(T)}^{1/2}}\right) - \frac{\widehat{E(T)} \cdot u}{\widehat{\text{Var}(T)}^{1/2}}$$

which is derived by a straightforward differential argument.

We considered three cases: the Gaussian MLE's with Gaussian data, and the Cauchy MLE's with both Gaussian and Cauchy data. For each of these cases the general saddlepoint approximation to the density—including the correction factor (4.7)—was calculated for $x = \pm 1, \pm 2$. The value 2 corresponds approximately to the 2% point of the distributions. Also, a random sample of size $n = 40$ was generated for each of the three cases considered, and used to approximate the ordinary cumulants, as in Section 5.

Table 1 summarizes the case of the Gaussian MLE with Gaussian data. The t-like statistic based on the Gaussian MLE is just $((n-1)/n)^{1/2}$ times the usual t-statistic; hence the known t-distribution was used to compute the density of T^* to assess the adequacy of the approximation here. The table also gives the results of the general saddlepoint approximation and the (left and right tail) bootstrap saddlepoint density approximations from the sample of size $n = 40$.

Table 2 concerns the Cauchy MLE with Cauchy distributed data. Here, in order to assess the adequacy of the approximation, the exact density of the T^* was approximated by a saddlepoint approximation. The cumulants of T^* for this approximation were estimated by means of a simple Monte Carlo of size 10,000. The errors from this saddlepoint approximation and computation of the moments are negligible compared with the variation resulting from the sample of size 40.

Table 1. Gaussian MLE, Gaussian Distribution

x	Exact Density	General SP App	Bootstrap SP Left	Bootstrap SP Right
0	0.4070	0.4064	0.4081	0.4081
1	0.2387	0.2392	0.2337	0.2439
2	0.0522	0.0533	0.0555	0.0509

Table 2. Cauchy MLE, Cauchy Distribution

x	Estimated Density	General SP App	Bootstrap SP Left	Bootstrap SP Right
1	0.2457	0.2427	0.2450	0.2219
2	0.0463	0.0545	0.0453	0.0690

Table 3. Cauchy MLE, Gaussian Distribution

x	Estimated Density	General SP App	Bootstrap SP Left	Bootstrap SP Right
1	0.2349	0.2374	0.2378	0.1876
2	0.0487	0.0539	0.0068	0.0798

Finally, Table 3 summarizes the case of the Cauchy MLE under Gaussian data. The exact density and distribution of the T^* statistic were estimated by simple Monte Carlo of size 10,000.

The results of Tables 1–3 show that the general saddlepoint approximation based on approximated and estimated cumulants for sample size $n = 40$ is exceedingly good for the Gaussian MLE and Gaussian distribution. For the Cauchy MLE and Cauchy distribution, the approximation has a useful accuracy for approximating p-values where typically less than one significant digit is required. (Real experiments—at least those involving human subjects—typically involve experimental biases of sampling and observation at least of this order.) The case of the Cauchy MLE and Gaussian data was selected here because it is known to be a more difficult approximation problem. (In this case the mismatch between the M-estimator and data distribution leads to an inherent inefficiency—in the Fisher sense—and hence to a lower effective sample size.) Here the approximation is at the edge of usefulness for this sample size. For this case, the estimation of third and fourth order cumulants is imprecise and larger samples sizes are required.

The saddlepoint approach for the estimation of densities may be paralleled for the estimation of tail areas. The Lugannani and Rice (1980) formula may be used with approximated and estimated cumulants replacing exact cumulants. The similarity in the order of accuracy of the saddlepoint density approximation and Lugannani-Rice formulae for exact cumulants and the known connection between these approximations (see, for example, Daniels, 1987) suggest that similar absolute precision will be achieved for approximated cumulants. Likewise, the methods of this paper may be adapted to implement saddlepoint approximation formulae like those of Tierney, Kass and Kadane (1989) or Diciccio and Martin (1991).

Acknowledgements. This work was supported in part by grants from the Natural Sciences and Engineering Research Council of Canada.

REFERENCES

Andrews, D.F. and Stafford, J.E. (1993). Tools for the symbolic computation of asymptotic expansions. *J. Roy. Statist. Soc. Ser.* B 55, 613–622.

Andrews, D.F. and Stafford, J.E. (2000). *Symbolic Computation for Statistical Inference.* Oxford Statist. Sci. Ser., vol. 21, Oxford Univ. Press, Oxford.

Barndorff-Nielsen, O. and Cox, D.R. (1989). *Asymptotic Techniques for use in Statistics.* Monogr. Statist. Appl. Probab. Chapman and Hall, New York.

Daniels, H.E. (1987). Tail probability approximations. *Internat. Statist. Rev.* 55, 37-48.

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

DiCiccio, T.J. and Martin, M.A. (1991). Approximation of marginal tail probabilities for a class of smooth functions with applications to Bayesian and conditional inference. *Biometrika* 78, 891–902.

DiCiccio, T.J., Martin, M.A. and Young, G.A. (1992a). Fast and accurate approximate double bootstrap confidence intervals. *Biometrika* 79, 285–295.

DiCiccio, T.J., Martin, M.A. and Young, G.A. (1992b). Analytical approximation for iterated bootstrap confidence intervals. *Stat. Comput.* 2, 161–171.

Easton, G.S. and Ronchetti, E. (1986). General saddlepoint approximations with applications to L statistics. *J. Amer. Statist. Assoc.* 81, 420–430.

Field, C. and Ronchetti, E. (1990). *Small Sample Asymptotics.* IMS Lecture Notes Monogr. Ser., vol. 13, IMS, Hayward, CA.

Hall, P. (1992) *The Bootstrap and Edgeworth Expansion.* Springer Ser. Statist. Springer-Verlag, New York.

Hu, F. and Kalbfleisch, J.D. (2000). The estimating function bootstrap (with discussion). *Canad. J. Statist.* 28, 449–499

Huber, P.J. (1977). *Robust Statistical Procedures.* CBMS-NSF Regional Conf. Ser. in Appl. Math., vol. 27, SIAM, Philadelphia, PA.

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

Lugannani, R. and Rice, S. (1980). Saddlepoint approximation for the distribution of a sum of independent random variables. *Adv. in Appl. Probab.* 12, 475–490.

McCullagh, P. (1987). *Tensor Methods in Statistics.* Monogr. Statist. Appl. Probab. Chapman and Hall, New York.

Tierney, L., Kass, R.E., and Kadane, J.B. (1989). Approximate marginal densities of nonlinear functions. *Biometrika*, 76, 425–433; Correction (1991), 78, 233–234.

Reid, N. (1988). Saddlepoint methods in statistical inference (with discussion). *Statist. Sci.* 3, 213–227.

Wang, S. (1992). General saddlepoint approximations in the bootstrap. *Statist.Probab. Lett.* 13, 61–66.

Wolfram, S. (1988). *Mathematica: A System for Doing Mathematics by Computer.* Addison-Wesley, New York.

D.F. Andrews
Department of Statistics
University of Toronto
100 St. George St.
Toronto, ON M5S 3G3
Canada
david@utstat.toronto.edu

A. Feuerverger
Department of Statistics
University of Toronto
100 St. George St.
Toronto, ON M5S 3G3
Canada
andrey@utstat.toronto.edu