Accurate approximation of tail probabilities using the likelihood function

N. Reid
Department of Statistics
University of Toronto
Toronto, Canada M5S 1A1

D.A.S. Fraser
Department of Mathematics
York University
North York, Canada M3J 1P3

In exponential families, the approximation to the cumulative distribution function of the sample mean can be simply expressed in terms of likelihood quantities. This means that the likelihood function can be converted to an approximate tail area or significance level. In more general contexts, conditional or marginal likelihood functions can be similarly used to compute tail areas. The conversion is described and illustrated, and extensions to more complex models are briefly considered.

KEY WORDS: asymptotic methods, exponential family, likelihood, saddlepoint approximation, tail area approximation

1. Introduction

In this paper we are concerned with accurate approximation of tail probabilities, i.e. probabilities of the form $\Pr_\theta(T \geq t^0)$, where T is a random variable whose distribution depends on θ, and t^0 is some fixed value. If the distribution of T is continuous, this is equivalent to approximating the cumulative distribution function of T, and in applications to significance testing we may often want to compute $F_\theta(t^0) = \Pr_\theta(T \leq t^0)$. Typically T is a statistic derived from a sequence $(X_1, \ldots, X_n)$ and t^0 is the observed value of T for a sample $(x_1, \ldots, x_n)$. Approximating the tail probability is then equivalent to approximating the observed significance level or p-value for testing the value θ. In the case that T has a lattice distribution, some continuity correction may be needed to convert the cumulative distribution function into a p-value. This is discussed in Daniels (1987), and Barndorff-Nielsen and Cox (1989, Ch. 4), and several examples are considered in Pierce and Peters (1991).

The approximation we will consider is best illustrated for a sample mean. Let $X_1, \ldots, X_n$ be independent, identically distributed random variables, and $\bar{X} = n^{-1} \sum X_i$. Define the cumulant generating function of X_i by $K(t) = \log[E\{\exp(tX_i)\}]$. The saddlepoint approximation to the density of $\bar{X}$ is given by

$$\hat{f}_{\bar{X}}(\bar{x}) = c_n \{n/|K''(\hat{t})|\}^{1/2} \exp\{nK(\hat{t}) - n\hat{t}\bar{x}\}, \tag{1}$$

where $\hat{t} = \hat{t}(\bar{x})$ is called the saddlepoint and is defined by

$$K'(\hat{t}) = \bar{x}. \tag{2}$$

Approximation (1) has a relative error of $O(n^{-3/2})$, i.e.

$$f_{\bar{X}}(\bar{x}) = \hat{f}_{\bar{X}}(\bar{x})\{1 + O(n^{-3/2})\} \tag{3}$$

and although the magnitude of the relative error depends on $\bar{x}$, the approximation is often very accurate even in the extreme tails of the density. The normalizing constant c_n

defined in (1) can be shown to satisfy

$$c_n = \{1/\sqrt{(2\pi)}\}^p\{1 + O(n^{-1})\}, \tag{4}$$

when X_i is a p-dimensional vector. (In the case that X_i is a vector, $K''(t)$ is the $p \times p$ matrix of partial derivatives of $K(t)$ with respect to the p components of t.)

The saddlepoint approximation to the density of $\bar{X}$ is derived in Daniels (1954), by two different methods, and illustrated in several examples. A detailed discussion is provided there of the regularity conditions required on the density of X_i, and of the behaviour of the error term for values of $\bar{x}$ tending to the endpoints of the support of $f_{\bar{X}}(\bar{x})$. An overview of the density approximation, some applications of it to problems in statistical inference, and several further references, are given in Reid (1988). Good textbook references are Barndorff-Nielsen and Cox (1989, Ch.4), and McCullagh (1987, Ch. 6).

Approximation (1) can be integrated to give an approximation to the cumulative distribution function of $\bar{X}$ when X_i and hence $\bar{X}$ are scalars. The key substitution is to let $nK(\hat{t}) - n\hat{t}K'(\hat{t}) = -r^2/2$, from which $r\mathrm{d}r = n\hat{t}K''(\hat{t})\mathrm{d}\hat{t} = n\hat{t}\mathrm{d}\bar{x}$. The resulting approximation is

$$\hat{F}_{\bar{X}}(\bar{x}) = \Phi(r) + \phi(r)(r^{-1} - q^{-1}) \tag{5}$$

where

$$\begin{aligned} r &= \pm\sqrt{n}[2\{-K(\hat{t}) + \hat{t}K'(\hat{t})\}]^{1/2}, \\ q &= \sqrt{n}\hat{t}\{K''(\hat{t})\}^{1/2}. \end{aligned} \tag{6}$$

The sign of r is chosen to match the sign of q. Approximation (5) is the basis for tail area approximations for more general statistics. It has absolute error $O(n^{-3/2})$, i.e.

$$F_{\bar{X}}(\bar{x}) = \hat{F}_{\bar{X}}(\bar{x}) + O(n^{-3/2}), \tag{7}$$

for $\bar{x}$ in $O(n^{-1/2})$ neighbourhoods of the mean, and error $O(n^{-1})$ for $\bar{x}$ in $O(1)$ neighbourhoods of the mean. It was derived by Lugannani and Rice (1980) and is often called the "Lugannani and Rice" approximation. It is derived by several routes, and illustrated, in Daniels (1987). A somewhat simpler derivation is given in Barndorff-Nielsen and Cox (1989, Ch.4), adapting an argument due to Temme (1982). Note that after making the change of variable from $\bar{x}$ to $\hat{t}$ to r, the integral of (1) is $\int_{-\infty}^{r_0} c_n \exp(-r^2/2)(r/q)\mathrm{d}r = c_n' \int_{-\infty}^{r_0} \phi(r)(r/q)\mathrm{d}r$, where $c_n' = 1 + O(n^{-1})$. Working with the unnormalized form of the integral, we have

$$\begin{aligned}\int_{-\infty}^{r_0} \phi(r)(r/q)\mathrm{d}r &= \int_{-\infty}^{r_0} \phi(r)(1 + r/q - 1)\mathrm{d}r \\ &= \Phi(r_0) + \int_{-\infty}^{r_0} \phi(r)r\left(\frac{1}{q} - \frac{1}{r}\right)\mathrm{d}r \\ &= \Phi(r_0) - \phi(r_0)\left(\frac{1}{q_0} - \frac{1}{r_0}\right) + \int \phi(r)\mathrm{d}\left(\frac{1}{q} - \frac{1}{r}\right).\end{aligned}$$

We now apply the results

$$\begin{aligned}\left(\frac{1}{r} - \frac{1}{q}\right) &= O(n^{-1/2}) \\ \mathrm{d}\left(\frac{1}{r} - \frac{1}{q}\right) &= \frac{a}{n}\mathrm{d}r + O(n^{-3/2})\end{aligned} \tag{8}$$

to get

$$\int_{-\infty}^{r_0} \phi(r)(r/q)\mathrm{d}r = \Phi(r_0)(1 + \frac{a}{n}) + \phi(r_0)\left(\frac{1}{r_0} - \frac{1}{q_0}\right)$$

and hence

$$(1 + a/n)^{-1} \int_{-\infty}^{r_0} \phi(r)(r/q)\mathrm{d}r = \Phi(r_0) + \phi(r_0)\left(\frac{1}{r_0} - \frac{1}{q_0}\right) + O(n^{-3/2}).$$

Since the right hand side approaches 1 as $r_0 \to \infty$, we have verified that a/n is the $O(n^{-1})$ term in c_n', and obtained the Lugannani and Rice result (5). The main work lies in verifying (8), which is done by examining in detail the expansion of q in terms of r. Note from (6) that $nr^2/2$ gives the first two terms in the Taylor expansion of $K'(0) = 0$ about $\hat{t}$, and $nq^2/2$ is the third term. From this it can be shown that $r - q = O(n^{-1})$.

The particular integration by parts step above is given in Fraser and Reid (1991); see also DiCiccio, Field and Fraser (1990). There are many different ways of establishing (8), or directly relating the $O(n^{-1})$ term in the expansion for c_n' to the difference between r

and q. In Fraser and Reid (1990), expansions for r and q are obtained from the fourth order expansions of either the cumulant generating function or the likelihood function. In Daniels (1987), q is expanded as a function of r directly. Barndorff-Nielsen and Cox (1989, Ch. 4) verify directly that $r/q = 1 + O(n^{-1})$. McCullagh's results (1987, Ch. 6) can be used to show that the $O(n^{-1})$ term in c_n' is equal to a in (8).

Note that as derived, the right hand side of (5) is an approximation to the cumulative distribution function of r, but since r is a one-to-one function of $\bar{X}$, this is equivalent to approximating $F_{\bar{X}}(\cdot)$.

2. The exponential family

Suppose each X_i comes from an exponential family density, which we write as

$$f(x;\theta) = \exp\{\theta x - k(\theta) + d(x)\}.$$

Define the likelihood function $L(\theta) = L(\theta; x_1, \ldots, x_n) = \Pi f(x_i;\theta)$, and the Fisher information function $j(\theta) = -\partial^2 \log\{L(\theta)\}/\partial\theta^2$. Then from (6) we obtain

$$\begin{aligned} r &= \operatorname{sign}(\hat{\theta} - \theta)[2\log\{L(\hat{\theta})/L(\theta)\}]^{1/2} \\ q &= (\hat{\theta} - \theta)|j(\hat{\theta})|^{1/2}, \end{aligned} \tag{9}$$

where $\hat{\theta}$ is the maximum likelihood estimate of θ, and satisfies the equation $k'(\hat{\theta}) = \bar{x}$. Note that $K'(t) = k'(\theta + t)$, showing that $\hat{t} = \hat{\theta} - \theta$, and that $K''(\hat{t}) = k''(\hat{\theta})$. Since $\hat{\theta}$ is a one-to-one function of $\bar{x}$, we can write

$$F_{\hat{\Theta}}(\hat{\theta};\theta) = \Phi(r) + \phi(r)\left(\frac{1}{r} - \frac{1}{q}\right) + O(n^{-3/2}), \tag{10}$$

with r and q defined in (9). This shows that the distribution of the maximum likelihood estimate can be accurately approximated by a relatively simple function of the signed square root of the likelihood ratio statistic, r, and the Wald statistic, q.

The density approximation (1) also has a likelihood formulation for exponential families, which is

$$\hat{f}_{\hat{\Theta}}(\hat{\theta}) = c_n|j(\hat{\theta})|^{1/2}\{L(\theta)/L(\hat{\theta})\} \tag{11}$$

and is often referred to as Barndorff-Nielsen's approximation (Barndorff-Nielsen, 1980, 1983). Barndorff-Nielsen and others have showed that (11), appropriately interpreted, also provides a valid approximation in a great many models outside the exponential family. A survey of some of these generalizations is given in Reid (1988). Similarly, the likelihood version of Lugannani and Rice's approximation can also be used in more general models. This is the basis for several recent developments in tail area approximations.

An important feature of the likelihood function in the exponential family is that is depends on the data $x = (x_1, \ldots, x_n)$ only through the maximum likelihood estimator, which is minimal sufficient. Thus there is little ambiguity in writing $L(\theta)/L(\hat{\theta})$ for $L(\theta; x)/L(\hat{\theta}; x)$, because $L(\theta; x)/L(\hat{\theta}; x) = L(\theta; \hat{\theta})/L(\hat{\theta}; \hat{\theta})$.

3. The location family

Suppose each X_i comes from a location family density, which we write as

$$f(x; \theta) = f(x - \theta).$$

The likelihood function is $L(\theta; x) = \Pi f(x_i - \theta)$. Although in this model the likelihood function depends on the entire vector of observations, x, there is a one-to-one transformation from x to $(\hat{\theta}, a)$, where $a_i = x_i - \hat{\theta}$, a is a vector in R^{n-1}, and a is *ancillary*, i.e. the marginal density of a does not depend on θ. Thus we can write

$$f(x; \theta) \propto f_1(\hat{\theta}|a; \theta) f_2(a)$$

and

$$L(\theta; x) \propto L(\theta; \hat{\theta}, a) \propto f_1(\hat{\theta}|a; \theta).$$

It is fairly straightforward to show that the conditional distribution of $\hat{\theta}$, given a, is obtained by renorming the likelihood function. We can write this result, which is originally due to Fisher (1934), in the form

$$f(\hat{\theta}|a;\theta) = c|j(\hat{\theta})|^{1/2} L(\theta;\hat{\theta},a)/L(\hat{\theta};\hat{\theta},a) \tag{12}$$

which makes the analogy with (11) explicit (although in fact in the location model $j(\hat{\theta})$ depends only on a and not on $\hat{\theta}$). Result (12) holds for any transformation model (Barndorff-Nielsen, 1980).

We might thus expect that a result analogous to the tail area approximation (11) also holds in the location model, which is indeed the case. The result is

$$F(\hat{\theta}|a;\theta) = \Phi(r) + \phi(r)\left(\frac{1}{r} - \frac{1}{s}\right) + O(n^{-3/2}), \tag{13}$$

where r is the signed square root of the likelihood ratio statistic, as before, and s is the Rao statistic:

$$s = \frac{\partial}{\partial\theta} \log L(\theta)|j(\hat{\theta})|^{-1/2}. \tag{14}$$

Note that (13) provides an approximation to the conditional cumulative distribution function even though (12) is exact. This version of the tail area approximation was obtained in DiCiccio, Field and Fraser (1990) and Fraser (1990) as a special case of more general parametrization invariant versions of (10) due to Barndorff-Nielsen (1988, 1990) and Fraser (1990). Even though (10) was derived for exponential families, the similarity of the density appproximations (11) and (12) suggested that (10) could hold more generally.

4. General scalar parameter families

We assume that X_i is from a one-parameter density $f(x;\theta)$, but that there may not be a sufficiency reduction to $\hat{\theta}$, as in the exponential family case, or an ancillary reduction to $(\hat{\theta}, a)$, as in the transformation family case. We write $L(\theta) = L(\theta; x)$ for the likelihood function and $l(\theta)$ for the log-likelihood function. Define the quantity $t = t(\theta)$ by

$$t = \{\dot{l}(\hat{\theta}) - \dot{l}(\theta)\} k^{-1}(\hat{\theta}) |j(\hat{\theta})|^{1/2}, \tag{15}$$

where $\dot{l}(\theta) = \partial l(\theta; x)/\partial\hat{\theta}$, and $k(\theta) = \partial^2 l(\theta; x)/\partial\theta\partial\hat{\theta}$. Then a parametrization invariant version of the tail area approximation is

$$\hat{F}_{\hat{\boldsymbol{\theta}}}(\hat{\theta}; \theta) = \Phi(r) + \phi(r) \left(\frac{1}{r} - \frac{1}{t} \right), \tag{16}$$

where r is as usual the signed square root of the likelihood ratio statistic. In the partial differentiation of $l(\cdot)$ with respect to $\hat{\theta}$ in (15), some remaining $n-1$ coordinates in the sample space must be held fixed, and the notation doesn't indicate how to do this. In the location family, the ancillary coordinates a are held fixed, and the resulting distribution is conditional on a. In the exponential family, $l(\cdot)$ depends on x only through $\hat{\theta}$, so the choice of coordinates is not necessary. In general models, a reasonable strategy is to reduce by sufficiency, and then look for approximately ancillary statistics to condition on. This strategy is used in Barndorff-Nielsen (1990) in establishing the validity of (12) to $O(n^{-1})$ in general models. Tail areas computed this way are described in Barndorff-Nielsen (1991). An alternative approach for choosing sample space directions geometrically is described in Fraser and Reid (1988). None of the solutions suggested to date are particularly easy to implement.

The choice of coordinates is more important in multiparameter models, to be discussed in the next section. In exponential and transformation models, nuisance parameters can be eliminated by sufficiency or ancillarity arguments, but in general multiparameter models, some further reduction is needed and the methods mentioned above may provide some guidelines.

5. Nuisance parameters in exponential families

In this section we assume that the multiparameter density takes the form

$$f(x;\psi,\phi) = \exp\{\psi x_1 + \phi^T x_{(2)} - k(\psi,\phi) - d(x)\} \tag{17}$$

where x is the vector of sufficient statistics, $x_{(2)}$ is the subvector of the last $p-1$ components of x, ψ is the parameter of interest, and ϕ is a vector of nuisance parameters. The nuisance parameters can be eliminated by conditioning, because

$$f(x_1|x_{(2)};\psi) = \exp\{\psi x_1 - k_2(\psi) - d_2(x)\},$$

where k_2 and d_2 depend on $x_{(2)}$ in general, and are obtained from k and d in (17). The conditional log-likelihood function for ψ is thus

$$l^c(\psi) = \psi x_1 - k_2(\psi)$$

and for fixed $x_{(2)}$ this depends on the data only through x_1. That is, conditionally, the model is a scalar parameter exponential family. Applying the result (10) in Section 2, then, we have

$$\hat{F}_{\hat\psi}(\hat{\psi}^c;\psi) = \Phi(r^c) + \phi(r^c)\left(\frac{1}{r^c} - \frac{1}{q^c}\right), \tag{18}$$

where $\hat{\psi}^c$ maximizes $l^c(\cdot)$ over ψ, and

$$r^c = \pm[2\{l^c(\hat{\psi}^c) - l^c(\psi)\}]^{1/2}$$

$$q^c = (\hat{\psi}^c - \psi)| - l^{c\prime\prime}(\hat{\psi}^c)|^{1/2}$$

are the standardized likelihood ratio and Wald statistics for the conditional model. Equation (18) is an approximation to the conditional cumulative distribution function of $\hat{\psi}^c$, given $x_{(2)}$, and also approximates the conditional cumulative distribution function of r^c, x_1, or q^c, since these are one-to-one functions of each other for fixed $x_{(2)}$.

In applications, the explicit form for $k_2(\psi)$ may be difficult to obtain, particularly if $d(x)$ in (16) is not available in closed form. However, a saddlepoint approximation to the marginal density of $x_{(2)}$ gives

$$l^c(\psi) \doteq l(\psi,\hat{\phi}_\psi) + \frac{1}{2}\log|j_{\phi\phi}(\psi,\hat{\phi}_\psi)| \tag{19}$$

where $l(\psi, \hat{\phi}_\psi)$ is the log-likelihood for the full model (16), $\hat{\phi}_\psi$ is the maximum likelihood estimate of ϕ for fixed ψ, and $j_{\phi\phi} = -\partial^2 l(\psi, \phi)/\partial\phi\partial\phi^T$ is the nuisance parameter submatrix of the observed information matrix. This type of approximation is discussed in Barndorff- Nielsen and Cox (1979), Cox and Reid (1987), and Skovgaard (1987). It is accurate to $O_p(n^{-3/2})$, in $\sqrt{n}$ neighbourhoods of $\hat{\psi}^c$ (Fraser and Reid, 1991).

With $l^c(\psi)$ approximated by (19), (18) provides an approximation to the cumulative distribution function of $\hat{\psi}^c$ also accurate to $O(n^{-3/2})$, and to this order $\hat{\psi}^c$, r^c, and q^c can all be computed from the approximate conditional likelihood. Numerical examples of this approximation are discussed in Fraser, Reid and Wong (1991).

In many problems it will be difficult to obtain explicit expressions for $l^{c\prime\prime}(\hat{\psi}^c)$, or other components of the approximation. In Fraser, Reid and Wong (1991) first and second derivatives of $l^c(\cdot)$ are obtained numerically. Further analytical approximation to (18) is discussed in DiCiccio and Martin (1991). Some numerical examples appear in Pierce and Peters (1991), and Butler, Huzurbazar and Booth (1991). Jensen (1991) shows that (18) is closely related to the more general r^* approximation of Barndorff-Nielsen (1986, 1990) in this context.

6. Nuisance parameters in transformation families

We illustrate these results using a regression model: the general transformation case is discussed in DiCiccio, Field and Fraser (1990). Assume that we have the model

$$x = Z\beta + e \tag{20}$$

where x is $n \times 1$, Z is a $n \times p$ matrix of known constants, β is a $p \times 1$ regression vector, and e follows a distribution $f_0(e)$. This is a generalization of the model given in Section 3, and again there is a statistic describing the configuration of the sample that is ancillary, i.e. its distribution doesn't depend on β. This vector, of dimension $n-p$, can be expressed as

$$a = x - Z\hat{\beta}$$

where $\hat{\beta}$ is the maximum likelihood estimate of β. As in the location case, we use the conditional density $f(\hat{\beta}|a;\beta)$, which can be obtained exactly by renormalizing the likelihood function.

If we take β_1 as the parameter of interest, and $\beta_{(2)}$ as the nuisance parameter, we need a further reduction to a one-dimensional distribution. In general, this will be obtained by integrating out the nuisance parameters; in fact it can be shown that $\hat{\beta}_1$ has a conditional distribution free of any nuisance parameters, and this distribution is obtained as

$$f(\hat{\beta}_1|a;\beta_1) = \int \cdots \int f(\hat{\beta}|a;\beta)\mathrm{d}\hat{\beta}_2 \cdots \mathrm{d}\hat{\beta}_p \tag{21}$$

where the integrand in (21) is obtained from the likelihood function. This type of inference for regression is discussed in detail in Fraser (1979, Ch.4). Note that the marginalization involved in (21) is similar in form to that required for obtaining marginal posterior densities in Bayesian inference. The integration cannot usually be carried out exactly or numerically, but the integral can be approximated using Laplace's method (Tierney and Kadane, 1986; Kass, Tierney and Kadane, 1988). The result is that the log-likelihood function for β_1 is approximated by

$$l^m(\beta_1) \doteq l(\beta_1, \hat{\phi}_{\beta_1}) - \frac{1}{2} \log |j_{\phi\phi}(\beta_1, \hat{\phi}_{\beta_1})| \tag{22}$$

and this approximation is accurate to $O_p(n^{-3/2})$. In (22) we have written $\phi = \beta_{(2)}$ for the nuisance parameter. The superscript m is to indicate marginalization over the components $\hat{\beta}_{(2)}$. Using (22) and the location family result (13) we have

$$\hat{F}^m(\hat{\beta}_1; \beta_1) = \Phi(r^m) + \phi(r^m)\left(\frac{1}{r^m} - \frac{1}{q^m}\right) \tag{23}$$

where

$$r^m = \pm[2\{l^m(\hat{\beta}_1) - l^m(\beta_1)\}]^{1/2},$$

$$s^m = l^{m\prime}(\beta_1)| - l^{m\prime\prime}(\hat{\beta}_1)|^{-1/2}$$

are the standardized likelihood ratio and score statistics computed from the marginal log-likelihood. Approximation (23) is also accurate to $O(n^{-3/2})$ in $\sqrt{n}$ neighbourhoods $\hat{\beta}_1$.

The argument applies equally to any component of $\hat{\beta}$. Extensions given in DiCiccio, Field and Fraser (1990) allow introduction of an unknown scale parameter; i.e. (20) is replaced by $x = Z\beta + \sigma e$.

7. Discussion

In most applications, the parameter of interest will not be a component of the canonical parameter in an exponential model, nor a component of the transformation parameter in a transformation model, so a simple reduction to a marginal or conditional likelihood will not be available. In such cases some extension is needed, for example of the general one-parameter version (16) given in Section 4. This would involve conditioning and/or marginalizing to sufficient and/or ancillary statistics as a preliminary reduction, and then a further dimension reduction obtained by choosing a direction in which to differentiate on the sample space. Dimension reduction by conditioning on approximate ancillary statistics is discussed in Barndorff-Nielsen (1990, 1991). The approximate ancillary statistics can be difficult to compute, but in curved exponential families general results are presented there, and also in Jensen (1991).

8. References

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**, 343–65.

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. (1990). Approximate interval probabilities. *J. R. Statist. Soc.* B **52**, 485–96.

Barndorff-Nielsen, O.E. (1991). Modified signed log likelihood ratio. *Biometrika* **78**, ,to appear.

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

Barndorff-Nielsen and Cox (1989). *A*.symptotic Techniques for Use in Statistics, Chapman and Hall, London.

Butler, R.W., Huzurbazar, S. and Booth, J.G. (1990). Saddlepoint approixmations to the generalized variance and Wilks' statistic. *Biometrika* **78**, ,to appear.

Cox, D.R. and Reid, N. (1987). Parameter orthogonality and approximate conditional inference (with discussion). *J. R. Statist. Soc.* B **49**, 1–39.

Daniels, H.E. (1954). Saddlepoint approximations in statistics. *Ann. Statist.* **25**, 631–50.

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

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

DiCiccio, T.J. and Martin, M.A. (1990). Approximations of marginal tail probabilities for a class of smooth functions with applications to Bayesian and conditionals inference. *Technical Report No. 346*. Department of Statistics, Stanford University.

Fisher, R.A. (1934). Two new properties of mathematical likelihood. *Proc. Roy. Soc.* A **144**, 285–307.

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

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

Fraser, D.A.S. and Reid, N. (1988). On conditional inference for a real parameter: A differential approach on the sample space. *Biometrika* **75**, 251–64.

Fraser, D.A.S. and Reid, N. (1991). Simple asymptotic connection between density and cumulant generating functions leading to accurate approximations for distribution functions. submitted for publication.

Fraser, D.A.S., Reid, N. and Wong, A. (1991). From observed likelihood to tail area: a two pass approximation. *J. R. Statist. Soc.* B **52**, 483–92.

Jensen, J.L. (1991). The relation of r^* to saddlepoint approximations. *Technical report No. 223*, Department of Theoretical Statistics, Aarhus.

Kass, R.E., Tierney, L.J. and Kadane, J.B. (1988). Asymptotics in Bayesian computation. in *Bayesian Statistics 3*, eds. J.M. Bernardo, M.H. DeGroot, D.V. Lindley, A.F.M. Smith, Oxford University Press, 263-78.

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

McCullagh, P. (1987). *T.*ensor methods in statistics. Chapman and Hall, London.

Pierce, D.A. and Peters, D. (1991). Practical use of higher order asymptotics for multiparameter exponential families. *Technical Report*, Department of Statistics, Oregon State University.

Reid, N. (1988). Saddlepoint methods and statistical inference. *Statist. Sci.* **3**, 213-37.

Skovgaard, I.M. (1987). Saddlepoint expansions for conditional distributions. *J. Appl. Prob.* **24(4)**, 875-887.

Skovgaard, I.M. (1988). Saddlepoint expansions for directional test probabilities. *J. R. Stat. Soc. B* **50**, 269-280.

Temme, N.M. (1982). The uniform asymptotic expansion of a class of integrals related to cumulative distribution functions. *SIAM J. Math. Anal.* **3**, 221–49.