# Edgeworth and Saddle-point Approximations with Statistical Applications

By O. BARNDORFF-NIELSEN and D. R. COX

*Aarhus University* *Imperial College, London*

[Read before the ROYAL STATISTICAL SOCIETY, at a meeting organized by the RESEARCH SECTION, on Wednesday, April 25th, 1979, Professor J. F. C. KINGMAN in the Chair]

SUMMARY

A simple exposition is given of Edgeworth and saddle-point approximations for some univariate, multivariate and conditional distributions. The application of these approximations to some problems of conditional statistical inference within the exponential family is illustrated. Examples connected with the time-dependent Poisson process, the von Mises distribution and with bioassay are among those studied in a little detail. Some general results are derived about conditional likelihoods and about the distribution of the maximum likelihood ratio test statistic. An appendix discusses regularity conditions.

*Keywords*: ASYMPTOTIC EXPANSION; CIRCULAR NORMAL DISTRIBUTION; CONDITIONAL INFERENCE; CONDITIONAL MAXIMUM LIKELIHOOD; EDGEWORTH SERIES; EXPONENTIAL FAMILY; HERMITE POLYNOMIALS; LARGE DEVIATION; MAXIMUM LIKELIHOOD RATIO TEST; SADDLE-POINT; STEEPEST DESCENT; TIME-DEPENDENT POISSON PROCESS; VON MISES DISTRIBUTION

## 1. INTRODUCTION

THIS paper has two objects. One is to give a simple statement of methods for obtaining asymptotic expansions for the densities of sums of independent vector random variables. The expansions are of the Edgeworth and saddle-point types and relate to conditional as well as to unconditional distributions. The second object is to apply the methods to study conditional densities arising in inference within the exponential family.

Section 2 discusses univariate results and is intended largely, although not entirely, as an introduction to Sections 3 and 4 dealing with bivariate and multivariate distributions. For conditional distributions the most important expansions are what we call the single and the double saddle-point expansions; these are discussed in Sections 3.3 and 4.3. Section 5 develops some illustrative examples and Section 6 gives two rather more general applications concerning conditional likelihoods and maximum likelihood ratio test statistics.

The discussion in the main part of the paper is deliberately informal without attention to regularity conditions. The appendix gives the required conditions both for continuous random variables, with which the paper is primarily concerned, and for the probability functions of discrete random variables. For a general account of asymptotic expansions connected with sums of independent random variables, see Bhattacharya and Rao (1976). We make no attempt to consider here problems, such as those arising in time series theory, which involve functions other than sums of independent random variables.

## 2. UNIVARIATE RESULTS

### 2.1. *Direct Edgeworth Expansion*

Let $U_1, \ldots, U_n$ be independent and identically distributed random variables with density $f(\cdot)$, moment generating function $M(\xi) = E(e^{-\xi U})$ and cumulant generating function $K(\xi) = \log M(\xi)$. The moment generating function is assumed to converge in a strip in the complex $\xi$-plane with the origin in its interior; of course, the more common definition of $M(\cdot)$ is recovered by changing the sign of $\xi$. Denote the cumulants of $U$ by $\{\kappa_j\}$, and the standardized

cumulants by $\rho_l = \kappa_l/\kappa_2^{\frac{1}{2}}$ $(l = 3, \ldots)$. Let $R_n = U_1 + \ldots + U_n$, $X_n = (R_n - n\kappa_1)/(a\sqrt{n})$, where $a$ is a scaling constant possibly, but not necessarily, the standard deviation $\sqrt{\kappa_2}$ of $U$.

The most immediate asymptotic expansion for the density of $X_n$ is by an Edgeworth series, derived formally by expanding the cumulant generating function $K_{X_n}(\xi)$ of $X_n$, and thereby the moment generating function $M_{X_n}(\xi)$, in powers of $1/\sqrt{n}$ and inverting. This gives what we shall call the direct Edgeworth expansion for the density of $X_n$,

$$f_{X_n}(x) = g(x;\, \kappa_2/a^2)\left\{1 + \frac{\rho_3}{6\sqrt{n}} H_3(ax/\sqrt{\kappa_2}) + \frac{\rho_4}{24n} H_4(ax/\sqrt{\kappa_2}) + \frac{\rho_3^2}{72n} H_6(ax/\sqrt{\kappa_2})\right\} + O(n^{-\frac{3}{2}}), \tag{2.1}$$

where $g(\cdot\,;\, \sigma^2)$ is the normal density of zero mean and variance $\sigma^2$ and where $H_l(\cdot)$ is the Hermite polynomial of degree $l$ defined by

$$(d^l/dx^l)\, g(x;\, 1) = (-1)^l H_l(x)\, g(x;\, 1). \tag{2.2}$$

In (2.1) the term in $1/\sqrt{n}$ corrects for skewness whereas the term in $1/n$ is an even function of $x$ and essentially corrects for kurtosis.

### 2.2. *Indirect Edgeworth Expansion*

An important feature of (2.1) is that the terms in odd powers of $1/\sqrt{n}$ depend on Hermite polynomials of odd degree, all of which vanish at $x = 0$. Thus if we wish to approximate to $f_{X_n}(0)$, the density at the mean, an expansion in powers of $1/n$, rather than in powers of $1/\sqrt{n}$, is obtained. A different aspect of the same thing is that (2.1) may give rather bad, and indeed negative, values in the tails.

Indirect Edgeworth expansion (Daniels, 1954) hinges on the idea of an Edgeworth expansion originating not from the density $f(\cdot)$ but rather from a suitable member of the exponential family or conjugate family (Khinchin, 1949)

$$f(u;\, \lambda) = e^{-u\lambda} f(u)/M(\lambda), \tag{2.3}$$

which can be written also in the more familiar form

$$f(u;\, \lambda) = \exp\{-u\lambda - \alpha(u) - \beta(\lambda)\}. \tag{2.4}$$

It has moment and cumulant generating functions

$$M(\xi;\, \lambda) = M(\xi + \lambda)/M(\lambda), \quad K(\xi, \lambda) = K(\xi + \lambda) - K(\lambda) = \beta(\xi + \lambda) - \beta(\lambda). \tag{2.5}$$

Consider the approach of Section 2.1 applied to the distribution $f(u;\, \lambda_0)$. We then define $X_n = (R_n - n\kappa_{1(0)})/(a\sqrt{n})$, where $\kappa_{1(0)} = E(U;\, \lambda_0)$. Of course we recover the previous situation exactly by putting $\lambda_0 = 0$. Suppose that we are interested in the density $f_{X_n}(x;\, \lambda_0)$ of $X_n$ for some fixed $x$. Then for any $\lambda$ we have from (2.4), on comparing $f_{X_n}(x;\, \lambda)$ with $f_{X_n}(x;\, \lambda_0)$ and using (2.5), that

$$f_{X_n}(x;\, \lambda_0) = \exp\{nK(\lambda) - nK(\lambda_0) + r(\lambda - \lambda_0)\} f_{X_n}(x;\, \lambda), \tag{2.6}$$

where $x = (r - n\kappa_{1(0)})/(a\sqrt{n})$. Equation (2.6) is central to the discussion. An alternative less direct derivation is *via* the relevant moment generating functions, using the "displacement rule" for the inversion of Laplace transforms.

Thus an approximation for $f_{X_n}(x;\, \lambda_0)$ can be obtained *via* an Edgeworth expansion of $f_{X_n}(x;\, \lambda)$ in (2.6) for any choice of $\lambda$ such that as $n \to \infty$ the value of $x$ deviates from the mean of $f_{X_n}(\cdot\,;\, \lambda)$ by a bounded multiple of the standard deviation; the direct Edgeworth expansion of Section 2.1 corresponds to the choice $\lambda = \lambda_0$.

An important special choice of $\lambda$ for a particular $x$ and $n$ is $\hat{\lambda} = \hat{\lambda}_n(x)$ defined equivalently by requiring that

$$E(X_n;\, \hat{\lambda}) = x, \quad E(U;\, \hat{\lambda}) = ax/\sqrt{(n)} + \kappa_{1(0)}, \tag{2.7}$$

or that $\hat{\lambda}$ is the formal maximum likelihood estimate for $\lambda$ based on the observation $x$. From either property we have the equivalent forms that

$$ax+\sqrt{(n)}\{K'(\hat{\lambda})+\kappa_{1(0)}\}=0, r+nK'(\hat{\lambda})=0. \tag{2.8}$$

The essential point is that $x$ is at the mean of the distribution of $X_n$ under $\hat{\lambda}$. Then, on taking zero argument in (2.1) and denoting cumulants and standardized cumulants under $\hat{\lambda}$ by $\{\hat{\kappa}_j\}$ and $\{\hat{\rho}_j\}$, we have from (2.6) that

$$f_{X_n}(x;\lambda_0)=\frac{\exp\{nK(\hat{\lambda})-nK(\lambda_0)+(n\kappa_{1(0)}+ax\sqrt{(n)})(\hat{\lambda}-\lambda_0)\}}{(2\pi\hat{\kappa}_2/a^2)^{\frac{1}{2}}}\left\{1+\frac{m_X(\hat{\rho})}{24n}+O(n^{-2})\right\}; \tag{2.9}$$

here $m_X(\rho)=3\rho_4-5\rho_3^2$ and we have used the facts that $H_4(0)=3$, $H_6(0)=-15$. As noted in Section 2.1 it will often be natural to take $a^2=\kappa_{2(0)}=\mathrm{var}(U;\lambda_0)$.

The leading term of (2.9) is central to the paper and it is therefore worth giving some alternative forms. Thus from (2.8) the leading term of (2.9) is

$$\frac{\exp\{nK(\hat{\lambda})-nK(\lambda_0)-nK'(\hat{\lambda})(\hat{\lambda}-\lambda_0)\}}{(2\pi\hat{\kappa}_2/a^2)^{\frac{1}{2}}}; \tag{2.10}$$

also, re-expressing the result as a density for $R_n$, we have that

$$f_{R_n}(r;\lambda_0)=\frac{\exp\{nK(\hat{\lambda})-nK(\lambda_0)+r(\hat{\lambda}-\lambda_0)\}}{(2\pi n\hat{\kappa}_2)^{\frac{1}{2}}}\{1+O(n^{-1})\} \tag{2.11}$$

for values of $r$ corresponding to bounded $x$.

The expansion using $\hat{\lambda}$ is the best indirect Edgeworth expansion in giving the best order of error. An alternative name is saddle-point approximation: the result was derived in the pioneering paper of Daniels (1954) by applying the saddle-point technique of asymptotic analysis to the inversion of the Laplace transform $M_{X_n}(\cdot)$. Equation (2.8) is easily seen to arise in defining the saddle point.

A valuable modification of the leading term of (2.11) is obtained by multiplication by a constant chosen so that the total integral of the resulting function is one; we call this the renormalized saddle-point approximation. It typically has error $O(n^{-\frac{3}{2}})$.

### 2.3. *A Simple Example*

About the simplest example is to take

$$f(u)=e^{-u}\quad(u\geqslant 0),\qquad K(\xi)=-\log(1+\xi). \tag{2.12}$$

The exact density of $R_n$ is, of course, of the gamma form, namely $r^{n-1}e^{-r}/(n-1)!$. The direct Edgeworth expansion gives a normal approximation for $(R_n-n)/\sqrt{n}$, supplemented by correction terms. From (2.8), $\hat{\lambda}=n/r-1$, so that $K(\hat{\lambda})=\log(r/n)$, $\hat{\kappa}_2=r^2/n^2$ and the leading term (2.11) with $\lambda_0=0$ is thus

$$\frac{r^{n-1}e^{-r}}{\sqrt{(2\pi)}\,n^{n-\frac{1}{2}}e^{-n}}, \tag{2.13}$$

which differs from the exact density only by the use of an approximation to $(n-1)!$ given by Stirling's formula. Thus by renormalizing the saddle-point approximation (2.13), the exact density is recovered, a remarkable fact noted by Daniels (1954).

Had the results been expressed in terms of $X_n=(R_n-n)/\sqrt{n}$ the corresponding linear transform of (2.13) would have been obtained. If we include the second term of (2.9), we have $\hat{\rho}_3=2$, $\hat{\rho}_4=6$ for the exponential density and the correction factor is $1-(12n)^{-1}$, the second term in the asymptotic expansion associated with Stirling's formula.

Finally, the exponential family (2.4) for the problem is $(1+\lambda)e^{-u(1+\lambda)}$, an exponential density of mean $(1+\lambda)^{-1}$. In this particular situation, the results for general $\lambda_0$ corresponds merely to a scale transformation of the results for $\lambda_0 = 0$.

For a detailed numerical study, see Pagurova (1965).

### 2.4. *Discussion*

As is common in studies of asymptotic expansions, the results can be put in a number of different forms; for example, the saddle-point approximations can be further expanded using the fact that $\hat{\lambda}-\lambda_0$ is typically $O(1/\sqrt{n})$. Such further expansions are not likely to be useful unless $K(\cdot)$ is complicated, but if they are made it is important to keep appropriate accuracy. Thus if in (2.9)–(2.11) $\hat{\kappa}_2$ were replaced by $\kappa_2$ the error committed would typically be $O(1/\sqrt{n})$. A normal approximation can be developed from (2.9) by expanding (2.8) to give approximately, with $a^2 = \kappa_2$, $\hat{\lambda} \sim -x/\sqrt{(n\kappa_2)}$, $K(\hat{\lambda}) \sim -\frac{1}{2}\hat{\lambda}^2\kappa_2$, $\hat{\kappa}_2 \sim \kappa_2$, when the leading term becomes the standard normal density; essentially the direct Edgeworth expansion is recovered by taking further terms. In general in (2.11), and subsequent similar expressions, $\hat{\lambda}$ may be replaced by an approximation $\tilde{\lambda}$ provided that $\hat{\lambda}-\tilde{\lambda} = O(n^{-1})$; in statistical terminology the maximum likelihood value is to be replaced by an asymptotically efficient value.

The proportional accuracy of (2.13) is constant for all $x$ in the example and by (2.9) constancy to order $1/n$ will apply in general, because $m_X(\hat{\rho})$ can be replaced by $m_X(\rho)$. Thus numerical renormalization of the leading term of the saddle-point approximation will typically produce an error that is $O(n^{-\frac{3}{2}})$. Sometimes, as in the example, an exact answer is produced by renormalization. A sufficient but not necessary condition for this is that the standardized cumulants should be independent of $\lambda$. A necessary and sufficient condition is that the leading term (2.10) is correct except possibly for a constant factor, even for $n = 1$. After use of the relation between variances and the cumulant generating function, the condition becomes

$$\alpha\{-\beta''(\xi)/\beta''(0)\}+\beta(\xi)-\xi\beta'(\xi)-\tfrac{1}{2}\log\beta''(\xi) = \text{const}; \tag{2.14}$$

we have not found a simpler form.

The leading term is exact even to the normalizing constant in only two cases, the normal distribution and the inverse normal (or Gaussian) distribution.

Note that while a given number of terms of the saddle-point approximation are usually more precise than the same number of terms of the direct Edgeworth expansion, the latter has the advantage in the continuous case of being easily integrated to give an expansion for the cumulative distribution function and also does not require the cumulant generating function $K(\cdot)$ to be available in explicit form. In the discrete case, the integral of an asymptotic expansion for the probability function, taken with the usual continuity correction, provides an asymptotic expansion for the distribution function only to order $1/\sqrt{n}$; to order $1/n$ an additional term is required, arising essentially from the discontinuous character of the distribution function.

### 2.5. *Law of Large Deviations*

The above expansions have been developed in the context of a fixed $x$ for which $\hat{\lambda} = \hat{\lambda}_n(x)$ differs from $\lambda_0$ by $O(1/\sqrt{n})$. The argument is, however, valid for any sequence $\{x_n\}$ such that $\{\hat{\lambda}_n(x_n)\}$ is bounded. In particular, we may take $\lambda_0 = 0$, $a^2 = \kappa_2$, $x_n = \sqrt{(n)}(\bar{r}-\kappa_1)/\sqrt{\kappa_2}$, so that when $X_n = x_n$, $R_n = U_1+\ldots+U_n = n\bar{r}$. Then $\hat{\lambda}_n(x_n) = \bar{\lambda}(\bar{r})$, say, satisfies by (2.8) $\bar{r}+K'(\bar{\lambda}) = 0$. The leading term of (2.11) now gives that

$$f_{\bar{R}_n}(\bar{r}) \sim [\exp\{nK(\bar{\lambda})+n\bar{r}\bar{\lambda}\}/(2\pi n\kappa_2)^{\frac{1}{2}}], \tag{2.15}$$

one version of a law of large deviations. For a more thorough study from this point of view, see Richter (1957), and for a general account Ibragimov and Linnik (1971, Chapter 6).

## 3. BIVARIATE RESULTS

### 3.1. *Preliminary Remarks*

We now develop bivariate and then multivariate versions of the results of Section 2 both for joint densities and for conditional densities. There are advantages in proceeding *via* the study of uncorrelated random variables, deriving the general case by linear transformation.

So far as is feasible, a notation is used that lends itself to easy generalization.

### 3.2. *Direct Edgeworth Expansion*

Let $(U_1, V_1), \ldots, (U_n, V_n)$ be independent and identically distributed vectors with density $f(\cdot, \cdot)$, cumulant generating function $K(\xi, \eta)$, cumulants $\kappa_{lm}$ and standardized cumulants $\rho_{lm}$. Write $R_n = U_1 + \ldots + U_n$, $S_n = V_1 + \ldots + V_n$ and introduce

$$X_n = (R_n - n\kappa_{10})/(a\sqrt{n}), \quad Y_n = (S_n - n\kappa_{01})/(b\sqrt{n}).$$

For a direct Edgeworth expansion (Mardia, 1970), we expand the moment generating function exactly as in the univariate case. It is convenient to use a notation in which a formal product $\mathbf{H}^{\mathrm{T}}\boldsymbol{\rho} = H_1\rho_1 + H_2\rho_2$ of $2\times 1$ vectors is expanded by the binomial theorem before the powers of $\rho_1$ and $\rho_2$ are replaced by appropriate individual standardized cumulants and the powers of $H_1$ and $H_2$ by Hermite polynomials. For example,

$$\left.\begin{aligned}(\mathbf{H}^{\mathrm{T}}\boldsymbol{\rho})^{[3]}(x,y) &= \rho_{30}H_3(x) + 3\rho_{21}H_2(x)H_1(y) + 3\rho_{12}H_1(x)H_2(y) + \rho_{03}H_3(y),\\ \{(\mathbf{H}^{\mathrm{T}}\boldsymbol{\rho})^{[3]}\}^2(x,y) &= \rho_{30}^2H_6(x) + \ldots + \rho_{03}^2H_6(y).\end{aligned}\right\} \tag{3.1}$$

Then in the special case $\kappa_{11} = 0$, i.e. with uncorrelated $U$ and $V$, and with $a = \sqrt{\kappa_{20}}$, $b = \sqrt{\kappa_{02}}$, we have that

$$f_{X_n,Y_n}(x,y) = g(x;1)g(y;1) \times \left[1 + \frac{1}{6\sqrt{n}}(\mathbf{H}^{\mathrm{T}}\boldsymbol{\rho})^{[3]}(x,y) + \frac{1}{24n}(\mathbf{H}^{\mathrm{T}}\boldsymbol{\rho})^{[4]}(x,y) + \frac{1}{72n}\{(\mathbf{H}^{\mathrm{T}}\boldsymbol{\rho})^{[3]}\}^2(x,y)\right] + O(n^{-\frac{3}{2}}). \tag{3.2}$$

If a general scaling is used the arguments $x$ and $y$ become $ax/\sqrt{\kappa_{20}}$ and $by/\sqrt{\kappa_{02}}$. If $U$ and $V$, instead of being uncorrelated, have correlation coefficient $\rho_{11} = \rho$ then (3.2) applies to the random variables

$$X_n,\ Y_n' = (Y_n - \rho X_n)/\sqrt{(1-\rho^2)}.$$

Write $\rho_{lm}'$ for the $(l,m)$ standardized cumulant of $X_n$ and $Y_n'$ and $\rho'$ for the corresponding generating vector in (3.1). On applying (3.2) to $(X_n, Y_n')$ and then transforming back to $(X_n, Y_n)$, we have for the general version of (3.2) that

$$f_{X_n,Y_n}(x,y) = \frac{g(x;1)g(y';1)}{(1-\rho^2)^{\frac{1}{2}}} \times \left[1 + \frac{1}{6\sqrt{n}}(\mathbf{H}^{\mathrm{T}}\boldsymbol{\rho}')^{[3]}(x,y') + \frac{1}{24n}(\mathbf{H}^{\mathrm{T}}\boldsymbol{\rho}')^{[4]}(x,y') + \frac{1}{72n}\{(\mathbf{H}^{\mathrm{T}}\boldsymbol{\rho}')^{[3]}\}^2(x,y')\right] + O(n^{-\frac{3}{2}}). \tag{3.3}$$

The leading term is a bivariate normal density in $(x,y)$ of zero mean and will be written $g_2(x,y;\boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is the covariance matrix of $(X_n, Y_n)^{\mathrm{T}}$.

The correction terms in (3.3) are expressed in terms of standard Hermite polynomials with arguments the orthonormal variables and coefficients the cumulants of these variables. This is probably the simplest form for appreciating the structure of the formula. It is clear, however,

that because of the linearity of the transformation from $(x, y)$ to $(x, y'$, the result could be expressed in terms of the vector $\boldsymbol{\rho}$ of cumulants of the original variables $(x, y)$ and some new polynomials. Thus, for example,

$$(\mathrm{H}^{\mathrm{T}}\boldsymbol{\rho}')^{[3]}(x,y') = (\mathscr{H}^{\mathrm{T}}\boldsymbol{\rho})^{[3]}(x,y)$$
$$= \rho_{30}H_{30}(x,y;\Delta)+3\rho_{21}H_{21}(x,y;\Delta)+3\rho_{12}H_{12}(x,y;\Delta)+\rho_{03}H_{03}(x,y;\Delta),$$

where $\Delta$ is conveniently taken to be the inverse covariance matrix of $(X, Y)$. With this and similar notation (3.3) and subsequent formulae can be rewritten with $\mathscr{H}^{\mathrm{T}}\boldsymbol{\rho}$ replacing $\mathrm{H}^{\mathrm{T}}\boldsymbol{\rho}'$ and $(x, y)$ replacing $(x, y')$. It turns out that the generalized Hermite polynomials required are exactly those introduced by Chambers (1967) by appropriate partial differentation in the general case of $\exp(-\frac{1}{2}\mathbf{z}^{\mathrm{T}}\Delta\mathbf{z})$. We shall not explore this connection further here; Barndorff-Nielsen and Pedersen (1979) give explicit formulae for the polynomials up to degree 6.

If we consider the density at $(0,0)$ many terms disappear and in fact

$$f_{X_n,Y_n}(0,0) = \frac{1}{2\pi D^{\frac{1}{2}}}\left\{1+\frac{m_{XY}(\rho')}{24n}+O(n^{-\frac{3}{2}})\right\}, \tag{3.4}$$

where $D = \det(\Sigma)$ and

$$m_{XY}(\rho') = 3\rho'_{40}+6\rho'_{22}+3\rho'_{04}-5\rho'^{2}_{30}-9\rho'^{2}_{21}-9\rho'^{2}_{12}-5\rho'^{2}_{03}-6\rho'_{30}\rho'_{12}-6\rho'_{03}\rho'_{21}. \tag{3.5}$$

An expansion for the conditional density of $Y_n$ given $X_n = x$ follows on dividing (3.3) by the corresponding expansion (2.1) for $f_{X_n}(x)$. There results

$$f_{Y_n|X_n}(y|x) = \frac{g(y';1)}{(1-\rho^2)^{\frac{1}{2}}}\left(1+\frac{1}{6\sqrt{n}}\{(\mathrm{H}^{\mathrm{T}}\boldsymbol{\rho}')^{[3]}(x,y')-\rho_{30}H_3(x)\}+\frac{1}{24n}\{(\mathrm{H}^{\mathrm{T}}\boldsymbol{\rho}')^{[4]}(x,y')-\rho_{40}H_4(x)\}\right.$$
$$+\frac{1}{72n}[\{(\mathrm{H}^{\mathrm{T}}\boldsymbol{\rho}')^{[3]}\}^2(x,y')-\rho_{30}^2H_6(x)]+\left.\frac{1}{36n}\rho_{30}H_3(x)\{\rho_{30}H_3(x)-(\mathrm{H}^{\mathrm{T}}\boldsymbol{\rho}')^{[3]}(x,y')\}\right)$$
$$+O(n^{-1}). \tag{3.6}$$

When $x = 0$, we have that

$$f_{Y_n|X_n}(y|0) = g(y;1-\rho^2)\left[1+\frac{1}{6\sqrt{n}}\{-3\rho'_{21}H_1(y')+\rho'_{03}H_3(y')\}+\frac{m_{Y.X}(\rho';y')}{24n}\right]+O(n^{-\frac{3}{2}}), \tag{3.7}$$

where

$$m_{Y.X}(\rho';y') = -6\rho'_{22}H_2(y')+\rho'_{04}H_4(y')+9\rho'^{2}_{21}H_2(y')-3\rho'^{2}_{12}H_4(y')$$
$$+\tfrac{1}{3}\rho'^{2}_{03}H_6(y')+6\rho'_{30}\rho'_{12}H_2(y')-2\rho'_{21}\rho'_{03}H_4(y').$$

If further $y = 0$, there results

$$f_{Y_n|X_n}(0|0) = \frac{1}{(2\pi D)^{\frac{1}{2}}}\left\{1+\frac{m_{Y.X}(\rho')}{24n}+O(n^{-2})\right\}, \tag{3.8}$$

where

$$m_{Y.X}(\rho') = 6\rho'_{22}+3\rho'_{04}-3(3\rho'^{2}_{21}+3\rho'^{2}_{12}+2\rho'_{30}\rho'_{12}+2\rho'_{03}\rho'_{21})-5\rho'^{2}_{03}.$$

### 3.3. *Indirect Edgeworth Expansion*

As in the one-dimensional case, we introduce the exponential family

$$f(\mathbf{w};\boldsymbol{\theta}) = f(u,v;\lambda,\psi)$$
$$= e^{-\lambda u-\psi v}f(u,v)/M(\lambda,\psi)$$
$$= \exp\{-\boldsymbol{\theta}^{\mathrm{T}}\mathbf{w}-\alpha(\mathbf{w})-\beta(\boldsymbol{\theta})\}, \tag{3.9}$$

where $\mathbf{w}^{\mathrm{T}} = (u, v)$, $\boldsymbol{\theta}^{\mathrm{T}} = (\lambda, \psi)$ are observation and parameter vectors. The cumulant generating function is $K(\xi, \eta) = \log M(\xi, \eta)$.

The "maximum likelihood" point $\hat{\boldsymbol{\theta}}$ for given $\mathbf{t}^{\mathrm{T}} = (r, s)$ is defined most directly via the analogue of (2.8), namely

$$\mathbf{t} + n \operatorname{grad} K(\hat{\boldsymbol{\theta}}) = 0, \tag{3.10}$$

where $\operatorname{grad} K(\hat{\boldsymbol{\theta}})$ is the column vector of partial derivatives of $K(\boldsymbol{\theta})$, evaluated at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$. Equation (3.10) can, if required, be expressed in terms of the standardized variables $(x, y)$. The two-dimensional extension of (2.6) is that with $\boldsymbol{\theta} = \boldsymbol{\theta}_0$

$$X_n = (R_n - n\kappa_{10(0)})/(a\sqrt{n}), \quad Y_n = (S_n - n\kappa_{01(0)})/(b\sqrt{n}),$$

$$f_{X_n, Y_n}(x, y; \boldsymbol{\theta}_0) = \exp\{nK(\boldsymbol{\theta}) - nK(\boldsymbol{\theta}_0) + \mathbf{t}^{\mathrm{T}}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)\} f_{X,Y}(x, y; \boldsymbol{\theta}). \tag{3.11}$$

Hence, on taking $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ and applying (3.4), we have that

$$f_{X_n, Y_n}(x, y; \boldsymbol{\theta}_0) = \frac{\exp\{nK(\hat{\boldsymbol{\theta}}) - nK(\boldsymbol{\theta}_0) + \mathbf{t}^{\mathrm{T}}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)\}}{2\pi\{\hat{D}/(a^2 b^2)\}^{\frac{1}{2}}} \left\{1 + \frac{m_{XY}(\hat{\rho}')}{24n} + O(n^{-2})\right\}, \tag{3.12}$$

where $\hat{D} = D(\hat{\boldsymbol{\theta}})$ is the determinant of second derivatives of $K(\cdot, \cdot)$, evaluated at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$, i.e. the generalized variance at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$, and the standardized cumulants required to calculate $m_{XY}(\cdot)$ are those of $U$ and of $V$ orthogonalized with respect to $U$ at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$. If the leading term of (3.12) is renormalized numerically then, as in Section 2.4, an error that is $O(1/n^3)$ results.

To obtain an expansion for $f_{Y_n|X_n}(y|x; \boldsymbol{\theta}_0)$, we have to take approximations for $f_{X_n, Y_n}(x, y; \boldsymbol{\theta}_0)$ and $f_{X_n}(x; \boldsymbol{\theta}_0)$. There are thus a number of possibilities depending on the combinations of approximations used. If $\boldsymbol{\theta}_0 = (\lambda_0, \psi_0)^{\mathrm{T}}$, the cumulant generating function of $U$ is $K(\lambda_0 + \xi, \psi_0) - K(\lambda_0, \psi_0)$. It thus follows from (2.6) and (3.11) that

$$f_{X_n}(x; \boldsymbol{\theta}_0) = \exp\{nK(\lambda', \psi_0) - nK(\lambda_0, \psi_0) + r(\lambda' - \lambda_0)\} f_{X_n}(x; \lambda', \psi_0), \tag{3.13}$$

$$f_{X_n, Y_n}(x, y; \boldsymbol{\theta}_0) = \exp\{nK(\lambda'', \psi) - nK(\lambda_0, \psi_0) + \mathbf{t}^{\mathrm{T}}(\boldsymbol{\theta}'' - \boldsymbol{\theta}_0)\} f_{X_n, Y_n}(x, y; \lambda'', \psi), \tag{3.14}$$

where $\boldsymbol{\theta}'' = (\lambda'', \psi)$, so that approximations for $f_{Y_n|X_n}(y|x; \boldsymbol{\theta}_0)$ can be derived via Edgeworth expansions of the final factors. Note that the ratio of (3.14) to (3.13) does not depend on $\lambda_0$, in line with considerations of sufficiency.

Other than direct Edgeworth expansion, as in Section 3.2, there are two main possibilities. One is to apply separate saddle-point approximations to (3.13) and (3.14), thus taking $\lambda' = \hat{\lambda}_{(0)}$ in (3.13) and $(\lambda'', \psi)^{\mathrm{T}} = \hat{\boldsymbol{\theta}}$ in (3.14), where $\lambda'_{(0)}$ is the "maximum likelihood" value for $\lambda$ when $\psi = \psi_0$. This choice gives

$$f_{Y_n|X_n}(y|x; \boldsymbol{\theta}_0) = \frac{\exp\{nK(\hat{\lambda}, \hat{\psi}) - nK(\hat{\lambda}_{(0)}, \psi_0) + r(\hat{\lambda} - \hat{\lambda}_{(0)}) + s(\hat{\psi} - \psi_0)\}}{\{2\pi\hat{D}/(\hat{\kappa}_{20(0)} b^2)\}^{\frac{1}{2}}}$$

$$\times \left\{1 + \frac{m_{XY}(\hat{\rho}') - m_X(\hat{\rho}_{(0)})}{24n} + O(n^{-2})\right\}, \tag{3.15}$$

where $\hat{\kappa}_{20(0)}$ is calculated at $(\hat{\lambda}_{(0)}, \psi_0)$ and $m_X(\hat{\rho}_{(0)})$ is determined from (2.9). To the order indicated it is enough to calculate the correction terms in (3.15) at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. We call (3.15) the double saddle-point approximation. It can be re-expressed simply in terms of $f_{S_n|R_n}(s|r; \boldsymbol{\theta}_0)$. Approximations of this kind were considered by Daniels (1958).

A second possibility is to take $\lambda' = \lambda'' = \lambda$, $\psi = \psi_0$, when the exponential terms in (3.13) and (3.14) cancel, and we are lead to take a direct Edgeworth expansion for the conditional density under $(\lambda, \psi_0)$ for any $\lambda$ such that $(r, s)$, and hence $(x, y)$, correspond to bounded standardized deviations from $(\lambda, \psi_0)$. The particular choice $\lambda = \hat{\lambda}_{(0)}$ ensures that after restandardization

the conditioning variable is zero, so that (3.7) is applicable. We call this the single saddle-point approximation or the mixed Edgeworth saddle-point approximation.

The leading term from (3.7) is such that, given $R_n = r$, $S$ is normal with mean $nE(V; \hat{\lambda}_{(0)}, \psi_0)$ and variance

$$n\{\mathrm{var}(V; \hat{\lambda}_{(0)}, \psi_0) - \mathrm{cov}^2(U, V; \hat{\lambda}_{(0)}, \psi_0)/\mathrm{var}(U; \hat{\lambda}_{(0)}, \psi_0)\}. \tag{3.16}$$

The correction terms are given from (3.7) with the $\rho'$'s being standardized cumulants of $U$ and $V'$, evaluated at $(\hat{\lambda}_{(0)}, \psi_0)$ and with $y'$ the standardized orthogonalized deviate

$$\{s - nE(V; \hat{\lambda}_{(0)}, \psi_0)\}/\{n\hat{\kappa}_{20(0)}(1-\hat{\rho}^2_{(0)})\}^{\frac{1}{2}}, \tag{3.17}$$

where $\hat{\rho}_{(0)} = \rho(\hat{\lambda}_{(0)}, \psi_0) = \mathrm{corr}(U, V; \hat{\lambda}_{(0)}, \psi_0)$.

To compute the correction terms we have, as noted above, the possibilities of direct use of (3.7) in terms of orthonormal variables, and of re-expressing (3.7) in terms of the cumulants of the original variables and the generalized polynomials, specified by the operator $\mathscr{H}$.

In statistical applications we shall very often want the cumulative probability corresponding to (3.16), the leading term being $\Phi(y')$, where $\Phi(\cdot)$ is the standard normal integral. The correction terms, obtained by integrating the single saddle-point expansion, have the general form

$$\{h_2(y')\phi(y')/\sqrt{n}\} - \{h_5(y')\phi(y')/n\}, \tag{3.18}$$

where $h_j(\cdot)$ is a polynomial of degree $j$ and $\phi(y) = g(y; 1)$ is the standard normal density. Pedersen (1979) has given an algorithm for the computation of this, once the relevant originating moments are known. He gives also further details about single saddle-point approximations.

### 3.4. *Two Simple Examples*

First suppose that $V_1, \ldots, V_n$ are independently normally distributed with mean $\mu$ and variance $\sigma^2$ and let $U_i = V_i^2$. Then $R_n = \Sigma V_i^2$, $S_n = \Sigma V_i$ and direct calculation shows that for fixed $(\mu, \sigma^2)$

$$f_{S_n|R_n}(s|r) = \exp(s\mu/\sigma^2)(r - s^2/n)^{\frac{1}{2}n-\frac{3}{2}} \times \left(2\sqrt{(n)}\, r^{\frac{1}{2}n-1}\int_0^1 \cosh\{x\sqrt{(nr)}\,\mu/\sigma^2\}(1-x^2)^{\frac{1}{2}n-\frac{1}{2}}\,dx\right)^{-1}. \tag{3.19}$$

The denominator of (3.19) is a normalizing constant. The result could be expressed in terms of the standardized variables

$$X_n = \{R_n - n(\mu^2+\sigma^2)\}/(a\sqrt{n}), \quad Y_n = (S_n - n\mu)/(b\sqrt{n}),$$

where a special choice is $a = \sigma_U = \{2\sigma^2(\sigma^2+2\mu^2)\}^{\frac{1}{2}}$, $b = \sigma$.

If we apply the direct Edgeworth expansion we recover a normal approximation to the conditional distribution, with correction terms; for the latter we calculate the joint standardized cumulants of the orthogonalized variables

$$(V^2 - \mu^2 - \sigma^2)/\sigma_U, \quad (V-\mu)/\sigma - \rho(V^2 - \mu^2 - \sigma^2)/\sigma_U,$$

where $\rho = \mathrm{corr}(V, V^2)$. As will be usual when the exact density can be written in reasonably explicit form, the expansion can be obtained directly, here by expanding the numerator about its maximum.

To apply the double saddle-point approximation (3.15) it is simplest to take $\boldsymbol{\theta} = 0$ as corresponding to the standard normal distribution, when the exponential family (3.9) is normal with mean $\mu$ and variance $\sigma^2$, where

$$\lambda = \tfrac{1}{2}\sigma^{-2} - \tfrac{1}{2}, \quad \psi = -\mu/\sigma^2,$$

$$M(\xi, \eta) = \exp\{\tfrac{1}{2}\eta^2/(1+2\xi)\}(1+2\xi)^{-\frac{1}{2}}.$$

The "maximum likelihood" values $(\hat{\lambda},\hat{\psi})$ are derived via the usual maximum likelihood estimates of mean and variance; the quantity $\hat{\lambda}_{(0)}$ derived when $\psi=\psi_0$ is a function of $r$ alone and not of $s$. It follows on substituting into (3.15) that the leading term is

$$k_n(r,\psi_0)\exp(-\psi_0 s)(r-s^2/n)^{\frac{1}{2}n-\frac{1}{2}}, \tag{3.20}$$

where $k_n(r,\psi_0)$ is a complicated but elementary function. Thus, except for the normalizing constant, the exact conditional density is recovered.

As a second example, consider the special inverse Gaussian density

$$\pi^{-\frac{1}{2}}u^{-\frac{1}{2}}\exp(2-u-1/u) \tag{3.21}$$

which generates the exponential family

$$\{(1+\psi)/\pi\}^{\frac{1}{2}}\exp\{2(1+\lambda)^{\frac{1}{2}}(1+\psi)^{\frac{1}{2}}\}u^{-\frac{1}{2}}\exp\{-(1+\lambda)u-1(+\psi)v\}, \tag{3.22}$$

with $v=1/u$. The rather clumsy parameterization is, of course, to fit in with the general formulation. If $U_1,\ldots,U_n$ are independent and identically distributed with the above density and $R_n=\Sigma U_i$, $S_n=\Sigma(1/U_i)$, we can, for example, use (3.12) to approximate to the joint density $f_{R_n,S_n}(r,s;\boldsymbol{\theta}_0)$.

The cumulant generating function of (3.21) is

$$K(\xi,\eta)=-\tfrac{1}{2}\log(1+\eta)-2\{(1+\xi)(1+\eta)\}^{\frac{1}{2}}+2.$$

The formal "maximum likelihood" equations are, with $\bar{r}=r/n$, $\bar{s}=s/n$,

$$(\bar{s}-1/\bar{r})^{-1}=2(1+\hat{\psi}),\quad\{\bar{r}^2(\bar{s}-1/\bar{r})\}^{-1}=2(1+\hat{\lambda}),$$

and the determinant of the matrix of second derivatives of $K(\cdot,\cdot)$ is $\frac{1}{4}\{(1+\xi)(1+\eta)\}^{-\frac{1}{2}}$.

It follows after some calculation that

$$\begin{aligned}f_{R_n,S_n}(r,s;\boldsymbol{\theta}_0)=(n\pi)^{-1}2^{\frac{1}{2}(n-3)}e^{\frac{1}{2}n}(1+\psi_0)^{\frac{1}{4}n}\exp\{2n(1+\psi_0)^{\frac{1}{2}}(1+\lambda_0)^{\frac{1}{2}}\}\\ \times\bar{r}^{-\frac{1}{2}}\exp\{-n(1+\lambda_0)\bar{r}-n(1+\psi_0)\bar{s}^{-1}\}(\bar{s}-\bar{r}^{-1})^{\frac{1}{2}(n-3)}\\ \times\exp\{-n(1+\psi_0)(\bar{s}-\bar{r}^{-1})\},\end{aligned} \tag{3.23}$$

which again is exact except for the normalizing constant (Tweedie, 1957).

## 4. Multivariate Results

### 4.1. *Preliminary Remarks and Notation*

The discussion of Section 3 has been put in a form for fairly easy generalization. Let $\mathbf{W}=(W^{(1)},\ldots,W^{(d)})^{\mathrm{T}}$ be a $d$-dimensional random variable with density $f_{\mathbf{W}}(\mathbf{w})$, cumulants $\kappa_{l_1,\ldots,l_s}$ and standardized cumulants $\rho_{l_1,\ldots,l_s}$. Write the associated exponential family

$$f(\mathbf{w})\exp(-\mathbf{w}^{\mathrm{T}}\boldsymbol{\theta})/M(\boldsymbol{\theta}) \tag{4.1}$$

with cumulant generating function $K(\boldsymbol{\xi})=\log M(\boldsymbol{\xi})$.

Let $\mathbf{W}_1,\ldots,\mathbf{W}_n$ be independent and identically distributed with the above density, $\mathbf{T}_n=\mathbf{W}_1+\ldots+\mathbf{W}_n=(T_n^{(1)},\ldots,T_n^{(d)})^{\mathrm{T}}$ and denote the standardized sum by $\mathbf{Z}_n$ with components

$$\{T_n^{(1)}-nE(W^{(1)})\}/(c_1\sqrt{n}),\ldots,\{T_n^{(d)}-nE(W^{(d)})\}/(c_d\sqrt{n}), \tag{4.2}$$

where a natural choice is $c_j=\sqrt{\operatorname{var}(W^{(j)})}$ $(j=1,\ldots,d)$.

Note that in the bivariate case $\mathbf{W}=(U,V)^{\mathrm{T}}$, $\mathbf{T}_n=(R_n,S_n)^{\mathrm{T}}$, $\mathbf{Z}_n=(X_n,Y_n)^{\mathrm{T}}$, $\boldsymbol{\theta}=(\lambda,\psi)^{\mathrm{T}}$, $\boldsymbol{\zeta}=(\xi,\eta)^{\mathrm{T}}$ and $c_1=a$, $c_2=b$. When we consider general conditional distributions it is convenient to write $\mathbf{W}=(\mathbf{U},\mathbf{V})^{\mathrm{T}}$, etc., where now $\mathbf{U}$ and $\mathbf{V}$ are respectively $d_1\times 1$ and $d_2\times 1$, with $d_1+d_2=d$.

It is again convenient to work with uncorrelated components and we write $\mathbf{W}' = \mathbf{C}(\boldsymbol{\theta}_0)\mathbf{W}$, $\mathbf{Z}'_n = \mathbf{C}(\boldsymbol{\theta}_0)\mathbf{Z}_n$, where $\mathbf{C}(\boldsymbol{\theta}_0)$ is a lower triangular matrix such that the components of $\mathbf{W}'$ are uncorrelated and of unit variance when $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. Note that with the convention that in (4.2) $c_1 = \ldots = c_d = 1$, we can use the same transformation to produce an orthonormal $\mathbf{Z}'_n$. The reason for considering a lower triangular transformation is to ensure that when $\mathbf{W}$ is partitioned the first $d_1$ components of $\mathbf{W}'$ depend only on the first $d_1$ components of $\mathbf{W}$. If $\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$ is the covariance matrix of $\mathbf{W}$ when $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, assumed non-singular, then

$$\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0) = \mathbf{C}^{\mathrm{T}}(\boldsymbol{\theta}_0)\mathbf{C}(\boldsymbol{\theta}_0). \tag{4.3}$$

We write $\rho'_{i_1,\ldots,i_s}(\boldsymbol{\theta}_0)$ for the standardized cumulants of $\mathbf{W}'$.

Finally, we write $g_d(\mathbf{w}; \boldsymbol{\Sigma})$ for the density of the $d$-dimensional multivariate normal distribution of zero mean and covariance matrix $\boldsymbol{\Sigma}$.

### 4.2. *Direct Edgeworth Expansion*

For the direct Edgeworth expansion applied to orthonormal components (3.2) generalizes immediately, with obvious extensions of notation. We again consider expansions at $\boldsymbol{\theta} = \mathbf{0}$, without loss of generality. We apply (3.2) to $\mathbf{Z}'_n$ and then transform back to obtain, instead of (3.3),

$$f_{\mathbf{Z}_n}(\mathbf{z}) = g_d(\mathbf{z}; \boldsymbol{\Sigma})\left[1 + \frac{1}{6\sqrt{n}}(\mathbf{H}^{\mathrm{T}}\boldsymbol{\rho}')^{[3]}(\mathbf{z}') + \frac{1}{24n}(\mathbf{H}^{\mathrm{T}}\boldsymbol{\rho}')^{[4]}(\mathbf{z}') + \frac{1}{72n}\{(\mathbf{H}^{\mathrm{T}}\boldsymbol{\rho}')^{[3]}\}^2(\mathbf{z}')\right] + O(n^{-\frac{3}{2}}). \tag{4.4}$$

To find $f_{\mathbf{Y}_n|\mathbf{X}_n}(\mathbf{y}|\mathbf{x})$, we divide (4.4) by a corresponding $d_1$-dimensional expansion for $f_{\mathbf{X}_n}(\mathbf{x})$. The ratio of the leading terms gives another multivariate normal density, because of a well-known property of conditional densities in the multivariate normal distribution. In fact, partitioning $\boldsymbol{\Sigma}$ into components in the usual notation, we have that

$$\begin{aligned} f_{\mathbf{Y}_n|\mathbf{X}_n}(\mathbf{y}|\mathbf{x}) = g_{d_2}(\mathbf{y} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\mathbf{x};\, \boldsymbol{\Sigma}_{22.1})\Big(1 &+ \frac{1}{6\sqrt{n}}\{(\mathbf{H}^{\mathrm{T}}\boldsymbol{\rho}')^{[3]}(\mathbf{z}') - (\mathbf{H}^{\mathrm{T}}\boldsymbol{\rho}'_1)^{[3]}(\mathbf{x}')\} \\ &+ \frac{1}{24n}\{(\mathbf{H}^{\mathrm{T}}\boldsymbol{\rho}')^{[4]}(\mathbf{z}') - (\mathbf{H}^{\mathrm{T}}\boldsymbol{\rho}'_1)^{[4]}(\mathbf{x}')\} + \frac{1}{72n}[\{(\mathbf{H}^{\mathrm{T}}\boldsymbol{\rho}')^{[3]}\}^2(\mathbf{z}') \\ &- \{(\mathbf{H}^{\mathrm{T}}\boldsymbol{\rho}'_1)^{[3]}\}^2(\mathbf{x}')] + \frac{1}{36n}(\mathbf{H}^{\mathrm{T}}\boldsymbol{\rho}'_1)^{[3]}(\mathbf{x}')\{(\mathbf{H}^{\mathrm{T}}\boldsymbol{\rho}'_1)^{[3]}(\mathbf{x}') - (\mathbf{H}^{\mathrm{T}}\boldsymbol{\rho}')^{[3]}(\mathbf{z}')\}\Big) \\ &+ O(n^{-\frac{3}{2}}), \end{aligned} \tag{4.5}$$

where $\boldsymbol{\Sigma}_{22.1} = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$, and where the notation $(\mathbf{H}^{\mathrm{T}}\boldsymbol{\rho}'_1)^{[s]}(\mathbf{x}')$ refers to expressions of the type (3.1) calculated from the $d_1$-dimensional variable $\mathbf{x}'$.

As noted in Section 3.2, (4.5) can be rewritten in terms of unorthogonalized variables by replacing $\mathbf{H}^{\mathrm{T}}\boldsymbol{\rho}'$ and $\mathbf{H}'^{\mathrm{T}}\boldsymbol{\rho}'_1$ by $\mathscr{H}^{\mathrm{T}}\boldsymbol{\rho}$ and $\mathscr{H}_1^{\mathrm{T}}\boldsymbol{\rho}$, and replacing $\mathbf{z}'$ and $\mathbf{x}'$ by $\mathbf{z}$ and $\mathbf{x}$, where $\mathscr{H}$ and $\mathscr{H}_1$ specify generalized polynomials determined by the inverse covariance matrices $\boldsymbol{\Delta} = \boldsymbol{\Sigma}^{-1}$ and $\boldsymbol{\Delta}_1 = \boldsymbol{\Sigma}_{11}^{-1}$. Thus to calculate correction terms there are two possibilities. One is to use the original variables and their cumulants together with the generalized polynomials. The second possibility is first to find the orthonormal variables for the particular problem, then to calculate the necessary cumulants and finally to use formulae involving standard Hermite polynomials. For theoretical discussions, we have used the second approach, although for the examples of Section 5 involving two variables and the formulae of Section 3 the first approach has been used.

Again (4.5) specializes first when $\mathbf{x} = \mathbf{0}$ and further when $\mathbf{z} = \mathbf{0}$ so that, in the latter case,

$$f_{\mathbf{Y}_n|\mathbf{X}_n}(\mathbf{0}|\mathbf{0}) = \frac{1}{(2\pi)^{d_2/2} D_{22.1}^{\frac{1}{2}}}\left[1 + \frac{m_{Y.X}(\rho')}{24n} + O(n^{-1})\right], \tag{4.6}$$

where $D_{22.1} = \det(\boldsymbol{\Sigma}_{22.1})$ and $m_{\mathbf{Y}.\mathbf{X}}(\rho')$ is easily written down from (4.5) in a form immediately generalizing (3.8).

### 4.3. *Indirect Edgeworth Expansion*

We can now use the results of Section 4.2 to obtain immediate generalization of the bivariate approximations of Section 3.3. The "maximum likelihood" value is again given by (3.10) and for the joint density, we have that

$$f_{\mathbf{Z}_n}(\mathbf{z};\, \boldsymbol{\theta}_0) = \frac{\exp\{nK(\hat{\boldsymbol{\theta}}) - nK(\boldsymbol{\theta}_0) + \mathbf{t}^T(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)\}}{(2\pi)^{d/2}\{D(\hat{\boldsymbol{\theta}})/(c_1^2 \dots c_d^2)\}^{\frac{1}{2}}}\left\{1 + \frac{m_Z(\hat{\rho}')}{24n} + O(n^{-2})\right\}. \tag{4.7}$$

For the double saddle-point approximation to $f_{\mathbf{Y}_n|\mathbf{X}_n}(\mathbf{y}|\mathbf{x};\, \boldsymbol{\theta}_0)$ we combine (4.7) with an approximation of the same form for $f_{\mathbf{X}_n}(\mathbf{x};\, \boldsymbol{\theta}_0)$, thereby giving

$$f_{\mathbf{Y}_n|\mathbf{X}_n}(\mathbf{y}|\mathbf{x};\, \boldsymbol{\theta}_0) = \frac{\exp\{nK(\hat{\boldsymbol{\theta}}) - nK(\hat{\boldsymbol{\theta}}_{(0)}) + \mathbf{r}^T(\hat{\boldsymbol{\lambda}} - \hat{\boldsymbol{\lambda}}_{(0)}) + \mathbf{s}^T(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}_0)\}}{(2\pi)^{d_2/2}\{D(\hat{\boldsymbol{\theta}})/D_{11}(\hat{\boldsymbol{\theta}}_{(0)})\}^{\frac{1}{2}}(c_{d_1+1} \dots c_d)^{-1}}$$

$$\times\left\{1 + \frac{m_Z(\hat{\rho}') - m_{\mathbf{X}}(\hat{\rho}'_{(0)})}{24n} + O(n^{-2})\right\}, \tag{4.8}$$

where $\hat{\boldsymbol{\theta}}_{(0)} = (\hat{\boldsymbol{\lambda}}_{(0)}, \boldsymbol{\psi}_0)$ is the "maximum likelihood" value for $\boldsymbol{\theta}$ when $\boldsymbol{\psi} = \boldsymbol{\psi}_0$, and $D_{11}(\hat{\boldsymbol{\theta}}_{(0)})$ is the generalized variance of $\mathbf{X}_n$ at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_{(0)}$.

For the single saddle-point or mixed Edgeworth saddle-point approximation we take a direct Edgeworth expansion at $\hat{\boldsymbol{\theta}}_{(0)}$. The standardized conditioning variable is zero at that point and the leading term for $\mathbf{S}_n$ is multivariate normal with mean $nE(\mathbf{V};\, \hat{\boldsymbol{\lambda}}_{(0)}, \boldsymbol{\psi}_0)$ and covariance matrix $n\boldsymbol{\Sigma}_{22.1}(\hat{\boldsymbol{\theta}}_{(0)})$. The correction terms are calculated from (4.5) with $\mathbf{x}' = \mathbf{0}$ and $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_{(0)}$. They are thus directly calculated from the necessary cumulants and special polynomials. Alternatively the transformation to orthonormal variables can be carried out algebraically, the cumulants of the new variables found and the required approximation then evaluated from standard Hermite polynomials; see Section 4.2.

A preassigned number of terms of (4.8) will typically be more accurate than the same number of terms of the single saddle-point approximation but the latter has the advantages, noted in Section 2.4, of not requiring the cumulant generating function in explicit form and of being more easily integrated.

## 5. SOME SPECIAL CASES OF STATISTICAL INTEREST

### 5.1. *Preliminary Remarks*

We now discuss in outline a number of particular problems of statistical inference within the exponential family to which the results of Sections 3 and 4 can be applied. In the first three examples, a test of the adequacy of an exponential family model is obtained by adding one or more extra parameters to the model and testing whether the data are consistent with zero values for these. Unfortunately in none of these examples is the cumulant generating function of the extended model available in useful form so that the double saddle-point approximation for the required conditional distribution is not available.

In discussing the examples we have used a notation appropriate to the special cases and this is in inessential conflict with the notation of Sections 3 and 4.

Pedersen (1979) gives some further examples of the use of the single saddle-point approximation.

### 5.2. *Time-dependent Poisson Process*

For a time-dependent Poisson process whose rate function is $e^{\alpha+\beta t}$ a test of the log-linearity of the rate is obtainable by employing the wider model with rate

$$\exp(\alpha+\beta t+\psi t^2) \tag{5.1}$$

and assessing the hypothesis $\psi = 0$. Suppose that in the fixed interval $(0, T)$ a total of $n$ events has occurred at the time points $0 < t_1 < \ldots < t_n < T$. Then, under the extended model (5.1), the observed data follow an exponential family model with $(n, \Sigma t_i, \Sigma t_i^2)$ as the canonical statistic and $(\alpha, \beta, \psi)$ as the canonical parameter. To perform the conditional test of $\psi = 0$ given $n$ and $\Sigma t_i$ it is simplest first to condition on $n$, thereby reducing the problem to two dimensions and rendering formulae (3.16) and (3.17) applicable with $r = \Sigma t_i$ and $s = \Sigma t_i^2$.

As a numerical example we consider the record of major freezes of Lake Constance discussed by Steinijans (1976) and reproduced in Table 1. Taking the year of the first recorded major freeze, i.e. 875, as the origin of the time axis (thus making a standard data modification) and choosing 1974−875 = 1099 as the time unit we are left with $n = 37$ events and we have

TABLE 1

*Years of major freezes in Lake Constance,* A.D. 875–A.D. 1974

| |
|---|
| 875, 895 |
| 928 |
| 1074, 1076 |
| 1108 |
| 1217, 1227, 1277 |
| 1323, 1325, 1378, 1379, 1383 |
| 1409, 1431, 1435, 1460, 1465, 1470, 1479, 1497 |
| 1512, 1553, 1560, 1564, 1565, 1571, 1573 |
| 1684, 1695 |
| 1763, 1776, 1788, 1796 |
| 1830, 1880 |
| 1963 |

$r/n = 0{\cdot}5394$ and $s/n = 0{\cdot}3443$. The table suggests that the rate function has a peak in the mid-fifteenth century which would correspond to a negative value of $\psi$ in (5.1), and the approximations to the $P$-value for the hypothesis $\psi = 0$ given by the single saddle-point expansion are 0·008594 (normal approximation), 0·008370 (one correction term) and 0·008180 (two correction terms). That is, the first number is the approximate normal test probability from (3.16) whereas the other two numbers have been calculated from (3.17). The test of $\psi = 0$ performed by Steinijans, which showed an enormous significance, must be in error, presumably because of the use of a bivariate normal approximation in an inappropriate range.

### 5.3. *Circular Normal Distribution*

As another instance of the derivation of a test by expansion of the exponential family, consider the testing of consistency with the circular normal distribution written in canonical form with density for the random angle $A$ proportional to

$$\exp(\lambda_1 \cos a + \lambda_2 \sin a) = \exp\{\kappa \cos(a-\phi)\}, \tag{5.2}$$

say. One natural expansion is to augment (5.2) by terms in $\cos 2a$ and $\sin 2a$, i.e. to consider the density proportional to

$$\exp(\lambda_1 \cos a + \lambda_2 \sin a + \psi_1 \cos 2a + \psi_2 \sin 2a). \tag{5.3}$$

Unfortunately the normalizing constant and hence the cumulant generating function cannot be written in useful explicit form.

To examine the hypothesis $\psi_1 = \psi_2 = 0$ on the basis of $n$ independent observations, we consider the conditional distribution of $S = (\Sigma \cos 2A_i, \Sigma \sin 2A_i)$ given $R = (\Sigma \cos A_i, \Sigma \sin A_i) = r$. This requires the discussion of Section 4 with $d_1 = d_2 = 2$. The leading term of the single saddle-point approximation is that $S$ is bivariate normal. This was stated on the basis of a rather ill-specified argument by Cox (1975); Professor K. V. Mardia has pointed out some errors in the detailed formulae.

It is simplest to begin by calculating the covariance matrix of $(\cos A, \cos 2A, \sin A, \sin 2A)$ from (5.2) with $\phi = 0$. On introducing the Bessel functions

$$I_m(\kappa) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \cos mx\, e^{\kappa \cos x}\, dx, \tag{5.4}$$

we have that the sine terms are uncorrelated with the cosine terms and that conditional variances from the least squares regression of $\cos 2A$ on $\cos A$ and of $\sin 2A$ on $\sin A$ are respectively

$$\left.\begin{aligned} v^c_{2.1}(\kappa) &= \frac{I_0^2 + I_0 I_4 - 2I_2^2}{2I_0^2} - \frac{(I_0 I_2 + I_0 I_1 - 2I_1 I_2)^2}{2I_0^2(I_0^2 + I_0 I_2 - 2I_1^2)}, \\ v^s_{2.1}(\kappa) &= \frac{(I_0 - I_4)(I_0 - I_2) - (I_1 - I_3)^2}{2I_0(I_0 - I_2)} \end{aligned}\right\} \tag{5.5}$$

where the Bessel functions have arguments $\kappa$.

Now the maximum likelihood estimates of $(\lambda_1, \lambda_2)$, or equivalently $(\kappa, \phi)$ in (5.2) satisfy

$$n^{-1}\Sigma \sin(a_i - \hat{\phi}) = 0, \quad n^{-1}\Sigma \cos(a_i - \hat{\phi}) = \hat{I}_1/\hat{I}_0,$$

where the Bessel functions now have argument $\hat{\kappa}$. It follows that the leading term in the approximation to the required conditional distribution is such that

$$\Sigma \cos 2(A_i - \hat{\phi}) - n\hat{I}_1/\hat{I}_0 \quad \text{and} \quad \Sigma \sin 2(A_i - \hat{\phi}) \tag{5.6}$$

are independently normally distributed with variances $nv^c_{2.1}(\hat{\kappa})$ and $nv^s_{2.1}(\hat{\kappa})$. A chi-squared statistic with two degrees of freedom can thus be formed.

### 5.4. *Dispersion Test for Gamma Distribution*

As a further example we mention briefly the construction of a dispersion test for conformity with the gamma family, both shape and scale being unknown. If $U$ denotes a typical observation, we consider $\mathbf{W} = (U, \log U, U^2)$. The gamma family is the exponential family generated by $(U, \log U)$ and it follows that a similar test most powerful against one-sided alternatives in the exponential family associated with $\mathbf{W}$ is provided from $n$ independent and identically distributed observations by considering the conditional distribution of $\Sigma U_i^2$ given $(\Sigma U_i, \Sigma \log U_i) = (\Sigma u_i, \Sigma \log u_i)$. Note that the extended exponential family is defined only for non-negative values of the associated parameter. This in particular precludes the use of the double saddle-point approximation.

Thus in the notation of Section 4, we identify the vector $\mathbf{U}$ with $(U, \log U)$ and $\mathbf{V}$ with $U^2$ and apply the single saddle-point approximation. It is convenient to write the gamma family in the form

$$(\beta/\mu)(\beta u/\mu)^{\beta-1} e^{-\beta u/\mu}/\Gamma(\beta),$$

the "maximum likelihood" equations being

$$\hat{\mu} = \Sigma u_i/n, \quad \psi(\hat{\beta}) - \log \hat{\beta} = \log\{(u_1 \ldots u_n)^{1/n}/(\Sigma u_i/n)\},$$

where $\psi(\beta) = d\log\Gamma(\beta)/d\beta$. The mean of $(U, \log U, U^2)$ is $(\mu, \log(\mu/\beta)+\psi(\beta), \mu^2+\mu^2/\beta)$ and the covariance matrix is

$$\begin{bmatrix} \mu^2/\beta & \mu/\beta & 2\mu^2(\beta+1)/\beta^2 \\ \cdot & \psi'(\beta) & \mu\,(2\beta+1)/\beta^2 \\ \cdot & \cdot & 2\mu^4(\beta+1)(2\beta+3)/\beta^3 \end{bmatrix}.$$

Thus the required conditional variance is

$$\sigma^2_{U\,.\,U,\log U}(\mu,\beta) = \frac{\mu^4}{\beta^4}\left\{2(\beta+1)\beta - \frac{1}{\psi'(\beta)-1/\beta}\right\} = \mu^2\sigma^2(\beta),$$

say.

Finally the result given by the single saddle-point approximation is that

$$\frac{\sum U_j^2 - n(\hat\mu^2+\hat\mu^2/\hat\beta)}{\sqrt{(n)}\,\sigma_{U^2.U,\log U}(\hat\mu,\hat\beta)} = \frac{\{n^{-1}\sum(U_j-\hat\mu)^2/\hat\mu^2 - 1/\hat\beta\}}{\sigma(\hat\beta)/\sqrt{n}} \tag{5.7}$$

has conditionally and unconditionally a standard normal distribution. The second form of the statistic shows most clearly its basis of a comparison of observed and predicted squared coefficients of variation.

A brief simulation study by Mr D. Pregibon suggests that the test based on (5.7) is conservative and that the limiting distribution is approached rather slowly. We shall not discuss the test further here and in particular shall not give correction terms. A practical drawback to the test in some contexts is its sensitivity to recording errors in the very small values.

### 5.5. *Binomial–Logistic Bioassay Model*

We now discuss a discrete example and illustrate the use of the double saddle-point approximation (3.15).

In a bioassay with binomial response variates and logistic response probability $1/\{1+\exp(\lambda+\psi d)\}$, where $d$ denotes the dose level, inference on the slope parameter $\psi$ is appropriately performed conditionally on the total number of individuals responding. To take the simplest yet practically relevant case, suppose that only the three dose levels $d = -1$, 0 and 1 are employed and that there are the same number of individuals $n$ at each level. If $a_d$ denotes the number of individuals responding to dose $d$ then the probability of the observation $(a_{-1}, a_0, a_1)$ is

$$\prod_{d=-1}^{1}\binom{n}{a_d}\{1+\exp(-\lambda-\psi d)\}^{-n}e^{-r\lambda-s\psi},$$

where $r = a_{-1}+a_0+a_1$ and $s = a_1-a_{-1}$. Table 2 illustrates the accuracy of the double saddle-point approximation (3.14). It gives the exact values of the conditional probability $p(s;\psi|r)$ of $s$ given $r$ together with the corresponding values of the leading term of (3.15), for $n = 16$, $\psi = 1$, $r = 24$ and 32, and $s \geqslant 0$. Note that the saddle-point approximation is undefined for $(r,s) = (24,16)$ and $(32,16)$ in which cases $(r,s)$ lies on the boundary of the convex support of its distribution, whence the maximum likelihood estimates of $\lambda$ and $\psi$ do not exist. The relative error of the double saddle-point approximation is between 2 and 7 per cent, except for $s = 15$, but could be considerably reduced by renormalizing the approximate conditional distribution to have total mass 1.

TABLE 2

*Logistic–binomial bioassay model. Exact value and double saddle-point approximation of the conditional probability $p(s; \psi \mid r)$, in the case $n=16$, $\psi=1$.*

| | $p(s; 1 \mid r)$ | | | |
|---|---|---|---|---|
| | $r=24$ | | $r=32$ | |
| $s$ | *Exact* | *Saddle-point approximation* | *Exact* | *Saddle-point approximation* |
| 0 | 0·002742 | 0·002797 | 0·004555 | 0·004655 |
| 1 | 0·007025 | 0·007166 | 0·011588 | 0·011844 |
| 2 | 0·015973 | 0·016299 | 0·025804 | 0·026382 |
| 3 | 0·032198 | 0·032870 | 0·050189 | 0·051342 |
| 4 | 0·057411 | 0·058645 | 0·084967 | 0·086992 |
| 5 | 0·090251 | 0·092268 | 0·12454 | 0·12765 |
| 6 | 0·12452 | 0·12745 | 0·15687 | 0·16101 |
| 7 | 0·14990 | 0·15363 | 0·16808 | 0·17282 |
| 8 | 0·15623 | 0·16040 | 0·15106 | 0·15565 |
| 9 | 0·13953 | 0·14358 | 0·11171 | 0·11537 |
| 10 | 0·10537 | 0·10873 | 0·066127 | 0·068461 |
| 11 | 0·066061 | 0·068435 | 0·030086 | 0·031221 |
| 12 | 0·033525 | 0·034915 | 0·009873 | 0·010271 |
| 13 | 0·013263 | 0·013925 | 0·002099 | 0·002203 |
| 14 | 0·003852 | 0·004103 | 0·000243 | 0·000262 |
| 15 | 0·000735 | 0·000814 | 0·000012 | 0·000013 |
| 16 | 0·000070 | — | 0·000000 | — |

## 6. SOME MORE THEORETICAL APPLICATIONS

### 6.1. *Preliminary Remark*

In the previous section we discussed in detail some very particular applications of the general results of Sections 2–4. Now we apply the results to two general problems of statistical theory, one the calculation of conditional likelihood functions and conditional maximum likelihood estimates, and the other the calculation of improved approximations to the null hypothesis distribution of maximum likelihood ratio test statistics.

### 6.2. *Approximations to Conditional Likelihood Functions*

The Edgeworth and saddle-point approximations yield approximations to conditional likelihood functions in the obvious way. Here we illustrate this for the leading term of the double saddle-point approximation (3.15), assuming that the basic statistical model is the bivariate exponential family (3.9) and that an approximation to the conditional likelihood function of $\psi$ given $R_n = r$ is sought. We suppose for simplicity that $\psi$ and $\lambda$ are one-dimensional although the results can be generalized by using (4.8) rather than (3.15).

The required approximation from the leading term of (3.15), obtained by taking the contributions that depend upon $\psi$, is for the conditional log-likelihood

$$l(\psi; s \mid r) \simeq \tfrac{1}{2} \log \kappa_{20(\psi)} - nK(\hat{\lambda}_\psi, \psi) - r\hat{\lambda}_\psi - s\psi, \tag{6.1}$$

where $\hat{\lambda}_\psi$ denotes the maximum likelihood estimate of $\lambda$ under the hypothesis that the second parameter has the value $\psi$ and where $\kappa_{20(\psi)}$ is $\kappa_{20}$ evaluated at $(\hat{\lambda}_\psi, \psi)$.

From (6.1) an approximate conditional likelihood equation for the determination of $\hat{\psi}_r$, the conditional maximum likelihood estimate of $\psi$ given $r$, may be obtained by differentiation. We employ the relations

$$n\kappa_{10} = r, \quad \kappa_{20}(\partial\hat{\lambda}_\psi/\partial\psi) + \kappa_{11} = 0,$$

where here and below the $\kappa$'s are to be evaluated at $(\hat{\lambda}_\psi, \psi)$. Then the approximate conditional likelihood equation may be written as

$$n\kappa_{01} = s - \tfrac{1}{2}(\kappa_{11}\kappa_{30} - \kappa_{20}\kappa_{21})/\kappa_{20}^2, \tag{6.2}$$

which has the form of the unconditional likelihood equation $n\kappa_{01} = s$ with a correction term. Note that the correction term is $O_p(1)$, whereas the other terms are $O_p(n)$, so that conditional and unconditional maximum likelihood estimates differ by $O_p(1/n)$; this difference is typically small compared with the formal large-sample standard error, which is $O_p(1/\sqrt{n})$.

To illustrate formula (6.1) we consider the $2\times 2$ table

| | | |
|---|---|---|
| $x_{11}$ | $x_{12}$ | $x_{1.}$ |
| $x_{21}$ | $x_{22}$ | $x_{2.}$ |
| $x_{.1}$ | $x_{.2}$ | $n$ |

with cell probabilities $\pi_{ij}$, and we take

$$\mathbf{t} = (x_{1.}, x_{.1}), \quad s = x_{11}$$

and

$$\lambda = \left(\log\frac{\pi_{22}}{\pi_{12}}, \log\frac{\pi_{22}}{\pi_{21}}\right), \quad \psi = \log\frac{\pi_{12}\pi_{21}}{\pi_{11}\pi_{22}}.$$

Note that the $x_{ij}$ here are not values of the standardized variable $X$ of Section 2.

The approximate conditional likelihood is then

$$\tfrac{1}{2}\log\{\pi_{11}\pi_{22}(\pi_{12}+\pi_{21}) + \pi_{12}\pi_{21}(\pi_{11}+\pi_{22})\} - x_{1.}\log\frac{\pi_{22}}{\pi_{12}} - x_{.1}\log\frac{\pi_{22}}{\pi_{21}} - x_{11}\psi + n\log\pi_{22}, \tag{6.3}$$

where $\pi_{ij}$ is to be evaluated at $(\hat{\lambda}_\psi, \psi)$. The required value of $\pi_{ij}$ is obtained by solving the set of equations

$$n\pi_{1.} = x_{1.}, \quad n\pi_{.1} = x_{.1}, \quad e^\psi = \frac{\pi_{12}\pi_{21}}{\pi_{11}\pi_{22}}.$$

The third equation may be written as

$$e^\psi = \frac{(\pi_{1.} - \pi_{11})(\pi_{.1} - \pi_{11})}{\pi_{11}(1 - \pi_{1.} - \pi_{.1} + \pi_{11})}$$

and hence computation of the conditional likelihood approximation essentially requires nothing more than the solution of a second degree equation.

For a numerical illustration we take the data on twins of criminals discussed by Fisher (1935, 1962) and reproduced in Table 3.

TABLE 3

*Convictions of like-sex twins of criminals*

| | *Convicted* | *Not convicted* | *Total* |
|---|---|---|---|
| *Dizygotic* | 2 | 15 | 17 |
| *Monozygotic* | 10 | 3 | 13 |
| Total | 12 | 18 | 30 |

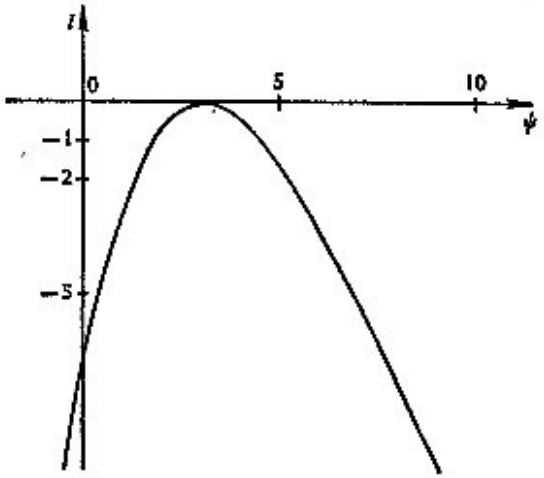


FIG. 1. The conditional log-likelihood function for Fisher's twins and criminality data. The double saddle-point approximation to this function coincides with the curve to within the drawing accuracy.

Fig. 1 shows the (exact) conditional log-likelihood function for $\psi$, the logarithm of the cross-product ratio, given the marginals of the table. The approximation to this function given by (6.3) coincides with the curve in the figure to within the drawing accuracy. The conditional maximum likelihood estimate is 3·06 compared with the unconditional maximum estimate of $\log(150/6) = 3{\cdot}22$.

### 6.3. *Null Distribution of Maximum Likelihood Ratio Test Statistic*

The saddle-point approximations of Sections 2.2, 3.3 and 4.3 lead to a direct proof of a property of the null hypothesis distribution of the maximum likelihood ratio test statistic, Wilks's statistic.

Consider first a one-parameter problem with observations $U_1, \ldots, U_n$ independently and identically distributed with the exponential family density (2.3), with $\lambda$ unknown. Let the null hypothesis be $\lambda = \lambda_0$. Then $P_n$, the maximum likelihood ratio test statistic, has value $p$ given in the notation of Section 2.2 by

$$p = -2r(\hat{\lambda} - \lambda_0) - 2n\{K(\hat{\lambda}) - K(\lambda_0)\}, \tag{6.4}$$

where $R_n = U_1 + \ldots + U_n$ and $\hat{\lambda}$ satisfies

$$r + nK'(\hat{\lambda}) = 0. \tag{6.5}$$

The saddle-point approximation for $f_{R_n}(r; \lambda_0)$ is, by (2.11) and (2.9),

$$f_{R_n}(r; \lambda_0) = \frac{e^{-p/2}}{\{2\pi n K''(\hat{\lambda})\}^{\frac{1}{2}}} \left\{1 + \frac{c}{n} + O(n^{-2})\right\}, \tag{6.6}$$

for a constant $c$, depending on $\lambda_0$. Now by (6.4) and (6.5) $dp/dr = -2(\hat{\lambda} - \lambda_0)$, so that, on transforming from $R_n$ to $P_n$,

$$f_{P_n}(p; \lambda_0) = \frac{e^{-p/2}}{2(2\pi n)^{\frac{1}{2}}} \Sigma \frac{1}{\{(\hat{\lambda} - \lambda_0)^2 K''(\hat{\lambda})\}^{\frac{1}{2}}} \left\{1 + \frac{c}{n} + O(n^{-2})\right\},$$

where the sum is over those $\hat{\lambda}$ leading to a given $p$; at least locally there are two such values.

Elimination of $r$ between (6.4) and (6.5), and Taylor expansions of the resulting equation and of $K''(\hat{\lambda})$, lead to

$$\{(\hat{\lambda}-\lambda)^2 K''(\hat{\lambda})\}^{-\frac{1}{2}} = \left(\frac{n}{p}\right)^{\frac{1}{2}}\left\{1 \pm c_1\frac{p^{\frac{1}{2}}}{n^{\frac{1}{2}}} + (c_2 \pm c_3)\frac{p}{n} + O(n^{-\frac{3}{2}})\right\},$$

where $c_1, c_2, c_3$ depend on $K''(\lambda_0), K^{(3)}(\lambda_0), K^{(4)}(\lambda_0)$. Therefore,

$$f_{P_n}(p;\, \lambda_0) = \frac{e^{-p/2}}{(2\pi p)^{\frac{1}{2}}}\left\{1 + \frac{c + c_2 p}{n} + O(n^{-2})\right\}. \tag{6.7}$$

For (6.7) to integrate to one, we have $c_2 = -c$. It is easily shown that (6.7) implies that $P_n/(1-c/n)$ has a chi-squared distribution with one degree of freedom, to order $1/n^{\frac{1}{2}}$; for this to happen it is necessary and sufficient that the correction factor in (6.7) is linear in $p$.

Now consider the two-parameter problem using the notation and results of Section 3.3. With null hypothesis $(\lambda, \psi) = (\lambda_0, \psi_0)$, and observed vector $\mathbf{W} = (W_1, \ldots, W_n)^{\mathrm{T}}$, we have that the observed value of the test statistic is

$$\begin{aligned} p &= \log \frac{f_{\mathbf{W}}(\mathbf{w};\, \hat{\lambda}, \hat{\psi})}{f_{\mathbf{W}}(\mathbf{w};\, \lambda_0, \psi_0)} \\ &= \log \frac{f_{\mathbf{W}}(\mathbf{w};\, \lambda_0, \hat{\psi}_0)}{f_{\mathbf{W}}(\mathbf{w};\, \lambda_0, \psi_0)} + \log \frac{f_{\mathbf{W}}(\mathbf{w};\, \hat{\lambda}, \hat{\psi})}{f_{\mathbf{W}}(\mathbf{w};\, \lambda_0, \hat{\psi}_0)} \\ &= p^{(2)} + p^{(1.2)}, \end{aligned}$$

say. The previous one-parameter discussion applies directly to the random variable $P_n^{(2)}$ and, slightly less directly, using the conditional density of $R_n = U_1 + \ldots + U_n$ given $S_n = V_1 + \ldots + V_n$, to $P_n^{(1.2)}$. To order $1/n^{\frac{1}{2}}$, $P_n^{(2)}/(1-c^{(2)}/n)$ and $P_n^{(1.2)}/(1-c^{(1.2)}/n)$ are independently distributed as chi-squared with one degree of freedom, because the second factor has the stated distribution for a given value of the first factor. It follows that to the same order

$$P_n = (P_n^{(2)} + P_n^{(1.2)})/(1-c/n),$$

with $c = \frac{1}{2}(c^{(2)} + c^{(1.2)})$, is distributed as chi-squared with two degrees of freedom, all under the null hypothesis; note that to sufficient accuracy both $c^{(2)}$ and $c^{(1.2)}$ are functions only of $\theta_0$.

The general $d$-parameter result follows in the same way, applying either to the simple null hypothesis $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ or to the null hypothesis $\boldsymbol{\psi} = \boldsymbol{\psi}_0$ with $\boldsymbol{\lambda}$ unknown. In the latter case the test statistic is preferably calculated from the appropriate conditional likelihood.

The idea of "improving" the asymptotic chi-squared distribution of the statistic $P_n$ is due to Bartlett (1937), who divided $P_n$ by a factor designed to produce the same expectation as chi-squared. Bartlett (1954) gave the appropriate factor for a number of standard multivariate tests. Lawley (1956), by an extremely lengthy calculation of cumulants, appeared to prove the remarkable result that the adjustment to order $1/n$ of the expected value produces a corresponding adjustment for all order cumulants of $P_n$. However, the recent results of Hayakawa (1976, 1977) show that in general the correction is of a more complicated form, although Lawley's simpler form is applicable to simple hypotheses and to hypotheses about canonical parameters in exponential family problems. Hayakawa proceeds by direct asymptotic expansion of the derivatives of the likelihood function and the arguments are complicated. The results of the present paper amount to a concise derivation of Theorem 1, Remark 2 of Hayakawa (1977).

Unfortunately the arguments here apply only to hypotheses about canonical parameters of exponential family distributions. We have not investigated the possibility of generalizing the results by local exponential family approximations.

ACKNOWLEDGEMENT

It is a pleasure to acknowledge the helpful discussions and the assistance with the numerical calculations which have been afforded us by M. Weis Bentzon, M. Frydenberg and B. V. Pedersen.

## REFERENCES

BARNDORFF-NIELSEN, O. and PEDERSEN, B. V. (1979). The bivariate hermite polynomials up to order six. *Scand. J. Statist.*, to appear.

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

—— (1954). A note on the multiplying factors for various $\chi^2$ approximations. *J. R. Statist. Soc.* B, **16**, 296–298.

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

CHAMBERS, J. M. (1967). On methods of asymptotic approximation for multivariate distributions. *Biometrika*, **54**, 367–384.

COX, D. R. (1975). Contribution to the discussion of Mardia, K. V.: Statistics of directional data. *J. R. Statist. Soc.* B, **37**, 380–381.

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

—— (1958). Contribution to discussion of Cox, D. R.: The regression analysis of binary sequences. *J. R. Statist. Soc.* B, **20**, 236–238.

FELLER, W. (1966). *An Introduction to Probability Theory and its Applications*, Vol. II. New York: Wiley.

FISHER, R. A. (1935). The logic of inductive inference. *J. R. Statist. Soc.* A, **98**, 39–54.

—— (1962). Confidence limits for a cross-product ratio. *Aust. J. Statist.*, **4**, 41.

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

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

HAYAKAWA, T. (1976). Asymptotic expansion of the distribution of the likelihood ratio criterion for homogeneity of parameters In *Essays in Probability and Statistics* (S. Ikeda *et al.*, eds), pp. 265–285. Tokyo: Shinko Tsusho.

—— (1977). The likelihood ratio criterion and the asymptotic expansion of its distribution. *Ann. Inst. Statist. Math.*, **29**, 359–378.

IBRAGIMOV, I. A. and LINNIK, Y. V. (1971). *Independent and Stationary Sequences of Random Variables.* Groningen: Walters-Noordhoff.

KHINCHIN, A. I. (1949). *Mathematical Foundations of Statistical Mechanics.* New York: Dover.

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

MARDIA, K. V. (1970). *Families of Bivariate Distributions.* London: Griffin.

PAGUROVA, V. I. (1965). On the evaluation of quantiles of the $\Gamma$-distribution. *Theory Prob. Appl.*, **10**, 677–680.

PEDERSEN, B. V. (1979). Approximating conditional distributions by the mixed Edgeworth-saddlepoint expansion. (To appear.)

RICHTER, W. (1957). Local limit theorems for large deviations. *Theory Prob. Appl.*, **2**, 206–219.

STEINIJANS, V. W. (1976). A stochastic point-process model for the occurrence of major freezes in Lake Constance. *Appl. Statist.*, **25**, 58–61.

TWEEDIE, M. C. K. (1957). Statistical properties of inverse Gaussian distributions. I. *Ann. Math. Statist.*, **28**, 362–377.

## APPENDIX

### *Regularity Conditions*

We give here a set of regularity conditions which ensure the validity of the Edgeworth and saddle-point expansions discussed in the main part of the paper. The precise versions of the theorems given below are taken, essentially, from unpublished lecture notes by P. Martin-Löf, but a detailed discussion of the core of the results can be found in the book of Bhattacharya and Rao (1976); see also Good (1957, 1961).

As in Section 4, let $\mathbf{W}$ denote a $d$-dimensional random variable and let $\mathbf{T}_n$ be the sum of $n$ independent variates $\mathbf{W}_1, \ldots, \mathbf{W}_n$ each with the same distribution as $\mathbf{W}$. We assume an exponential family model for $\mathbf{W}$

$$f_{\mathbf{W}}(\mathbf{w};\, \boldsymbol{\theta}) = \exp\{-\boldsymbol{\theta}^{\mathrm{T}}\mathbf{w} - \alpha(\mathbf{w}) - \beta(\boldsymbol{\theta})\}, \tag{A.1}$$

where now $f_{\mathbf{W}}(\mathbf{w};\,\boldsymbol{\theta})$ denotes the density of the distribution of $\mathbf{W}$ with respect to some dominating measure $\mu$ which will typically but not necessarily be either Lebesgue measure on $R^d$ or counting measure on $Z^d$, where $Z$ denotes the set of integers. The exponential representation (A.1) is assumed to be minimal; in other words the order of the exponential family is $d$, and the domain of variation $\Theta$ of the parameter $\boldsymbol{\theta}$ is supposed to be the full canonical parameter domain. Let $S$ denote the support of the distribution of $\mathbf{W}$, and let $\operatorname{int}\Theta$ denote the interior of $\Theta$. The random vector $\mathbf{T}_n$ follows an exponential model of the form

$$f_{\mathbf{T}_n}(\mathbf{t};\,\boldsymbol{\theta}) = \exp\{-\boldsymbol{\theta}^{\mathrm{T}}\mathbf{t} - \alpha_n(\mathbf{t}) - n\beta(\boldsymbol{\theta})\} \tag{A.2}$$

for some function $\alpha_n(\mathbf{t})$.

Henceforth we assume that one of the following two regularity conditions is satisfied:

[$d$] (discrete case) the support $S$ is contained in $Z^d$ but not in any sublattice of $Z^d$;

[$c$] (continuous case) for each $\boldsymbol{\theta}\in\operatorname{int}\Theta$ there exists a positive integer $n_0$ such that $\mathbf{T}_n$ possesses a bounded density with respect to Lebesgue measure for every $n\geqslant n_0$.

These conditions are, separately, equivalent to the two more technical conditions given below, which are more useful for the proofs of the validity of the expansions. We denote by $\zeta_{\boldsymbol{\theta}}(\boldsymbol{\tau})$ the characteristic function of $\mathbf{W}$. The new conditions are

[$d$]′ $|\zeta_{\boldsymbol{\theta}}(\boldsymbol{\tau})| \neq 1$ for $\boldsymbol{\tau}\in(-\pi,\pi]^d$;

[$c$]′ for each $\boldsymbol{\theta}\in\operatorname{int}\Theta$ there exists a positive number $\nu$ such that $|\zeta_{\boldsymbol{\theta}}(\boldsymbol{\tau})|^{\nu}$ is integrable with respect to $\boldsymbol{\tau}$.

An example where (A.1) is neither of discrete nor continuous type but where condition [$c$] is satisfied, for $n_0 = 2$, is provided by the normal distribution. Let $U$ be normally distributed and set $\mathbf{W} = (U, U^2)$. Here the support of the distribution of $\mathbf{W}$ is a parabola in $R^2$ while, of course, $\mathbf{Z}_n$ has a bounded density for $n\geqslant 2$.

In accordance with conditions [$d$] and [$c$], we take $f_{\mathbf{T}_n}(\mathbf{t};\,\boldsymbol{\theta})$ in (A.2) to be the point probability function in the discrete case and the density function with respect to Lebesgue measure in the continuous case, whenever the density exists.

For each $\boldsymbol{\theta}\in\operatorname{int}\Theta$ the variate $\mathbf{W}$ with distribution (A.1) has moments of all orders. Denoting the variance matrix of $\mathbf{W}$ by $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta})$ and the $l$th cumulant of $\mathbf{W}$ by $\kappa_l = \kappa_l(\boldsymbol{\theta})$, $l = (l_1, \ldots, l_d)$, and setting

$$\mathbf{z} = n^{-\frac{1}{2}}\{\mathbf{t} - nE(\mathbf{W};\,\boldsymbol{\theta})\}$$

we have the following result.

*Theorem* (direct Edgeworth expansion). Suppose that either of the conditions [$d$] or [$c$] is fulfilled. Then for any $r = 0, 1, 2, \ldots$

$$f_{\mathbf{T}_n}(\mathbf{t};\,\boldsymbol{\theta}) = g_d(\mathbf{z};\,\boldsymbol{\Sigma})\left\{1 + \sum_{j=1}^{r} Q_j(\mathbf{z};\,\boldsymbol{\theta})\,n^{-j/2}\right\} + O(n^{-(r+1)/2}) \tag{A.3}$$

uniformly in $\mathbf{t}$, and in $\boldsymbol{\theta}$ on every compact subset of $\operatorname{int}\Theta$. Here $g_d(\mathbf{z};\,\boldsymbol{\Sigma})$ is the density of the $d$-dimensional normal distribution with mean $\mathbf{0}$ and variance matrix $\boldsymbol{\Sigma}$, and $Q_j(\mathbf{z};\,\boldsymbol{\theta})$ is a polynomial in $\mathbf{z} = (z_1, \ldots, z_d)$ whose coefficients depend on $\boldsymbol{\theta}$ through the cumulants $\kappa_l(\boldsymbol{\theta})$. In particular, $Q_j(\mathbf{0};\,\boldsymbol{\theta}) = 0$ for $j$ odd.

The precise definition of the polynomials $Q_j$ will emerge in the following sketch of the proof of the theorem.

Let $\boldsymbol{\mu} = E(\mathbf{W};\,\boldsymbol{\theta})$ and let

$$\tilde{\mathbf{T}} = (n\boldsymbol{\Sigma})^{-\frac{1}{2}}(\mathbf{T}_n - n\boldsymbol{\mu}).$$

The characteristic function of $\hat{\mathbf{T}}$ is

$$E\{\exp(i\tau^{\mathrm{T}}\hat{\mathbf{T}});\ \theta\} = \exp(-in^{\frac{1}{2}}\mu^{\mathrm{T}}\Sigma^{-\frac{1}{2}}\tau)\,\zeta_\theta(n^{-\frac{1}{2}}\Sigma^{-\frac{1}{2}}\tau)$$

which may be rewritten as

$$E\{\exp(i\tau^{\mathrm{T}}\hat{\mathbf{T}});\ \theta\} = \exp\left\{-\tfrac{1}{2}\tau^{\mathrm{T}}\tau + \sum_{|l|=3}^{\infty} i^{|l|}\frac{(\Sigma^{-\frac{1}{2}}\tau)^l}{l!}\kappa_l n^{-\frac{1}{2}|l|+1}\right\}$$

$$= \exp(-\tfrac{1}{2}\tau^{\mathrm{T}}\tau)\left\{1+\sum_{j=1}^{\infty}P_j(\tau;\theta)\,n^{-j/2}\right\},$$

where $P_j(\tau;\theta)$, a polynomial in $\tau$ whose coefficients depend on $\theta$ through the cumulants $\kappa_l(\theta)$, is determined by coefficient identification for the power series in $n^{-\frac{1}{2}}$. The approximation given by (A.3) is obtained by summing only up to $r$ in the last expression above and then making a Fourier inversion. Thus $Q_j$ is determined by

$$Q_j(\mathbf{z};\theta)\,\phi(\Sigma^{-\frac{1}{2}}\mathbf{z}) = (2\pi)^{-d/2}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}\exp(i\tau^{\mathrm{T}}\Sigma^{-\frac{1}{2}}\mathbf{z})\,P_j(\tau;\theta)\,\phi(\tau)\,d\tau,$$

where $\phi$ denotes the probability density function of the $k$-dimensional standardized normal distribution. More explicit expression for $Q_1$ and $Q_2$ are available from Section 4.2. In view of this definition of the polynomials $Q_j$ it is obvious that the difference between the left-hand side and the main term on the right-hand side of (A.3) is expressible as a Fourier integral and the main step in showing that the difference is $O(n^{-(r+1)/2})$, uniformly, consists in splitting the domain of integration in two, namely according to whether $|\tau| < c\sqrt{n}$ or not, where $c$ is a suitably chosen constant. The technique is an elaboration of Feller's (1966) method for verifying the Edgeworth expansion.

Now, setting $\theta$ equal to $\hat{\theta}$ in the expansion (A.3) we obtain, using $E(\mathbf{T}_n;\hat{\theta}) = \mathbf{t}$ and $Q_j(0;\theta) = 0$, for $j$ odd:

*Corollary* (saddle-point expansion). Suppose either of the conditions [*d*] or [*c*] is fulfilled. Then, for any $m = 0, 1, 2, \ldots$

$$f_{\mathbf{T}_n}(\mathbf{t};\theta) = \frac{\exp\{n\beta(\hat{\theta})-n\beta(\theta)+\mathbf{t}^{\mathrm{T}}(\hat{\theta}-\theta_0)\}}{(2\pi n)^{d/2}\,D(\hat{\theta})^{\frac{1}{2}}}\left\{1+\sum_{\nu=1}^{m}Q_\nu(0;\hat{\theta})\,n^{-\nu}+O(n^{-(m+1)})\right\} \qquad \text{(A.4)}$$

uniformly in $\mathbf{t}$, provided $\hat{\theta}(\mathbf{t})$ belongs to a given, but arbitrary, compact subset of $\operatorname{int}\Theta$.

In formula (A.4), $D(\hat{\theta})$ denotes the determinant of $\Sigma(\hat{\theta})$.

It appears from Section 7 of Daniels (1954) that in the continuous case the saddle-point approximation will in many cases hold uniformly over the entire domain of values of $\mathbf{t}$, and not just under a compactness restriction as in the Corollary, and that the error will, in fact, often tend to zero towards the boundary of that domain.

Finally, precise conditions for the validity of the direct Edgeworth and double and single saddle-point approximations to conditional probability functions, as defined in Sections 3.2, 3.3, 4.2 and 4.4, are easily established from the Theorem and Corollary above.

## Discussion of the Paper by Professors Barndorff-Nielsen and Cox

Professor H. E. Daniels (Statistical Laboratory, Cambridge): This excellent paper is welcome as another example of the growth of interest in so-called "small sample asymptotics". A typical feature of these approximations to the probability density of estimators is that not only do they work well for remarkably small sample sizes, but often the relative error is bounded over the whole range of the parameter. I am sure that today's paper will generate many new developments and applications.

Let me start with a quibble. The authors work with the Laplace transform, which is natural in the context of the exponential family. However, I wish they had not called it the moment generating

function. There is only a trivial difference and I know the usage is current in some quarters, but it makes comparison with other literature unnecessarily awkward. We have been through all this before with the transfer function.

That said, let me turn to other matters. The theoretical development of the paper is based on the Edgeworth expansion applied to the "conjugate" family of densities. (The use of the word by Khinchin predates its use by the Bayesians for quite a different purpose.) My own preference is to apply the complex variable method of steepest descents to the inverse transform to get the saddle-point approximation. There are advantages in both approaches, though in the end one's choice is a matter of personal taste.

It is interesting to look at the historical background. In his 1949 book, *Statistical Mechanics*, Khinchin developed essentially the approach used in the present paper as a substitute for the method of steepest descent used by Darwin and Fowler. He pays them a generous tribute for having given the first rigorous derivation of the necessary asymptotic computations. But, he says, they develop "a special and very abstract analytical apparatus" instead of using "the known limit theorem of the theory of probability" (the central limit theorem applied to conjugate distributions). His own presentation is aimed at "many of those readers who are frightened by the complicated formalistics of the Darwin–Fowler method". This reads rather oddly to a traditional British applied mathematician like myself who tends to be frightened by the complicated formalistics of probability theorists. On the other hand, in the 1950s Linnik, perhaps because of his background of analytic number theory, and his student Richter seized on the method of steepest descent as the most natural way of developing local limit theorems of just the kind being discussed, so the distinction cannot be a national one.

While I agree that the conjugate density approach gives more probabilistic insight, I find the saddle-point method a more practical tool, in the sense that it can be applied without too much detailed thought to a wider range of problems. After all, that is why we use mathematics—it does our work for us. For example, suppose one wants the p.d.f. $g(r_n)$ of the length $r_n = |\Sigma_1^n \mathbf{x}_r| = n\bar{r}$ of the sum of $n$ i.i.d. vectors. In two dimensions the analogue for $r = |\mathbf{x}|$ of the Fourier transform and its inverse is the Hankel transform

$$\gamma(\rho) = \int_0^\infty g(r)\, J_0(r\rho)\, dr, \quad g(r) = r\int_0^\infty \gamma(\rho)\, J_0(r\rho)\, \rho\, d\rho.$$

The required density is the inverse transform of $\gamma^n(\rho)$. There is a saddle-point of the integrand within $O(n^{-1})$ of $-i\hat{R}$ where $\hat{R}$ is the real root of $G'(\hat{R})/G(\hat{R}) = \bar{r}$ and $G(i\rho) = \gamma(\rho)$. The path of integration can then be suitably deformed to pass through it. I cannot see how the conjugate density approach can be readily applied here.

One of the remarkable features of the renormalized saddle-point approximation is that in some cases it yields the exact formula for the density—an extreme case of getting more than we deserve out of this sort of approximation. There is a statement in Section 2.4 of the paper that the normal and the inverse normal are the only exact examples, without renormalization, in the univariate case. No proof is given and I wonder if the authors have one. I have managed to produce a proof that the only univariate densities where the approximation, renormalized or not, is exact are the three known ones—the gamma, the normal and the inverse normal. The multivariate situation is more difficult; it is known that there are exact cases but I have not so far managed to characterize them.

I am sure that the single-saddle point approximation will turn out to be a valuable new device. Although less elegant and less accurate than the double saddle-point approximation it can be applied to a wider variety of problems. But I wonder whether in the example of Section 5.3 the authors may have given up too easily. The density (5.2) can be got from the special bivariate normal density $(2\pi)^{-1}\exp -\frac{1}{2}\{(x-\lambda_1)^2+(y-\lambda_2)^2\}$ by conditioning on $r = (x^2+y^2)^{\frac{1}{2}}$. The augmented density (5.3) is similarly obtained from a general bivariate normal density with a suitable reparameterization. It might therefore be possible to use the double saddle-point approximation on the bivariate normal density with an extra conditioning on $r$.†

There are many insights in this stimulating paper which others will no doubt comment on. I propose that we accord Professors Barndorff-Nielsen and Cox a hearty vote of thanks.

† Dr Kent's comment makes me less optimistic about it.

Professor J. DURBIN (London School of Economics): I congratulate the authors on an important and stimulating paper. They have given an interesting new interpretation of the saddle point approximation, they have developed new methods for obtaining approximations to conditional distributions and they have applied their results to a range of problems of considerable interest.

In my remarks I will indicate how the authors' approach can be extended to a wider class of sufficient estimators including time-series applications. The paper's starting point is Professor Daniels' saddle point approximation. Daniels' idea was to approximate the density of a statistic $x$ by inverting the moment generating function, but instead of integrating along the imaginary axis as in conventional Fourier inversion he chose the contour of integration to pass through the saddle point. This gives a series expansion in powers of $n^{-1}$ instead of $n^{-\frac{1}{2}}$ as in the Edgeworth series. My first point is that one can obtain a comparable gain relative to Edgeworth by taking the contour through the observed point $x$, thus eliminating the need to find the saddle-point.

Take the typical case in which $x$ is standardized so that $E(x) = 0$, $V(x) = 1$ and the remaining cumulants satisfy $\kappa_r = O(n^{-\frac{1}{2}r+1})$, $r = 3, 4, \ldots$. If $M(z)$ is the moment generating function and $K(z) = \log M(z)$ is the cumulant generating function then the saddle-point $x_0$ is the solution of the equation

$$x = K'(x_0) = x_0 + \frac{1}{2}\kappa_3 x_0^2 + \frac{1}{3!}\kappa_4 x_0^3 + \ldots.$$

Thus $x - x_0 = O(n^{-\frac{1}{2}})$ and the difference between the observed value $x$ and the saddle point is small relative to the difference between $x_0$ and zero. Taking a contour through $x$ and integrating, we obtain the approximate density of $x$ in the form

$$g(x) = e^{-x^2} M(x) \operatorname{Edg}[\kappa_r(x)], \tag{1}$$

where $\operatorname{Edg}[\kappa_r(x)]$ is the Edgeworth series with $\kappa_r$ replaced by the $r$th derivative $\kappa_r(x)$ of $K(z)$ evaluated at $z = x$ instead of at $z = 0$, i.e.

$$\kappa_r(x) = \kappa_r + \kappa_{r+1} x + \tfrac{1}{2}\kappa_{r+2} x^2 + \ldots$$

for $r = 2, 3, \ldots$, and with mean $x - \kappa_1(x) = \frac{1}{2}\kappa_3 x + \frac{1}{3}\kappa_4 x^2 + \ldots$. Like the saddle point approximation, the first term of (1) has an error of order $n^{-1}$. If cumulants of orders up to $r = s$ are available one could replace $e^{-x^2} M(x)$ in (1) by

$$\exp\left(-\frac{1}{2}x^2 + \frac{1}{3!}\kappa_3 x^3 + \ldots + \frac{1}{s!}\kappa x^s\right),$$

truncate the series $\operatorname{Edg}[\kappa_r(x)]$ at $r = s$ and integrate numerically but I do not know how this would perform relatively to the usual Edgeworth approximation.

To extend the Barndorff-Nielsen and Cox approach, suppose that $\mathbf{y}$ is an observation matrix with density $f(\mathbf{y}, \boldsymbol{\theta})$ where $\boldsymbol{\theta}$ is an $m$-dimensional parameter and that $\mathbf{t}$ is a sufficient estimator of $\boldsymbol{\theta}$. We therefore have

$$f(\mathbf{y}, \boldsymbol{\theta}) = g(\mathbf{t}, \boldsymbol{\theta})\, h(\mathbf{y}) \tag{2}$$

by the factorization theorem where $g(\mathbf{t}, \boldsymbol{\theta})$ is the unknown density of $\mathbf{t}$. Suppose that we wish to approximate this at parameter point $\boldsymbol{\theta}_0$. From (2) we have

$$f(\mathbf{y}, \boldsymbol{\theta}_0) = g(\mathbf{t}, \boldsymbol{\theta}_0)\, h(\mathbf{y}) \tag{3}$$

so on dividing (3) by (2) we obtain

$$g(\mathbf{t}, \boldsymbol{\theta}_0) = \frac{f(\mathbf{y}, \boldsymbol{\theta}_0)}{f(\mathbf{y}, \boldsymbol{\theta})}\, g(\mathbf{t}, \boldsymbol{\theta}), \tag{4}$$

which is a generalization of Barndorff-Nielsen and Cox's formula (2.6).

Suppose that $E(\mathbf{t}) = \boldsymbol{\alpha}(\boldsymbol{\theta})$ and that $\hat{\boldsymbol{\theta}}$ is the solution of the equation $\mathbf{t} = \boldsymbol{\alpha}(\hat{\boldsymbol{\theta}})$, where $\mathbf{t}$ here denotes the observed sample point. Putting $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ in (4) and replacing $g(\mathbf{t}, \hat{\boldsymbol{\theta}})$ by an Edgeworth series $\hat{g}(\mathbf{t}, \hat{\boldsymbol{\theta}})$ we obtain

$$g(\mathbf{t}, \boldsymbol{\theta}_0) = \frac{f(\mathbf{y}, \boldsymbol{\theta}_0)}{f(\mathbf{y}, \hat{\boldsymbol{\theta}})}\, \hat{g}(\mathbf{t}, \hat{\boldsymbol{\theta}}), \tag{5}$$

which is a generalization of Barndorff and Nielsen's (2.9). However, for the reasons I gave earlier in the discussion it is often easier and just as effective to replace $\theta$ in (4) by $\mathbf{t}$ rather than $\hat{\theta}$ (of course they will often be the same) giving the slightly different form

$$g(\mathbf{t}, \theta_0) = \frac{f(\mathbf{y}, \theta_0)}{f(\mathbf{y}, \mathbf{t})} \hat{g}(\mathbf{t}, \mathbf{t}), \tag{6}$$

where $\hat{g}(\mathbf{t}, \mathbf{t})$ is the Edgeworth series for $g(\mathbf{t}, \theta)$. Effectively, this gives a series in powers of $n^{-1}$.

The first term of (6) gives

$$g(\mathbf{t}, \theta_0) = \left[\frac{n}{2\pi}\right]^{\frac{1}{2}m} |\mathbf{D}(\mathbf{t})|^{-\frac{1}{2}} \frac{f(\mathbf{y}, \theta_0)}{f(\mathbf{y}, \mathbf{t})} [1 + O(n^{-1})], \tag{7}$$

where $\mathbf{D}(\theta)$ is the limit of $n$ times the variance matrix of $\mathbf{t}$. This approximation has the advantage that unlike the saddle point approximation one does not need an explicit knowledge of the moment generating function of $\mathbf{t}$ to construct it. Like the saddle point approximation one can usually reduce the error to order $n^{-2}$ by renormalization. Of course, I appreciate that there is a close relation between sufficiency and the exponential family, but my own preference is to deal with the problem in terms of sufficiency, even in the independent case. Of course, when the observations are dependent, the relation between sufficiency and exponentiality breaks down.

*Example* 1

Suppose that $y_1, \ldots, y_n$ are independent $N(\mu, \sigma^2)$. Take $\theta = [\mu, \sigma^2]'$ and $\mathbf{t} = [\bar{y}, s^2]'$ where $\bar{y}$, $s^2$ are the sample mean and variance. Substituting in (7) gives for the joint density of $\bar{y}$ and $s^2$,

$$\sqrt{\left(\frac{n}{2\pi\sigma^2}\right)} \exp\left\{-\frac{n}{2\sigma^2}(\bar{y}-\mu)^2\right\} \frac{\sqrt{n} \exp\{-\frac{1}{2}(n-1)\}}{2\sqrt{\pi\sigma^2}} \left[\frac{s^2}{\sigma^2}\right]^{\frac{1}{2}(n-3)} \exp\left\{-\frac{(n-1)s^2}{2\sigma^2}\right\} [1 + O(n^{-1})].$$

The first-term approximation is therefore exact for this case apart from the substitution of Stirling's approximation for $\Gamma[\frac{1}{2}(n-1)]$.

*Example* 2

Here $y_1, \ldots, y_n$ are generated by the circular autoregression

$$y_t = \rho y_{t-1} + \varepsilon_t \quad t = 1, \ldots, n, \quad y_0 \equiv y_n,$$

where $\varepsilon_1, \ldots, \varepsilon_n$ are independent $N(0, \sigma^2)$ and suppose that we want an approximation for the density of the lag-1 coefficient $r = \Sigma y_t y_{t-1} / \Sigma y_t^2$. On applying (7) to the joint density of $\Sigma y_t y_{t-1}$ and $\Sigma y_t^2$ and integrating out $\Sigma y_t^2$ we obtain after renormalization for the density of $r$,

$$g(r, \rho) = \left\{2^n B\left(\frac{n+1}{2}, \frac{n+1}{2}\right)\right\}^{-1} \frac{(1-r^2)^{\frac{1}{2}n-1}}{(1+\rho^2-2\rho r)^{\frac{1}{2}n}} [1 + O(n^{-1})],$$

which is Leipnik's (1947) approximation. In fact the first term is exact apart from a term which is exponentially small. The same approximation was obtained by Daniels (1956) by the saddle point method. Further details are given in Durbin (1980).

It is also rather interesting that this technique may be used to derive an approximate density for the estimate of the coefficient of the circular first-order moving average. The rather surprising result is that exactly the same approximation is obtained as for the autoregression; in other words, to order $n^{-2}$ the estimate of the first-order moving average coefficient is distributed with the same density as the corresponding autoregressive coefficient.

I have not tried to extend this to conditional distributions in the way that is done in the paper, but I imagine there might be some possibilities there.

One final comment is about the theorem that has been quoted by the authors in the Appendix. As far as I am aware, this is the first example of a validation of the Edgeworth expansion for a parametric family of densities. All the classical theorems are for a specific density. For the applications in this paper, however, because the statistic is being substituted for the parameter, expansions are needed which are valid for families of densities. I want to ask the authors, first, is Martin-Löf's theorem specific to the exponential families of the kind that they consider in the paper? Secondly, is a proof of the theorem going to be published? I think it is important that there should be a proof of this theorem available in the literature.

It is clear from my comments that I have found this an extremely stimulating paper. I have great pleasure in seconding the vote of thanks.

The vote of thanks was carried by acclamation.

Dr H. W. Peers (Leeds University): The indirect Edgeworth expansions discussed by the authors give us a valuable way of side-stepping the $O(n^{-\frac{1}{2}})$ terms occurring in theoretical investigations involving higher order corrective terms. For instance, we are often interested in the distribution of quantities like

$$t_\theta = (\hat{\theta} - \theta)\sqrt{-\partial^2 L/\partial\theta^2}, \tag{i}$$

where $\hat{\theta}$ is the m.l.e. of $\theta$, $L$ is the log-likelihood function and $-\partial^2 L/\partial\theta^2$ is the observed Fisher information. $t_\theta$ is an asymptotically pivotal $N(O, 1)$ quantity but the question arises as to how far the pivotal nature persists if a more refined approximation to its distribution is made. Such approximations then involve terms which are $O(n^{-\frac{1}{2}})$ and higher.

An alternative way of gaining remission from $O(n^{-\frac{1}{2}})$ terms is by initial parameterization. Taking $\phi = \phi(\theta)$ to be a monotone function of $\theta$ we have that

$$t_\phi = (\hat{\phi} - \phi)\sqrt{-\partial^2 L/\partial\hat{\phi}^2} \tag{ii}$$

is also asymptotically $N(O, 1)$ and we can sometimes remove the $O(n^{-\frac{1}{2}})$ terms by appropriate choice of $\phi$. Whichever procedure we adopt we still have the $O(n^{-1})$ terms to contend with, and it would be nice to have a way of dealing with these in an algebraically more efficient manner.

Returning to (i) it is easy to see that to $O(1)$, $t_\theta$ can be written as a standardized sum $X_n$ (based on $\partial L/\partial\theta$) to which the indirect Edgeworth expansion of Section 2 can be applied. Going to higher order terms, however, involves increasingly complex non-linear functions of derivatives of $L$ for which there is no obvious cumulant generating function $K$. We would then, presumably, be compelled to use a truncated series approximation to $K$. Would this mean that an indirect approach would then necessarily be vitiated by expansion of quantities to say $O(n^{-1})$ (thereby reproducing essentially the direct expansion) or do the authors have a way of circumventing this obstacle in non-exponential family problems?

On a point of minor detail, it is of interest to note that the correction term at (6.2) involves a particular cumulant function which also occurs as an invariant in the study of curved exponential families.

Professor D. V. Lindley (Somerset): The inferential methods developed in this paper for the exponential family (4.1) depend on the fact that the distribution of $T_n^{(s)}$, given $T_n^{(1)}, \ldots, T_n^{(s-1)}$, involves only the single parameter $\theta_s$ and can therefore be used to make inferences about $\theta_s$ irrespective of the values of the nuisance parameters $\theta_1, \ldots, \theta_{s-1}$. A major application of the saddle-point method is in obtaining these distributions. The coherent approach is different because the likelihood does not factor and the discarded distribution of $T_n^{(1)}, \ldots, T_n^{(s-1)}$ typically involves $\theta_s$. It is thus more complicated; but, on the other hand, it is simpler because no distribution theory is involved. Nevertheless, saddle-point methods have their place in the coherent approach. They are not restricted to the exponential family and the following exposition is fairly general.

Let $L(\theta)$ be the logarithm of the likelihood for $\theta$ given a random sample of size $n$, and so $O(n)$; let $\rho(\theta)$ be the logarithm of the prior density for $\theta$. Then we are interested in the ratio of integrals of the form

$$\int u(\theta) \exp\{L(\theta) + \rho(\theta)\}\, d\theta \Big/ \int \exp\{L(\theta) + \rho(\theta)\}\, d\theta$$

being the posterior expectation of $u(\theta)$. Since $L$ is $O(n)$, the integrals may be expanded in power series in $n^{-1}$, the individual terms of which are complicated. However, on calculating the ratio, many of the terms vanish and the terms of order 1 and $n^{-1}$ are rather straightforward. I content myself with quoting the results for $u(\theta) = \theta_a$ where we have the posterior mean to compare with the authors' result. Asymptotically

$$E(\theta_a) - \hat{\theta}_a = \sum_s \rho_s \sigma_{as} + \tfrac{1}{2} \sum_{s,t,k} L_{stk}\, \sigma_{as}\, \sigma_{tk} + O(n^{-2}),$$

where a subscript $i$, say, denotes differentiation with respect to $\theta_i$, $\sigma_{ij}$ are the elements of the matrix inverse to that with elements $-L_{ij}$, and all expressions are evaluated at the maximum likelihood value $\hat{\theta}$. Two points of interest here are that the fourth derivatives of $L$ do not appear, and that $\rho$ is absent from the second correction term. There has not been time to compare results obtained this way with the authors', but such a comparison might shed light on why the two approaches, whilst apparently so different, often yield closely similar answers.

Professor K. V. MARDIA (University of Leeds): First of all, let me join the others in congratulating the authors on a stimulating paper. Indeed we have already been inspired by the working of Section 5.3 relating to the test of circular normality. Various Bessel functions in (5.5) make it complicated to use. In fact, by approximating the von Mises distribution by the wrapped normal distribution we have the approximation

$$I_p(\hat{\kappa})/I_0(\hat{\kappa}) \doteq \bar{R}^p,$$

where $\bar{R}$ is the mean resultant length. If the approximation is used in $v^2_{2,1}(\hat{\kappa})$ and $v^2_{3,1}(\hat{\kappa})$, the approximation should work well for small $\kappa$ and large $\kappa$. However, this simplification does not extend to the von Mises–Fisher case. Indeed this difficulty exists in general. Consider the family

$$\exp\{a(\boldsymbol{\theta}) + b_0(\mathbf{x}) + \mathbf{b}_1^{\mathrm{T}}(\mathbf{x})\,\boldsymbol{\theta}_1 + \mathbf{b}_2^{\mathrm{T}}(\mathbf{x})\,\boldsymbol{\theta}_2\},$$

where

$$\boldsymbol{\theta}^{\mathrm{T}} = (\boldsymbol{\theta}_1^{\mathrm{T}}, \boldsymbol{\theta}_2^{\mathrm{T}}), \quad \mathbf{b}^{\mathrm{T}}(\mathbf{x}) = \{\mathbf{b}_1^{\mathrm{T}}(\mathbf{x}), \mathbf{b}_2^{\mathrm{T}}(\mathbf{x})\}, \quad \boldsymbol{\theta}_1 : p \times 1, \quad \boldsymbol{\theta}_2 : q \times 1.$$

Let $H$ be $\boldsymbol{\theta}_2 = \mathbf{0}$. The test-statistic is

$$G_1 = (\bar{\mathbf{b}}_2(\mathbf{x}) - \hat{\boldsymbol{\mu}})^{\mathrm{T}}\, \hat{\boldsymbol{\Sigma}}_{2.1}^{-1} (\bar{\mathbf{b}}_2(\mathbf{x}) - \hat{\boldsymbol{\mu}}),$$

where $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}_{2.1}$ are obtained under $H$. This is asymptotically $\chi^2_q$. If $\hat{\boldsymbol{\Sigma}}_{2.1}$ is replaced by its sample counterpart $\mathbf{S}_{2.1}$ we obtain an asymptotically equivalent statistic

$$G_2 = (\bar{\mathbf{b}}_2(\mathbf{x}) - \hat{\boldsymbol{\mu}})^{\mathrm{T}}\, \mathbf{S}_{2.1}^{-1} (\bar{\mathbf{b}}_2(\mathbf{x}) - \hat{\boldsymbol{\mu}})$$

which has the advantage of being much easier to calculate since $\hat{\boldsymbol{\Sigma}}_{2.1}$ is usually more complicated than $\mathbf{S}_{2.1}$. Besides, the derivation of $\hat{\boldsymbol{\Sigma}}_{2.1}$ is usually tedious. In particular, my expression for $G_1$ for testing the von Mises–Fisher distribution is found to be much more complicated than $G_2$ (which is not surprising in view of (5.5)). The same remark applies to $G_1$ and $G_2$ in testing dependence for the following exponential model

$$\exp\{a(\boldsymbol{\theta}) + b_0(\mathbf{x}, \mathbf{y}) + \mathbf{b}_1^{\mathrm{T}}(\mathbf{x})\,\boldsymbol{\theta}_1 + \mathbf{b}_2^{\mathrm{T}}(\mathbf{y})\,\boldsymbol{\theta}_2 + \mathbf{b}_1^{\mathrm{T}}(\mathbf{x})\,\boldsymbol{\theta}_3\,\mathbf{b}_2(\mathbf{y})\},$$

where we test $\boldsymbol{\theta}_3 = \mathbf{0}$ against $\boldsymbol{\theta}_3 \neq \mathbf{0}$. Various simulation studies of these models show that there is no significant difference in size between $G_1$ and $G_2$, presumably because the discussion of Section 2.4 applies. However, is there any reason to believe that there would be a difference in power by using $G_2$ rather than $G_1$? I should add that the approach of Section 5.3 is far-reaching. Indeed, it has already provided a measure of correlation robust against "scale" on the circle/sphere/Stiefel manifold, etc. It has also provided various results for a distributional model in catastrophe theory.

The lognormal distribution is again coming into the limelight because of its predominance in geostatistics. This is a member of the exponential family but one cannot obtain the m.g.f. of the distribution in a closed form, even when the density is given by

$$f(x) = \{\sqrt{(2\pi)}\,\sigma x\}^{-1} \exp\{-(2\sigma^2)^{-1} \log^2 x\}.$$

Can the device in (2.4) be extended to cover this case or here are we stuck with the direct Edgeworth expansion? For the above distribution, perhaps the sufficient statistic $\Sigma (\log x_i)^3$ can be used in tonight's argument of Professor Durbin.

Professor R. SIBSON (University of Bath): The authors point out (a) that the low-order approximations discussed in their paper may be of little use in the tails of the distribution, and (b) that the indirect Edgeworth expansion is not conveniently integrable to yield tail probabilities. It may therefore be of interest for me to add a footnote to their paper by reporting on the successful use of high-order direct Edgeworth expansions for the calculation of tail probabilities. I have carried out expansions of this kind as far as the term in $(1/\sqrt{N})^{80}$; this involves the use of moments or cumulants up to order 32, and of Hermite polynomials up to order 90. There is no question of writing down explicit formulae for any but the first few terms of the Edgeworth expansion—the algebraic complexity is too great, and in any case the rounding errors arising from numerical substitution in such explicit expansions would quickly swamp the calculations. The appropriate technique is to use the computer as a device for symbolic manipulation on truncated power series, with numerical substitutions being carried out at whatever stage is most effective in retaining

accuracy. This approach, combined with the use of standard numerical analysis techniques for the accurate evaluation of polynomials when such a substitution is carried out, and the use of variable-precision arithmetic as a check on the balance between rounding and truncation errors, makes it possible to obtain useful results even in extreme cases. As an example, I consider the case of the tenfold convolution of the uniform distribution. The exact distribution function can be calculated for comparison in this case; I am indebted to Professor J. A. Campbell for doing this calculation for me exactly using his symbolic computation package. The example has some independent interest because of the occasional (although wholly unnecessary) use of such a convolution to simulate approximately a normal random variable. To three significant figures, the following values are obtained.

| | *s.d. from mean* | | |
|---|---|---|---|
| | 1·96 | 4·23 | 4·76 |
| Normal | 2·50 E-2 | 1·16 E-5 | 9·67 E-7 |
| 30-term Edgeworth | 2·45 E-2 | 1·00 E-6 | 4·19 E-9 |
| Exact | 2·45 E-2 | 1·00 E-6 | 3·99 E-9 |

At 1.96, the truncation error is lost in the rounding error. At 4.23, it is about one part in $10^5$. At 4.76 it is, as is visible in the above table, about one part in 20. Considering that the normal approximation which is being "improved" is inaccurate by factors of about 10 and 250 respectively in these latter two cases, this performance is quite encouraging, although it must be pointed out that the symmetry of the uniform distribution is a great help. However, other experiments confirm that even in the non-symmetrical case, high-order Edgeworth expansions of this kind are a useful general-purpose tool for the calculation of tail probabilities, being especially reliable in cases where the normal approximation is already of the right order of magnitude. The context in which my interest in this problem arose was that of controlling a nuclear reactor; I was asked to find, numerically, $10^{-6}$ tail points for $N$-fold convolutions of truncated exponential distributions, with $N$ in the range 20 to 1000, and the use of 30-term Edgeworth expansions achieved this entirely satisfactorily.

Dr A. C. Atkinson (Imperial College, London): The problem of finding good approximations to the distribution of test statistics is important and interesting in both statistical theory and practice. There can be no doubt about the theoretical interest of the results in tonight's paper. However, in the presentation and discussion the point was made that the results are intended to provide practical tools. I wonder whether the methods may not turn out to require too high a degree of mathematical expertise to become widely used.

One way of approximating distributions which requires less expertise is to simulate the system and to build up the empirical distributions of the relevant statistics. The significance of the observed value can then be estimated by ranking in the results of, for example, 999 simulations. Unfortunately, simulation may become more complicated when conditional distributions are required. An example is the distribution of $M$ estimates where Professor Daniels' still, alas, unpublished results obtained by saddle-point approximations agree closely with results from a careful simulation as described in the Princeton Robustness Study (Andrews *et al.*, 1972).

My questions to the authors are concerned with exposition. If it is easier to simulate the unconditional distribution, how much is lost by so doing? For example, in the analysis of Section 5.2 what is the quantitative effect of not conditioning on $\Sigma t_i$? Can the authors give any guidance on the simulation of conditional distributions? Can simulation be usefully combined in any way with their approximations, perhaps as an alternative to numerical integration for renormalization? And, underlying all this, how accurate do we need to be anyway?

Dr J. T. Kent (University of Leeds): I would like to direct my remarks to Section 5.3 where the exponent of the von Mises density is augmented by the addition of second-order trigonometric terms. This distribution is more interesting in higher dimensions especially on $S_2$, the unit sphere in $R^3$. Here the full eight-parameter density takes the form

$$f(\mathbf{x}) \propto \exp\left\{\alpha \boldsymbol{\mu}^{\mathrm{T}}\mathbf{x} + \sum_{i=1}^{6}\beta_i(\boldsymbol{\gamma}_i^{\mathrm{T}}\mathbf{x})^2\right\}, \quad \mathbf{x} \in S_2, \tag{*}$$

where $\gamma_1, \gamma_2, \gamma_3$ are orthonormal vectors, $\mu$ is a unit vector, $\alpha \geqslant 0$, and we can suppose $\beta_1 \geqslant \beta_2 \geqslant \beta_3 = 0$. This density is the product of a Fisher-type factor and a Bingham-type factor.

The distribution takes a slightly simpler form if

$$\beta_1 = \beta_2 \quad \text{and} \quad \mu = \gamma_1. \tag{**}$$

Then the Bingham factor represents a girdle distribution and the pole of the Fisher factor lies along the equator of the girdle. This density has ellipse-shaped probability contours about the pole. If the three-parameter Fisher distribution is considered to be the spherical analogue of the isotropic bivariate normal, then this five-parameter distribution is analogous to the general bivariate normal.

The distribution (*) has been introduced by several authors but deeper study has been hampered by the lack of tractability of the normalization constant. Of course, one prefers to use the Fisher distribution if it is adequate, and the approach of Section 5.3 can be used to provide straightforward tests for the adequacy of the Fisher distribution against either the full alternative (*) or the alternative restricted by (**).

Mr A. J. Mayne (University College London): I would like to present a brief summary of two sets of saddle-point formulae for distribution functions, as opposed to density functions or probability masses, that I obtained several years ago. Both sets of formulae are derived for the case of an integer-valued random variable, using Good's (1957) arguments as a starting point and adopting his notation, but the results for a continuous random variable follow as a limiting case.

Let $X_t$ be the sum of $t$ independent integer-valued random variables, each with p.g.f. $f(z)$, and let $c(N, t)$ be the probability that $X_t = N$. Good (1957, p. 868) showed that, if $c(N, t) \neq 0$, the equation

$$t\rho \frac{d}{d\rho} f(\rho) = Nf(\rho)$$

has a unique non-negative real root and (Good, 1957, p. 869) that

$$c(N, t) = \frac{1}{2\pi\rho^N} \int_{-\pi}^{\pi} [f\{\rho \exp(i\theta)\}]^t \exp(-Ni\theta)\, d\theta.$$

He derived his saddle-point approximation for $c(N, t)$ from this formula. In my own work, I considered formulae for

$$a(N, t) = \Pr(X_t \leqslant N) = \sum_{n=-\infty}^{N} c(N, t)$$

and

$$b(N, t) = \Pr(X_t \geqslant N) = \sum_{n=N}^{\infty} c(N, t).$$

My first set of saddle-point formulae was derived from the equations

$$a(N, t) = \frac{1}{2\pi\rho^N} \int_{-\pi}^{\pi} \frac{[f\{\rho \exp(i\theta)\}]^t}{1 - \exp(-i\theta)}\, d\theta$$

and

$$b(N, t) = \frac{1}{2\pi\rho^N} \int_{-\pi}^{\pi} \frac{[f\{\rho \exp(i\theta)\}]^t}{1 - \exp(i\theta)}\, d\theta,$$

which led to the same type of expansion as Good obtained for $c(N, t)$, by expanding the integrand as

$$\exp(-\tfrac{1}{2} t\kappa_2 \theta^2) \times (\text{power series in } \theta),$$

where $\kappa_2 = (\partial/\partial u)^2 (\log(f\rho e^u))\big|_{u=0}$.

My second set of formulae (Mayne, 1959) applied Good's saddle-point approximation for $c(N, t)$ directly, with $f(z)$ replaced by some suitable function $g(z)$, and $t$ set equal to one. For calculating $a(N, t)$, $g(z)$ was defined as $\{f(z)\}^t/(1-z)$ and, for calculating $b(N, t)$, $g(z)$ was defined as $\{f(z)\}^t/(z-1)$.

Both sets of approximations are accurate in the tails, but the first diverges near the mean. The second set converges at the mean, but is less satisfactory in its neighbourhood than in the tails. Numerical tests for the second type of approximation for the Poisson distribution, using computer programs and taking the first three terms, show that it is accurate to within about 1 per cent near the mean and performs much better in the tails. The second set of saddle-point approximations for the distribution thus seems likely to have more practical importance for the general case.

For the first set of approximations, the results for the binomial distribution are identical with those of Brockwell (1964), derived by a different method. For the limiting case of the Poisson distribution, they have, as far as I know, not previously been published explicitly, and are as follows. If $P_m$ denotes a Poisson random variable with parameter $m$, then, for $x<m$,

$$\frac{\Pr(P_m < x)}{\Pr(P_m = x)} \sim \frac{m}{m-x} + \sum_{k=1}^{\infty} (-1)^k x m f_k(m) (m-x)^{-2k-1}$$

and, for $x>m$,

$$\frac{\Pr(P_m > x)}{\Pr(P_m = x)} \sim \frac{x}{x-m} + \sum_{k=1}^{\infty} (-1)^k x m f_k(m) (x-m)^{-2k-1},$$

where

$$f_1(m) = 1, \quad f_2(m) = x+2m, \quad f_3(m) = x^2+8xm+6m^2,$$

and, in general,

$$f_k(m) = \{x+(2k-1)m\} f_{k-1}(m) + m(x-m)\frac{d}{dm} f_{k-1}(m).$$

These approximations are very good in the tails, but diverge for $x = m$ and are poor for $x$ near $m$.

Numerically, these results seem to be very largely complementary to the Cornish–Fisher expansion, so that one technique can usually be applied when the other gives unsatisfactory results. I also found, more recently, that it is possible, in at least some cases, to derive *uniform approximations* that are accurate for the whole range of the random variable. I have obtained such approximations for the Poisson distribution, and numerically tested them with successful results, and also obtained some results for the binomial distribution, but these have not yet been published. Temme (1975) obtained some formulae that were in some respects similar.

Professor R. N. Bhattacharya (University of Arizona): Professors Barndorff-Nielsen and Cox have made use of important ideas due to Cramér and Khinchin to give a novel and elegant derivation of an asymptotic expansion of the density of the likelihood ratio statistic for the exponential family. Recently such an expansion has been obtained by T. Chanda and J. K. Ghosh (to appear in *Sankhyā*) using a more direct method due to Ghosh and myself (1978). The last derivation goes beyond the exponential case.

Dr P. J. Bickel (University of California at Berkeley): It may be worth noting that the indirect Edgworth expansion is formally applicable to any statistic $X_n$ with moment-generating function $M_n$ such that, $M_n(t) \cong \exp(-\frac{1}{2}t^2)$ by writing $f_{X_n}(x, \lambda) = \exp(-u\lambda) f_{X_n}(x)/M_n(\lambda)$ whose moment-generating function is $M_n(t+\lambda)/M_n(\lambda)$ selecting $\hat{\lambda}$ so that $E(x_n; \hat{\lambda}) = x$, etc. Thus this approach could be applied to various statistics arising in non-parametric contexts, such as those considered in Albers *et al.* (1976), Bickel and Van Zwet (1978). A related approach via a different type of exponential family to approximating the power of one-sample rank tests may be found in Chow and Hodges (1975) and Chow (1976).

Professor J. K. Ghosh (Indian Statistical Institute): What interests me most is the result in Section 6 on the distribution of the maximum likelihood ratio statistic. Recently in related work Tapas Chandra and myself have proved the validity of the Hayakawa's formal expansion for the distribution function of $P_n$. (The results for the null distribution will appear in *Sankhyā* A, 1979; the results under local alternatives are being written up.) Even in the special case of Section 6, where stronger results are available, we found it essential to work with the Edgeworth expansion for the (normalized) MLE, *vide* Theorem 2 or Theorem 3 of Bhattacharya and Ghosh (1978), in

place of the (direct) Edgeworth expansion for $(R_n, S_n)$. I wonder if there is any simple relation between the indirect Edgeworth expansion for $(R_n, S_n)$ and the Edgeworth expnasion for the MLE.

The Appendix on validity is adequate if one seeks to expand densities. But surely (A.3) is not the right result to use if one also wishes to integrate the expansion over an unbounded region. In the context of example 5.2 this is necessary if the alternative is one-sided. I anticipate similar difficulties in respect of Section 6.3; here a complete proof will involve integrating out certain variables over a region which does not seem to be bounded.

Professor F. HAMPEL (ETH Zürich, Switzerland): The authors are to be congratulated for their interesting paper on how to apply some asymptotic approximations for the arithmetic mean to multivariate and conditional distributions. I should only like to add a few notes and references on the comparison, accuracy and interpretation of various asymptotic expansions.

The relationship between Edgeworth expansions, large deviations and the saddle-point method has already been discussed in Hampel (1973), and in more detail in Field and Hampel (1978), using a unifying new variant of the saddle-point method. By the way, this variant yields naturally the renormalized saddle-point approximation and also adds a new interpretation to the formalism of saddle-point methods, which furthermore suggests new methods of proof for the central limit problem, reducing it essentially to one of smoothing. While the interpretation of the saddle-point approximation as a "recentred" Edgeworth approximation is also contained in the present paper and already in the cited basic paper by Daniels (1954), it may be pointed out that the non-applicability of the saddle-point method for long-tailed distributions without moment generating functions, which also causes a rather "narrow" solution for the central limit problem, is connected with the non-robustness of the arithmetic mean and ceases to be a problem if the method is extended to robust $M$-estimators (cf. Hampel, 1973; Field and Hampel, 1978). Given either short tails of the distribution or robustness of the estimator (and a certain smoothness of both), it appears empirically to be both necessary and sufficient for a very good approximation down to very small $n$ to use the first two terms of the saddle-point expansion (formula (2.11)). By contrast, the full infinite sequence of the usual "large deviation" expansion (as in the cited works by Richter or Feller) only recovers the first term without recentring and thus fails badly in two ways (cf. the example at the end of Field and Hampel, 1978). There is nothing wrong with the formal asymptotics; the problem lies in the way infinity is approached. If no recentring is to be (or can be) used, then the Edgeworth approximation is quite good in some central region, though the "Edgeworth expansion put back into the exponent" (Field and Hampel, 1978) appears to be still slightly better; both have (different) problems in the tails.

Professor D. V. HINKLEY (University of Minnesota): The authors are to be congratulated on rekindling the excitement of Daniels' pioneering paper. I wonder if they have seriously considered generalization to curved exponential family models, where problems of ancillarity and goodness of fit can be discussed with relative ease. Suppose that the model is equation (3.9), with special case $\theta^{\mathrm{T}} = \{\lambda(\xi), \psi(\xi)\}$ indexed by the single parameter $\xi$. If we write $\mu(\xi) = E(W; \xi)$, then (3.10) is replaced by

$$(\partial\beta/\partial\xi)^{\mathrm{T}}_{\hat{\xi}}\{t - \mu(\hat{\xi})\} = 0.$$

In testing the goodness of fit of the reduced model, we are then concerned with the conditional distribution of $(R_n, S_n)$ given a locally determined linear combination of $R_n$ and $S_n$. As is apparent from Efron and Hinkley (1978), the likelihood goodness of fit measure is their asymptotic ancillary $Q$ derived from the observed information $I$; see also Peers (1978). It would be nice to see the authors derive asymptotic expansions for such problems, along the lines of their Section 6.3.

As far as I can tell, the results in Section 5 are obtainable by standard score-statistic methods, although the authors do isolate single degrees of freedom. Recently I have been studying a bivariate generalization of the circular normal model (5.2) with a single correlation parameter,

$$f(a; \theta) = c^{-1} \exp\{\kappa \cos(a_1 - \phi_1) + \kappa \cos(a_2 - \phi_2) + \psi \cos(a_1 - a_2 - \phi_1 + \phi_2)\}$$

where $c = I_0^2(\kappa)\, I_0(\psi) + 2 \sum_1^\infty I_j^2(\kappa)\, I_j(\psi)$. This appears to be as difficult as the authors' model (5.3). In some applications I find it more appropriate to consider a "fixed-effects" model for pairs of angles, in which case part of the analysis will involve conditioning on single-degree-of-freedom

estimates of the fixed effects. Have the authors gone beyond Section 6.2 and considered the problem of infinitely-many nuisance parameters?

The results and ideas in this paper will surely prove useful, and will certainly help to simplify derivations of asymptotic results.

Mr B. JØRGENSEN and Mr B. V. PEDERSEN (Aarhus University): In the present note we discuss the approximate conditional likelihood introduced in Section 6.2. One notes that (6.1) is simply the partially maximized log-likelihood plus a correction term $\frac{1}{2}\ln \kappa_{\hat{\theta}\hat{\theta}(\psi)}$. If the full domain of variation for the parameter $\theta = (\lambda, \psi)$ is an open proper subset of $\mathscr{R}^k$ then $\partial^2 K/\partial\lambda^2$ will tend to infinity as $\psi$ tends to a finite boundary value, and one might suspect that the approximation will not in general be as good as in the example considered by the authors, where the parameters vary freely. We illustrate this by two examples.

First we consider inference about the shape parameter of the gamma distribution. Taking the density in the form

$$\exp\{-\lambda x+(\psi-1)\ln x-\ln\Gamma(\psi)+\psi\ln\lambda\},$$

we find the exact conditional log-likelihood for $\psi$ given $r = \Sigma x_i$ to be

$$l(\psi; s \mid r) = (\psi-1)s-(n\psi-1)\ln r-n\ln\Gamma(\psi)+\ln\Gamma(n\psi), \tag{1}$$

where $s = \Sigma \ln x_i$.

To obtain the approximation (6.1) to the conditional log-likelihood we need the estimate of $\lambda$ for fixed $\psi$ and the variance of $x$ evaluated at $(\lambda, \psi) = (\hat{\lambda}_\psi, \psi)$. These are

$$\hat{\lambda}_\psi = n\psi/r \quad \text{and} \quad \kappa_{\hat{\theta}\hat{\theta}(\psi)} = r^2/n^2\psi.$$

Inserting in (6.1) and reducing, the approximation is found to be

$$l(\psi; s \mid r) \simeq (\psi-1)s-(n\psi-1)\ln r-n\ln\Gamma(\psi)+\{-n\psi+(n\psi-\tfrac{1}{2})\ln n\psi-\tfrac{1}{2}\ln n\}. \tag{2}$$

Except for an additive constant this differs from (1) only by the use of Stirling's formula to approximate $\ln\Gamma(n\psi)$. For $\psi$ small the difference between (1) and (2) is thus of the order $-\frac{1}{2}\ln\psi$, whereas the difference between (1) and the partially maximized log-likelihood is of the order $-\ln\psi$. We conclude that for $\psi$ small the double saddle-point approximation (2) gives a better approximation to the true conditional likelihood than the partially maximized likelihood function, though in both cases the error tends to infinity at the boundary.

As a second example we consider a non-regular exponential family, namely the generalized inverse Gaussian distribution (see, for instance, Barndorff-Nielsen, 1977). The probability density function is

$$\frac{(\psi/\chi)^{\frac{1}{2}\lambda}}{2K_\lambda(\sqrt{(\chi\psi)})}\, x^{\lambda-1}\exp\{-\tfrac{1}{2}(\chi x^{-1}+\psi x)\},$$

where $x>0$ and $K_\lambda$ is the modified Bessel function of the third kind with index $\lambda$.

We consider inference about $\psi$ for $\lambda>1$ fixed, in which case the domain of variation for the parameters is $(\chi, \psi) \in [0, \infty)\times(0, \infty)$, and the resulting family is exponential of order two.

To calculate the approximation to the conditional likelihood for $\psi$ given $r = \Sigma x_i^{-1}$ we need to estimate $\chi$ for fixed $\psi$. The likelihood equation becomes

$$n\tau_\psi(\chi)-r = 0,$$

where $\tau_\psi(\chi)$ denotes the mean of $x^{-1}$, i.e.

$$\tau_\psi(\chi) = \frac{K_{\lambda-1}\{\sqrt{(\chi\psi)}\}}{K_\lambda\{\sqrt{(\chi\psi)}\}}\sqrt{\frac{\psi}{\chi}}.$$

Using the relations

$$\lim_{\chi\to\infty}\tau_\psi(\chi) = 0 \quad \text{and} \quad \lim_{\chi\to 0}\tau_\psi(\chi) = \frac{\psi}{2(\lambda-1)}$$

(obtained from asymptotic relations for the Bessel functions) and the fact that $\tau_\psi(\cdot)$ is monotone it follows that the likelihood equation has a solution $\hat{\chi}_\psi>0$ if

$$\frac{r}{n}<\frac{\psi}{2(\lambda-1)}.$$

In the opposite case the likelihood is a decreasing function of $\chi$ and hence $\hat{\chi}_\psi = 0$. Thus, for $\psi$ in the interval

$$0 < \psi \leqslant 2(\lambda - 1)\, r/n \tag{3}$$

we have that $\kappa_{20(\psi)}$ is the variance of the reciprocal of a gamma variate, which is infinite for $\lambda \leqslant 2$. Since the partially maximized likelihood is always finite, it follows that in the case $1 < \lambda \leqslant 2$ the double saddle-point approximation (6.1) is infinite in the interval (3).

Professor D. A. Sprott (University of Waterloo): This is an impressive and complex paper that will take some time to examine and absorb. However, I have one query that comes immediately to mind. This concerns Section 6.3. I would like to see an example of the calculation of $c$ in (6.6) in a particular case such as $\hat{\theta} = 5$, in a sample of size $n = 4$ from an exponential distribution $(1/\theta)\exp(-x/\theta)$. It would then be interesting to compare the application of $P_u/(1 - c/n)$ in such a case with other methods of improvement cited by the authors. Modification to produce the same expectation appears to require division by 1·041 in the above example. If so, this gives significance levels of 0·017, 0·007 to $\theta = 24{\cdot}301$, 1·9910, for which the exact levels are 0·01. It would be of interest also to compare these with a different sort of improvement, applicable to the case of 1 d.f., in which an additive constant is applied to $\sqrt{P_u}$, namely

$$\sqrt{P_u} \pm \{(\tfrac{1}{4})\,(F_3(\hat{\theta})) + \tfrac{1}{2}I^{-\frac{1}{2}}(\hat{\theta})\, I'(\hat{\theta})\} \sim N(0, 1)$$

described by Sprott (1973, (11)), the plus or minus being used as $\theta < \hat{\theta}$, $\theta > \hat{\theta}$. Applied to the above numerical example this is

$$\sqrt{P_u} \pm \tfrac{1}{8}\sqrt{n} \sim N(0, 1)$$

which gives significance levels (0·0095, 0·0097) to the above 1 per cent values of $\theta$. This correction can be applied to the estimation of a single parameter, with or without nuisance parameters, but does not seem more general.

Mr I. M. S. White (Forestry Commission): I would like to clear up a small point in connection with the example of Section 5.2. The reason for the gross discrepancy between the result of Steinijans' test and the saddle-point approximation is probably due to an erroneous formula in Lewis (1972, p. 33). Lewis gives a standard error for the test statistic $n^{-1}\Sigma t_i^2$ as $T/\sqrt{(192n)}$, whereas the correct value is $T^2/\sqrt{(180n)}$. Steinijans appears to have used the incorrect value. Since he scales his data so that $T = 10{\cdot}99$, his standard error is too small by a factor of about 10. With the correct standard error the test gives a (one-sided) $P$-value of 0·010, close to the values 0·008–0·009 given by the saddle-point approximation.

The Authors replied briefly at the meeting and subsequently more fully in writing as follows.

We are most grateful to all contributors for their constructive and encouraging comments.

Professor Daniels has given an interesting review of the relevant history. We agree that there is a rôle both for complex variable methods and for the method of conjugate distributions, although the relative advantages of the latter seem more marked in the multivariate case. A proof of the unique property of the normal and inverse normal distributions in giving exact saddle-point results without renormalization follows from the special case $n = 1$ to which any exact result must apply. We hope that Professor Daniels will investigate further the angular distribution of Section 5.3; Dr Kent's elegant contribution brings up the difficulty, which we have found in many examples, of the evaluation of the normalizing constant other than by series expansion.

Professor Durbin's results are most valuable and we look forward to publication of a full account of his work. Of course, for independent observations from an exponential family distribution a sufficient statistic is a function of a sum of independent random variables and his results are a re-expression of ours. His time series generalization is especially important. The possibility of expanding the equation for the "maximum likelihood" point is discussed in Section 2.4 although some care is needed in retaining appropriate precision; this applies also to Professor Bickel's interesting applications in non-parametric theory. Professor Hampel's contribution is important both mathematically and for the connections with robut estimation: his remarks are relevant to Professor Mardia's question about the log normal distribution for which the moment generating function is not analytic at the origin. We agree with Mr Mayne's implicit point that often the

distribution function rather than the density is required and reinforce the rather cryptic remark in our paper that in the discrete case care is needed in passing from the latter to the former.

Our work since the present paper has been aimed at the matters which Dr Peers and Professor Hinkley mention; a central difficulty, for example in generalizing the results of Section 6.3, is concerned with notions of approximate ancillarity and the need for conditioning. Professor Sprott's result highlights the point that confidence regions based on the maximum likelihood ratio have good "two-sided" properties but may not be so good when separate upper and lower confidence limits are needed. The higher-order asymptotic theory of the distribution of the maximum likelihood ratio statistic has been rather a puzzle for many years and it is good to learn of Professor Bhattacharya and Dr Ghosh's important as yet unpublished work.

Mr Jørgensen and Mr Pedersen's remarks clarify considerably the behaviour of the approximations on the boundary of the natural parameter space.

We agree with Professor Lindley that asymptotic theory can clarify the similarities and differences between alternative approaches; the work of Welch and Peers (1963) is particularly relevant here.

Professor Sibson's remarks are an important reminder of the role of the computer. As we mention in our paper, Mr B. V. Pedersen has prepared an algorithm for implementing the main results of our paper. Incidentally, the generalized Hermite polynomials of Section 3.2 seem first to have been studied by Appell and Kampé de Feriet (1926). We certainly agree that if one were to go to appreciably more complicated problems use of computerized algebra would be highly desirable if not essential; such methods have been used successfully also in asymptotic expansions arising in fluid mechanics. In general, although not perhaps in Professor Sibson's example, the choice of stopping point in the expansion will be critical; of course, asymptotic expansions may well be divergent.

Dr Atkinson raises a number of central practical issues. In most "tail area" calculations high precision is quite unnecessary. On the other hand, in any application of asymptotic results the question "how good is the approximation?" should always be raised, even if a detailed answer is often unnecessary. One main "practical" rôle of the theoretical discussion is to increase one's feel for when ordinary asymptotic theory is adequate; for critical specific applications simulation may indeed be the simplest and therefore the best approach. Conditional simulation was discussed by Trotter and Tukey (1956). If standard asymptotic theory is not an adequate approximation, it will normally be important to distinguish between conditional and unconditional results. For examples, see Efron and Hinkley (1978) and Cobb (1978).

We are very glad of Mr White's comments sorting out the background to the example of Section 5.2.

### References in the Discussion


Albers, W., Bickel, P. J. and Van Zwet, W. R. (1976). Asymptotic expansions for the power of distribution free tests in the one-sample problem. *Ann. Statist.*, **4**, 108–156.

Andrews, D. F., Bickel, P. J., Hampel, F. R., Huber, P. J., Rogers, W. H. and Tukey, J. W. (1972). *Robust Estimates of Location.* Princeton: University Press.

Appell, P. and Kampé de Feriet, J. (1926). *Fonctions Hypergéometrique et Hypersphérique. Polynomes d'Hermite.* Paris: Gauthier Villars.

Barndorff-Nielsen, O. (1977). Exponentially decreasing distributions for the logarithm of particle size. *Proc. Roy. Soc. London*, A, **353**, 401–419.

Bickel, P. J. and Van Zwet, W. (1978). Asymptotic expansions for the power of distribution free tests in the two-sample problem. *Ann. Statist.*, **6**, 937–1004.

Brockwell, P. J. (1964). An asymptotic expansion for the tail of a binomial distribution and its application in queueing theory. *J. Appl. Prob.*, **1**, 163–169.

Chow, W. (1976). Ph.D. Dissertation, University of California at Berkeley, Department of Statistics.

Chow, W. and Hodges, J. L., Jr (1975). An approximation to the distribution of the Wilcoxon one-sample statistic. *J. Amer. Statist. Ass.*, **70**, 648–655.

Cobb, G. W. (1978). The problem of the Nile: conditional solution to a changepoint problem. *Biometrika*, **65**, 243–251.

Daniels, H. E. (1956). The approximate distribution of serial correlation coefficients. *Biometrika*, **43**, 169–185.

Durbin, J. (1980). Approximations for densities of sufficient estimators.

Efron, B. and Hinkley, D. V. (1978). Assessing the accuracy of the maximum likelihood estimator: observed versus expected Fisher information. *Biometrika*, **65**, 457–482.

Field, C. A. and Hampel, F. R. (1978). Small sample asymptotic distributions of *M*-estimators of location. Research Report No. 17, Fachgruppe für Statistik, ETH, Zürich.

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

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

HAMPEL, F. R. (1973). Some small sample asymptotics. *Proc. Prague Symp. on Asymptotic Statistics*, September 3rd–6th, 1973.

LEWIS, P. A. W. (1972). Recent results in the statistical analysis of univariate point processes. In *Stochastic Point Processes; Statistical Analysis, Theory and Applications* (P. A. W. Lewis, ed.), pp. 1–54. New York: Wiley.

MAYNE, A. J. (1959). A generalized saddle-point method for approximating distributions. *Bull. Int. Statist. Inst.*, **43** (Book 2), 378–380.

PEERS, H. W. (1978). Second order sufficiency and statistical invariants. *Biometrika*, **65**, 489–496.

SPROTT, D. A. (1973). Normal likelihoods and their relation to large sample theory of estimation. *Biometrika*, **60**, 457–465.

TEMME, N. M. (1975). Uniform asymptotic expansions of the incomplete gamma functions and the incomplete beta function. *Math. Comp.*, **29**, 1109–1114.

TROTTER, H. F. and TUKEY, J. W. (1956). Conditional Monte Carlo for normal samples. In *Symposium on Monte Carlo Methods* (H. A. Meyer, ed.), pp. 64–79. New York: Wiley.

WELCH, B. L. and PEERS, H. W. (1963). On formulae for confidence points based on integrals of weighted likelihood. *J. R. Statist. Soc.* B, **25**, 318–329.