Statistical Science
1988, Vol. 3, No. 2, 213–238

Saddlepoint Methods and Statistical Inference

N. Reid

Abstract. This paper reviews Daniels' saddlepoint approximation to the distribution of the mean of a random sample, and the many aspects of second order asymptotic inference that have been developed from it. These include Barndorff-Nielsen's approximation to the distribution of the maximum likelihood estimate, Bartlett factors for the likelihood ratio statistic and approximations to predictive and conditional likelihood. The emphasis is on statistical applications of the saddlepoint method. The intention is to provide fairly broad coverage of the literature and to indicate possibilities for future development. An annotated bibliography is included.

Key words and phrases: Asymptotic expansion, Bartlett factor, conditional inference, Edgeworth series, exponential families, likelihood ratio statistic, maximum likelihood estimate, saddlepoint approximation, score statistic, second order approximations.

1. INTRODUCTION

A very accurate approximation to the density of the mean of a sample of independent, identically distributed observations was derived by Daniels (1954) using the saddlepoint technique of asymptotic analysis. The resulting approximation, often more accurate than the normal approximation or even the one- or two-term Edgeworth series approximation, is generally called the saddlepoint approximation. Occasional examples of its use have appeared regularly in the statistical literature (Cox, 1948; Good, 1957, 1961; Daniels, 1956, 1958; Keilson, 1963; Blackwell and Hodges, 1959), but with the appearance of a discussion paper by Barndorff-Nielsen and Cox (1979) its importance and usefulness in statistics began to be more widely appreciated. Since that paper several statistical applications of the saddlepoint approximation have been developed, many of them representing substantial contributions to the asymptotic theory of statistics. Examples include Barndorff-Nielsen's formula for the distribution of the maximum likelihood estimate, development and interpretation of the Bartlett factor for the likelihood ratio test, highly accurate approximations to the distribution of M estimates, improvements in goodness of fit tests and residual analysis for generalized linear models, accurate formulae for approximating tail probabilities and approximations to posterior means and densities in Bayesian analysis using the closely related Laplace approximation.

It is the purpose of this paper to summarize these recent developments and show how they are all related to Daniels' original saddlepoint expansion. In Section 2, Daniels' (1954) derivation from the Edgeworth expansion is reviewed and discussed. The next three sections develop from this approximations to the densities of the maximum likelihood estimator, the likelihood ratio statistic and the score statistic, respectively. In these sections, we assume that a sample of independent, identically distributed observations is to be used for inference about a vector parameter θ and that inference is required for all components of θ. Section 6 considers various extensions and recent developments. Inference in the presence of nuisance parameters is discussed in Section 6.1 from a Bayesian point of view and in Section 6.2 from a conditional point of view. The nonindependent, identically distributed case is considered briefly in Section 6.5. Other extensions include saddlepoint expansions for von Mises functionals, saddlepoint-type expansions derived from general orthogonal series expansions and tail area approximations.

2. THE SADDLEPOINT APPROXIMATION

Let $X_1, \ldots, X_n$ be independent, identically distributed random vectors from a density $f_X(\cdot)$ on $\mathbf{R}^k$. Denote the moment generating function by $\mathrm{M}(\phi) = E \exp(\phi^T X)$ and the cumulant generating function by $\mathrm{K}(\phi) = \log \mathrm{M}(\phi)$. The moment generating function is

N. Reid is Associate Professor, Department of Statistics, University of Toronto, Toronto, Ontario, Canada M5S 1A1.

assumed to exist in an open neighborhood around the origin. The saddlepoint expansion of the density of $\bar{X} = n^{-1} \sum X_i$ is

$$f_{\bar{X}}(\bar{x}) = (2\pi)^{-k/2}\{n/|\,\mathrm{K}''(\hat{\phi})\,|\}^{1/2} \cdot \exp[n\{\mathrm{K}(\hat{\phi}) - \hat{\phi}^T\bar{x}\}](1 + R_n). \tag{1}$$

The righthand side of (1), excluding the factor $(1 + R_n)$, will be called the *saddlepoint approximation* to the density of $\bar{X}$. The value $\hat{\phi} = \hat{\phi}(\bar{x})$ is called the saddlepoint and is defined by

$$\mathrm{K}'(\hat{\phi}) = \bar{x}, \tag{2}$$

where $\mathrm{K}'(\phi)$ is the vector $(\partial \mathrm{K}(\phi)/\partial\phi_1, \ldots, \partial \mathrm{K}(\phi)/\partial\phi_k)^T$. The $k \times k$ matrix $\mathrm{K}''(\phi)$ has (i, j) component $(\partial^2 \mathrm{K}(\phi)/\partial\phi_i\partial\phi_j)$ and determinant $|\,\mathrm{K}''(\phi)\,|$. The remainder R_n has an expansion in powers of n^{-1} and will be discussed in more detail later.

To express $\hat{\phi}$ explicitly as a function of $\bar{x}$, define the function

$$\begin{aligned}\mathrm{K}^*(\bar{x}) &= \hat{\phi}^T\bar{x} - \mathrm{K}(\hat{\phi}) \\ &= \sup_{\phi}\{\phi^T\bar{x} - \mathrm{K}(\phi)\},\end{aligned} \tag{3}$$

called the Legendre transform of the function $\mathrm{K}(\phi)$. Differentiating both sides of (3), and applying (2), gives $\mathrm{K}^{*\prime}(\bar{x}) = \hat{\phi}$ and $\mathrm{K}^{*\prime\prime}(\bar{x}) = \{\mathrm{K}''(\hat{\phi})\}^{-1}$, so (1) can be re-expressed as

$$f_{\bar{X}}(\bar{x}) = (2\pi)^{-k/2}\{n\,|\,\mathrm{K}^{*\prime\prime}(\bar{x})\,|\}^{1/2}\exp\{-n\mathrm{K}^*(\bar{x})\}\{1 + R_n\}.$$

This version of the saddlepoint expansion is derived and discussed in McCullagh (1984b, 1987, Chapter 6). Conditions ensuring the existence of a unique solution to (2) can be derived from conditions on the existence of the Legendre transform of $\mathrm{K}(\phi)$; see Barndorff-Nielsen (1978, Chapter 5) and Rockafellar (1970, Chapter 26). In the scalar case Daniels (1954, Section 7) shows that a unique real root of the saddlepoint equation (2) exists under very broad conditions. In particular, the random variables X_i may have a discrete distribution.

Approximation (1) is usually derived by one of two arguments, both of which were treated in Daniels (1954) in the scalar case. The first, from which the approximation takes its name, is as an application of the saddlepoint method of asymptotic analysis. The density for $\bar{X}$ is expressed as the inversion integral of its characteristic function,

$$f_{\bar{X}}(\bar{x}) = \frac{n}{2\pi i}\int_{\tau-i\infty}^{\tau+i\infty} \exp[n\{\mathrm{K}(\phi) - \phi\bar{x}\}]\,d\phi,$$

where τ is any point in the open interval around 0 in which $\mathrm{M}(\cdot)$ exists. The contour of integration in the complex plane can be deformed (within the strip of convergence of $\mathrm{M}(\cdot)$) to ensure that the leading contribution to the value of the integral comes from a small region near one point, the saddlepoint. A Taylor series expansion of the integrand and term-by-term integration leads to (1). For a more detailed summary of this argument, see Daniels (1987). For all the details, see Daniels (1954, Section 2). The saddlepoint method is explained in many books on asymptotic analysis; concise summaries are provided in Courant and Hilbert (1950, Chapter VII, Section 6.3), De Bruijn (1970, Chapter 5) and Jeffreys and Jeffreys (1961, Chapter 17).

A more statistical version of the derivation begins by embedding $f_X(x)$ in a conjugate exponential family

$$f_X(x;\,\phi) = \exp\{\phi x - \mathrm{K}(\phi)\}f_X(x). \tag{4}$$

The term conjugate exponential family is used in the literature on large deviations, and does not refer to the Bayesian meaning of conjugate posterior and prior distributions. The technique is usually attributed to Cramér (1938), but was also used by Esscher in 1932. Efron (1981) and others call the operation of (4) "exponential tilting." Note that $\mathrm{K}(\phi)$ is the cumulant generating function for the original density $f_X(\cdot)$.

From (4),

$$f_{\bar{X}}(\bar{x}) = f_{\bar{X}}(\bar{x};\,\phi)\exp[n\{\mathrm{K}(\phi) - \phi\bar{x}\}].$$

The Edgeworth expansion for the conjugate density $f_{\bar{X}}(\bar{x};\,\phi)$ is now used to derive an expansion for the density of interest $f_{\bar{X}}(\bar{x})$. This expansion is

$$\begin{aligned}&f_{\bar{X}}(\bar{x};\,\phi) \\ &= \psi(z)(\mathrm{var}_\phi\bar{X})^{-1/2}\{1 + \rho_3(\phi)h_3(z)/6 \\ &\qquad + \rho_4(\phi)h_4(z)/24 + \rho_3^2(\phi)h_6(z)/72 + \cdots\},\end{aligned} \tag{5}$$

where $\psi(z)$ is the standard normal density, and $z = (\bar{x} - E_\phi\bar{X})/(\mathrm{var}_\phi\bar{X})^{1/2}$ is the standardized version of $\bar{x}$. Under the conjugate exponential family, $E_\phi\bar{X} = \mathrm{K}'(\phi)$ and $\mathrm{var}_\phi\bar{X} = \mathrm{K}''(\phi)$. The standardized cumulants $\rho_3(\phi)$ and $\rho_4(\phi)$ are given by

$$\rho_3(\phi) = \mathrm{K}^{(3)}(\phi)/\{\mathrm{K}''(\phi)\}^{3/2},$$

$$\rho_4(\phi) = \mathrm{K}^{(4)}(\phi)/\{\mathrm{K}''(\phi)\}^2.$$

The Hermite polynomials $h_j(z)$ are defined by

$$h_j(z) = (-1)^j\frac{\partial^j\psi(z)}{\partial z^j}\Big/\psi(z);$$

explicit expressions for those appearing in (5) are

$$h_3(z) = z^3 - 3z,$$

$$h_4(z) = z^4 - 6z^2 + 3,$$

$$h_6(z) = z^6 - 15z^4 + 45z^2 - 15.$$

Because $\bar{X}$ is the mean of n independent, identically distributed random variables, ρ_3 is $O(n^{-1/2})$ and ρ_4 is

$O(n^{-1})$, and the $(\cdots)$ appearing in (5) is $O(n^{-3/2})$. The main difficulty with the Edgeworth approximation using the three correction terms in (5) is that as $|z| \to \infty$, i.e., $\bar{x}$ is in the tail of the distribution, the Hermite polynomials become unbounded. However, because ϕ is arbitrary, it can be chosen so that for each value of $\bar{x}$ the Edgeworth series is evaluated at the mean of the conjugate density. As noted above, this mean is precisely $\hat{\phi} = \hat{\phi}(\bar{x})$ defined by (2). Combining (4) with (5), we have

$$f_{\bar{X}}(\bar{x}) = \left\{\frac{n}{2\pi \mathrm{K}''(\hat{\phi})}\right\}^{1/2} \exp[n\{\mathrm{K}(\hat{\phi}) - \hat{\phi}\bar{x}\}]\{1 + O(n^{-1})\},$$

which is the scalar version of the saddlepoint expansion (1). For a rigorous derivation of the univariate Edgeworth expansion see Feller (1971, Section XVI.6); the multivariate expansion is discussed in the appendix of Barndorff-Nielsen and Cox (1979) and in Bhattacharya and Rao (1976).

Daniels (1954) presented expansion (1) for the scalar case and gave explicitly the form of the $O(n^{-1})$ term. Good (1957) gave the form of the $O(n^{-2})$ term in the scalar case, and derived the bivariate ($k = 2$) version up to and including the $O(n^{-1})$ term. For general k the difficulty is not in deriving the leading term given in (1), but rather in devising suitable notation for recording the remainder term. The bivariate and multivariate approximations are derived in Barndorff-Nielsen and Cox (1979, Sections 3 and 4), but the notation introduced in McCullagh (1984b) is somewhat more convenient.

The coefficients in the expansion of the remainder R_n depend on the higher order derivatives of $\mathrm{K}(\phi)$. For $k = 1$, the coefficient of the $O(n^{-1})$ term is

$$\{3\rho_4(\hat{\phi}) - 5\rho_3^2(\hat{\phi})\}/24. \tag{6}$$

In the multidimensional case, this term takes the form

$$\{3\rho_4(\hat{\phi}) - 3\rho_{13}^2(\hat{\phi}) - 2\rho_{23}^2(\hat{\phi})\}/24, \tag{7}$$

where ρ_4 is Mardia's measure of multivariate kurtosis and ρ_{13}^2, ρ_{23}^2 are the two invariant skewness measures for multivariate distributions (Mardia, 1970; McCullagh, 1987, Chapter 2). Formulas for these invariants are provided in the Appendix.

In order to study the relative error R_n in the saddlepoint approximation, it is necessary to examine the behavior of the saddlepoint $\hat{\phi}$ as a function of $\bar{x}$, or more conveniently, as a function of $\bar{x} - \mu$, where $\mu = \int x f_X(x)\, dx$. It is not difficult to verify that $R_n = R_n(\bar{x} - \mu)$ is uniformly bounded and has uniformly bounded derivatives for $|\bar{x} - \mu| < \delta$, where δ does not depend on n (cf. Feller, 1971, Section XVI.7). In fact this property is ensured by the method of construction of the conjugate family. The importance of this is that the saddlepoint approximation is accurate for relatively large values of $\bar{x}$. In contrast, the relative error in the Edgeworth series approximation for $f_{\bar{X}}(\bar{x})$ is bounded only for shrinking neighborhoods $|\bar{x} - \mu| < \delta n^{-1/2}$, and Edgeworth series approximations tend to be very poor in the tails. Furthermore, in a wide class of continuous univariate densities, Daniels (1954, Section 7) showed that $R_n(\bar{x} - \mu)$ is uniformly bounded for all values of $\bar{x}$ in the support of $f_{\bar{X}}(\cdot)$. This is further discussed in Jensen (1988).

The leading term in (1) will not in general integrate exactly to 1, and the approximation can be improved by renormalizing it to do so. Then the approximation is written

$$f_{\bar{X}}(\bar{x}) \cong c\{n/|\mathrm{K}''(\hat{\phi})|\}^{1/2} \exp[n\{\mathrm{K}(\hat{\phi}) - \hat{\phi}^T\bar{x}\}] \tag{8}$$

and is called the renormalized saddlepoint approximation. For some special densities $f_X(\cdot)$ the invariants appearing in (6) or (7) do not depend on $\hat{\phi}$. In that case the $O(n^{-1})$ term will be absorbed into the normalizing constant, and the error in the renormalized version will be $1 + O(n^{-2})$.

In the one-dimensional case, there are just three families for which (8) is exact; the normal, gamma and inverse Gaussian. Exactness requires not only that $3\rho_4(\phi) - 5\rho_3^2(\phi)$ is independent of ϕ, but also that the coefficients in the n^{-2}, n^{-3}, $\cdots$ terms are also free of ϕ. An elegant proof that this holds only for the three families above is given in Blaesild and Jensen (1985), where they also give Good's (1957) expression for the n^{-2} coefficient in the one parameter case, an expression for the n^{-2} coefficient in the bivariate case, and discuss exactness cases for $k \geq 2$. Daniels (1980) also proves the exactness result, by a somewhat lengthier argument. The fact that (8) is exact for the normal and gamma was shown in Daniels (1954).

Even if the invariants appearing in (6) do depend on $\hat{\phi}$, the order of the approximation can often be improved by renormalization. This is suggested by the interpretation of $\hat{\phi}$ as the maximum likelihood estimate in the conjugate exponential family. The argument usually presented (e.g., Barndorff-Nielsen and Cox, 1979, Section 2.4; Daniels, 1956, Section 7) is that (6) can be replaced by $3\rho_4(\phi) - 5\rho_3^2(\phi)/24$, thus incurring an error of order $n^{-1/2}$. This term is absorbed into the normalizing constant, and the relative error is now $O(n^{-3/2})$. To make this argument precise it is necessary to truncate the range of integration to $\{\bar{x}: |\hat{\phi} - \phi| < \delta n^{-1/2}\}$ and show that the error incurred is negligibly small. In Durbin (1980a, Section 2.7) this is verified for the case that $f_X(\cdot)$ is itself a member of the exponential family. (Durbin also considers renormalized saddlepoint approximations outside the independent, identically distributed setting; cf. also Section 6.6.) I am not aware of any detailed discussion

of renormalization in the k-dimensional case, where the leading term of R_n is given by (7).

A referee has suggested that the renormalization issue can be clarified by choosing the simpler renormalization of dividing by $f(\mu)$. Using $\hat{f}(\cdot)$ to denote the saddlepoint approximation to $f(\cdot)$, we can write

$$\frac{f(\bar{x})}{f(\mu)} = \frac{\hat{f}(\bar{x})}{\hat{f}(\mu)}\left\{1 + \frac{(\bar{x} - \mu)R_n'(\zeta)/n}{1 + R_n'(0)/n}\right\}, \tag{9}$$

for some ζ between $\bar{x}$ and μ, by the mean value theorem. Because R_n and R_n' are uniformly bounded in a neighborhood of μ, the renormalized approximation has a relative error of $O(n^{-1})$ for fixed $|\bar{x} - \mu|$ and $O(n^{-3/2})$ for $|\bar{x} - \mu| = O(n^{-1/2})$.

Many of the statistical applications of (1) derive from the fact that the saddlepoint $\hat{\phi} = \hat{\phi}(\bar{x})$ is the maximum likelihood estimate of the "parameter" ϕ in the density (5) and that $n\{\mathrm{K}(\hat{\phi}) - \hat{\phi}^T\bar{x}\}$ is the maximized "log likelihood ratio" for an independent, identically distributed sample from that density. The density is an artificial construct, however, and the true parameter of interest will be the parameter of the original density $f_X(\cdot)$, which has been suppressed in the present notation. An important exception is the case when $f_X(\cdot)$ is itself an exponential family density with canonical parameter θ, in which case there is a simple relationship between $\hat{\theta}$ and $\hat{\phi}$, and between $n\{\mathrm{K}(\hat{\phi}) - \hat{\phi}^T\bar{x}\}$ and the maximized log likelihood ratio statistic for θ. This connection will be exploited in each of the next three sections.

3. MAXIMUM LIKELIHOOD ESTIMATION

Suppose that the density of X_i takes the exponential family form

$$f_X(x; \theta) = \exp\{\theta^T t(x) - \psi(\theta) - d(x)\} \tag{10}$$

where θ is the canonical parameter and $t(x) = (t_1(x), \cdots, t_k(x))^T$ is the minimal sufficient statistic. The conjugate family is again of the form (10), with canonical parameter $\theta + \phi$, and $\mathrm{K}(\phi) = \psi(\theta + \phi) - \psi(\theta)$ is the cumulant generating function for $f_X(x; \theta)$. The density of $T = \sum t(X_i)$ is given by

$$f_T(t; \theta) = \exp\{\theta^T t - n\psi(\theta) - h(t)\},$$

and we now approximate $\exp\{-h(t)\}$ using the saddlepoint approximation. The saddlepoint equation is simply $n\psi'(\phi) = t$, so that the saddlepoint occurs at the maximum likelihood estimate $\hat{\theta}$, and (1) gives

$$f_T(t; \theta) = (2\pi)^{-k/2}\,|\,n\psi''(\hat{\theta})\,|^{-1/2}\exp\{(\theta - \hat{\theta})^T t - n\psi(\theta) + n\psi(\hat{\theta})\} \cdot \{1 + O(n^{-1})\}. \tag{11}$$

As was first pointed out in Daniels (1958), equation (11) has a very simple likelihood formulation:

$$f_T(t; \theta) = (2\pi)^{-k/2}\,|\,j(\hat{\theta})\,|^{-1/2}\{\mathrm{L}(\theta)/\mathrm{L}(\hat{\theta})\}\{1 + O(n^{-1})\}. \tag{12}$$

In (12) $\mathrm{L}(\theta)$ is the joint likelihood for the sample $(x_1, \cdots, x_n)$ and $j(\theta) = -\partial^2 \log \mathrm{L}(\theta)/\partial\theta\partial\theta^T$ is the observed Fisher information, in this case equal to $n\psi''(\theta)$. Both L and j should properly be written $\mathrm{L}(\theta, t)$ and $j(\theta, t)$, to emphasize their dependence on the data. The transformation from t to $\hat{\theta}$ is one-to-one, with Jacobian $|\,j(\hat{\theta})\,|$, giving

$$f_\Theta(\hat{\theta}; \theta) = c\,|\,j(\hat{\theta})\,|^{1/2}\{\mathrm{L}(\theta)/\mathrm{L}(\hat{\theta})\}\{1 + O(n^{-3/2})\}. \tag{13}$$

We have replaced $(2\pi)^{-k/2}$ by c to indicate that an improvement via renormalization is incorporated into (13). In this case the renormalization does reduce the error to $O(n^{-3/2})$, because the region of integration can be truncated to $\sqrt{n}\,|\,\hat{\theta} - \theta\,| < \varepsilon$ and the error incurred is exponentially small (Durbin, 1980a, Section 2.3).

The righthand side of (13) is often called Barndorff-Nielsen's approximation, as Barndorff-Nielsen has investigated extensively its application outside the exponential family. This will be discussed in detail below, but for the moment we simply point out that it is readily obtained from the saddlepoint approximation, in full exponential families. The argument outlined above is given in Barndorff-Nielsen (1983).

Example 1. Gamma density with unknown shape. We write $f_X(x) = (\nu/\mu)^\nu x^{\nu-1}e^{-x\nu/\mu}\Gamma^{-1}(\nu)$. The maximum likelihood estimate $\hat{\theta} = (\hat{\mu}, \hat{\nu})$ is given by $\hat{\mu} = t_1/n$ and $\psi(\hat{\nu}) - \log \hat{\nu} = (t_2/n) - \log(t_1/n)$, where $t_2 = \Sigma \log x_i$, $t_1 = \Sigma x_i$, and $\psi(\nu)$ is the digamma function $d \log \Gamma(\nu)/d\nu$. It is easily shown that (13) is

$$f(\hat{\mu}, \hat{\nu}; \mu, \nu) \cong c g_1(\hat{\mu}; \mu, \nu) g_2(\hat{\nu}; \nu)$$

where

$$g_1(\hat{\mu}; \mu, \nu) = (\nu/\mu)^{n\nu}\hat{\mu}^{n\nu-1}\exp(-n\nu\hat{\mu}/\mu)$$

and

$$g_2(\hat{\nu}; \nu) = \{\Gamma^n(\hat{\nu})\Gamma^n(\nu)\}\{\hat{\nu}\psi'(\hat{\nu}) - 1\}^{1/2} \cdot \exp[n\{(\hat{\nu} - \nu)\psi(\hat{\nu}) + \hat{\nu} - \nu \ln \hat{\nu}\}],$$

showing that $(\hat{\mu}, \hat{\nu})$ are independent to the order considered, and that the approximation to the density of $\hat{\mu}$ is exact after renormalization. The renormalized version of $g_2(\hat{\nu}; \nu)$ is displayed in Figure 1, for $n = 10$, $\nu = 1$. Also shown are the "exact" density of $\hat{\nu}$ estimated from 10,000 simulations, and the approximating normal density with mean ν and variance $j_{22}(\nu)$.

The gamma example is further discussed in Jensen (1986b), where an alternative approximation to the density of $\hat{\nu}$ and approximations to the similar test for μ are derived.

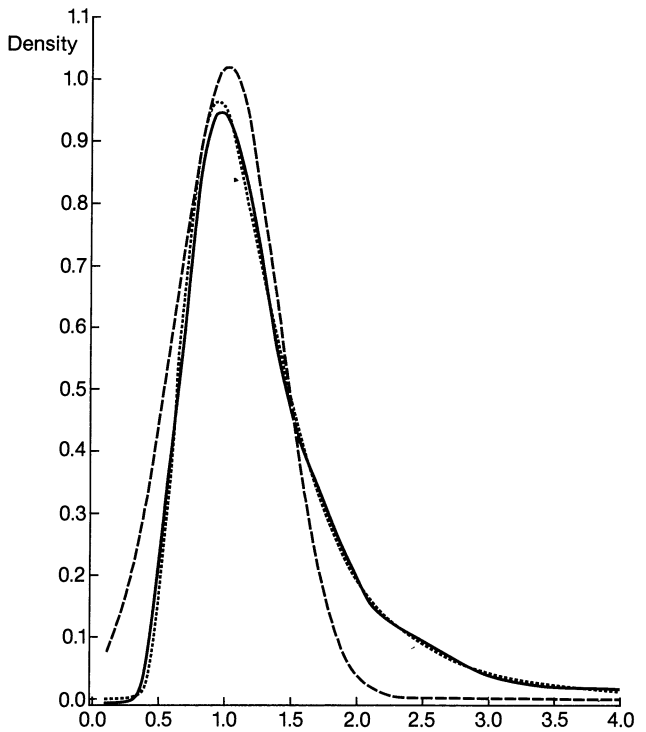

FIG. 1. *Saddlepoint approximation (dotted line) to the density of the maximum likelihood estimator of the shape parameter of a gamma distribution, based on a sample of size* 10. *The "exact" density (solid line) was estimated from* 10,000 *simulations. The normal approximation (dashed line) is shown for comparison.*

To construct a confidence interval for ν it is necessary either to carry out repeated numerical integration or to expand (13) and invert it algebraically. For the general one-parameter case, the details of the expansion are carried out in Barndorff-Nielsen (1985b) and McCullagh (1984a). Unfortunately, the relatively simple approximations to tail probabilities discussed in Daniels (1987) (cf. also Section 6.3) cannot be directly applied to this example because $\hat{\theta}$ is not a one-to-one function of a sample average. It may be possible to adapt the conditional probability tail approximation of Skovgaard (1988b) to this example.

Outside the exponential family setting, the maximum likelihood estimator will not be a one-to-one function of the minimal sufficient statistic, so even if we contemplated using the righthand side of (13) we would not be able to write $L(\theta; x)$ for example, as a function only of θ and $\hat{\theta}$. However, (13) does continue to provide an approximation to a conditional density of $\hat{\theta}$, as is illustrated in the next example.

Example 2. Location-scale family. Suppose $f_X(x; \theta)$ is an arbitrary continuous density on R^1, with θ as a two-dimensional location-scale parameter (μ, σ), so that for an independent, identically distributed sample,

$$f(x_1, \cdots, x_n; \mu, \sigma) = \prod_{i=1}^{n} \sigma^{-1} f_X\{(x_i - \mu)/\sigma\}.$$

Without further assumptions about f, the minimal sufficient statistic is the order statistic $(x_{(1)}, \cdots, x_{(n)})$. It can be separated into two components, $\hat{\theta}$, the maximum likelihood estimate of θ, and $a = (a_1, \cdots, a_n)$, where $a_i = (x_{(i)} - \hat{\mu})/\hat{\sigma}$. The vector a has $n - 2$ independent components, and is ancillary; i.e., its distribution does not depend on θ. The conditional distribution of $\hat{\theta}$, given a, is

$$f_{\hat{\Theta}|A}(\hat{\theta} \mid a; \theta) = c_0(a)\hat{\sigma}^{n-2}\Pi f_X(\hat{\sigma}a_i + \hat{\mu}; \mu, \sigma)$$

which can be re-expressed as

$$(14) \qquad f(\hat{\theta} \mid a; \theta) = c(a)\,|j(\hat{\theta})|^{1/2}\{L(\theta)/L(\hat{\theta})\},$$

using the fact that $|j(\hat{\mu}, \hat{\sigma})| = \hat{\sigma}^{-4}d(a)$, where $d(\cdot)$ depends on the derivatives of $\log f$.

Note the similarity of (14) to approximation (13), and also that (14) is the exact conditional density of $\hat{\theta}$, given the maximal ancillary a. Fisher (1934) derived (14) and argued that inference for θ should be based on this conditional distribution; see also Cox and Hinkley (1974, page 115). Different versions of formula (14) have been derived by several authors, including Pitman (1938), Fraser (1968), Efron and Hinkley (1978), and Barndorff-Nielsen (1980, 1983). Barndorff-Nielsen (1983) emphasized the similarity of (13) and (14), and showed further that (14) provides an expression for the conditional density of the maximum likelihood estimate in any transformation model, i.e., any model generated by a group.

That the same formula provides either a highly accurate approximation or an exact expression for the distribution of the maximum likelihood estimator in full exponential families or transformation families is rather surprising. Exponential families and transformation families are usually considered to be quite different statistical objects, but this suggests that there may be a close connection between them. McCullagh (1987, Chapter 8) has investigated to what extent an arbitrary family of densities can be made to "look like" an exponential family, by conditioning on some approximately distribution-free statistic. Also relevant is Mitchell (1988) in which the geometry of a subclass of transformation models, the elliptic families, is studied. This geometry has some striking similarities to the geometry of exponential families outlined in Amari (1982; 1985, Chapter 2) and Efron (1978).

What about densities that are not members of exponential or transformation families? Remarkably, the same formula continues to provide an approximation to the conditional distribution of the maximum likelihood estimate, conditioned on an approximately ancillary statistic a. Approximately ancillary is taken to mean that the distribution of a depends on θ only in terms of $O(n^{-1})$ or higher, for θ

within $O(n^{-1/2})$ of the true value. We write

$$f_{\hat\theta|A}(\hat\theta \mid a; \theta) = c(\theta, a)\,|j(\hat\theta)|^{1/2}\left\{\frac{L(\theta; \hat\theta, a)}{L(\hat\theta; \hat\theta, a)}\right\} \times \{1 + O(n^{-1})\}. \tag{15}$$

In (15) we have supposed that the sufficient statistic based on our sample is a one-to-one function of $(\hat\theta, a)$, where a is to be specified. In fact what is required is that $(\hat\theta, a)$ is approximately sufficient to the same order that a is approximately ancillary. The accuracy of this approximation has been extensively investigated, particularly by Barndorff-Nielsen (1980, 1983, 1984, 1985a, 1985b, 1986a, 1986b), but also by McCullagh (1984a, 1987, Chapter 8), Cox (1980), Hinkley (1980), Durbin (1980a) and Barndorff-Nielsen and Cox (1984a). Many of these papers restricted attention to curved exponential families, for which particular approximate ancillaries could often be defined. A (k, d) curved exponential family is an exponential family with k sufficient statistics but only $d < k$ parameters of interest. This typically arises when the components of the vector of canonical or natural parameters of a full exponential family are constrained by some nonlinear relationship. The parameter space is thus confined to a curve in the natural parameter space for the full family. Curved exponential families and their properties are extensively discussed in Amari (1985). However, McCullagh (1984a, 1987, Chapter 8) was able to show that (15) is valid in arbitrary families, to $O(n^{-1})$, using any approximate ancillary statistic.

The renormalization implied by the way we have written (15) might be expected to reduce the error to $O(n^{-3/2})$. Barndorff-Nielsen (1985a, 1985b, 1986a, 1986b) showed that there is a unique approximate ancillary for which this is the case. The 1985 papers considered curved exponential families but in 1986a (Section 1.3) the result was extended to a more general setting involving nuisance parameters.

Example 3. A (2, 1) curved exponential family. Suppose we are observing failure times following an exponential (θ) density, but censored at a fixed time T. The likelihood based on the n observations $(x_1, \delta_1), \cdots, (x_n, \delta_n)$ is

$$L(\theta) = \theta^{n_0} e^{-\theta t}$$

where $n_0 = \sum \delta_i$ is the number of failures and $t = \sum x_i$ is the total time on text. Then $\hat\theta = n_0/t$, $j(\hat\theta) = n_0/\hat\theta^2$, and (15) becomes

$$f(\hat\theta \mid a; \theta) \cong c n_0^{1/2} \hat\theta^{-1} \exp\{t(\hat\theta - \theta) - n_0 \log(\hat\theta/\theta)\} \tag{16}$$

where t and n_0 must be expressed as functions of $\hat\theta$ and a. There have been a variety of approximately ancillary statistics suggested in the literature. Three such are the Efron-Hinkley ancillary, defined in Efron and Hinkley (1978, equation 5.6), the local ancillary defined in Cox (1980, equation 8) and Barndorff-Nielsen's affine ancillary (1986a, Section 2.2), which is related to the likelihood ratio statistic. In this example the computation of any of these approximate ancillaries is not entirely straightforward, which somewhat limits the usefulness of (16) for inference. (In this example the exact density for $\hat\theta$ has a point mass at $\hat\theta = 0$, so in fact (12) cannot be uniformly accurate to $O(n^{-1})$, but this is in a sense a technical difficulty.)

Example 4. $N(\theta, b^2\theta^2)$, b^2 known. The sufficient statistic based on n independent, indentically distributed observations from this density is (t_1, t_2), where $t_1 = n^{-1}\Sigma x_i$ and $t_2 = n^{-1}\Sigma x_i^2$. There is an exact ancillary for this problem, which is a function of t_1^2/t_2 (Hinkley, 1977; Amari, 1982). Expression (15) becomes

$$f(\hat\theta \mid a; \theta) \cong c\left\{\frac{2b^2+1}{t_2 - t_1\hat\theta}\right\}^{1/2}\left(\frac{\hat\theta}{\theta}\right)^n \cdot \exp -\frac{n}{2b^2}\left\{t_2\left(\frac{1}{\theta^2} - \frac{1}{\hat\theta^2}\right) - 2t_1\left(\frac{1}{\theta} - \frac{1}{\hat\theta}\right)\right\}, \tag{17}$$

where again (t_1, t_2) depend on $(\hat\theta, a)$ through the equations $t_1\hat\theta - t_2 + b^2\hat\theta^2 = 0$ and $(t_1^2/t_2) = h(a)$, where $h(a) = (1 + ka)^2/(1 + b^2 + ka)$, and $k = \sqrt{2}b^2/(2b^2 + 1)$. In fact the $N(\theta, b^2\theta^2)$ density is a transformation model, so from Barndorff-Nielsen (1983) we know that (17) is exact. For further discussion of this example, see L. Skovgaard (1984).

As can be seen from Examples 3 and 4, formulae (13) through (15), while intriguing, can be rather difficult to use for inference about θ. The saddlepoint approximation is also very useful in approximating the distribution of the likelihood ratio statistic and the score statistic, as we now describe.

4. THE LIKELIHOOD RATIO STATISTIC

In the exponential family there is a simple relationship between the sufficient statistic, t, and the maximum likelihood estimate, $\hat\theta$. This is the basis for the derivation of Barndorff-Nielsen's formula from the saddlepoint approximation; but as described in Section 3, the formula holds much more widely. There is also a simple relationship between t and the likelihood ratio statistic; this leads to an approximation to the distribution of the likelihood ratio statistic that also holds outside the exponential family.

Let $W = W(\theta) = -2\log\{L(\theta)/L(\hat\theta)\}$ be the likelihood ratio statistic for testing the hypothesized value θ.

Standard large sample theory shows that W is asymptotically distributed as χ^2_k. Furthermore, in wide generality $EW = k\{1 + b(\theta)/n + O(n^{-2})\}$, so that $W' = W/\{1 + b(\theta)/n\}$ has expected value k to $O(n^{-2})$. The correction $b(\theta)$ is called a Bartlett factor. Bartlett (1937, 1953) suggested such a rescaling to improve the χ^2 approximation to the distribution of W. In 1956, Lawley showed that in fact all cumulants of W' agree with those of a χ^2_k, to the same order, but the proof was exceedingly complicated. A much simpler derivation is obtainable from the saddlepoint approximation, as was first discussed in Barndorff-Nielsen and Cox (1979, Section 6.3).

From the saddlepoint approximation (13), with the normalizing constant explicitly included,

$$f_W(w;\theta) = (2\pi)^{-k/2}\{1 + d(\theta)/n\}e^{-(1/2)w} \cdot \int 1\{\hat\theta \in \mathrm{R}(w)\}\,|j(\hat\theta)|^{1/2}\,d\hat\theta \cdot \{1 + O(n^{-3/2})\}, \tag{18}$$

where $1\{\cdot\}$ is an indicator function, $\mathrm{R}(w)$ is the set of $\hat\theta$ values giving the same value of w and $d(\theta)$ is computed from (7). To first order $w = (\hat\theta - \theta)^T j(\hat\theta)(\hat\theta - \theta)$, but to keep accuracy of $O(n^{-1})$ it is necessary to expand j and w as functions of $\hat\theta$ to higher order. Expression (18) simplifies to

$$f_W(w;\theta) = c \cdot w^{-k/2}e^{-(1/2)w}\{1 + h(\theta)w/n + O(n^{-3/2})\}.$$

Using the fact that $f_W(w;\theta)$ must integrate to 1, and $wf_W(w;\theta)$ must integrate to $k\{1 + b/n + O(n^{-2})\}$, gives $b = -2d/k$ and

$$f_{W'}(w;\theta) = c_k w^{(1/2)k-1}e^{-(1/2)w}\{1 + O(n^{-3/2})\},$$

where c_k is the normalizing constant for χ^2_k density.

This result is derived in Barndorff-Nielsen and Cox (1979, Section 6) for $k = 1$, and Barndorff-Nielsen and Cox (1984a) for general k. A more direct proof starting from the Legendre transform version of the saddlepoint approximation (4) is given in McCullagh (1987, Chapter 6). For an interesting application of this result, see Eriksen (1987).

In fact it is not necessary to restrict attention to full exponential families. It was only used in (18) to replace the $O(n^{-1})$ term $d(\hat\theta)/n$ by $d(\theta)/n$, and to ensure that the transformation from the minimal sufficient statistic to the maximum likelihood estimate is one-to-one. Barndorff-Nielsen and Cox (1984a) derived the same result from the more general approximation (15): If

$$f_{\hat\theta|A}(\hat\theta \mid a;\theta) = c(\theta, a)\,|j(\hat\theta)|^{1/2}\{\mathrm{L}(\theta)/\mathrm{L}(\hat\theta)\} \cdot \{1 + O(n^{-3/2})\}$$

then

$$c(\theta, a) = (2\pi)^{-k/2}\{1 + d(\theta)/n + O(n^{-3/2})\}, \tag{19}$$

$$b(\theta) = -2d(\theta)/k \tag{20}$$

and

$$f_{W'|A}(w \mid a) = c_k w^{(1/2)k-1}e^{-(1/2)w}\{1 + O(n^{-3/2})\}. \tag{21}$$

In other words, the Bartlett factor for W is a one-to-one function of the norming constant in (13), and the rescaled version of the log likelihood ratio statistic has all cumulants agreeing with those of a χ^2_k random variable, to $O(n^{-3/2})$. It is essential for this argument that (13) be accurate to $O(n^{-3/2})$; which is of course true for exponential families and transformation models. Outside these cases the result of Barndorff-Nielsen (1986a) that a unique approximate ancillary exists for which (13) is accurate to $O(n^{-3/2})$ is important. Note, however, that result (21) holds unconditionally as well: To the order considered W' is independent of the ancillary.

Several examples of Bartlett factors are provided in Barndorff-Nielsen and Cox (1984a,b). Depending on the problem, it may be easier to compute the Bartlett factor from the norming constant, or from the expected value of the likelihood ratio statistic. Both of these require integrations over the sample space. A general formula for the Bartlett factor for $k = 1$ is given in Lawley (1956) and Cox and Hinkley (1974, page 339) and for $k > 1$ in McCullagh and Cox (1986). DiCiccio (1984, 1986, 1988) provides several alternative formulae: In location-scale and regression models he has obtained expressions that depend only on observed log likelihood derivatives, thus avoiding the need for high dimensional integration. This should make approximations (15) and (19) much more useful in practice. At the present time, general correction factors are not built in to available software such as GLIM, although this is to be expected in the near future.

Of possibly more relevance for inference, at least for a scalar parameter θ, are related results on the signed square root of W, the sign taken being that of $\hat\theta - \theta$. Barndorff-Nielsen (1986a) and McCullagh (1984a) show that $R = \pm W^{1/2}$ can be recentered and rescaled to follow a $N(0, 1)$ distribution to $O_p(n^{-3/2})$. The recentering is of a particularly simple form: $R^+ = R + \rho_{30}/6$, where ρ_{30} is the standardized skewness of the score statistic, follows a $N(0, 1)$ distribution to $O_p(n^{-1})$. Pierce and Schafer (1986) exploit this result to study normal approximations to deviance residuals in generalized linear models. The multivariate version of R, called the directed log likelihood ratio, is discussed in McCullagh (1984b), Barndorff-Nielsen (1986a) and Bickel and Ghosh (1987). For inference about scalar θ in the presence of nuisance parameters

(cf. Section 6.2), the recentering and rescaling constants are somewhat more complicated, and may both be equally important (Barndorff-Nielsen, 1986a).

5. THE SCORE STATISTIC

An alternative to the likelihood ratio statistic or the maximum likelihood estimate is the score statistic

$$U(\theta) = \sum U_i(\theta) = \sum \partial \log f(X_i; \theta)/\partial\theta. \tag{22}$$

To first order U follows a normal distribution and it has exact mean 0 and exact variance $i(\theta) = E\{j(\theta); \theta\}$. In exponential families $U(\theta) = T - n\psi'(\theta)$, so the saddlepoint approximation to the density of T is effectively also the saddlepoint approximation to the score statistic.

Noting that (22) is a sum of independent, identically distributed random variables, a saddlepoint approximation for $U(\theta)$ follows directly from (1), whether or not the observations are from an exponential family. Let $\mathrm{K}(t, \theta) = \log E[\exp\{tU_i(\theta)\}; \theta]$ be the cumulant generating function for the score. Then (1) becomes

$$f_{\bar{U}}(\bar{u}; \theta) \cong (2\pi)^{-k/2}\{n/|\mathrm{K}''(t_0, \theta)|\}^{1/2} \cdot \exp[n\{\mathrm{K}(t_0, \theta) - t_0\bar{u}\}], \tag{23}$$

where $\bar{U}. = n^{-1}U$, and t_0 is the saddlepoint, defined by $\mathrm{K}'(t_0, \theta) = \bar{u}$. In general the function K and the saddlepoint t_0 may be difficult to compute. In exponential families, $\mathrm{K}(t, \theta)$ is simply $\psi(t + \theta) - \psi(\theta) - t\psi'(\theta)$ and $t_0 = \hat{\theta} - \theta$.

In the case that θ is a scalar, the approximation (23) can be used to approximate both the density and the distribution function of the maximum likelihood estimate, as shown in Daniels (1983). Writing $U_i(a) = \partial \log f(X_i; a)/\partial a$ and $\mathrm{K}_\theta(t, a)$ for its cumulant generating function under the density $f(x; \theta)$, a generalization of (23) is

$$f_{\bar{U}(a)}(\bar{u}; \theta) \cong \{n/2\pi\mathrm{K}_\theta''(t_0, a)\}^{1/2} \cdot \exp[n\{\mathrm{K}_\theta(t_0, a) - t_0\bar{u}\}]. \tag{24}$$

Then if U is monotone decreasing in θ we have

$$\begin{aligned} &\mathrm{pr}(\hat{\theta} > a; \theta) \\ &= pr\{\bar{U}(a) > 0; \theta\} = \int_0^\infty f_{\bar{U}}(u; \theta)\, du \\ &\cong \int_{t_0^*}^{t_m} \{n\mathrm{K}_\theta''(t, a)/(2\pi)\}^{1/2} \cdot \exp[n\{\mathrm{K}_\theta(t, a) - t\mathrm{K}_\theta'(t, a)\}]\, dt \end{aligned} \tag{25}$$

where t_0^* and t_m are defined by $\mathrm{K}_\theta'(t_0^*, a) = 0$ and $\mathrm{K}_\theta'(t_m, a) = \infty$. Version (25), which is equation (3.6) of Daniels (1983), avoids the need for repeated solution of the saddlepoint equation. Note that (25) provides an approximation to the tail area for the distribution of the maximum likelihood estimate. Except for exponential families, it will not be equivalent to integrating Barndorff-Nielsen's approximation (13), because (24) is not conditional on an exact or approximate ancillary. Numerical comparison of (25) to the tail area approximation of Lugannani and Rice (1980) (cf. Section 6.3) is given in Daniels (1983).

The derivative of (25), with respect to a, provides a saddlepoint approximation to the unconditional density of $\hat{\theta}$. (Explicit expressions for this density in the case of curved exponential families are provided in Hougaard (1985) and Skovgaard (1985b).) An alternative approximation is also derived by Daniels (1983); the differentiation is carried out in the inversion integral and then the saddlepoint method is applied. This gives

$$f_{\hat{\theta}}(\hat{\theta}; \theta) \cong \left\{\frac{n}{2\pi\mathrm{K}_\theta''(t_0, \hat{\theta})}\right\}^{1/2} \cdot \left\{-\frac{\dot{\mathrm{K}}_\theta(t_0, \hat{\theta})}{t_0}\right\}\exp\{n\mathrm{K}_\theta(t_0, \hat{\theta})\}, \tag{26}$$

where $\dot{\mathrm{K}}_\theta(t, a) = \partial\mathrm{K}_\theta(t, a)/\partial a$, and t_0 depends on $\hat{\theta}$ through the saddlepoint equation. This version of the approximate distribution of the maximum likelihood estimate is equivalent to one derived by Field and Hampel (1982), as shown in the appendix of Daniels (1983). Field and Hampel's result is extended to arbitrary M estimates and multivariate parameter θ in Field (1982).

Note that in the above, the correspondence between $\bar{U}.$ and $\hat{\theta}$ given by the first equality in (25) is exact, as opposed to the usual large sample approximation $(\hat{\theta} - \theta)i(\theta) \cong n\bar{U}.(\theta)$. By expanding $U(\theta)$ to higher order as a function of $\hat{\theta}$, Barndorff-Nielsen's approximation (15) can be used to provide an approximation to the density of $\bar{U}$, conditional on the exact or approximate ancillary in (15). This has been carried out by Barndorff-Nielsen (1987b).

Neither approximation (24) nor (26) has been much used in practice, presumably because of the difficulty of computing the cumulant generating function.

Recall that in the case of the likelihood ratio statistic, the saddlepoint approximation essentially led to an improvement in the χ^2 approximation by the Bartlett factor rescaling. No such Bartlett factor exists for the score statistic (Bartlett, 1953), i.e., there is no scaling constant that uniformly improves the χ^2 approximation to $U^T(\theta)i^{-1}(\theta)U(\theta)$. Harris (1985) has investigated alternative asymptotic approximations to the score statistic by adjusting the χ^2 distribution rather than the statistic itself. His method proceeds from the Edgeworth expansion; it may be that an alternative improvement can be derived from the saddlepoint expansion. Improved sample-based inference

from score statistics using the saddlepoint approximation to tail areas (Section 6.3) is discussed in Tingley and Field (1986).

6. FURTHER DEVELOPMENTS

6.1 Approximations in Bayesian Inference

The saddlepoint method is a technique for approximating integrals over the complex plane by deforming the curve of integration. A closely related technique for approximating integrals over $\mathbf{R}^k$ is Laplace's method, and an important application of Laplace's method is to evaluating integrals arising in Bayesian analysis.

Tierney and Kadane (1986) consider approximations to the posterior mean of a function $g(\theta)$,

$$Eg(\theta) = \int g(\theta)e^{l_n(\theta,x)}\pi(\theta)\,d\theta \Big/ \int e^{l_n(\theta,x)}\pi(\theta)\,d\theta,$$

where the dependence of the log-likelihood function on the sample size is indicated explicitly. As n increases, with $\pi(\theta)$ and $g(\theta)$ fixed, the main contribution to the integrands in the numerator and denominator come from a neighborhood of the respective maxima, leading to a very accurate approximation to the posterior mean. In fact the maxima for the two integrands differ by a term of $O(n^{-1})$ only, so that the relative error in the ratio of the approximations is $O(n^{-2})$ and hence is more accurate than the individual approximations to the numerator and denominator.

Tierney and Kadane (1986) also discuss approximations to the marginal posterior density for some components of θ obtained by two applications of Laplace's approximation. In this case the relative error is not reduced by an order of magnitude, essentially because the maxima of the two integrands are not sufficiently close. Further developments on second-order approximations in Bayesian inference are reported in Tierney, Kass and Kadane (1987) and Kass, Tierney and Kadane (1987, 1988). The last includes some discussion of the important practical issue of implementation.

A similar development in Davison (1986) is used to approximate the posterior density of θ. The resulting approximation,

$$f_{\Theta|X}(\theta \mid x) = (2\pi)^{-k/2}\left\{\frac{f_{X|\Theta}(x \mid \theta)}{f_{X|\Theta}(x \mid \hat\theta)}\right\}|j(\hat\theta)|^{1/2}\{1 + O(n^{-1})\}, \tag{27}$$

is in form identical to formula (13), with the important difference that now θ is considered to vary and $\hat\theta$ is fixed. The prior has been eliminated in (27) by assuming that it is flat in the neighborhood of $\hat\theta$. The approximate posterior for nonflat priors is given by (11) of Davison or (4.1) of Tierney and Kadane.

There is also a formal equivalence between predicting an observation z from a sample from a density $f(x; \theta)$ and estimating a scalar parameter in the presence of nuisance parameters. Various approximate predictive likelihoods proposed in Butler (1986) and Davison (1986) are thus very closely related to modified likelihoods proposed by Barndorff-Nielsen (1983, 1986a) and Cox and Reid (1987) that are obtained by applying the formula for the distribution of the maximum likelihood estimate to the conditional setting.

6.2 Conditional Inference

Suppose that the canonical parameter θ in the exponential family (10) has p components of interest, $\theta_{(1)}$, and $k - p$ nuisance parameters $\theta_{(2)}$. A similar test for the hypothesis $\theta_{(1)} = \theta^0_{(1)}$, say, is constructed from the conditional density of $t_{(1)}$, given $t_{(2)}$, where the partition of t corresponds to that of θ. The saddlepoint approximations to the density of $(t_{(1)}, t_{(2)})$ and the marginal density of $t_{(2)}$ are readily obtained, and their ratio is usually called the double saddlepoint approximation. The saddlepoint for the numerator is the maximum likelihood estimate $\hat\theta$ and for the denominator is the restricted maximum likelihood estimate $\tilde\theta_{(2)}$, with $\theta_{(1)}$ fixed at $\theta^0_{(1)}$. Details of the double saddlepoint approximation in exponential families and generalized linear models with unknown scale are provided in Davison (1988). The double saddlepoint approximation is not the same as the saddlepoint approximation to the exact conditional density, and can be more accurate; see McCullagh (1987, Chapter 6). Jensen (1986a) shows that the resulting similar test is well approximated by the likelihood ratio test, after Bartlett correction.

Outside the exponential family, it is still possible to use approximation (15) for the density of the maximum likelihood estimate $\hat\theta$ and the restricted maximum likelihood estimate $\tilde\theta_{(2)}$. This leads to results for the generalized likelihood ratio test essentially the same as those described in Section 4, and the discussion in Barndorff-Nielsen and Cox (1984a) includes a treatment of the nuisance parameter setting. The asymptotic normality of the directed log likelihood ratio statistic in the presence of nuisance parameters is considered in Barndorff-Nielsen (1986a) and Bickel and Ghosh (1987). Skovgaard (1988a) finds saddlepoint approximations to the conditional distribution of the score statistic, and compares that to inference based on the likelihood ratio statistic.

However, although both W and the appropriately defined one-sided version of W can be adjusted so that their nominal asymptotic distributions are close to exact, and they are to the order considered independent of any ancillary statistics, it is not the case in general that they provide optimal, or even correct

inference for $\theta_{(1)}$. Maximizing out the nuisance parameters via $\tilde{\theta}_{(2)}$ does not make allowance for the errors of estimation of $\theta_{(2)}$ in the inference for $\theta_{(1)}$; the resulting "profile" likelihood may give misleadingly precise inference for $\theta_{(1)}$. This problem takes its most extreme form when the dimension of $\theta_{(2)}$ increases with the sample size (Neyman and Scott, 1948). (See also Cox and Hinkley, 1974, page 329; Kalbfleisch and Sprott, 1970; Fraser, 1979, Chapter 5.3.)

Conditional versions of the likelihood ratio test that are proposed in Barndorff-Nielsen (1983, 1987a) and Cox and Reid (1987) attempt to adjust for the estimation of the nuisance parameters. Both constructions rely on applying (15) twice, as described in the previous paragraph. Substantial difficulties remain in evaluating the inferential properties of these procedures, essentially because the ancillary statistic, a_2 say, required for the validity of the approximation to the density of $\tilde{\theta}_{(2)}$, may not be the same as the ancillary statistic a used in approximating the density of $\hat{\theta}$, and in general a_2 may depend on the parameter of interest, $\theta_{(1)}$.

6.3 Approximating Tail Areas

In many applications, it will be of interest to compute approximate tail areas or cumulative distribution functions, rather than densities. Integrating the saddlepoint expansion will usually require a lot of numerical calculation, although Daniels' simplification outlined in Section 5 may be applicable. In contrast to this the Edgeworth expansion is easily integrated term by term and the expansion for the cumulative distribution function is of the same form as the expansion for the density.

A variety of tail area approximations for the sample mean derived by the saddlepoint method are reviewed in Daniels (1987). Special prominence there is given to the approximation of Lugananni and Rice (1980), because it is quite simple to use, and accurate over the range of $\bar{x}$. The approximation takes the form

$$1 - F_{\bar{X}}(\bar{x}) = [1 - \Phi(y) + \phi(y)\{z^{-1} - y^{-1}\}]\{1 + O(n^{-1})\} \tag{28}$$

where $y = \pm\sqrt{2n}\{\hat{\phi}\bar{x} - K(\hat{\phi})\}^{1/2}$ and $z = \hat{\phi}\{n\mathrm{K}''(\hat{\phi})\}^{1/2}$. If $\bar{x} = \mu$, (28) is replaced by

$$1 - F_{\bar{X}}(\mu) \cong (1/2) - \rho_3/\{6(2\pi n)^{1/2}\} + O(n^{-3/2}).$$

This tail area approximation is extremely accurate: for numerical examples see Daniels (1983, 1987) and Davison and Hinkley (1988).

For evaluating tail areas in conditional densities, a technique suggested by Barndorff-Nielsen and Cox (1979, Section 4) is to use the saddlepoint expansion for the marginal density in the denominator, and the Edgeworth expansion in the numerator. This is called the single saddlepoint expansion, and has the virtue of being easily integrated, term by term. Skovgaard (1988a) extends the Lugannani and Rice (1980) formula to conditional tail probabilities of the form $\mathrm{pr}(\bar{X}_1 \geq \bar{x}_1 \mid \bar{X}_2 = \bar{x}_2)$.

Note that for the normal approximation to the signed square root of the likelihood ratio statistic discussed in Section 4, quite accurate tail areas are easily computed from the normal cumulative distribution function. This is discussed in more detail in McCullagh (1987, Chapter 6.2.6).

6.4 Approximating More Complicated Statistics

Many statistics can be approximated by sums of independent, identically distributed random variables. A wide class of examples is provided by the von Mises functionals, satisfying

$$T(F_n) \cong T(F) + n^{-1}\Sigma\alpha_F(X_i)$$

where F_n is the empirical cumulative distribution function, $T(F_n)$ is the statistic to be approximated, and $\alpha_F(x)$ is the influence function for the statistic T. Examples of von Mises functionals are the L estimates of robust estimation theory. Another class of examples are statistics obtained by Hajek projection (Efron and Stein, 1981). Linear rank statistics can be accurately approximated this way.

As in Section 5, it is possible to compute the saddlepoint approximation to the distribution of the linear term. However, as pointed out in Davison and Hinkley (1988), this does not give an approximation to the distribution of $T(F_n)$ with relative error $O(n^{-1})$: the next term in the von Mises expansion must also be accounted for. The relevant calculations are illustrated in several examples in their paper, but the accuracy and general applicability of such techniques remain to be investigated.

A general technique for approximating the cumulant generating function is proposed in Easton and Ronchetti (1986). This may prove quite useful in examples such as described above, although at present it is not clear whether or not the resulting approximation is more accurate than the Edgeworth expansion.

6.5 Other Series Expansions

The Edgeworth expansion approximates a density by a normal density and correction terms based on derivatives of the normal density. If the basic approximating density is something else, for example gamma, then the correction terms depend on the associated system of orthogonal polynomials. For the gamma density, these are the Laguerre polynomials. This can be exploited to give a better finite sample

approximation to the true density. Several examples appear in Firth (1987).

This would also seem appropriate in problems such as those arising in extreme value theory where the limiting distribution is not normal. Presumably there is an analogue of the saddlepoint technique in this context, although whether or not it is more useful for special problems than the regular saddlepoint technique is an open question. The only application of the technique that I know of appears in Jensen (1986b). A series expansion for Barndorff-Nielsen's formula is given in Barndorff-Nielsen (1986b).

6.6 Sums of Nonindependent, Identically Distributed Random Variables

The derivation of the saddlepoint approximation outlined in Section 2 suggests that if an Edgeworth expansion is available for a conjugate density of a statistic, with terms of successively higher order in $n^{-1/2}$, then a saddlepoint approximation can be derived. Thus the assumption that the variable of interest is a sum or average of independent, identically distributed random variables would not be essential.

This observation is made precise in Robinson (1982) (see also Daniels, 1955) and applied to the one- and two-sample permutation test statistics, where the randomization distribution induces some dependence between the components. The same technique could be used for estimation in finite populations such as arise in sample surveys, although in this application the sample sizes are likely to be large enough that the saddlepoint refinement to the asymptotic distribution is somewhat academic. It could also be applied to approximate the distribution of linear rank statistics, such as the Wilcoxon statistic, as an alternative to the Hajek projection argument suggested in Section 6.4. The difficulty again is the need to evaluate the cumulant generating function.

Durbin (1980a) showed that Barndorff-Nielsen's formula given in equation (13) also applies if the maximum likelihood estimate $\hat{\theta}$ is a one-to-one function of the sufficient statistic, t. In the independent, identically distributed case this is only true for exponential families, but in the nonindependent, identically distributed setting some nonexponential models will satisfy this requirement. In particular the distribution of the serial correlation coefficient in an AR(1) model can be approximated this way (Daniels, 1956; Durbin, 1980b; Phillips, 1978).

In the probability and stochastic process literature, the computation of accurate tail probabilities is usually phrased in terms of laws of large deviations. The conjugate exponential family plays an important role in obtaining these results, and the embedding defined by (5) is also called exponential tilting. Accurate approximations to boundary crossing probabilities can be obtained this way, and these play a fundamental role in sequential analysis. Related references are Woodroofe (1982), Siegmund (1985), Asmussen (1985) and Sorensen (1986).

7. CONCLUSION

There are now available saddlepoint approximations for very many statistical problems. The emphasis in this paper has been on likelihood-based parametric inference, and the relationship between second order asymptotic results and the saddlepoint expansion, as this seems to be an area where substantial progress is now being made.

The saddlepoint approximations have not yet had much impact on statistical practice, though. This is partly because the computation involved needs specialized software, and partly because problems with fairly complex structure cannot yet be readily fitted in to the available framework. This can be expected to change rather quickly in the near future.

It is encouraging that an elegant mathematical technique has the scope to provide new insights for both applied and theoretical statistics.

APPENDIX: THE MULTIVARIATE SADDLEPOINT APPROXIMATION

This derivation follows McCullagh (1987, Chapter 6) and is included here for completeness. The random variable or vector X to be approximated is assumed to have density $f(x; \kappa)$, where the cumulants of X are denoted κ to conform with McCullagh. The normal density is denoted by ψ as in Section 2.

The univariate Edgeworth expansion, in its unstandardized form, is

$$f(x;\kappa) = \psi(x;\kappa)\left\{1 + \frac{\kappa_3 h_3(x;\kappa)}{6} + \frac{\kappa_4 h_4(x;\kappa)}{24} + \frac{\kappa_3^2 h_6(x;\kappa)}{72} + \cdots\right\}, \tag{A.1}$$

where $\psi(x; \kappa)$ is the normal (κ_1, κ_2) density and the Hermite polynomials $h_j(x; \kappa)$ are defined by

$$h_j(x;\kappa) = (-1)^j \frac{\partial^j \psi(x;\kappa)}{\partial x^j} \Big/ \psi(x;\kappa).$$

In the application of this to $f_{\bar{X}}(\bar{x}; \phi)$ outlined in Section 2, the cumulants κ are all functions of ϕ. The standardized version of (A.1), appropriate when X is a sum or average of independent, identically distributed random variables, is given by (5) of Section 2. It is an expansion in powers of $n^{-1/2}$. As noted in Section

2, at $x = \kappa_1$, or $z = 0$, the right hand side of (5) becomes

$$\text{(A.2)} \qquad \frac{1}{\sqrt{2\pi\kappa_2}}\left(1 + \frac{3\rho_4}{24} - \frac{15\rho_3^2}{72} + \cdots\right)$$

and the skewness term in the expansion has disappeared. In fact all odd powers of $n^{-1/2}$ are 0 at $x = \kappa_1$, and the expansion is in powers of n^{-1} for standardized random variables. The saddlepoint approximation exploits this by appropriately choosing the parameter of the conjugate density.

The multivariate Edgeworth expansion is most easily expressed in the version analogous to (A.1),

$$\text{(A.3)} \qquad \begin{aligned} &f(x;\kappa) \\ &= \psi(x;\kappa)\left\{1 + \frac{\kappa^{i,j,k}h_{ijk}(x;\kappa)}{3!} + \frac{k^{i,j,k,l}h_{ijkl}(x;\kappa)}{4!}\right. \\ &\qquad \left. + \frac{k^{i,j,k}\kappa^{l,m,n}h_{ijklmn}(x;\kappa)[10]}{6!} + \cdots\right\}. \end{aligned}$$

The approximating $\psi(x;\kappa)$ is now the multivariate normal density of dimension p with mean vector and covariance matrix matching the mean and covariance of X; the components of the vector x are written with superscripts $x = (x^1, \cdots, x^p)$. The Hermite polynomials are generated from ψ by differentiating:

$$h_{ijk}(x;\kappa) = (-1)^3\frac{\partial^3\psi(x;\kappa)}{\partial x^i\partial x^j\partial x^k}\Big/\psi(x;\kappa),$$

$$h_{ijkl}(x;\kappa) = (-1)^4\frac{\partial^4\psi(x;\kappa)}{\partial x^i\partial x^j\partial x^k\partial x^l}\Big/\psi(x;\kappa).$$

The multivariate cumulant arrays are easier to describe than to express in terms of moments. First note that since the components of X are written with superscripts, it is consistent to write the components of EX the same way: $E(X^1, \cdots, X^p) = (\kappa^1, \cdots, \kappa^p)$. The $p \times p$ covariance matrix of X is $(\kappa^{i,j})$, where the element $\kappa^{i,j}$ is the second cumulant, or covariance of X^i and X^j. Then $k^{i,j,k}$ is the joint third cumulant of X^i, X^j, X^k (in fact equal to $E\{(X^i - \kappa^i)(X^j - \kappa^j)(X^k - \kappa^k)\}$), and $\kappa^{i,j,k,l}$ is the joint fourth cumulant of X^i, X^j, X^k, X^l. These are arrays of size $p \times p \times p$ and $p \times p \times p \times p$, respectively. The summation convention is used in (A.3) and all indices run from 1 to p. The [10] in the final term is shorthand for the sum of 10 similar terms, each corresponding to one of the partitions of the six indices $i, \cdots, n$ into two sets of three. If the third cumulant array is symmetric with respect to the components of X, this can be replaced by the factor 10. In the univariate case each index appears only once, so it makes sense to replace $\kappa^{1,1,1}$ by κ_3 and h_{111} by h_3, etc., in which case (A.1) is recovered.

In order to standardize the variables to obtain a multivariate version of (5), with the basic normal density having covariance matrix I, some work is needed. For the bivariate case the standardization is explicitly presented in Barndorff-Nielsen and Cox (1979, Section 3). Beyond the bivariate case the standardization is better described than actually presented. All that is needed for the saddlepoint approximation is the multivariate version of ρ_4 and ρ_3^2. Dividing by κ_2 corresponds to multiplying by $(\kappa^{i,j})^{-1}$, written $\kappa_{i,j}$. It turns out (McCullagh, 1987, Chapter 2) that there is just one affinely invariant scalar obtainable from the fourth cumulant array, and exactly two affinely invariant scalars obtainable from the outer product of two third cumulant arrays. They are written

$$\rho_4 = \kappa^{i,j,k,l}\kappa_{i,j}\kappa_{k,l},$$

$$\rho_{13}^2 = \kappa^{i,j,k}\kappa^{l,m,n}\kappa_{i,j}\kappa_{k,l}\kappa_{m,n},$$

$$\rho_{23}^2 = \kappa^{i,j,k}\kappa^{l,m,n}\kappa_{i,l}\kappa_{j,m}\kappa_{k,n}$$

and the multivariate version of (A.2) is

$$\text{(A.4)} \qquad (2\pi)^{-p/2}|\kappa_{i,j}|^{1/2}\left(1 + \frac{3\rho_4 - 3\rho_{13}^2 - 2\rho_{23}^2}{24} + \cdots\right)$$

where the $(\cdots)$ in (A.4) is $O(n^{-2})$ under independent, identically distributed sampling.

ACKNOWLEDGMENTS

Much of the background material was developed in conjunction with a working seminar at the Center for Statistical Sciences, Austin, Texas, and I am grateful to the participants: J. Bjornstadt, A. C. Davison, D. V. Hinkley, C. N. Morris and C. L. Tsai. Many recent and current developments in asymptotic theory were discussed at an S.E.R.C. workshop held in Edinburgh in June 1986. I am grateful to all the participants for making their unpublished work available and to P. R. Fisk for assembling a summary of the presentations. I would particularly like to thank D. R. Cox, A. C. Davison, D. A. S. Fraser, J. L. Jensen and R. Kass for helpful discussion, and P. McCullagh for providing the manuscript of his book. I am grateful to the editors and referees for thorough and helpful comments on the original draft, and particularly to one referee for helpful comments on the role of renormalization.

REFERENCES AND ANNOTATED BIBLIOGRAPHY

AMARI, S.-I. (1982). Geometrical theory of asymptotic ancillarity and conditional inference. *Biometrika* **69** 1–17.

AMARI, S.-I. (1985). *Differential-Geometrical Methods in Statistics. Lecture Notes in Statist.* **28**. Springer, New York.

ASMUSSEN, S. (1985). Conjugate processes and the simulation of ruin problems. *Stochastic Process. Appl.* **20** 213–229.

BARNDORFF-NIELSEN, O. E. (1978). *Information and Exponential Families in Statistical Theory*. Wiley, New York.

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

First discussion by Barndorff-Nielsen of the formula for the distribution of the m.l.e. Proof sketched for curved exponential families, using the affine ancillary as conditioning statistic, when parameter is one-dimensional. Proof sketched for transformation models.

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

Further discussion of the formula. Derivation from the saddlepoint in full exponential families. Several examples. Derivation of modified profile likelihood for inference with nuisance parameters. Less restrictive proof of exactness for transformation models.

BARNDORFF-NIELSEN, O. E. (1984). On conditionality resolution and the likelihood ratio for curved exponential models. *Scand. J. Statist.* **11** 157–170.

Error in the formula is $O(n^{-3/2})$ using a particular, unique, ancillary in $(k, k-1)$ families.

BARNDORFF-NIELSEN, O. E. (1985a). Correction to Barndorff-Nielsen (1984). *Scand. J. Statist.* **12** 19.

Further modification of the ancillary to get $O(n^{-3/2})$ in (k, d) families.

BARNDORFF-NIELSEN, O. E. (1985b). Confidence limits from $c|\hat{j}|^{1/2}\overline{L}$ in the single parameter case. *Scand. J. Statist.* **12** 83–87.

BARNDORFF-NIELSEN, O. E. (1986a). Inference on full or partial parameters based on the standardized signed log likelihood ratio. *Biometrika* **73** 307–322.

The signed square root of the likelihood ratio statistic can be recentered and rescaled to follow a $N(0, 1)$ distribution to $O(n^{-3/2})$, even in the presence of nuisance parameters.

BARNDORFF-NIELSEN, O. E. (1986b). Likelihood and observed geometries. *Ann. Statist.* **14** 856–873.

BARNDORFF-NIELSEN, O. E. (1987a). Discussion of "Parameter orthogonality and approximate conditional inference" by D. R. Cox and N. Reid. *J. Roy. Statist. Soc. Ser. B* **49** 18–20.

BARNDORFF-NIELSEN, O. E. (1987b). Likelihood, ancillarity and strings. *Proc. First World Congress of the Bernoulli Soc.* **2** 205–213.

BARNDORFF-NIELSEN, O. E. and COX, D. R. (1979). Edgeworth and saddlepoint approximations with statistical applications (with discussion). *J. Roy. Statist. Soc. Ser. B* **41** 279–312.

Contains much of the material relating Edgeworth and saddlepoint techniques to statistical applications. Derivation of the saddlepoint approximation from the Edgeworth expansion in a conjugate density. Bivariate and multivariate approximations, leading to saddlepoint approximations for conditional distributions. Illustrative examples. Brief discussion of Bartlett factor.

BARNDORFF-NIELSEN, O. E. and COX, D. R. (1984a). Bartlett adjustments to the likelihood ratio statistic and the distribution of the maximum likelihood estimator. *J. Roy. Statist. Soc. Ser. B* **46** 483–495.

Relationship between Barndorff-Nielsen's formula and likelihood ratio tests established and illustrated. Very important and interesting result.

BARNDORFF-NIELSEN, O. E. and COX, D. R. (1984b). The effect of sampling rules on likelihood statistics. *Internat. Statist. Rev.* **52** 309–326.

Extension of above results to data subject to censoring or sequential stopping. In general the Bartlett factor for the likelihood ratio depends on the stopping rule, even though the likelihood itself does not.

BARTLETT, M. S. (1937). Properties of sufficiency and statistical tests. *Proc. Roy. Soc. London Ser. A* **160** 268–282.

BARTLETT, M. S. (1953). Approximate confidence intervals. *Biometrika* **40** 12–19.

BHATTACHARYA, R. N. and GHOSH, J. K. (1978). On the validity of the formal Edgeworth expansion. *Ann. Statist.* **6** 434–451.

BHATTACHARYA, R. N. and RANGA RAO, R. (1976). *Normal Approximations and Asymptotic Expansions*. Wiley, New York.

BICKEL, P. and GHOSH, J. K. (1987). A decomposition of the likelihood ratio statistic and the Bartlett correction. Technical Report 111, Dept. Statistics, Univ. California, Berkeley. *IMS Bull.* **16** 40.

BLACKWELL, D. and HODGES, J. L. (1959). The probability in the extreme tail of a convolution. *Ann. Math. Statist.* **30** 1113–1120.

BLAESILD, P. and JENSEN, J. L. (1985). Saddlepoint formulas for reproductive exponential models. *Scand. J. Statist.* **12** 193–202.

The saddlepoint approximation to the density of the sample mean is exact for a class of exponential models. A new and simple proof of the exactness for scalar random variables is provided. Characterizing exactness in the multivariate setting is more difficult.

BLEISTEIN, N. and HANDELSMAN, R. A. (1975). *Asymptotic Expansions of Integrals*. Holt, Rinehart and Winston, New York.

Useful and fairly detailed discussion of approximation techniques, including saddlepoint and Laplace approximations. A standard reference text.

BUTLER, R. (1986). Predictive likelihood inference with applications (with discussion). *J. Roy. Statist. Soc. Ser. B* **48** 1–38.

COURANT, R. and HILBERT, D. (1950). *Methods of Mathematical Physics* **1**. Wiley, New York.

COX, D. R. (1948). A note on the asymptotic distribution of the range. *Biometrika* **35** 310–315.

Method of steepest descents used to approximate the distribution of $x_{(n)} - x_{(1)}$, for x_i's iid from a symmetric unimodal density. The idea of renormalizing the saddlepoint approximation to improve its accuracy introduced.

COX, D. R. (1980). Local ancillarity. *Biometrika* **67** 279–286.

COX, D. R. and HINKLEY, D. V. (1974). *Theoretical Statistics*. Chapman and Hall, London.

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

CRAMÉR, H. (1938). Sur un nouveau theoreme-limite des probabilites. *Actualités Sci. Indust.* **736** 5–23.

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

The fundamental paper. The saddlepoint approximation to the density of the mean derived by the method of steepest descents, and via the Edgeworth expansion for the conjugate density. Existence and uniqueness of saddlepoints discussed. Validity of expansion over its range discussed. Application to lattice random variables.

DANIELS, H. E. (1955). Discussion of "Permutation theory in the derivation of robust criteria and the study of departures from assumption" by G. E. P. Box and S. L. Andersen. *J. Roy. Statist. Soc. Ser. B* **17** 27–28.

DANIELS, H. E. (1956). The approximate distribution of the serial correlation coefficient. *Biometrika* **43** 169–185.

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

DANIELS, H. E. (1960). Approximate solutions of Green's type for univariate stochastic processes. *J. Roy. Statist. Soc. Ser. B* **22** 376–401.

Discussion of a related approximation technique also known as WKB method.

DANIELS, H. E. (1980). Exact saddlepoint approximations. *Biometrika* **67** 59–63.

DANIELS, H. E. (1982). The saddlepoint approximation for a general birth process. *J. Appl. Probab.* **19** 20–28.

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

Alternative derivations of the results of Field and Hampel, leading to approximations to distribution function and density for score statistic, maximum likelihood estimator, and other (one-dimensional) M-estimators. Brief discussion of the tail area approximation of Lugannani and Rice.

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

Adaptation of saddlepoint method to provide two tail area approximations, one due to Lugannani and Rice. Discussion of relative error over the range of x. Application to lattice random variables.

DAVISON, A. C. (1986). Approximate predictive likelihood. *Biometrika* **73** 323–332.

Application of Laplace's method to predictive likelihood and Bayes posterior predictive density. Leads to a Bayesian version of Barndorff-Nielsen's formula, now as an approximation to the posterior density of the parameter.

DAVISON, A. C. (1988). Approximate conditional inference in generalized linear models. *J. Roy. Statist. Soc. Ser. B.* **50**. To appear.

The double saddlepoint technique of Barndorff-Nielsen and Cox (1979) is applied to generalized linear models with canonical link, unknown scale.

DAVISON, A. C. and HINKLEY, D. V. (1988). Saddlepoint approximations in resampling methods. *Biometrika* **75**. To appear.

DE BRUIJN, N. G. (1970). *Asymptotic Methods in Analysis*, 2nd ed. North-Holland, Amsterdam. Reprinted by Dover, 1981.

More concise than Bleistein and Handelsman and nicely written.

DICICCIO, T. J. (1984). On parameter transformations and interval estimation. *Biometrika* **71** 477–485.

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

DICICCIO, T. J. (1988). Likelihood inference for linear regression models. *Biometrika* **75** 29–34.

These papers investigate computation of the Bartlett factor and other second-order corrections to standard techniques. This sometimes leads to approximations to Barndorff-Nielsen's formula that are relatively easy to compute.

DURBIN, J. (1980a). Approximations for densities of sufficient estimators. *Biometrika* **67** 311–333.

The saddlepoint approximation derived for the maximum likelihood estimator in the non-iid setting. See also Durbin's discussion to Barndorff-Nielsen and Cox (1979). Discussion of reduction of error to $O(n^{-3/2})$ by renormalization.

DURBIN, J. (1980b). The approximate distribution of serial correlation coefficients. *Biometrika* **67** 334–349.

EASTON, G. and RONCHETTI, E. (1986). General saddlepoint approximations. *J. Amer. Statist. Assoc.* **81** 420–430.

An Edgeworth expansion for an arbitrary symmetric statistic is used to approximate the derivative of the cumulant generating function, leading to an approximate saddlepoint expansion.

EFRON, B. (1978). The geometry of exponential families. *Ann. Statist.* **6** 362–376.

EFRON, B. (1981). Nonparametric standard errors and confidence intervals (with discussion). *Canad. J. Statist.* **9** 139–172.

EFRON, B. and HINKLEY, D. V. (1978). Assessing the accuracy of the maximum likelihood estimator: observed versus expected Fisher information. *Biometrika* **65** 457–487.

Introduction of an approximate ancillary constructed from the observed information. Approximate conditional density of the maximum likelihood estimator introduced; gives Barndorff-Nielsen's formula in location-scale models. No direct connection to saddlepoint approximations, but an important development for statistical theory.

EFRON, B. and STEIN, C. (1981). The jackknife estimate of variance. *Ann. Statist.* **9** 586–596.

EMBRECHTS, P., JENSEN, J. L., MAEJIMA, M. and TEUGELS, J. L. (1985). Approximations for compound Poisson and Polya processes. *Adv. in Appl. Probab.* **17** 623–637.

Application of saddlepoint approximation to a sum of a random number of iid random variables. See also Daniels (1982).

ERIKSEN, P. S. (1987). Proportionality of covariance matrices. *Ann. Statist.* **15** 732–748.

ESSCHER, F. (1932). The probability function in the collective theory of risk. *Skand. Aktuarietidskr.* **15** 175–195.

FELLER, W. (1971). *An Introduction to Probability Theory and Its Applications* **2**, 2nd ed. Wiley, New York.

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-estimators of location. *Biometrika* **69** 29–46.

An approximation to $d \log p_n(x)/dx$ is derived directly. When integrated it is equivalent to the saddlepoint approximation; see Daniels (1983).

FIRTH, D. (1987). On the efficiency of quasi-likelihood estimation. *Biometrika* **74** 233–246.

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

FRASER, D. A. S. (1968). *The Structure of Inference*. Wiley, New York.

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

GOOD, I. J. (1957). Saddlepoint methods for the multinomial distribution. *Ann. Math. Statist.* **28** 861–881.

GOOD, I. J. (1961). The multivariate saddlepoint method and chi-squared for the multinomial. *Ann. Math. Statist.* **32** 535–548.

HARRIS, P. (1985). An asymptotic expansion for the null distribution of the score statistic. *Biometrika* **72** 653–660.

HARRIS, P. and PEERS, H. W. (1980). The local power of the efficient score test statistic. *Biometrika* **67** 525–529.

HINKLEY, D. V. (1977). Conditional inference about a normal mean with known coefficient of variation. *Biometrika* **64** 105–108.

HINKLEY, D. V. (1980). Likelihood as approximate pivotal distribution. *Biometrika* **67** 287–292.

HOUGAARD, P. (1985). Saddlepoint approximations for curved exponential families. *Statist. Probab. Lett.* **3** 161–166.

Results of Field (1982) and Daniels (1983) simplify in this setting and more explicit formulas are provided. Nonlinear regression is discussed as a special case.

JEFFREYS, H. and JEFFREYS, B. S. (1961). *Methods of Mathematical Physics*. Cambridge Univ. Press, Cambridge.

Chapter 17 is a discussion of asymptotic expansions and the method of steepest descent.

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

JENSEN, J. L. (1986b). Inference for the mean of a gamma distribution with unknown shape parameter. *Scand. J. Statist.* **13** 135–151.

JENSEN, J. L. (1987). Standardized log-likelihood ratio statistics for mixtures of discrete and continuous distributions. *Ann. Statist.* **15** 314–324.

JENSEN, J. L. (1988). Uniform saddlepoint approximations. *Adv. in Appl. Probab.* **20**. To appear.

Thorough technical discussion of bounding the error for the whole range of the density.

KALBFLEISCH, J. D. and SPROTT, D. A. (1970). Application of likelihood methods to models involving large numbers of parameters. *J. Roy. Statist. Soc. Ser. B* **32** 175–208.

KASS, R. E., TIERNEY, L. and KADANE, J. B. (1987). The validity of posterior expectations based on Laplace's method. Technical Report, Dept. Statistics, Carnegie Mellon Univ.

KASS, R. E., TIERNEY, L. and KADANE, J. B. (1988). Asymptotics in Bayesian computation. In *Bayesian Statistics 3* (J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith, eds.). To appear.

KEILSON, J. (1963). On the asymptotic behaviour of queues. *J. Roy. Statist. Soc. Ser. B* **25** 464–476.

LAWLEY, D. N. (1956). A general method for approximating to the distribution of likelihood ratio criteria. *Biometrika* **43** 295–303.

The first, and most complicated, proof that the Bartlett factor adjusts all cumulants of the likelihood ratio statistic. The proof starts with a stochastic expansion of the maximum likelihood estimator.

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

A highly accurate formula for tail probabilities; its relation to the saddlepoint method explained in Daniels (1987).

MARDIA, K. V. (1970). Measures of multivariate skewness and kurtosis. *Biometrika* **57** 519–530.

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

A proof of Barndorff-Nielsen's formula is provided to $O(n^{-1})$, for scalar parameters. An important result is that any approximate ancillary can be used for the conditioning, and that the remaining component of the data vector, called locally sufficient, is related to the log-likelihood ratio statistic. Follows on Cox (1980), Durbin (1980a), Barndorff-Nielsen (1980) and Hinkley (1980). Further discussion and more general results in Chapter 8 of McCullagh (1987).

MCCULLAGH, P. (1984b). Tensor notation and cumulants of polynomials. *Biometrika* **71** 461–476.

Special notation is developed which is extremely useful for second order asymptotic calculations. Expanded development provided in Chapters 1 to 3 of McCullagh (1987).

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

Chapter 6 discusses the saddlepoint approximation, with emphasis on its derivation from the Edgeworth expansion (Chapter 5) and on the Legendre transform.

MCCULLAGH, P. and COX, D. R. (1986). Invariants and likelihood ratio statistics. *Ann. Statist.* **14** 1419–1430.

MITCHELL, A. F. S. (1988). Statistical manifolds of univariate elliptic distributions. *Internat Statist. Rev.* **56** 1–16.

NEYMAN, J. and SCOTT, E. L. (1948). Consistent estimates based on partially consistent observations. *Econometrica* **16** 1–32.

PEERS, H. W. (1971). Likelihood ratio and associated test criteria. *Biometrika* **58** 577–587.

Second order calculations indicate that none of the usual procedures dominates the others uniformly in terms of local power.

PHILLIPS, P. C. B. (1978). Edgeworth and saddlepoint approximations in first order noncircular autoregression. *Biometrika* **65** 91–98.

There are several papers on higher order asymptotics in the econometric literature: This paper provides a good introduction to those. See also Sargan (1976).

PIERCE, D. A. and SCHAFER, D. W. (1986). Residuals in generalized linear models. *J. Amer. Statist. Assoc.* **81** 977–986.

PITMAN, E. J. G. (1938). The estimation of the location and scale parameters of a continuous population of any given form. *Biometrika* **30** 391–421.

ROBINSON, J. (1982). Saddlepoint approximations for permutation tests and confidence intervals. *J. Roy. Statist. Soc. Ser. B* **44** 91–101.

ROCKAFELLAR, R. T. (1970). *Convex Analysis*. Princeton Univ. Press, Princeton, N. J.

SARGAN, J. D. (1976). Econometric estimators and the Edgeworth approximation. *Econometrica* **44** 421–448.

SIEGMUND, D. (1985). *Sequential Analysis: Tests and Confidence Intervals*. Springer, New York.

SKOVGAARD, I. (1981). Edgeworth expansions of the distribution of maximum likelihood estimators in the general (non-iid) case. *Scand. J. Statist.* **8** 227–236.

SKOVGAARD, I. (1985a). A second order investigation of asymptotic ancillarity. *Ann. Statist.* **13** 534–551.

SKOVGAARD, I. (1985b). Large deviation approximations for maximum likelihood estimators. *Probab. Math. Statist.* **6** 89–107.

SKOVGAARD, I. (1988a). Saddlepoint expansions for conditional distributions. *J. Appl. Probab.* **25**. To appear.

An Edgeworth expansion for the conditional distribution of the maximum likelihood estimator and investigation of the Efron-Hinkley (1978) results on observed Fisher information.

SKOVGAARD, I. (1988b). Saddlepoint expansions for directional test probabilities. *J. Roy. Statist. Soc. Ser. B.* **50**. To appear.

SKOVGAARD, L. (1984). A Riemannian geometry of the multivariate normal model. *Scand. J. Statist.* **11** 211–223.

SORENSEN, M. (1986). On sequential maximum likelihood estimation for exponential families of stochastic processes. *Internat. Statist. Rev.* **54** 191–210.

TIERNEY, L. and KADANE, J. B. (1986). Accurate approximations for posterior moments and marginal densities. *J. Amer. Statist. Assoc.* **81** 82–87.

Applications of Laplace's approximation in Bayesian inference.

TIERNEY, L., KASS, R. E. and KADANE, J. B. (1987). Approximation of expectations and variances using Laplace's method. Technical Report, Dept. Statistics, Carnegie Mellon Univ.

TINGLEY, M. and FIELD, C. A. (1986). Robust bootstrap confidence intervals. Technical Report 86-3, Dalhousie Univ.

WOODROOFE, M. (1982). *Nonlinear Renewal Theory in Sequential Analysis*. SIAM, Philadelphia.