Exponential linear models: a two pass procedure for saddlepoint approximation

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

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

Summary

For an exponential linear model, the saddlepoint method gives accurate approximations for the density of the minimal sufficient statistic or maximum likelihood estimate, and for the corresponding distribution functions. In this paper we describe a simple numerical procedure that constructs such approximations for a real parameter in an exponential linear model, using only a two-pass calculation on the observed likelihood function for the original data. Simple examples of the numerical procedure are discussed, but we take the general accuracy of the saddlepoint procedure as given.

An immediate application of this is to exponential family models, where inference for a component of the canonical parameter is to be based on the conditional density of the corresponding component of the sufficient statistic, given the remaining components. This conditional density is also of exponential family form, but its functional form and cumulant generating function may not be accessible. The procedure is applied to the corresponding likelihood, approximated as the full likelihood divided by an approximate marginal likelihood obtained from Barndorff-Nielsen's formula as in Cox and Reid (1987). A double saddlepoint approximation provides another means of bypassing this difficulty; implementation of this for some linear exponential models is discussed in Davison (1988).

The computational procedure is also examined as a numerical procedure for obtaining the saddlepoint approximation to the Fourier inversion of a characteristic function. As such it is a two pass calculation on a table of the cumulant generating function.

Some key words: Barndorff-Nielsen's approximation, conditional inference, exponential family, observed information, saddlepoint approximation, tail area approximations.

1. Introduction

For a real parameter exponential linear model $f(y;\theta)dy = \exp\{\theta t(y) - K(\theta)\}H(y)dy$ the minimal sufficient statistic $t = t(y)$ has density

$$f(t;\theta)dt = \exp\{\theta t - K(\theta)\}h(t)dt. \tag{1.1}$$

It may be difficult to calculate the function $h(t)$ but the density $f(t;\theta)$ and the distribution function $F(t;\theta)$ can be accurately approximated by the saddlepoint method; the corresponding density and distribution function for the maximum likelihood estimate $\hat{\theta}$ are then easily derived. In this paper we exploit the simple structure of both (1.1) and the saddlepoint approximation for exponential families to approximate the density and distribution function using only the observed likelihood function from the observed data y^{obs}. The procedure is a numerical implementation that uses a fine tabulation of the observed log likelihood function $l^{\text{obs}}(\theta) = l(\theta; y^{\text{obs}})$. This converts an observed log likelihood to a density or distribution function, and will have saddlepoint accuracy when the model is close to exponential family form. In contrast, the usual χ^2 approximation for twice the difference of log likelihoods is typically accurate only when the model is close to the normal.

A general vector parameter linear exponential family model can be expressed in canonical form as

$$f(y;\theta)dy = \exp\{\lambda' t_1 + \psi' t_2 - K(\lambda,\psi)\}H(y)dy \tag{1.2}$$

where λ and ψ, of dimensions r and k, respectively, provide a reparametrization of the original parameter θ. If the parameter of interest is a linear function of the canonical components of θ, then the family can be expressed in the form (1.2) with ψ as the parameter of interest and λ as a nuisance parameter.

The conditional density of t_2, given t_1,

$$f(t_2|t_1;\psi) \doteq \exp\{\psi' t_2 - K_{t_1}(\psi)\} H_{t_1}(t_2), \tag{1.3}$$

is free of the nuisance parameter, and would typically be used for inference about ψ in the absence of knowledge of λ. This conditional density has the largest sample space range of any that is free of λ; see, for example, Fraser (1979, p. 81). Except for special cases, the functions $K_{t_1}(\psi)$ and $h_{t_1}(t_2)$ will be available only as integrals requiring numerical evaluation.

By approximating the marginal density of t_1 by the saddlepoint approximation, we can obtain an approximation to the conditional observed log likelihood for (1.3), say l_c^{obs} (Cox & Reid , 1987). We then apply the numerical procedure mentioned above, and obtain approximate conditional density and distribution functions for arbitrary parameter values, and hence conditional confidence intervals for the parameter of interest. We call this a *sequential* saddlepoint procedure.

Another possibility for approximating $f(t_2|t_1;\psi) \doteq f(t_2, t_1;\psi,\lambda)/f(t_1;\psi,\lambda)$ is to apply the saddlepoint approximation separately to the joint and marginal densities in the numerator and the denominator. This double saddlepoint approximation is discussed in Davison (1988) for discrete and continuous linear exponential models of the form (1.1). Davison also obtains a conditional likelihood for generalized linear models with canonical link and unknown scale parameter: this coincides with the approximate conditional likelihood of Cox and Reid (1987).

We restrict our derivation to models for continuous responses, although the numerical procedure may give reasonable results in discrete cases. In the remainder of this section we record the needed density and tail area approximations as applied to linear exponential models.

For the model (1.1) generalized to k dimensions, the usual saddlepoint approximation to the density of t (Daniels, 1954; Barndorff-Nielsen & Cox, 1979) can be written in the form

$$f(t;\theta) \approx (2\pi)^{-k/2} \exp\{-\mathrm{w}(\hat\theta;\theta)/2\} |j(\hat\theta)|^{-1/2} , \tag{1.4}$$

where $\mathrm{w}(\hat\theta;\theta) = 2\{l(\hat\theta;t) - l(\theta;t)\}$; this is a function of just the likelihood function at the point t of interest. The transformation from t to $\hat\theta(t)$ has Jacobian $|j(\hat\theta)|$ and gives

$$f(\hat\theta;\theta) \approx (2\pi)^{-k/2} \exp\{-\mathrm{w}(\hat\theta;\theta)/2\} |j(\hat\theta)|^{1/2} , \tag{1.5}$$

which is Barndorff-Nielsen's (1983) formula as specialized to the exponential model context.

Let $\eta = \eta(\theta)$ be a reparameterization that gives constant information determinant; for the scalar parameter case, $\eta = \eta(\theta) = \int^{\theta} j^{\frac{1}{2}}(\hat\theta)d\hat\theta$. The density function for $\hat\eta$ is then

$$f(\hat\eta;\eta) \approx (2\pi)^{-k/2} \exp\{-\mathrm{w}(\hat\theta;\theta)/2\} ; \tag{1.6}$$

for the scalar case, theoretical assessment (Fraser, 1988) and numerical evaluation indicate that $\hat\eta$ can have a density remarkably close to that of a normal location model with mean η and standard deviation 1.

For the case $k=1$, an approximation for the distribution function $F(t;\theta)$ or $F(\hat\theta;\theta)$, is available using a formula due to Lugannani and Rice (1980), and reviewed in Daniels (1987):

$$F(\hat\theta;\theta) = F(t;\theta) \approx \Phi(r) + \phi(r)\left\{\frac{1}{r} - \frac{1}{q}\right\} , \tag{1.7}$$

where Φ and ϕ are the standard normal distribution and density functions, r is the signed square root of the likelihood ratio statistic, and q is the standardized maximum likelihood estimate:

$$r = \operatorname{sgn}(\hat{\theta} - \theta)[\mathrm{w}(\hat{\theta};\theta)\}]^{1/2}$$
$$q = (\hat{\theta} - \theta)|j(\hat{\theta})|^{1/2} \; . \tag{1.8}$$

The formula becomes unstable for $\hat{\theta}$ near θ , and at $\hat{\theta}=\theta$ (1.7) should be replaced by

$$F(\hat{\theta};\hat{\theta}) \approx \frac{1}{2} + \frac{1}{6}(2\pi)^{-1/2}\frac{j_3(\hat{\theta})}{|j(\hat{\theta})|^{3/2}}$$

where $j_3(\hat{\theta}) = -\partial^3 l(\theta;t)/\partial\theta^3|_{\hat{\theta}}$ (Daniels,1987).

2. From observed likelihood to density and distribution functions

For an exponential model (1.1) with a k-dimensional canonical parameter θ and minimal sufficient statistic $t = t(y)$, we denote an observed data value by y^{obs}, and let $t^{\text{obs}} = t(y^{\text{obs}})$, $\hat{\theta}^{\text{obs}} = \hat{\theta}(y^{\text{obs}})$, and $l^{\text{obs}}(\theta) = l(\theta; y^{\text{obs}})$, where for $l^{\text{obs}}(\theta)$ any version, say $\log f(y^{\text{obs}};\theta)$, or $\log f(t^{\text{obs}};\theta)$, or $a + \log f(y^{\text{obs}};\theta)$ can be used.

The general log likelihood function for the model (1.1) can be expressed in terms of the observed log likelihood function as

$$l(\theta;t) = \theta'(t - t^{\text{obs}}) + l^{\text{obs}}(\theta) \; . \tag{2.1}$$

The corresponding log likelihood difference can then be written

$$\begin{aligned} l(\hat{\theta};t) - l(\theta;t) &= l^{\text{obs}}(\hat{\theta}) - l^{\text{obs}}(\theta) + (\hat{\theta} - \theta)'(t - t^{\text{obs}}) \\ &= l^{\text{obs}}(\hat{\theta}) - l^{\text{obs}}(\theta) - (\hat{\theta} - \theta)'S^{\text{obs}}(\hat{\theta}) \end{aligned} \tag{2.2}$$

where the general score function is $S(\theta;t) = (t - t^{\text{obs}}) + S^{\text{obs}}(\theta)$, and

$$t - t^{\text{obs}} \doteq - S^{\text{obs}}(\hat\theta) \; . \qquad (2.3)$$

The saddlepoint approximation for the densities of t and $\hat\theta$ in the continuous case are given by (1.4) and (1.5) with

$$\begin{aligned} \mathrm{w}(\hat\theta;\theta)/2 &\doteq \{l^{\text{obs}}(\hat\theta) - l^{\text{obs}}(\theta)\} - (\hat\theta - \theta)S^{\text{obs}}(\hat\theta) \; , \\ j(\hat\theta) &\doteq -\frac{\partial^2 l^{\text{obs}}(\theta)}{\partial\theta\partial\theta'} \Big|_{\hat\theta} \; . \end{aligned} \qquad (2.4)$$

For the case $k = 1$ the saddlepoint approximation for the distribution function of t or $\hat\theta$ is given by (1.7) using (1.8) and (2.4).

The approximations for the densities and distribution functions for general $\hat\theta$ and θ in the case $k = 1$ can be obtained from a simple two-pass calculation on a tabulation of the observed log likelihood function $l^{\text{obs}}(\theta)$. This requires only a numerical tabulation of $l^{\text{obs}}(\theta)$ over a fine grid of θ values in steps of $\pm\delta$, and successive divided differences

$$\begin{aligned} l_1(\theta) &\doteq \{l^{\text{obs}}(\theta+\delta) - l^{\text{obs}}(\theta)\}/\delta \; , \\ l_2(\theta) &\doteq \{l_1(\theta+\delta) - l_1(\theta)\}/\delta \; . \end{aligned} \qquad (2.5)$$

Saddlepoint approximations for $f(\hat\theta;\theta)$, and $F(\hat\theta;\theta)$ are then given by (1.5), and (1.7) with (1.8) using the formulas

$$\begin{aligned} \mathrm{w}(\hat\theta;\theta)/2 &\doteq l^{\text{obs}}(\hat\theta) - l^{\text{obs}}(\theta) - (\hat\theta - \theta)l_1(\hat\theta) \\ j(\hat\theta) &\doteq -l_2(\hat\theta - \delta) \; ; \end{aligned} \qquad (2.6)$$

with values only from the θ and $\hat\theta$ entries in the tabulation. The density $f(t;\theta)$ is also given by (1.4) at the t value corresponding to $\hat\theta$,

$$t \doteq t^{\text{obs}} - l_1(\hat\theta) \; ;$$

for plotting directly against t this needs t^{obs} in addition to the tabular values. The effect of the interval size δ can perhaps be reduced by replacing $l_1(\theta)$ by

$\{l_1(\theta) + l_1(\theta - \delta)\}/2$.

A confidence interval $(\hat{\theta}_L, \hat{\theta}_U)$ for θ at level $1 - \alpha$ is obtained from the observed $\hat{\theta}^{\text{obs}}$ by solving

$$F(\hat{\theta}^{\text{obs}}; \theta) = 1 - \alpha/2 \,, \quad F(\hat{\theta}^{\text{obs}}; \theta) = \alpha/2 \,. \tag{2.7}$$

We call the decreasing function $F(\hat{\theta}^{\text{obs}}; \theta)$ the confidence distribution function; $\hat{\theta}^{\text{obs}}$ is the observed maximum likelihood estimate, which could be available from separate calculations or by linear interpolation between the two values $l_1(\theta)$ and $l_1(\theta + \delta)$ that bracket zero. If only the confidence distribution function $F(\hat{\theta}^{\text{obs}}; \theta)$ is needed, then just a single pass computation is needed plus a directly computed $j(\hat{\theta}^{\text{obs}})$.

An interactive program written in C was developed for the examples in this paper; a copy is available on request from A. Wong at the University of Toronto. As a guide to more general use of the program, it should be emphasized that the use of the formulas (2.6) with the saddlepoint expressions assumes that θ is the canonical parameter of the exponential model; procedures for extracting the canonical parameter in applications, exactly or approximately, are being developed in current work.

The generalization to cover the case k=2, say, would be straightforward and would yield saddlepoint density approximations. However, our primary interest is in the tail probabilities supplied by the distribution approximation and no simple version of this seems to be available in the vector case.

Example 2.1. For a first example we choose an exponential model context far removed from normality, and consider a sample of $n = 1$ with data value $y^0 = 1$ from the shape gamma

$$\Gamma^{-1}(\theta) y^{\theta - 1} e^{-y} \tag{2.8}$$

on $(0, \infty)$. We view this as an extreme case having the special feature of a fixed left boundary at $y = 0$: the effectiveness of the procedure can be partly judged from the closeness to normality of the constant information variable $\hat{\eta}$ defined for (1.6); the change of variable attempts to remove this left boundary.

The maximum likelihood estimate of θ is $\hat{\theta}^{\text{obs}} = 1.46$. The density function for $\hat{\eta}$ is recorded in Figure 1 for the 5 values of η in Table 1 where $\hat{\eta}^{\text{obs}}$ is taken conveniently to be zero; note the very close approximation to normal form except perhaps for the left-hand curve with $\theta = 0.0983$. The parameter values are the confidence points obtained from the approximation to $pr(\hat{\eta} \leq 0)$ using the tail approximation formula (1.7); this function is labelled $F_a(0;\eta)$ in Table 1. The exact values are easily calculated from the gamma function and are given by $F_e(0;\eta)$. The values obtained from the observed log likelihood function by numerical tabulation are exceptionally close to the exact values. The numerical integration of the density is also recorded in Table 1 as $\hat{F}_a(0;\eta)$; this integration of a saddlepoint density seems to be less reliable than the saddlepoint distribution function.

Figure 1 about here

Table 1

The distribution function at the observed $\hat{\theta}^{\text{obs}}$ for five values of θ: F_a is the approximation (1.7); F_e is the exact, and $\hat{F}_a$ is an approximation obtained by numerically integrating the density (1.6).

θ	0.0983	0.5697	1.2676	2.2974	3.8138
η	-3.0046	-1.1873	-0.1993	0.7035	1.6545
$F_a(\hat{\theta}^0; \theta) \doteq F_a(0; \eta)$	0.975	0.810	0.500	0.190	0.025
$F_e(\hat{\theta}^0; \theta) \doteq F_e(0; \eta)$	0.976	0.815	0.519	0.191	0.025
$\hat{F}_a(\hat{\theta}^0; \theta) \doteq \hat{F}_a(0; \eta)$	0.978	0.816	0.517	0.189	0.025

3. Conditional inference in the exponential linear model

An approximation to the observed conditional log likelihood for ψ in the conditional context (1.3), is obtained as described in Section 1:

$$l_c^{\text{obs}}(\psi) \doteq l(\psi, \hat{\lambda}_\psi^{\text{obs}}; y^{\text{obs}}) + \tfrac{1}{2} \log |j_{\lambda\lambda}(\psi, \hat{\lambda}_\psi^{\text{obs}}; y^{\text{obs}})| \qquad (3.1)$$

where $\hat{\lambda}_\psi^{\text{obs}}$ is the observed maximum likelihood estimate of λ for fixed ψ and $j_{\lambda\lambda} \doteq -\partial^2 l(\psi, \lambda; y^{\text{obs}})/\partial\lambda\partial\lambda'$ is the corresponding observed information for λ with ψ fixed; all ingredients for this are available from the observed likelihood function. The expression is obtained from the quotient of the full likelihood $L(\psi, \lambda; y^{\text{obs}})$ from y^{obs} and the approximate marginal likelihood

$$L(\psi, \lambda; y^{\text{obs}}) L^{-1}(\psi, \hat{\lambda}_\psi^{\text{obs}}; y^{\text{obs}}) |j_{\lambda\lambda}(\psi, \hat{\lambda}_\psi^{\text{obs}}; y^{\text{obs}})|^{-1/2} \qquad (3.2)$$

for t_1^{obs} from (1.4) with fixed ψ.

We consider the analysis for the particular case of a real parameter ψ and use the results from Section 2 with $l_c^{\text{obs}}(\psi)$ in place of the earlier $l^{\text{obs}}(\theta)$. For this conditional context, the method can be described as a *sequential* saddlepoint procedure: a saddlepoint approximation is used for the marginal log likelihood (3.2) and the resulting conditional log likelihood (3.1) is used to approximate the conditional density and distribution functions by the method of Section 2 which uses a second saddlepoint application from (1.6) or (1.7) computed numerically. Work in progress indicates that this approximation is accurate to $O(n^{-3/2})$ in the sampling context.

The double saddlepoint approximation described in Davison (1988) requires an information calculation for each value (t_1^{obs}, t_2). The resulting likelihood function is the same as that above, but the saddlepoint norming of that likelihood function to provide a density and distribution function is different. The results in Fraser (1988) suggest that the saddlepoint approximation applied directly to the conditional likelihood from a density of the form (1.1) may be preferable to the double saddlepoint method on general grounds. Exceptions can occur, however: an example given in McCullagh (1987, p.195) describes the conditional distribution of one of two independent exponentials, given their sum. In this case the double saddlepoint method will reproduce the exact conditional density, if the numerator and denominator are normed separately.

Example 3.1. Cox and Snell (1981) examine some data from Proschan (1963) concerning the time interval between failures of air conditioning equipment on aircraft. The data consist of n_i independent failure times observed on each of $k = 10$ aircraft. Various levels of modelling were considered; here we examine the use of a gamma distribution for the jth failure on airplane i:

$$\Gamma^{-1}(\beta)(\beta/\mu_i)(\beta y_{ij}/\mu_i)^{\beta-1} \exp\{-\beta y_{ij}/\mu_i\}$$

where the mean failure time depends on i, but the groups have a common shape parameter β. The parameter of interest is taken to be the shape parameter β; the value $\beta = 1$

has special interest and corresponds to the exponential distribution.

The approximation (1.7) to the distribution function $F(\hat{\beta};\beta)$ at the observed $\hat{\beta}^{\text{obs}}$ is recorded as $F_a(\hat{\beta};\beta)$ and is plotted in Figure 2. Near the centre of the distribution where r and q are close to zero, the numerical procedure and the Lugannani and Rice formula (1.7) are unreliable; we do not directly address this as our main objective is to obtain confidence intervals and tail probabilities.

Five values of the parameter β were obtained from F_a corresponding to the values 0.975, 0.810, 0.500, 0.190, 0.025. Table 2 records the corresponding values of F_a as well $\hat{F}_a$ obtained by numerically integrating the density (1.6).

Figure 2 about here

Table 2

The distribution function at the observed $\hat{\beta}^{\text{obs}}$ for five values of β: F_a is the approximation (1.7); $\hat{F}_a$ is an approximation based on numerically integrating the density (1.6).

β	0.805	0.890	0.972	1.044	1.146
η	-2.065	-0.937	0.048	0.846	1.865
$F_a(\hat{\beta}^{\text{obs}};\beta) \doteq F_a(0;\eta)$	0.975	0.810	0.500	0.190	0.025
$\hat{F}_a(\hat{\beta}^{\text{obs}};\beta) \doteq \hat{F}_a(0;\eta)$	0.975	0.813	0.478	0.193	0.025

This example involves 199 observations covering 10 aircraft, so it is not surprising that the standardized densities in the constant information parameterization are found to be close to standard normal and that the approximate conditional confidence intervals are similar to those obtained by the usual first order asymptotic theory. The conditional maximum likelihood estimate $\hat{\beta}^{\text{obs}}$ is 0.9675 whereas the ordinary maximum likelihood estimate (Snell, 1987) is 1.006. The 95% conditional confidence interval is (0.805, 1.146) which can be compared with the ordinary large sample approximation (0.795, 1.140).

4. Numerical saddlepoint inversion of a cumulant generating function

Consider a real variable t with cumulant generating function $K(s)$:

$$\exp\{K(s)\} = \int_{-\infty}^{\infty} f(t) e^{st}\, dt \; .$$

We apply the computational procedure from Section 2 to obtain saddlepoint approximations to $f(t)$ and $F(t)$ from the generating function $K(s)$.

The embedding exponential model is

$$f(t;\theta) = f(t) e^{\theta t - K(\theta)} \; . \qquad (4.1)$$

For some conveniently chosen value of t, say t^0, the log likelihood function is

$$l^0(\theta) = \theta t^0 - K(\theta) \; .$$

The two-pass procedure on a tabulation of $l^0(\theta)$ gives $f(t)$ and $F(t)$ for $\theta = 0$ from (1.4) and (1.7) with

$$\begin{aligned} w(\hat\theta; 0)/2 &= l^0(\hat\theta) - l^0(0) - \hat\theta l_1(\hat\theta) \\ j(\hat\theta) &= -l_2(\hat\theta - \delta) \end{aligned} \qquad (4.2)$$

Example 4.1. We illustrate the inversion of a cumulant function by reconsidering Example 2.1.

Consider the extreme value density

$$f(u) = \exp\{-e^u + u\} \tag{4.3}$$

on $(-\infty, \infty)$; this is the density of the logarithm of a standard exponential variable. The cumulant generating function is $K(s) = \ln\Gamma(s+1)$. As a choice of a reference value for u we take $u^0 = 0$. Note that the embedding exponential model is

$$\begin{aligned} f(u;\alpha)du &= f(u)e^{\alpha u - K(\alpha)} \\ &= \exp\{-e^u + (\alpha+1)u - \ln\Gamma(\alpha+1)\}du \\ &= \Gamma^{-1}(\alpha+1)y^{\alpha+1}e^{-y}dy/y \end{aligned} \tag{4.4}$$

which is the gamma $(\alpha+1, 1)$ distribution for $y = e^u$.

The observed log likelihood for $y = y^0 = 1$ $(u^0 = 0)$ is obtained from Example 2.1: $l^0(\alpha) = \alpha u^0 - K(\alpha) = -\ln\Gamma(\alpha+1)$. The numerical tabulation procedure gives a saddle-point approximation to the density (4.3) which can be compared with the exact density (Figure 3): the difference is less than the drawing accuracy.

Figure 3 about here

5. Discussion

In Section 2 we described how a fine tabulation of an observed log likelihood function could be augmented by first and second differences to give highly accurate saddle-point tail probabilities and confidence points in exponential families. Daniels (1983,

1988) has suggested the possible use of numerical procedures for the saddlepoint approximation.

It seems likely that the technique can be applied more generally as an alternative to the chi-square evaluation of likelihood drop; for a model close to exponential form the result has been verified in current work to have saddlepoint accuracy. All that is needed is a tabulation of an observed log likelihood in terms of a parameter that corresponds to the canonical parameter of an exponential model. Some discussion of the choice for such a parameterization may be found in Fraser (1988, 1989) based on a sample space derivative of the likelihood function; refinements of this are discussed elsewhere.

As an example of the application of the numerical procedure in a more general context we consider inference for the mean of a gamma distribution. This is still an exponential family, but the parameter of interest is not a linear component of the canonical parameter. The approximate conditional likelihood of Cox and Reid (1987) is readily obtained; that likelihood and an adjusted version are compared to the profile likelihood and Barndorff-Nielsen's modified profile likelihood in Fraser and Reid (1989). Figure 4 shows the confidence distribution functions for the mean, obtained from a simulated sample of size 10 from a gamma (5,1) distribution, and constructed by applying the numerical method outlined above to the Cox and Reid conditional profile and the Fraser and Reid adjusted version. As noted in Example 3.1 the numerical procedure is unstable in the neighbourhood of the maximum likelihood estimate (r and q near 0 in (1.8)); this accounts for the bump in the plot corresponding to the conditional profile log likelihood. The 95% confidence intervals for the mean are (.654, 1.036) and (.636, 1.068) for the conditional profile and the adjusted profile, respectively. An exact conditional confidence interval can be obtained by a modification of arguments outlined in Jensen (1986); this is discussed separately, in Fraser, Reid and Wong (1989).

Figure 4 about here

References

Barndorff-Nielsen, O.E. (1983) On a formula for the distribution of the maximum likelihood estimator. *Biometrika*, **70,** 343-65.

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

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

Cox, D.R., and Snell, J. (1981) *Applied Statistics.* London: Chapman and Hall.

Daniels, H.E. (1954) Saddlepoint approximations in statistics. *Ann. Math. Statist.,* **29,** 631-50.

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

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

Daniels, H.E. (1988) Comment on ‘Saddlepoint methods and statistical inference’ (by N. Reid). *Statistical Science,* **2,** 229.

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

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

Fraser, D.A.S. (1988) Normed likelihood as saddlepoint approximation. *J. Mult. Anal.,* **27,** 181-193.

Fraser, D.A.S. (1990) Tail probabilities from observed likelihoods. *Biometrika*. **77,** to appear.

Fraser, D.A.S., and Reid, N. (1989) Adjustments to profile likelihood. *Biometrika,* **76,** 477-488.

Fraser, D.A.S., Reid, N., and Wong. A (1989) Inference for the gamma mean. Submitted for publication.

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

Lugannani, R., and Rice, S.O. (1980) Saddlepoint approximation for the distribution of the sums of independent random variables. *Adv. App. Prob.,* **12,** 475-90.

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

Proschan, F. (1963) Theoretical explanation of observed decreasing failure rate. *Technometrics,* **5,** 375-383.

Snell, E.J. (1987) *Applied Statistics: a handbook of BMDP analyses*. London: Chapman and Hall.

Figure Captions.

Figure 1. The density function $f(\hat{\eta};\eta)$ is recorded from left to right for the 5 values of θ and η in Table 1.

Figure 2. The confidence distribution function $F(0;\eta)$ from the air conditioning data in Example 3.1.

Figure 3. The exact (____) and saddlepoint (– – –) approximation to the extreme value density (4.3); the approximation is obtained by the numerical inversion of the cumulant generating function.

Figure 4. The confidence distribution functions for the gamma mean parameter, based on the approximate conditional likelihoods of Cox and Reid,(____) and Fraser and Reid(– – –).