Likelihood and higher order approximations to tail areas: a review and annotated bibliography

N. Reid

Abstract

Recent developments in higher order asymptotic theory for statistical inference have emphasized the fundamental role of the likelihood function in providing accurate approximations to cumulative distribution functions. This paper summarizes the main results, with an emphasis on classes of problems for which relatively easily implemented solutions exist. A survey of the literature indicates the large number of problems solved and solvable by this method. Generalizations and extensions, with suggestions for further development, are considered.

Key Words: approximation, Laplace expansion, parametric inference, nonparametric inference, saddlepoint method, tail area
AMS Subject Classification: 62-02, 62E20

1 Introduction

An active area of development in theoretical statistics is the area of so called 'higher order asymptotics', in which an asymptotic expansion for the density or distribution function of a statistic of interest is obtained. The purpose in obtaining such an expansion is to use the first two or three terms in the expansion to provide an approximation to the density or distribution function. Although it is not guaranteed that this will provide a better finite sample approximation to the true distribution function than that based on just the first term, in a great many examples some of these approximations have turned out to be surprisingly accurate.

Consider approximation to the density of the mean, $\bar{X}$, of n independent, identically distributed random variables X_i. Under mild conditions on the true distribution of the X_i's, the limiting distribution of $\bar{X}$, suitably normalized, is normal, and the normal density provides an approximation to the exact density function of $\bar{X}$ with relative error $O(n^{-1/2})$. In contrast to this, the *saddlepoint approximation* to the density of $\bar{X}$ provides an approximation to the exact density with relative error $O(n^{-3/2})$, which from one point of view means it incorporates the first three terms of an asymptotic expansion in powers of $n^{-1/2}$, rather than just the first term. The saddlepoint approximation to the density of $\bar{X}$ is derived in Daniels (1954), and reviewed in Reid (1988).

In both the normal and saddlepoint approximation, the relative errors are not uniform in $\bar{x}$, the value at which the density function is evaluated, but an advantage of the saddlepoint approximation beyond its improved relative error is that the error is nearly uniform in $\bar{x}$, which means that in practice the relative errors remain small relatively far out in the tails. The two-term Edgeworth expansion for the density of $\bar{X}$, by contrast, has a relative error of $O(n^{-3/2})$ that fluctuates substantially with $\bar{x}$, and in finite samples the Edgeworth approximation tends to perform poorly in the tails of the distribution.

Another feature of the saddlepoint approximation is that if the distribution of the X_i's is a member of an exponential family, then the saddlepoint approximation can be reformulated entirely in terms of the likelihood function. In its likelihood formulation, this approximation can be shown to apply to more general families of distributions, where it provides an approximation (which is sometimes exact) to the density of the maximum likelihood estimator.

For practical use in statistical inference, we are more often interested in approximating the cumulative distribution function of a statistic, in order to compute a p-value or confidence coefficient. The saddlepoint approximation to the cumulative distribution function of $\bar{X}$, derived by Lugannani and Rice (1980) and reviewed in Daniels (1987), also has a likelihood formulation in exponential families that turns out to apply much more generally.

By studying the likelihood based formulation of this approximation to the cumulative distribution function, we can obtain very accurate approximations in problems of practical importance, but in addition we gain considerable insight into the role of the likelihood function in statistical inference.

In Section 2 I summarize the recent results on using the likelihood function to approximate cumulative distribution functions, and provide an introduction to the literature on this topic. Section 3 illustrates the versatility of the likelihood approach by giving examples of problems solved and solvable by this approach. Section 4 discusses generalizations and extensions, with suggestions for further development. In the remainder of this section I briefly outline how the saddlepoint approximation for the density and distribution function of the sample mean can be re-expressed using the likelihood function.

Assume that $X_1, \ldots, X_n$ are independent and identically distributed from a density function $f(\cdot)$ on R^p. Denote by $K(\phi) = \log[E\{\exp(\phi^T X)\}]$ the cumulant generating function for X_i. The saddlepoint approximation to the density of $\bar{X}$ is

$$\hat{f}_{\bar{X}}(\bar{x}) = c\{\frac{n}{|K''(\hat{\phi})|}\}^{1/2} \exp\{nK(\hat{\phi}) - n\hat{\phi}^T \bar{x}\} \tag{1}$$

where $\hat{\phi}$ is defined by $K'(\hat{\phi}) = \bar{x}$, and c is a normalizing constant for the right hand side of (1) that can be shown to be $(2\pi)^{-p/2}\{1+O(n^{-1})\}$. The relative error in (1) is $O(n^{-3/2})$, meaning $f_{\bar{X}}(\bar{x}) = \hat{f}_{\bar{X}}(\bar{x})\{1 + O(n^{-3/2})\}$ for values of $\bar{x}$ in the normal deviation region $\sqrt{n}|\bar{x} - \mu| \leq \delta$ for some fixed constant δ. The relative error is not uniform in $\bar{x}$, but to $O(n^{-1})$ the relative error is uniform in a large deviation region $|\bar{x} - \mu| \leq \delta$. A detailed discussion of the error properties of the saddepoint approximation to the denisty of the sample mean is given in Kolassa (1994, Ch. 4). Uniformity is discussed in Daniels (1954) and Jensen (1988) as well.

When $p = 1$ the saddlepoint approximation to the distribution function of $\bar{X}$ is

$$\hat{F}_{\bar{X}}(\bar{x}) = \Phi(r) + \varphi(r)(\frac{1}{r} - \frac{1}{q}), \tag{2}$$

where Φ and φ are the standard normal distribution and density function, respectively, and

$$r = \mathrm{sign}(q)\{2nK(\hat{\phi})\}^{1/2} \tag{3}$$
$$q = \hat{\phi}\{nK''(\hat{\phi})\}^{1/2}. \tag{4}$$

The relative error in (2) is also $O(n^{-3/2})$. The derivation of (2) using saddlepoint techniques is reviewed in Daniels (1987). Uniformity properties of the approximation are discussed in Jensen (1988).

In the special case that the underlying density f is a canonical exponential family model,

$$f(x) = \exp\{\theta^T x - k(\theta) - d(x)\}, \tag{5}$$

then $\bar{X}$ is sufficient for θ, the maximum likelihood estimator $\hat{\theta}$ is a one to one function of $\bar{X}$ defined by $k'(\hat{\theta}) = \bar{X}$, and the cumulant generating function is $K(\phi) = k(\theta + \phi) - k(\theta)$. Putting these facts together and transforming (1) to an approximation for the density of $\hat{\theta}$ gives

$$\hat{f}(\hat{\theta}; \theta) = c|j(\hat{\theta})|^{1/2} \exp\{\ell(\theta) - \ell(\hat{\theta})\} \tag{6}$$

where $\ell(\theta) = \theta^T n\bar{x} - nk(\theta)$ is the log-likelihood function based on the sample $x_1, \ldots, x_n$, and $j(\theta)$ is the Fisher information function $-\ell''(\theta)$, in this case simply equal to $nc''(\theta)$.

The saddlepoint approximation to the cumulative distribution function of $\hat{\theta}$ is again given by (2), and the likelihood formulation of the quantities r and q is

$$r = \mathrm{sign}(q)[2\{\ell(\hat{\theta}) - \ell(\theta)\}]^{1/2} \tag{7}$$
$$q = (\hat{\theta} - \theta)\{j(\hat{\theta})\}^{1/2}. \tag{8}$$

Note that r is the signed square root of the log-likelihood ratio statistic, and q is the standardized maximum likelihood estimate, or Wald statistic. To first order both r and q have standard normal distributions, and the correction term involving $\varphi(r)$ in (2) provides the improvement from $O(n^{-1/2})$ to $O(n^{-3/2})$. Since r is a monotone function of $\hat{\theta}$ in this setting, (2) also provides an approximation to the distribution of r.

At the value $\hat{\theta} = \theta$, approximation (2) breaks down, as $r = q = 0$. It should be replaced at that point by its limiting value, $\hat{F}(\hat{\theta}) = (1/2) + \rho_3/(6\sqrt{2\pi n})$, where $\rho_3 = n^{-1}\ell'''(\hat{\theta})/\{n^{-1}j(\hat{\theta})\}^{3/2}$. In practical work this means that the approximation given by (2) can be unstable near $\hat{\theta} = \theta$, and some care must be taken if the distribution function is to be computed over its entire domain. For many applications it is only of interest to compute $\hat{F}$ in the left or right tail of the distribution, well away from θ.

The likelihood formulation of (1) was pointed out in Daniels (1958). Barndorff-Nielsen (1980, 1983) noted that the right hand side of (6) provides the exact distribution of the maximum likelihood estimate in transformation families. The right hand side of (6) is a special case of what has come to be known as Barndorff-Nielsen's approximation, or the p^* approximation, and was the main focus of the review in Reid (1988). It was introduced in Barndorff-Nielsen (1980, 1983) and has been discussed in many papers by Barndorff-Nielsen and others since then. A relatively complete list of references is given in Barndorff-Nielsen and Cox (1994).

The likelihood formulation of (2) (or more precisely of (3) and (4)) also leads to versions of the Lugannani and Rice approximation that apply in more general settings, and this is the focus of the next section.

2 Tail areas from likelihood

2.1 Scalar case

Assume first that X has a distribution $f(x;\theta)$, where both x and θ are scalar valued or one-dimensional. Although this is a very special case, the structure of the solution lends insight to generalizations of the approximation to more realistic problems. As in Section 1, the log-likelihood function will be denoted by $\ell(\theta)$, or $\ell(\theta;x)$ when it is necessary to indicate its dependence on the data.

Writing $F(x;\theta)$ for the cumulative distribution function of X, we have the approximation

$$F(x;\theta) \doteq \Phi(r) + \varphi(r)(\frac{1}{r} - \frac{1}{t}) \tag{9}$$

where

$$r = \text{sign}(t)[2\{\ell(\hat\theta) - \ell(\theta)\}]^{1/2} \tag{10}$$
$$t = \{\dot\ell(\hat\theta) - \dot\ell(\theta)\}k^{-1}(\hat\theta)\{j(\hat\theta)\}^{1/2}. \tag{11}$$

Approximation (9) is essentially the same as (2), and r is again the signed square root of the log-likelihood ratio statistic. In the definition of t, $\dot\ell(\theta) = \partial\ell(\theta;x)/\partial x$ is the derivative of the log-likelihood with respect to the data x, and $k(\theta) = \partial\dot\ell(\theta)/\partial\theta$ is the mixed derivative. As before $\hat\theta$ is the maximum likelihood estimate, and $j(\hat\theta)$ is the observed Fisher information. In this simple model, $\hat\theta$ is a one-to-one function of X, and differentiation of the likelihood equation shows that $d\hat\theta/dx = j^{-1}(\hat\theta)k(\hat\theta)$. Thus an alternative expression for t is given by

$$\{\ell_{;\hat\theta}(\hat\theta) - \ell_{;\hat\theta}(\theta)\}\{j(\hat\theta)\}^{-1/2} \tag{12}$$

where $\ell_{;\hat\theta}(\theta) = \partial\ell(\theta;x)/\partial\hat\theta$.

The expression given in (11) was derived in Fraser (1990), using a tangent exponential model approximation to a general one-dimensional model. The version using (12) is due to Barndorff-Nielsen (1988, 1990a), and was derived by integrating the p^* approximation directly. Actually Barndorff-Nielsen derived an alternate approximation, accurate to the same order of accuracy, that incorporates the correction term into the quantile of the normal cumulative distribution function, giving

$$F(x;\theta) \doteq \Phi(r^*) \tag{13}$$

where

$$r^* = r + \frac{1}{r}\log\frac{t}{r}. \tag{14}$$

This version is usually called the r^* approximation. The equivalence of (13) and (9) can be verified by expanding r^* about r, and was rigorously established in Jensen (1992) for exponential family models. It also follows from expansions given in Barndorff-Nielsen (1986), as indicated in DiCiccio and Martin (1993).

In a sample of size n from a one-parameter exponential family, as defined at (5) with $p = 1$, the one dimensional variable is $\bar{X}$, which is minimal sufficient, as mentioned in Section 1. In this case $\ell(\theta;\hat{\theta}) = n\{\theta k'(\hat{\theta}) - k(\theta)\}$, and t is readily computed as the standardized maximum likelihood estimate (8).

In a sample of size n from a one-parameter location family, $f(x;\theta) = f(x-\theta)$, the one dimensional variable can be taken as $\hat{\theta}$ (or any other location estimate of θ), and (9) gives an approximation to the conditional distribution of $\hat{\theta}$, given the location model ancillary statistic $a = (a_1, \ldots, a_n)$, $a_i = x_i - \hat{\theta}$. In this case $\ell(\theta;\hat{\theta},a) = \sum \log f(a_i + \hat{\theta} - \theta)$, $\partial\ell(\theta;\hat{\theta},a)/\partial\hat{\theta} = -\partial\ell(\theta)/\partial\theta$, and t simplifies to the standardized score statistic

$$\ell'(\theta)\{j(\hat{\theta})\}^{-1/2}. \tag{15}$$

Note that although approximation (9) was derived under the assumption that both the parameter and the data were one-dimensional, the applications we have in mind are those in which the model permits a reduction of the data to a one-dimensional sufficient statistic. In the exponential case the reduction is via minimal sufficiency, and in the location model case the reduction is via ancillarity.

Fraser (1991) provides a general introduction to these one parameter models and approximations, with particular emphasis on using the approximation to compute p-values for arbitrary θ values, thus generating a set of nested confidence intervals. Daniels (1987) gives an excellent discussion of the derivation of the exponential family version from the saddlepoint approximation. Barndorff-Nielsen (1990ab, 1991) discusses the derivation of (13) from the p^* approximation.

Since r and r^* are, asymptotically, monotone functions of x, approximations (9) and (13) also give approximations to the distribution of r and r^*. An alternative approximation to the distribution of r, that was derived earlier by Lawley (1956) (see also McCullagh, 1984), is the mean and variance corrected version

$$F(r) \doteq \Phi\{(r-m)/s\}$$

where $\mathrm{E}(r) = m + O(n^{-1})$ and $\mathrm{var}(r) = s^2 + O(n^{-3/2})$. This approximation also has relative error $O(n^{-3/2})$, but in numerical work versions (9) or (13) tend to be much more accurate.

Table 1: Approximations (9) and (13) for a sample of size 1 from the standard Cauchy density (see also Barndorff-Neilsen, 1991; Fraser, 1990).

$\hat{\theta}$	-100	-70	-10	-1
(9)	0.0028	0.0040	0.0281	0.2322
(13)	0.0015	0.0023	0.0214	0.2283
exact	0.0032	0.0045	0.0317	0.2500

Table 2: Approximation (9) for a sample of size 1 from the standard logistic density (taken from Fraser and Reid, 1993).

$\hat{\theta}$	2	3	4	5	6
(9)	0.2669	0.1177	0.0466	0.0179	0.0067
exact	0.2689	0.1192	0.0474	0.0180	0.0067

This approximation is very closely related to the Bartlett correction, discussed in Section 4.2.

There are a large number of examples in the literature of the accuracy of the tail area approximations in one parameter families. To give a flavour of these results, Tables 1 and 2 illustrate the tail area approximations for a sample of size 1 from the Cauchy and the logistic densities respectively.

Barndorff-Nielsen (1990a) compares the r^* and the Lugannani and Rice approximations for exponential, inverse Gaussian and vonMises distributions and Barndorff-Nielsen (1991) compares them for the Cauchy. Barndorff-Nielsen and Chamberlin (1991) consider the location log-Gamma, the inverse Gaussian, and a specially constructed (2,1) curved exponential family. Fraser (1990) treats the location log-Gamma, the Gamma, the logistic, and the Cauchy. DiCiccio, Field and Fraser (1990) compare the Lugannani and Rice approximation to the mean and variance corrected version for Fisher's hyerbola model which is a (2,1) curved exponential family. DiCiccio and Martin (1993) consider an exponential model with censoring; although an r^* approximation accurate to $O(n^{-3/2})$ is not available, a mean and variance corrected version due to Jensen (1987) works quite well. Butler, Huzurbazar and Booth (1992a) use a one parameter exponential family to approximate the distribution of the generalized variance statistic. Fraser and Reid (1993) compare Lugannani and Rice to r^* in a tilted and shifted logistic model (which has one parameter but is neither an exponential nor a location model). Wang (1992) considers the sample mean from a Beta distribution, and the bootstrap distribution of the estimate of ρ in an AR(1) model. Numerical illustrations of saddlepoint approximations appear in most of the papers on this topic; for tail area approximations Daniels (1987) is particularly interesting. Kolassa (1994), especially in the exercises, considers several numerical examples, and provides Mathematica code for computing many of them.

2.2 Nuisance parameters

The essential aspect of the simplification of the previous subsection is that the parameter of interest and the minimal sufficient statistic have the same dimension. As discussed in Reid (1988), this is required for general application of the p^* approximation. In canonical exponential models this is already the case. In location models, or more generally transformation models, a reduction in dimension is available by conditioning on the maximal ancillary statistic. A more detailed discussion of the role of ancillarity and sufficiency is given in Reid (1995b).

To use something like the Lugannani and Rice approximation in the multiparameter setting, it is necessary to isolate a scalar parameter of interest, as the approximation is inherently one-dimensional. In this Section we assume that $\theta = (\psi, \lambda)$, where θ is p-dimensional and λ is $(p-1)$-dimensional. Exponential and transformation models play a special role in this case as well, as exact elimination of nuisance parameters is available by conditioning and marginalizing, respectively.

Example 2.1: Exponential family models

Assume that X is the minimal sufficient statistic for a sample of size n from a canonical exponential family on R^p; i.e. $f(x;\theta) = \exp\{\theta^T x - k(\theta) - d(x)\} = \exp\{\psi x_1 + \lambda^T x_{(2)} - k(\psi,\lambda) - d(x)\}$. A one dimensional likelihood for ψ is available by conditioning, as $f(x_1|x_{(2)})$ depends only on ψ, and not on λ. It can be shown that

$$\begin{aligned} \ell_c(\psi) &= \log f(x_1|x_{(2)};\psi) & (16)\\ &= \ell(\psi,\hat{\lambda}_\psi) + \frac{1}{2}\log|j_{\lambda\lambda}(\psi,\hat{\lambda}_\psi)| + O_p(n^{-3/2}) & (17)\end{aligned}$$

where $\ell(\psi,\lambda) = \ell(\theta)$ is the log-likelihood function from the model, $\hat{\lambda}_\psi$ is the restricted maximum likelihood estimate of λ for fixed ψ, and $j_{\lambda\lambda}(\psi,\lambda) = -\partial^2\ell(\psi,\lambda)/\partial\lambda\lambda^T$ is the $(p-1)\times(p-1)$ submatrix of the observed Fisher information matrix. The exponential family version of the approximation to the cumulative distribution function can be applied to this conditional log-likelihood, and the result is

$$F(r_c) = F(\hat{\psi}_c) \doteq \Phi(r_c) + \varphi(r_c)(\frac{1}{r_c} - \frac{1}{t_c}) \qquad (18)$$

where r_c and t_c are respectively the signed square root of the log likelihood ratio statistic and the standardized maximum likelihood estimate, computed from the exact or approximate conditional log likelihood ℓ_c:

$$r_c = \mathrm{sign}(t_c)[2\{\ell_c(\hat{\psi}_c) - \ell_c(\psi)\}]^{1/2} \qquad (19)$$

$$t_c = (\hat{\psi}_c - \psi)\{-\ell_c''(\hat{\psi}_c)\}^{1/2}. \qquad (20)$$

The version given in (18) was derived in Fraser, Reid and Wong (1991), and called the sequential saddlepoint approximation there. The saddlepoint approximation is applied to the densities of X and $X_{(2)}$ to get the approximation to ℓ_c indicated in (17), and this conditional log-likelihood is converted to a cumulative distribution function by the method outlined in Section 1. A version called the double saddlepoint approximation, due to Skovgaard (1987), leads to an approximation of the form

$$F(r_p) = \Phi(r_p) + \varphi(r_p)(\frac{1}{r_p} - \frac{1}{t_p \rho}). \tag{21}$$

DiCiccio and Martin (1991) show that this approximation also can be derived directly from general formulae given in Barndorff-Nielsen (1986). Jensen (1986) shows that the mean and variance corrected version of r_p approximates the exact conditional distribution in this setting.

In (21) r_p and t_p are computed as at (19) and (20), but based on the profile log likelihood function $\ell_p(\psi) = \ell(\psi, \hat{\lambda}_\psi)$, and ρ is a correction for the nuisance parameter λ that takes the form

$$\rho = \{\frac{|j_{\lambda\lambda}(\hat{\psi}, \hat{\lambda})|}{|j_{\lambda\lambda}(\psi, \hat{\lambda}_\psi)|}\}^{1/2}. \tag{22}$$

The double saddlepoint approximation first applies the saddlepoint approximation to the densities of X and $X_{(2)}$, as in Barndorff-Nielsen and Cox (1979), and then applies the saddlepoint approximation to the cumulative distribution function of X_1 given $X_{(2)}$ following the method of Lugannani and Rice (1980).

Both (18) and (21) are accurate to $O(n^{-3/2})$, and the difference between them is essentially whether the likelihood function is adjusted for nuisance parameters and then converted to a tail area approximation, or whether the tail area approximation is adjusted for nuisance parameters after the likelihood has been converted. From a more technical point of view, the difference is whether an integrand is expressed as $\exp\{a(\psi) + \log b(\psi)\}$ or $\exp\{a(\psi)\}b(\psi)$ before the integral is approximated. The function $a(\psi)$ is $O(n)$, whereas $b(\psi)$ is $O(1)$, which is why the asymptotic properties of the two verisons are the same. As is often the case with asymptotic approximations, two versions equivalent to the same order may behave very differently in practice, and the available numerical evidence, summarized in Section 3, seems to indicate that version (18) is typically more accurate than version (21).

There are also r^* versions of both these approximations, so tail areas can also be computed using

$$\Phi(r_c^*) = \Phi\{r_c + \log(r_c/t_c)\} \tag{23}$$

$$\Phi(r_p^*) = \Phi\{r_p + \log(r_p/\rho t_p)\}. \tag{24}$$

Numerical differences between these and (18) or (21) tend to be slight, with no clear preference for one version over the other in general. The r^* version is often very convenient for other purposes, such as computing residuals or qq-plots, and as well the error bounds for the r^* version can sometimes be shown to be uniform in a large-deviation region (Barndorff-Nielsen, 1991).

Table 3: Tail areas based on approximations (18) and (23): the exact tail area is 0.05. From Reid (1995b). Sample size n.

(ψ,λ)	$(1/2,1)$	$(1,1)$	$(2,2)$	$(1,1/2)$	(1.1)	$(2,2)$
n	3	3	3	10	10	10
(18)	0.046	0.047	0.049	0.049	0.049	0.046
(23)	0.046	0.048	0.049	0.049	0.049	0.049

As an example, Table 3, taken from Reid (1995a) compares the approximations with simulated tail areas for inference about the shape parameter of the inverse Gaussian distribution $f(x;\psi,\lambda)=\{\psi/(2\pi)\}^{1/2}x^{-3/2}\exp\{-\frac{1}{2}(\psi/x+\lambda x)+\sqrt{(\psi\lambda)}\}$.

The application of these tail area approximations in exponential families is the subject of Pierce and Peters (1992), and that paper includes a detailed discussion of implementation in particular families, a derivation of the r^* version from the p^* approximation, and several examples of the performance of the approximations in specific examples. Particular attention is given to the role of continuity correction in applications of (18) to discrete models such as the Poisson and multinomial. Pierce and Peters (1992) also clarified the difference between (18) and (21), and the discussion here is largely adapted from their paper.

Example 2.2: Transformation models

The other class of models for which exact elimination of nuisance parameters is possible is the family of transformation models, i.e. models generated from a basic density via a group of transformations on the parameter space. In these models an exact ancillary statistic is available, so that the joint density of the data can be expressed as proportional to $f(\hat\theta|a;\theta)f(a)$, where neither the proportionality constant nor the marginal density of a depend on θ. The role of X in Section 2.1 is played by $\hat\theta$, and the density of this is to be interpreted as conditonal on a. In these models the nuisance parameter λ is eliminated by integrating:

$$\begin{aligned} \ell_m(\psi) &= \log f_m(\hat\psi;\psi|a) && (25)\\ &= \log\int f(\hat\psi,\hat\lambda;\psi,\lambda|a)d\hat\lambda && (26)\\ doteq &\quad \ell(\psi,\hat\lambda_\psi)-\frac{1}{2}\log|j_{\lambda\lambda}(\psi,\hat\lambda_\psi)| && (27)\end{aligned}$$

where the integral in (26) is over R^{p-1}, and (27) follows from an application of the Laplace approximation (see e.g. Tierney and Kadane, 1986).

The cumulative distribution function for $\hat\psi_m=\hat\psi$ provides a means of computing a tail area approximation, and the result is

$$F(r_m)\doteq\Phi(r_m)+\varphi(r_m)(\frac{1}{r_m}-\frac{1}{t_m}) \qquad (28)$$

where r_m and t_m are the signed square root of the likelihood ratio statistic and the standardized score statistic computed from the exact or approximate marginal log likelihood

ℓ_m:

$$r_m = \mathrm{sign}(t_m)[2\{\ell_m(\hat{\psi}) - \ell_m(\psi)\}]^{1/2} \tag{29}$$
$$t_m = \ell_m'(\psi)\{-\ell_m''(\hat{\psi})\}^{-1/2}. \tag{30}$$

This approximation was derived in DiCiccio, Field and Fraser (1990), and numerical examples of both this version and a "double saddlepoint" version are provided there. There is also a discussion of the use of the Laplace approximation to establish the $O_p(n^{-3/2})$ accuracy of (27) as an approximation to the exact marginal likelihood. The sequential saddlepoint version given in (28) corresponds to equation (19) of DiCiccio, Field and Fraser (1990). The double saddlepoint version is their equation (12) and in the present notation is

$$F(r_p) = \Phi(r_p) + \varphi(r_p)(\frac{1}{r_p} - \frac{\rho}{t_p}) \tag{31}$$

where ρ is defined by (22). There are also two corresponding r^* versions incorporating the adjustment into the normal quantile, as in the exponential family case. Fraser, Lee and Reid (1990) provide an alternate derivation of (28) via conditioning, and Pierce and Peters (1992) discuss transformation models briefly in their reply to the discussion.

There is a very close connection to the use of the Laplace approximation for computing marginal posterior densities, as the marginal likelihood is equal to the marginal posterior in a transformation model, if a flat prior is used. Laplace approximations for marginal posteriors are discussed in Tierney and Kadane (1986) and Tierney, Kass and Kadane (1989).

Outside the exponential and transformation family models, two problems prevent the direct application of the approximations discussed here. First, the dimension of the minimal sufficient statistic is typically greater than the dimension of θ. In the case that θ is a scalar, a natural approach to is find an approximately ancillary statistic a such that the conditional distribution given a is a one-dimensional distribution. Then the general formula for t given by (11) or (12) can be applied, with the understanding that $\partial\ell/\partial\hat{\theta}$ is taken with a held fixed.

When $\theta = (\psi, \lambda)$, the general approach is to reduce to the dimension of θ by conditioning on an approximately ancillary statistic for θ, and then to eliminate the nuisance parameter λ either by conditioning, as in the exponential family model, or by marginalizing, as in the location family model. While substantial progress has been made in general models, particularly in Barndorff-Nielsen (1986, 1991, 1994), McCullagh (1984, 1987, Ch. 8), Skovgaard (1990), Fraser and Reid (1993, 1995), there is not a definitive method available at this time, and the details of implementation tend to be rather difficult. I will not survey the available results in this paper: general accounts are available in Barndorff-Nielsen and Cox (1994) and Reid (1995b).

In the next Section I illustrate that even within the families of exponential and transformation models, for which definitive solutions are available, the likelihood approach has proved itself to be very versatile.

3 The likelihood approach is versatile

3.1 Regression

Two classes of regression models are rather directly handled using the approximations of Section 2: linear regression with non-normal errors and generalized linear models. Examples of the first appear in DiCiccio, Field and Fraser (1990) and Fraser, Lee and Reid (1990), and examples of the second are considered in detail in Pierce and Peters (1992) and in Davison (1988).

Suppose first that we have a regression model of the form

$$y = X\beta + \sigma e \tag{32}$$

where $y = (y_1, \ldots y_n)$ are independent, and e_i follows a known distribution $f(\cdot)$. This is a location-scale model, with p location parameters $(\beta_1, \ldots \beta_p)$ and scale parameter σ. The exact conditional distribution of the maximum likelihood estimate $(\hat\beta, \hat\sigma)$, given the ancillary vector $a = (y - X\hat\beta)/\hat\sigma$ is given by the p^* formula

$$f(\hat\beta, \hat\sigma | a; \beta, \sigma) = c|\hat{j}|^{1/2} L(\beta, \sigma; \hat\beta, \hat\sigma, a)/L(\hat\beta, \hat\sigma; \hat\beta, \hat\sigma, a) \tag{33}$$

where $L(\beta, \sigma; \hat\beta, \hat\sigma, a) \propto \prod \sigma^{-1} f[\{\hat\sigma a_i + x_i'(\hat\beta - \beta)\}/\sigma]$, with x_i' the ith row of the design matrix.

For inference about a component parameter, say β_1, it is convenient to re-express the right hand side of (33) as a function of the statistics $v_1, \ldots, v_p, s$, defined as $v_i = (\hat\beta_i - \beta_i)/\hat\sigma$, and $s = \hat\sigma/\sigma$. The 'nuisance parameters' $\beta_2, \ldots, \beta_p, \sigma$ are eliminated by integrating this expression with respect to $v_2, \ldots, v_p$ and s, leaving a marginal likelihood that depends on β_1 only. This approach is detailed, for example, in Fraser (1979, Ch. 4). As described in DiCiccio, Field and Fraser (1990), and sketched in Section 2.2 above, this integral can be approximated by Laplace's method, and then the p-value for testing a particular value of β_1 is computed using approximation (28) or the asymptotically equivalent version (31).

DiCiccio, Field and Fraser (1990) illustrate (31) on two examples of simple location-scale models, the normal and the Weibull. They also consider extreme value regression, and normal regression, with $n = 10$ and $p = 6$. For inference about σ version (28) performs much better than (31), although for the regression parameters the performance of both versions is quite comparable. It is relatively straightforward to adapt the transformation model analysis outlined here to the case of Type II censoring, as discussed in Lawless (1982, App. G), and both the Weibull and extreme value examples include this feature. Further examples of application of (28) in normal, t_5, and extreme value regression models is given in DiCiccio and Field (1991), where an approximation to (28) is suggested that is slightly easier to calculate.

Two generalizations of (32) that are of practical importance are to allow a nonlinear regression function, possibly of the generalized linear form $g(X\beta)$ for a known function g, and to allow the error distribution $f(e)$ to depend on additional unknown parameters (for example the degrees of freedom of a t distribution). These generalizations present two problems, one a problem of inference, and the other a problem of approximation. For the first, there is no exact ancillary statistic available to condition on, in order to reduce the dimension of the sufficient statistic to the dimension of the parameter, and no exact method of eliminating the 'nuisance parameters' (eliminated above by integration) to provide a one-dimensional distribution for inference about the parameter of interest. In the absence of an 'exact' solution, it is difficult to derive an approximate solution, since it is not clear what should be approximated. However, the more general work of Barndorff-Nielsen (1991, 1994), DiCiccio and Martin (1993), and Fraser and Reid (1995) suggests that given a suitable approximate method for dimension reduction (such as conditioning on an approximately ancillary statistic), the derivation of the tail area approximation is straightforward.

Generalized linear models provide an alternative class of regression models with non-normal error that have proved very useful in practice. Introductions to generalized linear models are given in McCullagh and Nelder (1989, Ch. 2) and Firth (1990). When the scale parameter is given, and the canonical link function is used, these are simply exponential family models with a regression structure

$$f(y; \beta) = \exp\{y'X\beta - c(\beta) - d(y)\} \tag{34}$$

and inference about one component of β is readily carried out using the conditional method outlined in Example 2.1. A number of discrete examples are considered in detail in Pierce and Peters (1992), including logistic regression, a regression model for a 4×4 contingency table, and inference about a common odds ratio in a collection of K independent pairs of binomial observations. They find that both r^* versions and the sequential saddlepoint version are of similar (very high) accuracy for computing p-values and confidence limits; the double saddlepoint version (21) is not quite as reliable in problems with large numbers of nuisance parameters. Davison (1988) also illustrates the use of (21) in generalized linear models. An example of particular interest is a test for an added variable in a regression model, discussed in Davison's Section 7.

Computation of the sequential saddlepoint approximation is more complicated than the double saddlepoint approximation, as it involves the second derivative, with respect to ψ, of the term $|j_{\lambda\lambda}(\psi, \hat{\lambda}_\psi)|$, which is itself already a second derivative of the original log-likelihood. These derivatives will typically have to be calculated numerically, and some preliminary smoothing may be required when the likelihood function itself is only available numerically. Some details are discussed in Fraser, Reid and Wong (1991). Numerical derivatives are also used in DiCiccio and Stern (1993). For testing a null hypothesis it is only necessary to compute the approximation once, whereas iteration will be needed to construct a confidence interval from the approximation. Fraser (1991) refers to the "confidence distribution function" as the function that computes the tail area for as a function of ψ. Generalized linear models with non-canonical link functions are often of particular importance in applications. Again, the difficulty is essentially an inferential one; there is no exact way to eliminate nuisance parameters in this setting. Applications of saddlepoint techniques to quasi-likelihood functions are considered in Nelder and Pregibon (1987), although I am not aware of any extension of this work to computing tail area approximations.

Saddlepoint approximations for estimates obtained by robust regression techniques, typically defined by minimizing something like $\sum \psi(y_i - x_i'\beta)$ for a suitably chosen function $\psi(\cdot)$, are considered in Daniels (1983), Field (1982) and Spady (1991). These approximations are for the density of the resultant estimator of β, but extensions to tail probabilities might be straightforward and would be of considerable practical interest. Saddlepoint approximations to the density of M-estimators in linear and nonlinear regression are discussed in Field and Ronchetti (1990, Ch. 4) and Ronchetti and Welsh (1994).

Software for analysing generalized linear models is widely available, and the profile likelihood and nuisance parameter information matrix are readily obtained from the output. Although it would be relatively straightforward to provide a set of routines that implemented the tail area approximations when canonical link functions were used, I am not aware of any systematic development along these lines.

Software for evaluating Laplace approximations to marginal posteriors is available in Lisp-Stat (Tierney, 1990), and this can be adapted to the non-normal regression problem very easily, as the marginal likelihood (27) simply corresponds to the marginal posterior with a flat prior for $(\beta, \log\sigma)$.

In the regression problems discussed here, I have emphasized inference for a single parameter, with the other parameters treated as 'nuisance' parameters, although in practice most, if not all, of the parameters in the model would be of interest. The tail area approximations can be used successively for inference about each component, but multiparameter techniques are needed to test 'omnibus' hypotheses such as $\beta = 0$. These are discussed briefly in Section 4.2.

3.2 Constructed exponential families

It is often possible to construct an exponential family that is not of interest in itself, but which introduces a canonical parameter component ψ, say, so that a test of $\psi = 0$ corresponds to a test of a particular null hypothesis of interest. An example of this is given in Barndorff-Nielsen and Cox (1979), in which the fit of a gamma regression model is assessed by introducing a quadratic term. Thus the full model is given by

$$f(y; \beta, \nu) \propto \Gamma(\nu)^{-n} \prod (\nu/\mu_i)^{\nu} \exp\{-(y_i/\mu_i)\} \tag{35}$$

where $1/\mu_i = \beta_1 + \beta_2 x_i + \beta_3 x_i^2$. A test of $\beta_3 = 0$, readily carried out using approximation (18) or (21), provides an assessment of the fit of the linear regression model for the canonical parameter $1/\mu_i$. Tests for added variables and case influence can be constructed similarly, and examples are given in Davison (1988, Section 7) and Carlin et al. (1992).

In Booth and Butler (1990) the distributions of four nonparametric test statistics are derived using constructed exponential family regression models. The examples considered are the bootstrap distribution for the sample mean, the permutation distribution of the two-sample t-test, the jackknife version of an M-estimate, and the Bayesian bootstrap. The tail area approximation using the double saddlepoint version (21) is used. The use of saddlepoint

Table 4: Tail areas based on approximations (18) and (23): the exact tail area is 0.05. From Butler, Huzurbazar and Booth (1992b). Sample size n, p nuisance parameters.

n	10	28	58	10
p	3	3	3	10
(18)	0.0510	0.0477	0.0507	0.0528
(23)	0.0510	0.0477	0.0508	0.0528

approximations for the bootstrap distribution has a large literature that is summarized in the next subsection.

Butler, Huzurbazar and Booth (1992ab) and Butler, Booth and Huzurbazar (1993) have used constructed exponential families to find approximations to the distribution of several test statistics used in multivariate analysis. In Butler, Huzurbazar and Booth (1992a) a one-dimensional exponential family is constructed for obtaining the distribution of the generalized variance statistic, and a multi-parameter exponential family is constructed for which the conditional distribution of one component given the others gives the distribution of Wilks' Λ statistic. There are several ways to carry out the construction of this multi-parameter exponential family, and three such are considered by the authors. In each case p-values are computed using the double saddlepoint version, which tends to break down when the dimension of the nuisance parameter increases. Numerical integration of the double saddlepoint density approximation is more reliable, but computationally more intensive. Butler, Huzurbazar and Booth (1992b) use the same technique to approximate the distribution of the trace statistic. The sequential and double saddlepoint methods are directly compared, as are the r^* and the Lugannani and Rice versions. The conclusions are consistent with Pierce and Peters (1992) that the double saddlepoint version is not reliable, but the other three approximations are very good. Butler, Booth and Huzurbazar (1993) use the same technique to approximate several further likelihood ratio statistics related to tests for particular forms of covariance matrices. Table 4 is taken from Butler, Huzurbazar and Booth (1992b), and compares the approximations using the sequential saddlepoint methods to the exact value (obtained by simulation) for the distribution of the trace statistic for testing equality of two covariance matrices in multivariate analysis of variance.

The constructed exponential family is also used in Kolassa and Tanner (1994) in conjunction with the Gibbs sampler. This gives an approximation to the exact p-value for testing row and column independence in $r \times c$ contingency tables.

3.3 Approximating nonparametric statistics

The nonparametric bootstrap distribution is a multinomial distribution with probability vector $(1/n, \ldots, 1/n)$ and sample size n; the value N_i, say, of the ith component of the vector counts how many times the observation x_i is repeated in the bootstrap sample. The multinomial distribution can be obtained by conditioning in a Poisson model, and this representation is used by Booth and Butler (1990) to derive a saddlepoint approximation to the bootstrap distribution of the sample mean. This provides an alternative derivation of the approximation of Davison and Hinkley (1988) who used the empirical cumulant generating function $K_n(t) = \log\{n^{-1} \sum \exp(tx_i)\}$ in (3) and (4) to approximate the bootstrap distribution of the sample mean.

The method of Davison and Hinkley (1988) requires the statistic of interest to be linear in the observations. Extensions to nonlinear statistics are discussed in Daniels and Young (1991) and DiCiccio, Martin and Young (1992, 1994). Davison and Hinkley's technique is justified in Wang (1990c) and extended to nonlinear statistics by a different method in Wang (1990a).

The amount of calculation involved in ordinary bootstrap sampling is not usually very large, and one might question the need for an asymptotic approximation in this setting. However, double bootstrapping usually does require prohibitive amounts of computing, and a method investigated in detail in DiCiccio, Martin and Young (1992) is to use the saddlepoint approximation in the inner bootstrap loop. The sequential saddlepoint version is substantially more reliable in this setting, although it is more difficult to compute. A simplification of the sequential version that still performs better than the double version is suggested in DiCiccio, Martin and Young (1994).

Other nonparametric statistics whose distribution is conditional on the observed sample, such as permutation and randomization statistics can be approximated similarly, and examples are given in Davison and Hinkley (1988) and in Booth and Butler (1990); the latter using the constructed exponential family described in the previous subsection. Wang (1991) applies the bootstrap technique of Davison and Hinkley (1988) to obtain approximate confidence bands in nonparametric regression. Robinson (1982) derived a different tail area approximation for use in permutation tests: this approach is discussed in detail in Kolassa (1994, Ch. 5).

3.4 Bayesian posterior intervals

The approximation used in transformation models can be directly applied to the construction of Bayesian posterior marginal distributions; in fact the Laplace approximation to ℓ_m appeared in Tierney and Kadane (1986) where it was studied in the Bayesian context. The marginal likelihood used in a transformation model is a Bayesian marginal posterior with a flat prior, but incorporation of an arbitrary prior is quite straightforward. There are both 'sequential' and 'double' versions of the approximations, depending on whether the prior is used to adjust the log-likelihood before integrating, or whether the adjustment due to the prior is incorporated into the approximation. A rather general result on the equivalence of Bayesian and likelihood inferences in transformation models is given in Chang and Villegas (1986).

Several results applying to transformation models and to Bayesian posteriors have been established by DiCiccio and Martin (1991, 1993). Their approach, following on from DiCiccio, Field and Fraser (1990) is to find approximations for integrals of the form $\int b(x) \exp\{\ell(x)\} dx$, where $\ell(x)$ is $O(n)$ and $b(x)$ is $O(1)$. This can represent a p^* type formula in a transformation model, as at (27), or a marginal Bayesian posterior with $b(\cdot)$ the prior and $\ell(\cdot)$ the likelihood, (and integration with respect to only some components of x), or a Bayesian posterior expectation of $b(\cdot)$ with $\ell(\cdot) = \log \text{likelihood} + \log \text{prior}$. The forms of the resulting approximations are of the type (28) or (31), depending on how the adjustment represented here by $b(\cdot)$ is incorporated into the argument. What this means in practice is that the statistic t that is used in the approximation is always a (modified) score statistic.

There is a substantial literature on using these types of approximations for various aspects of Bayesian inference. Relevant references include Kass, Tierney and Kadane (1988), Tierney, Kass and Kadane (1989), Kass and Slate (1992), Sweeting (1995ab), and the papers of DiCiccio and Martin cited above. An overview of the connections between likelihood and Bayesian asymptotics is given in Reid (1995a) and substantial further development can be found in Sweeting (1995b).

4 Generalizations and extensions

4.1 Dimension reduction

As mentioned above, when the dimension of the minimal sufficient statistic is greater than the dimension of the parameter, some means of dimension reduction must be found in order to obtain a p^* type formula for the distribution of $\hat{\theta}$. This is typically done by conditioning on an approximate ancillary statistic. When θ is a scalar, the resulting approximation can be integrated as in Section 2.1 to give a tail area approximation. In fact, from formulas (11) or (12) it can be seen that in order to obtain a tail area approximation it is only necessary to be able to differentiate the log-likelihood with the ancillary directions held fixed, and not to have an explicit expression for the ancillary statistic. Further, as explained in Skovgaard (1990), to obtain an approximation accurate to $O(n^{-3/2})$ it is sufficient to condition on a statistic ancillary to $O(n^{-1})$. These two simplifications are used in Fraser and Reid (1995) to derive a general tail area formula using a local location model.

As discussed briefly in Section 2.1, when $\theta = (\psi, \lambda)$, with λ a nuisance parameter, further conditioning or marginalizing is needed to eliminate, at least approximately, the effect of the nuisance parameter λ. A fairly general approach to this is to find an approximately ancillary statistic for λ in the model with ψ held fixed; i.e. a statistic whose marginal distribution depends only on ψ. This approach is discussed in fairly general terms in Barndorff-Nielsen (1986, 1991, 1994), Skovgaard (1990) and Fraser and Reid (1995).

It was shown by McCullagh (1984; 1987, Ch. 8) that the p^* approximation is accurate to $O(n^{-1})$ for any approximate ancillary, and as a result it is possible to obtain tail area approximations (or more precisely expressions for the statistic t) that do not require specification of an ancillary statistic. Two such are derived and illustrated in Barndorff-Nielsen and Chamberlin (1991, 1994) and another is derived in DiCiccio and Martin (1993) using a Bayesian approach. Skovgaard (1995) gives a particular choice of t that does not require specification of the ancillary statistic, that is relatively easily computed from the likelihood function, and for which the r^* approximation has relative error $O(n^{-1})$ in a large deviation region. Jensen

(1995) discusses use of a signed log-likelihood ratio statistic as an approximate ancillary, in curved exponential families.

4.2 Vector parameter of interest

There may be situations where simultaneous inference for several components of θ is more appropriate than sequential testing of individual components. The usual test statistic for this situation is the generalized likelihood ratio statistic

$$w(\psi) = 2\{\ell(\hat{\psi}, \hat{\lambda}) - \ell(\psi, \hat{\lambda}_\psi)\} \tag{36}$$

which has an asymptotic χ^2_q distribution, where q is the dimension of the parameter of interest ψ. There are asymptotically equivalent versions of $w(\psi)$ based on the maximum likelihood estimator or the score statistic (see, e.g. Cox and Hinkley, 1974, Ch. 9), but $w(\psi)$ has the following advantage in terms of its asymptotic distribution. It can be shown that $Ew(\psi) = q\{1 + b(\psi, \lambda)/n + O(n^{-2})\}$, and that

$$w'(\psi) = w(\psi)/(1 + \hat{b}/n) \simeq \chi^2_q\{1 + O(n^{-2})\}. \tag{37}$$

The technique of dividing by (an estimate of) $b(\theta)$ is known as Bartlett correction, and the validity of (37) to $O(n^{-3/2})$ was established by Lawley (1956) and to $O(n^{-2})$ by Barndorff-Nielsen and Hall (1988). The general expression for $b(\cdot)$ is relatively complicated, involving fourth order cumulants of derivatives of the log likelihood function (Lawley, 1956). Simplifications of various types are available in a number of papers, including DiCiccio (1986), Cordeiro (1983, 1987), Barndorff-Nielsen and Hall (1988), and DiCiccio and Stern (1993). Illustrations of Bartlett correction in generalized linear models are given in McCullagh and Nelder (1989, Ch. 11).

A rather clear derivation of (37) is given in Bickel and Ghosh (1990), who also consider a Bayesian version that adjusts the analogue of w based on the log-posterior. They emphasize in particular Lawley's construction of w as $\sum r_i^2$, where is r_i^2 is is a likelihood ratio statistic for testing one component of ψ, and r_i is a likelihood root of the form (10). Lawley's method of proof was to establish the joint normality of these r-statistics to sufficiently high order. Thus, even though in principle the limiting distribution for a vector parameter of interest is quite different than the tail area approximations discussed in previous sections, the tail area approximation plays a central role here as well.

The Bartlett adjustment b in the scalar parameter case is computed from $\mathrm{E}r^2$, which is $m^2 + s^2 + O(n^{-1})$, in the notation of Section 2.1. Expressions for m and s are given, for example, in DiCiccio, Field and Fraser (1990).

A quite different approach to the multiparameter problem is outlined in Cheah, Fraser and Reid (1994) in the special case of a canonical exponential family. A test is constructed for the length of a vector parameter ψ, given its direction. Since the length is a scalar quantity, a version of a tail area approximation can be derived using similar arguments to those leading to (18) and (21). The starting point is a saddlepoint-type approximation more difficult to integrate than the p^* formula, though, and the resulting tail area approximation is not as elegant nor as accurate as the one-parameter version of Example 3.1.

4.3 Derivations based on the sample mean

The third order approximations to tail probabilities discussed here are formulated in terms of the likelihood function, but in many ways the most natural starting point is the saddlepoint approximation to the density and distribution function of the sample mean in an i.i.d. sample, as described in the Introduction. This provides a rigorous mathematical derivation, as developed for example in Field and Ronchetti (1990) and in Kolassa (1994).

Skovgaard (1987) derived a saddlepoint approximation to the conditional distribution function of $\bar{X}$, say, given $\bar{Y}$, where $(X_1, Y_1), \ldots (X_n, Y_n)$ are independent bivariate random variables. Wang (1990a) generalized this to smooth functions $g(\bar{X}, \bar{Y})$. Jensen (1991c) considers saddlepoint approximations for sample means constructed from Markov processes. There are also some results available for randomly stopped sequences $\bar{X}_N$ of i.i.d. random variables (Homble and McCormick, 1993) I am not aware of any saddlepoint type results for sample means constructed from sequences that exhibit long range dependence, although this would certainly be of interest both theoretically and practically. Mykland (1993) and Gu (1992) have obtained Edgeworth expansions to second order for martingales, but I am not aware of any saddlepoint approximations available for martingales.

4.4 Verifying the accuracy of the approximation

One problem with all the approximations discussed here is that little is known about the accuracy of the approximation for any given fixed sample size n. Knowing that the error term in the approximation for $F(\bar{x})$, say, is, for example, $O(n^{-3/2})$ only states that the error is of the form $C/n^{3/2}$ for some unspecified C, which does not depend on n but may well depend on $\bar{x}$, and the exact form of this dependence is rarely known. There are conditions under which the error in the r^* approximation can be proved to be uniform (Barndorff-Nielsen, 1991; Jensen, 1991ab), which is one advantage of the r^* version over the Lugannani and Rice version.

One way of checking the accuracy of the approximation is to compute the next term in the asymptotic expansion. This is the approach taken in Wong and Li (1992) in the context of Laplace expansion for a marginal posterior density. Kolassa (1994, Ch. 7) has derived the next term in Skovgaard's approximation to the conditional distribution of $\bar{X}$ given $\bar{Y}$.

Although there is no guarantee that including an extra term in an asymptotic expansion will improve the approximation, if its inclusion gives quite different results from the simpler approximation, it is probably an indication that something is going wrong. The difficulty with this approach is that the fourth order term, i.e. the $O(n^{-2})$ term in the expansions we have been considering, is typically extremely complicated.

An alternative method of checking the accuracy of the tail area approximation in transformation models was proposed in Fraser, Lee and Reid (1990), but has not been widely tested. It involves Monte Carlo resampling using an importance sampling function, but the dependence of the method on the importance function and the applicability of the method to more general models is unknown.

Bootstrap methods can also be used to construct confidence intervals accurate to third order. In particular the so-called ABC intervals discussed in DiCiccio and Efron (1992) seem particularly closely related to intervals constructed for scalar interest parameters using the tail area approximations described here. The exact relationship between the two has not yet been obtained, as far as I am aware, although the question of the relationship was raised by Davison (1992) and a comparison of the confidence limits from various third order methods is described in DiCiccio and Martin (1993).

4.5 Other topics

There are of course many statistics, such as extreme order statistics, with a limiting distribution that is non-normal, and one might expect that a version of the tail area approximation that built on a non-normal kernel would provide a better approximation than one a normal distribution for the leading term. A Lugannani and Rice type formula based on non-normal kernels is developed in detail in Wood, Booth and Butler (1992), and several examples are presented. In general the normal-based versions perform as well as those based on a gamma kernel, even when the true limiting distribution is non-normal, although there are exceptions. Booth and Wood (1995) discuss an example where the non-normal kernel gives substantially better results in the far tail of the distribution of $\bar{X}$ ($\bar{x} \to \infty$) than (2), and in fact performs better than the r^* version, even though Barndorff-Nielsen (1991) showed that the r^* version has uniformly bounded error as $\bar{x} \to \infty$.

I am not aware of any higher order asymptotic theory for extreme value limits, although it would seem that the likelihood based approach could be useful in this setting as well.

There are now many types of pseudo-likelihoods that have been used in connection with particular complex problems, such as partial likelihood for semiparametric models, pseudo-likelihood for spatial processes, integrated likelihoods, and so on. It is relatively straightforward to compute r and u from these pseudo-likelihoods and use them in the tail probability formulas such as (18) or (21), but in general it is not clear what tail probability is being approximated, nor what the order of the error term is. An unpublished University of Toronto thesis by A. Tapia-Aguilar considers this type of analysis for inference about the regression parameters in the proportional hazards model. Although preliminary simulation results are promising, it seems very difficult to prove that the resulting expression approximates the relevant probability beyond first order. Such a proof seems to need a saddlepoint type approximation for martingales, which is not yet available.

5 Conclusion

The abstract and introduction refer to the "fundamental role of the likelihood function", and some comments on this claim are perhaps in order. From one point of view this is just the observation that the tail area approximations discussed here are constructed from likelihood based quantities, such as the log-likelihood ratio statistic, the maximum likelihood estimator, and the observed Fisher information. In addition to this, the development of the tail area approximation, from the Lugannani and Rice (1980) approximation for the distribution function of the sample mean, to the general one parameter version (9) and the nuisance parameter versions (18) and (28), is, in my opinion, much clearer, and in a sense almost obvious, as soon as the Lugannani and Rice approximation is reformulated in terms of likelihood.

But more importantly, the construction of the tail area approximations depend only on the likelihood function evaluated at the observed data point, that is they do not involve values of the likelihood function at other points in the sample space, with one exception. Outside exponential and transformation families, the statistic t requires differentiation of the likelihood function with respect to the minimal sufficient statistic, in a direction specified by an exact or approximate ancillary statistic. This has led to the observation that the third order approximations discussed here cannot be equivalent to Bayesian inference, which uses only the observed likelihood function, but that the lack of equivalence arises in a well defined and single calculation (Lindley, 1992; Pierce and Peters, 1992, 1994), and arises only in the improvement of the approximation from $O(n^{-1})$ to $O(n^{-3/2})$. This point is also made in DiCiccio and Martin (1991, 1993). This seems to have the potential for illuminating the connections between asymptotic theory for Bayesian and frequentist inference, although much work remains to be done.

I have tried to emphasize in the paper both the actual and potential applicability of recent theoretical developments in asymptotic theory, by concentrating on the role of the likelihood function in computing higher order approximations to tail areas. Some questions that remain are the following.

- How general is the notion of converting a likelihood or an approximate likelihood to a density or distribution function approximation?
- How do we go about constructing likelihoods for inference about particular parameters of interest? To what extent is the construction tied to the computation of a tail area?
- What is the scope for treating specialized, subject matter, applications?
- What is the scope for providing software for general, routine, application?
- How can the surprising accuracy of the tail area approximation in many examples be explained?
- How best should we compare asymptotically equivalent approximations?

Acknowledgements

This paper is based on an invited talk organized by the IMS New Researchers Committee at the Annual Joint Meetings in Boston in 1992. Background material also appears in Reid and Fraser (1990) and in Reid (1993). I am grateful to the Senior Associate Editor and two referees for very thoughtful comments, and to Tom DiCiccio, Don Fraser, Snehelata Huzurbazar, Elvezio Ronchetti and Suojin Wang for helpful comments on an earlier draft. This research was partially supported by the Natural Sciences and Engineering Research Council.

References and Annotated Bibliography

The following gives references to work cited in the text, and also attempts to provide a comprehensive list of papers published since 1987 on higher order approximations in statistics.

Barndorff-Nielsen, O.E. (1980) Conditionality resolutions. *Biometrika* **67**, 293–310.

Barndorff-Nielsen, O.E. (1983) On a formula for the distribution of the maximum likelihood estimator. *Biometrika* **70**, 343–365.

Barndorff-Nielsen, O.E. (1986) Inference on full or partial parameters based on the standardized signed log likelihood ratio. *Biometrika* **73**, 307–22.

Barndorff-Nielsen, O.E. (1988) Discussion of "Saddlepoint methods and statistical inference" by N. Reid. *Statist. Science* **3**, 228–229.

Barndorff-Nielsen, O.E. (1990a) Approximate interval probabilities. *J. R. Statist. Soc.* B **52**, 485–96.

Derives a parametrization invariant version of the tail area approximation of Lugannani and Rice for a one-dimensional parameter. The result was announced in Barndorff-Nielsen (1988). The Wald statistic in the usual Lugannani and Rice approximation is replaced by the function $u = \hat{j}^{-1/2}(\hat{l}_{;1} - l_{;1})$, *where* $l_{;1} = \partial l(\omega;\hat{\omega},a)/\partial\hat{\omega}$ *is a sample space derivative of the likelihood function. In this it is assumed that* $(\hat{\omega}, a)$ *is minimal sufficient and that* a *is an ancillary statistic, and finally that Barndorff-Nielsen's* p^* *formula is accurate to* $O(n^{-1})$. *The proof uses the result of Temme (1982), which enables a uniform error bound to be obtained. The relation to* r^* *of Barndorff-Nielsen (1986) is also discussed, as well as the result that* r^* *is standard normal to* $O(n^{-3/2})$. *(In this connection see Jensen, 1992). The approximation derived here is essentially the same as that of Fraser (1990), but Fraser's version does not require the derivative to be taken with respect to the maximum likelihood estimate: it incorporates a Jacobian effect to compensate for differentiating with respect to some other one-dimensional statistic.*

Barndorff-Nielsen, O.E. (1990b) A note on the standardized signed log likelihood ratio. *Scand. J. Statist.* **17**, 57–60.

This provides a very simple verification of the result of Barndorff-Nielsen (1986) that if p^* *is accurate to* $O(n^{-3/2})$, *then* r^* *has a standard normal distribution to the same order. The derivation is for the one-dimensional case (no nuisance parameters), although the 86 paper also covers the nuisance parameter case. The proof is adapted* *to the exponential family setting with nuisance parameters conditioned out in the appendix of Pierce and Peters (1992).*

Barndorff-Nielsen, O.E. (1990c) p^* and Laplace's method. *Rev. Brasil. de Prob. e Estatistica* **4**, 89–103.

Barndorff-Nielsen, O.E. (1991) Modified signed log likelihood ratio. *Biometrika* **78**, 557–64.

Provides a simple derivation of the r^* approximation, as well as an expression for the function u that is simpler than that in Barndorff-Nielsen (1986). An approximation to the expression for u is suggested, that avoids specification of an ancillary statistic, which is very close to a suggestion of DiCiccio and Martin (1993). Using this simplification the standard normal distribution for r^* is accurate to $O(n^{-1})$. A related simplification is discussed in Barndorff-Nielsen and Chamberlin (1991).

Barndorff-Nielsen, O.E. (1994) Adjusted versions of profile likelihood and likelihood roots, and extended likelihood. *J. R. Statist. Soc.* B **56**, 125–140.

Barndorff-Nielsen, O.E. and Chamberlin, S.R. (1991) An ancillary invariant modification of the signed log likelihood ratio. *Scand. J. Statist.* **18**, 341–52.

Barndorff-Nielsen, O.E. and Chamberlin, S.R. (1994) Stable and invariant adjusted directed likelihoods. *Biometrika* **81**, 485–499.

Barndorff-Nielsen, O.E. and Cox, D.R. (1979) Edgeworth and saddlepoint approximations with statistical applications (with discussion). *J. R. Statist. Soc.* B **41**, 279–312.

Barndorff-Nielsen, O.E. and Cox, D.R. (1990) *Asymptotic Techniques for Use in Statistics.* Chapman and Hall, London.

Detailed introduction to the methods of asymptotic analysis, with emphasis on applications to problems arising in statistics and stochastic processes. Discussion of Edgeworth expansions, saddlepoint expansions, Laplace expansions, in univariate and multivariate models. A wealth of detail on various asymptotic techniques.

Barndorff-Nielsen, O.E. and Cox, D.R. (1994) *Inference and Asymptotics.* Chapman and Hall, London.

Parametric statistical inference with particular emphasis on the use of asymptotic theory to approximate the distribution of statistics of interest, especially the use of the p^* and r^* approximations.

Barndorff-Nielsen, O.E. and Hall, P. (1988) On the level-error after Bartlett adjustment of the likelihood ratio statistic. *Biometrika* **75**, 374–378.

Bickel, P. J. and Ghosh, J. K. (1990) A decomposition for the likelihood ratio statistic and Bartlett correction – a Bayesian argument *Ann. Statist.* **18**, 1070–1090.

Booth, J.G. and Butler, R. W. (1990) Randomization distributions and saddlepoint approximations in generalized linear models. *Biometrika* **77**, 787–96.

Four nonparametric distributions are derived as conditional distributions in constructed exponential family regression models. This clever technique enables the distributions to be approximated by the saddlepoint technique. The tail area approximation of Skovgaard (1987) is used. In the special case that the constructed exponential family is a Poisson regression, so that the marginal density is free of the parameter of interest, the saddlepoint approximation can be directly applied to the conditional density. The conditional cumulant generating function, generally difficult to obtain, is not needed in this special case. Fraser, Reid and Wong (1991) also discuss direct approximation of the conditional distribution, but they need to approximate the cumulant generating function.

Booth, J.G. and Wood, A. (1995) An example in which the Lugannani-Rice saddlepoint formula fails. *Statist. and Probab. Letters*, to appear.

Butler, R.W., Booth, J.G. and Huzurbazar, S. (1992) Saddlepoint approximations for test of block independence, sphericity, and equal variances and covariances. *J. R. Statist. Soc.* B **55**, 171–184.

The technique is again to construct an exponential family and approximate the conditional distribution of a component of the sufficient statistic. The exponential family here is a gamma regression model, for which the conditional distribution is Dirichlet. The statistic of interest can be represented as a product of Dirichlet distributions. There are several possible representations of the test statistic, all using Dirichlet distributions. Some are more 'parsimonious' than others, in the sense that there are fewer 'nuisance parameters' to condition on in the constructed exponential model. Skovgaard's approximation typically works better for these more parsimonious models.

Butler, R.W., Huzurbazar, S., and Booth, J.G. (1992a) Saddlepoint approximations for the generalized variance and Wilks' statistics. *Biometrika* **79**, 157–70.

The distribution of these two multivariate test statistics under the relevant null hypotheses are approximated by embedding each of them in a full exponential family, with a constructed canonical parameter component, whose null value gives back the distribution of interest. The generalized variance statistic can be directly approximated by a one-dimensional exponential family. Wilks' Λ requires construction of a multi-parameter exponential family; the distribution of interest is obtained as the conditional distribution of one component of the sufficient statistic. There are several ways to carry out this construction; three such are considered. In each case p-values are computed using Skovgaard's version of the Lugannani and Rice formula. This tends to break down when the dimension of the nuisance parameter increases. The numerical integration of the double saddlepoint density approximation is more reliable, but more computationally intensive.

Butler, R.W., Huzurbazar, S., and Booth, J.G. (1992b) Saddlepoint approximations for the Bartlett-Nanda-Pillai trace statistic in multivariate analysis. *Biometrika* **79**, 705–16.

Compares the sequential saddlepoint of Fraser, Reid and Wong (1991), and the double saddlepoint of Barndorff-Nielsen and Cox (1979). In this setting the sequential saddlepoint is easier to compute than the double saddlepoint. Three versions of each approximation are computed: integrating the density approximation numerically, using a Lugannani and Rice version of the integral (in the double saddlepoint case this is Skovgaard's approximation) and using an r^* version of the integral, due to Barndorff-Nielsen. The sequential version with r^* was suggested in Pierce and Peters (1992). An interesting aspect is the discussion of consistency of the various approximations as $n \to \infty$; the limiting distribution of the test statistic is chi-squared, not normal. Although the Lugannani and Rice, and r^* versions are not consistent, this does not seem to matter for numerical work.

Carlin, B.P., Kass, R.E., Lerch, F.J., and Huguenard, B.R. (1992) Predicting working memory failure: a subjective Bayesian approach to model selection. *J. Am. Statist. Assoc.* **87**, 319–327

An approximation to the Bayes factor, derived from the Laplace approximation, is used to assess the sensitivity of inference to the choice of prior distribution, and to assess case influence.

Chang, T. and Villages, C. (1986) On a theorem of Stein relating Bayesian and classical inferences. *Canad. J. Statist.* **14**, 289–296.

Cheah, P.K., Fraser, D.A.S. and Reid, N. (1993) Some alternatives to Edgeworth. *Canad. J. Statist.* **21**, 131–139.

Cheah, P.K., Fraser, D.A.S. and Reid, N. (1994) Multiparameter testing in exponential models: third order approximations from likelihood. *Biometrika* **81**, 259–270.

Cordeiro, G. (1983) Improved likelihood ratio statistics for generalized linear models. *J. R. Statist. Soc.* B **45**, 404–413.

Cordeiro, G. (1987) On corrections to the likelihood ratio statistics. *Biometrika* **74**, 265–274.

Cox, D.R. and Hinkley, D.V. (1974) *Theoretical Statistics*. Chapman and Hall, London.

Daniels, H.E. (1954) Saddlepoint approximations in statistics. *Ann. Math. Statist.* **25**, 631–650.

Daniels, H.E. (1958) Discussion of "The regression analysis of binary sequences" by D.R. Cox *J. R. Statist. Soc.* B **20**, 236–238.

Daniels, H.E. (1983) Saddlepoint approximations for estimating equations. *Biometrika* **70**, 89–96.

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

Daniels, H.E. and Young, G.A. (1991) Saddlepoint approximation for the studentized mean, with an application to the bootstrap. *Biometrika* **78**, 169–79.

Davison, A.C. (1988) Approximate conditional inference in generalized linear models *J. R. Statist. Soc.* B **50**, 445–61

Applies the double saddlepoint density approximation of Daniels (1958) and Barndorff-Nielsen and Cox (1979) to the linear exponential model. At the end of Section 7, this technique is used to construct an added variable test in regression. Although a similar application was considered in Barndorff-Nielsen and Cox (1979), this hasn't been developed much further and probably has potential.

Davison, A.C. (1992) contribution to the discussion of "Practical use of higher order asymptotics for multiparameter exponential families" by Pierce and Peters. *J. R. Statist. Soc.* B **54**, 726–727.

Davison, A.C. and Hinkley, D.V. (1988) Saddlepoint approximations in resampling methods. *Biometrika* **75**, 417–31.

Shows how to use the empirical cumulant generating function to get saddlepoint approximations for statistics that condition on the observations, such as randomization or bootstrap statistics. A different derivation of their approximation to the multinomial bootstrap is given in Booth and Butler (1990). Their method requires the statistic to be linear in the observations. Methods of extending it to nonlinear statistics are discussed in Daniels and Young (1990) and DiCiccio, Martin and Young (1992).

DiCiccio, T.J. (1986) Approximate conditional inference for location families. *Canad. J. Statist.* **14**, 5–18.

DiCiccio, T.J. and Efron, B. (1992) More accurate confidence intervals in exponential families. *Biometrika* **79**, 231–45.

A method is proposed for approximating BC_a bootstrap confidence intervals analytically and its asymptotic properties are verified using Edgeworth expansion. The results on the parametric bootstrap restrict attention to exponential families. The results on the nonparametric bootstrap use the empirical cumulant generating function as in Davison and Hinkley (1988). Although saddlepoint methods are not used in this paper, it seems likely that their results are closely related to saddlepoint type approximations (as also suggested by Davison in the discussion to Pierce and Peters (1992).)

DiCiccio, T.J. and Field, C.A. (1991) An accurate method for approximate conditional and Bayesian inference about linear regression models from censored data. *Biometrika* **78**, 903–10.

Applies the double saddlepoint version of DiCiccio, Field and Fraser (1990) to some regression models. An argument is proposed to suggest that this version might more nearly approximate the sequential saddlepoint version if the scale parameter is transformed. This is the same as Pierce and Peters (1992) distinction between NP and INF, applied to transformation models. Three examples are approximated and compared to the exact value (obtained from simulation).

DiCiccio, T.J., Field, C.A., and Fraser, D.A.S. (1990) Approximations of marginal tail probabilities and inference for scalar parameters. *Biometrika* **77**, 77–95.

Develops a Lugannani and Rice type tail area approximation for transformation models. In the univariate case this is a special case of the parametrization invariant tail area approximation of Barndorff-Nielsen (1988) and Fraser (1990): here the derivation is based on a simple integration by parts argument. In the multivariate case the approximation is for the marginal density of one component. The multidimensional integration of the joint density is approximated by Laplace's method, as in Tierney and Kadane (1986) and then the univariate result is applied to the approximate marginal likelihood. The result is useful for conditional inference in location-scale and linear regression models, as well as for approximation of posterior cumulative distribution functions. An asymptotically equivalent expression in the multivariate case is presented which avoids the Laplace approximation, but is not quite as accurate in the examples, particularly for the scale parameter in regression-scale problems.

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

Applies the approximate integration method of DiCiccio, Field and Fraser (1990) to a Laplace approximation of Tierney, Kass and Kadane (1989) for the marginal posterior of a nonlinear function.

DiCiccio, T.J. and Martin, M.A. (1993) Simple modifications for signed roots of likelihood ratio statistics. *J. R. Statist. Soc.* B **55**, 305–316.

Applies the tail area approximation of DiCiccio and Martin (1991) to the case where there is a non-uniform prior, of the Stein-type (i.e. is a particular function of the information matrices, applicable when the interest parameter and nuisance parameters are orthogonal). The resulting tail area approximation uses a type of score statistic (modified by the prior). Under orthogonal parametrization, this statistic can be shown to approximate (to $O_p(n^{-1})$) the general statistic of Barndorff-Nielsen (1986, 1991), which requires sample space differentiation (and thus specification of an ancillary statistic). A sequential saddlepoint approximation using the Cox and Reid likelihood (now viewed as an approximate posterior, as in Sweeting (1987)) is also illustrated. Example 1 highlights the distinction made in Pierce and Peters (1992) of the corrections for nuisance parameters (NP) and the information correction (INF) and confirms that the NP correction is typically much more important.

DiCiccio, T.J. , Martin, M.A., and Young, G.A. (1994) Analytic approximation to bootstrap distribution functions using saddlepoint methods. *Statist. Sinica* **4**, 281–296.

Extends method of Daniels and Young (1990) to statistics that are smooth functions of sample means, using the approximation developed in DiCiccio and Martin (1991). The saddlepoint approximation is based on the multivariate cumulant generating function.

DiCiccio, T.J. , Martin, M.A., and Young, G.A. (1992) Fast and accurate approximation to double bootstrap confidence intervals. *Biometrika* **79**, 285–96.

DiCiccio, T.J. and Stern, S.E. (1993) On Bartlett adjustments for approximate Bayesian inference. *Biometrika* **80**, 731–740.

Feuerverger, A. (1989) On the empirical saddlepoint approximation. *Biometrika* **76**, 457–464.

Field, C.A. (1982) Small sample asymptotic expansions for multivariate M-estimates. *Ann. Statist.* **10**, 672–689.

Field, C.A. and Hampel, F.R. (1982) Small sample asymptotic distributions of M-estimates of location. *Biometrika* **69**, 29–46.

Field, C. and Ronchetti, E. (1990) *Small Sample Asymptotics.* Institute of Mathematical Statistics, Hayward.

Rigorous derivation of the saddlepoint density approximation, and derivation of saddlepoint approximations for estimates derived from estimating equations. The emphasis in the latter work is on the use of exponential tilting.

Firth, D. (1990) Generalized linear models. in *Statistical Theory and Modelling.* D.V. Hinkley, N. Reid and E.J. Snell, eds. 55–82. Chapman and Hall, London.

Fraser, D.A.S. (1979) *Inference and Linear Models.* McGraw Hill, New York.

Fraser, D.A.S. (1988) Normed likelihood as saddlepoint approximation. *J. Mult. Anal.* **27**, 181–93.

Discusses a local, but non-asymptotic, interpretation of Barndorff-Nielsen's p^* formula. The p^* formula can be derived by a Laplace approximation (see also Barndorff-Nielsen, 1990c), and can also be derived as a saddlepoint approximation for an approximating exponential model, called the tangent exponential model. The tangent exponential model is further exploited in Fraser (1990) and Fraser and Reid (1993).

Fraser, D.A.S. (1990) Tail probabilities from observed likelihoods. *Biometrika* **77**, 65–76.

Derives the parametrization invariant tail area approximation by first noting that an arbitrary exponential model is completely determined by the observed likelihood function, its derivative with respect to the parameter, and its derivative with respect to the sufficient statistic at the observed data point only. The canonical parameter is given by $\{\partial l(\phi; x_0)/\partial x\}\{\partial^2 l(\phi_0; x_0)/\partial\phi\partial x\}^{-1}$ *and the canonical variable or minimal sufficient statistic is given by* $\{\partial l(\phi_0; x)/\partial\phi\}$ *where* ϕ_0 *is the maximum likelihood value for* ϕ *when* $x = x_0$. *An exponential model using these definitions of the canonical parameter and sufficient statistic can be constructed from any initial density for which the parameter and the data have the same dimension. (This may have been obtained by a preliminary reduction by sufficiency or ancillarity.) This exponential model is called the tangent exponential model (tangent at the observed data point* x_0*). The Lugannani and Rice approximation applied to the tangent exponential model gives a parameterization invariant tail area approximation for any one parameter model.*

The approximation coincides with the version proposed in Barndorff-Nielsen (1988), although it is defined slightly differently here. Several numerical examples are given for one parameter location and exponential models.

Fraser, D.A.S. (1991) Statistical inference: likelihood to significance. *J. Am. Statist. Assoc.* **86**, 258–265.

Fraser, D.A.S., Lee, H.S. and Reid, N. (1990) Nonormal linear regression: an example of significance levels in high dimensions. *Biometrika* **77**, 333–41.

In nonnormal regression models inference about a component parameter is constructed from a marginal density that integrates out the remaining components. In this paper this marginal density is approximated by a conditional density that is constructed to be as nearly independent of the conditioning as possible. Applying the invariant tail area approximation of Fraser (1990) to the resulting approximate density gives the same approximation as that of DiCiccio, Field and Fraser (1990). An advantage of the approach in this paper is that a Monte Carlo procedure is developed that checks on the accuracy of the approximation.

Fraser, D.A.S., Reid, N. and Wong, A. (1991) Exponential linear models: a two-pass procedure for saddlepoint approximation. *J. R. Statist. Soc.* B **53**, 483–92.

Develops a sequential saddlepoint approximation for a component of the canonical parameter in an exponential family. It is also suggested that the first and second derivatives of the approximate conditional likelihood (which may be difficult to compute analytically) be approximated by finite differences or a suitably smoothed version of these. The approximation to the significance level starts with the same likelihood function as the double saddlepoint approximation, but gives a different tail area approximation. In particular examples one or the other may be more accurate. An explanation of this is given in Pierce and Peters (1992), and some numerical comparisons of the two methods are given there and in Butler, Huzurbazar and Booth (1992b).

Fraser, D.A.S. and Reid, N. (1993) Third order asymptotic models: likelihood functiions leading to accurate approximations to distribution functions. *Statist. Sinica* **3**, 67–82.

Shows that a fourth order expansion for a log-density function leads to a fourth order expansion for its cumulant generating function, and vice versa, with a simple correspondence between the coefficients. This provides a straightforward way to embed any model in an exponential model and provides an easy verification that the parametrization invariant tail area approximation has relative error $O(n^{-3/2})$.

Fraser, D.A.S. and Reid, N. (1995) Ancillaries and third order significance. *Util. Math.*, **47** 33–53.

Gu, M.-G. (1992) On the Edgeworth expansion and bootstrap approximation for the Cox regression model under random censorship. *Canad. J. Statist.* **20**, 399–414.

Homble, and McCormick, W. (1993) Tail areas for randomly stopped sums defined on a Markov chain. *Comm. Statist.* **9**, 563–584.

Jensen, J.L. (1986) Similar tests and the standardized log likelihood ratio statistic. *Biometrika* **73**, 567–572.

Jensen, J.L. (1988) Uniform saddlepoint approximations. *Adv. Appl. Probab.* **20**, 622–34.

Jensen, J.L. (1991a) Uniform saddlepoint approximations and log-concave densities. *J. R. Statist. Soc.* B **53**, 157–72.

Jensen, J.L. (1991b) A large deviation type approximation for the Box class of likelihood ratio criteria. *J. Am. Statist. Assoc.* **86**, 437–40.

Jensen, J.L. (1991c) Saddlepoint expansions for sums of Markov dependent variables on a continuous state space. *Probab. Theory and Rel. Fields* **89**, 181–199.

Jensen, J.L. (1992) The modified signed likelihood statistic and saddlepoint approximations. *Biometrika* **79**, 693–704.

If a Lugannani and Rice type approximation is available, it can be converted to an r^* type approximation while maintaining the same order of approximation. In exponential families with or without nuisance parameters, this order is $O(n^{-3/2})$ in shrinking neighbourhoods of the m.l.e. and $O(n^{-1})$ in compact neighbourhoods, and the error is uniform in compact neighbourhoods.

Jensen, J.L. (1995) A simple derivation of a natural large deviatio modified likelihood ratio statistic. Technical Report 328, Department of Theoretical Statistics, University of Aarhus, Denmark.

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

Kass, R.E. and Slate, E.H. (1992) Reparametrization and diagnostics of posterior nonnormality. in *Bayesian Statistics IV.*, Bernardo, J.M., Berger, J.O., Dawid, A.P. and Smith, A.F.M., eds. 289–305. Oxford University Press.

Kass, R.E., Tierney, L.J. and Kadane, J.B. (1988) Asymptotics in Bayesian computation. in *Bayesian Statistics III.*, Bernardo, J.M., DeGroot, M.H., Lindley, D.V. and Smith, A.F.M. eds. Clarendon Press, Oxford.

Kolassa, J.E. (1992) Saddlepoint approximations in the case of intractable cumulant generating functions. in *Selected Proceedings of the Sheffield Symposium on Applied Probability*, I.V. Basawa and R.L. Taylor, eds. IMS Lecture Notes Vol. 18, pp. 236-55.

Kolassa, J.E. (1994) *Series Approximation Methods in Statistics.* Springer-Verlag, New York.

Provides a very detailed derivation of the Edgeworth and saddlepoint expansions for the density of the sample mean, and of expansions for the cumulative distribution function of the sample mean, including the Lugannani and Rice approximation, Robinson's (1982) approximation, and Skates (1993) secant approximation, along with numerical comparisons.

Kolassa, J.E. and Tanner, M.A. (1994) Approximate conditional inference in exponential families via the Gibbs sampler. *J. Am. Statist. Assoc.* **89**, 697–705.

Lawless, J.A. (1982) *Statistical Models and Methods for Lifetime Data.* J. Wiley & Sons, New York.

Lawley, B. (1956) A general method for approximating to the distribution of the likelihood ratio criterion. *Biometrika* **43**, 295–303.

Lindley, D.V. (1992) contribution to the discussion of "Practical use of higher order asymptotics for multiparameter exponential families" by Pierce and Peters. *J. R. Statist. Soc.* B **54**, 728.

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

McCullagh, P. (1984) Local sufficiency. *Biometrika* **71**, 233–244.

McCullah, P. (1987) *Tensor Methods in Statistics.* Chapman and Hall, London.

McCullagh, P. and Nelder, J.A. (1989) *Generalized Linear Models.*, 2nd ed. Chapman and Hall, London.

Monti, A.C. and Ronchetti, E. (1993) On the relationship between empirical likelihood and empirical saddlepoint approximation for multivariate M-estimators. *Biometrika* **80**, 329–38.

The empirical saddlepoint density for M-estimators is given in Ronchetti and Welsh (1994). An expansion of the empirical cumulant generating function is given in Section 2. In Section 3 an expression is derived for Owen's empirical likelihood, and it is also expanded in powers of n. This enables a direct comparison of the empirical likelihood ratio statistic ($\hat{W}$ in their notation) with the empirical c.g.f. ($\hat{K}$ in their notation). There are interesting connections made to the BC_a constant (DiCiccio and Efron, 1992), and to results of DiCiccio, Field and Fraser (1990). There should be a tail area version of this result, and as well the results may well apply directly to the Cox partial likelihood estimator.

Mykland, P.A. (1993) Asymptotic expansions for martingales. *Ann. Probab.* **21**, 800–818.

Nelder, J.A. and Pregibon, D. (1987) An extended quasi-likelihood function. *Biometrika* **74**, 221–232.

Pierce, D.A. and Peters, D. (1992) Practical use of higher order asymptotics for multiparameter exponential families (with discussion). *J. R. Statist. Soc.* B **54**, 701–38.

Investigates the use of tail area approximations for computing p-values and confidence intervals for a scalar component of the canonical parameter in a full exponential family. The r^*, double and sequential saddlepoint approximations are all compared to first-order and exact methods. The r^* approximation is slightly better than the other two, and the double saddlepoint approximation is the least accurate of the three. All the methods are substantial improvements on first order methods and are reliable in all but the most extreme cases. Attention is restricted to discrete models (binomial, multinomial, and Poisson) and an excellent discussion of continuity corrections is provided in Section 4. A useful aspect of their development is separate consideration of the

role of an adjustment for nuisance parameters (typically through an adjustment to the likelihood function) and an adjustment for curvature (which they call an information adjustment). The accuracy of higher order approximations depends on the relative magnitudes of these two adjustments. The introductory and concluding sections are exceptionally clear summaries of their work and related literature, and the derivation of Barndorff-Nielsen's result that is sketched in the Appendix is also very helpful.

Pierce, D. A. & Peters, D. (1994) Higher order asymptotics and the likelihood principle: one parameter models. *Biometrika* **81**, 1–10.

Reid, N. (1988) Saddlepoint methods and statistical inference (with discussion). *Statist. Science* **3**, 213–38.

Reid, N. (1993) Making higher order asymptotics useful. in*Proceedings of a Conference on Statistical Inference and Biostatistics* ed. D.A. Sprott, CIMAT, Mexico. 227–246.

Reid, N. (1995a) Likelihood and Bayesian approximation methods. in *Bayesian Statistics V* eds. J. O. Berger, J. M. Bernardo, D.V. Lindley, and A.F.M. Smith. Oxford University Press., to appear

Reid, N. (1995b) The roles of conditioning in inference. *Statist. Sci.*, to appear.

Reid, N. and Fraser, D.A.S. (1990) Accurate approximation of tail areas using the likelihood function. in*Proceedings of IV Clapem* eds. E. Cabaña et al., Instituto Nacional de Estadística, Geografía e Informática, México. 36–50.

Robinson, J. (1982) Saddlepoint approximations for permuation tests and confidence intervals. *J. R. Statist. Soc.* B **67**, 91–101.

Ronchetti, E. (1990) Small sample asymptotics: A review with applications to robust statistics. *Comp. Stat. Data Anal.* **3**, 207–23.

Outlines the derivation of the saddlepoint approximation using the method of steepest descents (includes a picture of the path of steepest descent) and the derivation by Field and Hampel (1982) which constructs directly an expansion of $f'(t)/f(t)$. Includes a survey of one parameter examples from the literature, for means and M-estimates, and for the one-sample Wilcoxon statistic. Brief discussion of tail area approximation.

Ronchetti, E. and Welsh, A. (1994) Empirical saddlepoint approximations for multivariate M-estimators. *J. R. Statist. Soc.* B **56**, 313–326.

Feuerverger (1989) studied the empirical saddlepoint approximation with special emphasis on the sample mean. Further results were obtained in Feuerverger, Jing and Robinson (1994). This paper extends Feuerverger's work to multivariate M-estimators. The discussion of the order of the error and renormalization in Section 1 is helpful.

Skates, S.J. (1993) On secant approximations to cumulative distribution functions. *Biometrika* **80**, 223–36.

Discussion of the Lugannani and Rice approximation, including higher order terms. Uses the method of Temme (1982), but gets simpler expressions for the needed higher order terms. Also discusses uniformity of the approximation, and applies it to the Pearson Type IV family.

Skovgaard, I.M. (1987) Saddlepoint expansions for conditional distributions. *J. Appl. Probab.* **24**, 875–87.

Derives a result analogous to that of Lugannani and Rice (1980) for the conditional distribution of $\bar{X}$, given $\bar{Y}$. Mentions that using the cumulant generating function for the conditional density in the usual Lugannani and Rice formula may be more accurate. This is borne out in later numerical work comparing Skovgaard's method (now called the double saddlepoint method, although Skovgaard did not obtain it by saddlepointing numerator and denominator separately) with the sequential saddlepoint of Fraser, Reid and Wong (1991). Skovgaard also suggests in this paper that a combination of the double and sequential saddlepoint might be used: this seems not to have been followed up, and might prove useful in things like Gibbs samplers.

Skovgaard, I.M. (1990) On the density of minimum contrast estimators. *Ann. Statist.* **18**, 779–89.

Very helpful for clarifying the role of the approximate ancillary in Barndorff-Nielsen's p^* formula, as well as providing a very clear proof of p^*. Considers the exponential family case, the exact ancillary case, and the approximate ancillary case; this order of discussion is very useful. Also makes explicit what is needed for approximate ancillarity. Because the proof doesn't use any special properties of the likelihood function, the approximation can be obtained for a family of estimators that minimize a contrast function.

Skovgaard, I.M. (1995) A general large deviation approxiamtion to one-parameter tests. Technical Report, Department of Mathematics and Physics, The Royal Veterinary and Agricultural University, Denmark.

Spady, R.H. (1991) Saddlepoint approximation for regression models. *Biometrika* **78**, 879–89.

Develops a saddlepoint density formula for regression M-estimates, which is most conveniently applied when the cumulant generating function of the ψ-function is known. This is related to but slightly different than the formula derived in Daniels (1983). In order to compute tail areas high dimensional integration will be needed. Suggests that the technique of DiCiccio, Field and Fraser (1990) or Laplace approximation could be used, but this does not seem to have been followed up yet.

Sweeting, T.J. (1995a) Approximate Bayesian computation based on signed roots of log-density ratios. in *Bayesian Statistics V* eds. J. O. Berger, J. M. Bernardo, D.V. Lindley, and A.F.M. Smith. Oxford University Press., to appear.

Sweeting, T.J. (1995b) A framework for Bayesian and likelihood approximations in statistics. *Biometrika* **82**, 1–24.

Temme, N.M. (1982) The uniform asymptotic expansions of a class of integrals related to cumulative distribution functions. *SIAM J. Math. Anal.* **13**, 239–253.

Tierney, L.J. (1990) *Lisp-Stat: An Object-Oriented Environment for Statistical Computing and Dynamic Graphics.* J. Wiley & Sons, New York.

Tierney, L.J. and Kadane, J.B. (1986) Accurate approximation for posterior moments and marginal densities. *J. Am. Statist. Assoc.* **81**, 82–87.

Tierney, L.J., Kass, R.E., and Kadane, J.B. (1989) Approximation of marginal densities of nonlinear functions. *Biometrika* **76**, 425–33. with Correction (1991) *Biometrika* **78**, 233–4.

Wang, S. (1990a) Saddlepoint approximation for nonlinear statistics. *Canad. J. Statist.* **18**, 255–64.

Establishes a saddlepoint approximation for a nonlinear statistic which has a von Mises expansion to second order: the explicit form of the quadratic term in this expansion comes into the result. The approximation established is similar to that of Skovgaard (1987), as are the techniques of proof.

Wang, S. (1990b) Saddlepoint approximation for the bivariate distribution. *J. Appl. Probab.* **27**, 586–97.

Wang, S. (1990c) Saddlepoint approximations in resampling. *Ann. Inst. Statist. Math.* **42**, 115–31.

Provides the techinical details for justifying the saddlepoint approximation to the bootstrap distribution derived in Davison and Hinkley (1988). In the bootstrap case the relative error is no longer uniform in compact neighbourhoods.

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

Uses Easton and Ronchetti's (1986) idea of replacing the cumulant generating function by the first four terms of its Taylor series expansion. To avoid the problem that the saddlepoint equation based on this fourth degree polynomial may not have a unique root, an adjustment to the cumulant generating function is proposed that has an arbitrary constant b which is chosen to obtain a unique solution. May be related to Cheah, Fraser and Reid (1993). The technique is adapted to the bootstrap saddlepoint, where it is shown to work very well numerically.

Wang, S. (1993) Saddlepoint methods for bootstrap confidence bands in nonparametric regression. *Austr. J. Statist.* **35**, 93–101.

Wong, W.H. and Li, B. (1992) Laplace expansion for posterior densities of nonlinear functions of parameters. *Biometrika* **79**, 393–8.

Wood, A.T.A., Booth, J.G. and Butler, R.W. (1993) Saddlepoint approximations to the cdf of some statistics with non-normal limit distributions. *J. Am. Statist. Assoc.* **88**, 680–686.

A tail area approximation of the Lugannani and Rice type is derived following Section 4 of Daniels (1987), but using a non-normal kernel function. Both one-parameter

and multi-parameter (i.e. conditional) problems are treated. The gamma kernel (with somewhat ad hoc choices of shape parameter) provides excellent results, but so does the normal kernel. In the conditional case, double saddlepoint and sequential saddlepoint methods are both considered, as are r^* versions of them. The sequential methods are much better than the double methods, but the gamma based kernel is not really much better than the normal based kernel. The r^* version and the Lugannani and Rice version are about equally good.