CONVERTING OBSERVED LIKELIHOOD FUNCTIONS TO TAIL PROBABILITIES

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

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

ABSTRACT

The chi-square approximation for likelihood drop is widely used but may be inaccurate for small or medium sized samples; mean and variance corrections may help. The Lugannani and Rice tail probability formula provides high accuracy based on a cumulant generating function, which is readily available for exponential family models. This paper surveys extensions of this formula to more general contexts and describes a simple numerical procedure for testing real parameters in exponential linear models.

1. INTRODUCTION

The saddlepoint method (Daniels [5]; Barndorff-Nielsen and Cox, [4]) gives accurate approximations to a density function based on the corresponding cumulant generating function. For statistical contexts the cumulant generating function is typically available for an exponential family model. With such a model the approximate density of the minimal sufficient statistic can be expressed in terms of likelihood,

$$f(t;\theta) \approx (2\pi)^{-k/2}|j(\hat\theta)|^{-1/2}\exp\{-w(\hat\theta;\theta)/2\} \tag{1.1}$$

where $w(\hat\theta;\theta) \doteq 2\{l(\hat\theta;t)-l(\theta;t)\}$ is the likelihood ratio statistic, $j(\hat\theta)$ is the observed information for the canonical parameter, and k is the dimension.

For the real variable case, Lugannani and Rice [16] and Daniels [6] give a saddlepoint approximation for the distribution function,

$$F(t;\theta) \doteq F(\hat\theta;\theta) \approx \Phi(r)+\phi(r)\left\{\frac{1}{r}-\frac{1}{q}\right\} \tag{1.2}$$

where Φ and ϕ are the standard normal distribution and density functions, r is the signed square-root of the likelihood ratio statistic

$$r \doteq \operatorname{sgn}(\hat{\theta} - \theta)\{w(\hat{\theta}; \theta)\}^{1/2} , \tag{1.3}$$

and q is the standardized maximum likelihood estimate of the canonical parameter,

$$q \doteq (\hat{\theta} - \theta)|j(\hat{\theta})|^{1/2} . \tag{1.4}$$

For a sample of size n the approximation is accurate to $O(n^{-3/2})$ on an $O(n^{-1/2})$ region of the sample average about the mean. In the typical statistical context with a one parameter exponential linear model this gives an observed level of significance for testing the parameter value θ.

In Section 2 we describe a simple two-pass numerical procedure that for the real parameter case produces approximate density and distribution functions using only an observed likelihood function. The procedure extends to inference for a canonical parameter or ratio of canonical parameters in a k-dimensional exponential model.

In Section 3 we discuss a parametrization invariant version of the Lugannani and Rice formula that extends in more general models to give approximate observed levels of significance for testing real parameters. The method is based on tangent exponential models and uses an observed likelihood and a sample space derivative of that observed likelihood function.

In Section 4 we discuss briefly the close connections between density functions, likelihood functions, and cumulant generating functions and indicate some further extensions of the procedure to calculate tail probabilities.

2. THE NUMERICAL SADDLEPOINT

Consider an exponential model with a k-dimensional canonical parameter θ:

$$\exp\{\theta' t(y) - K(\theta)\} h(y) .$$

A general likelihood function can be calculated from an observed likelihood function

$$l(\theta; t) \doteq \theta'(t - t^0) + l^0(\theta) , \tag{2.1}$$

and the corresponding likelihood drop can be written as

$$l(\hat\theta; t) - l(\theta; t) \doteq l^0(\hat\theta) - l^0(\theta) - (\hat\theta - \theta)' S^0(\hat\theta) ; \tag{2.2}$$

for this the general score $S(\theta; t) \doteq (t - t^0) + S^0(\theta)$ gives $t - t^0 \doteq - S^0(\hat\theta)$.

As a numerical procedure on an observed likelihood function in the real parameter case, let

$$l_1(\theta) \doteq \{l^0(\theta + \delta) - l^0(\theta)\}\delta^{-1}, \; l_2(\theta) \doteq \{l_1(\theta + \delta) - l_1(\theta)\}\delta^{-1} \tag{2.3}$$

based on a fine grid tabulation with step size δ. Then the density (1.1) and distribution function (1.2) approximations in Section 1 are available with

$$w(\hat\theta; \theta)/2 \doteq l^0(\hat\theta) - l^0(\theta) - (\hat\theta - \theta) l_1(\theta) \tag{2.4}$$

and $j(\hat\theta) \doteq - l_2(\hat\theta - \delta)$. An observed level of significance is given by $F(\hat\theta^0; \theta)$ and a confidence interval (θ_L, θ_U) by solving

$$F(\hat\theta^0; \theta) \doteq 1 - \alpha/2, \quad F(\hat\theta^0; \theta) \doteq \alpha/2 .$$

For details and numerical examples see Fraser, Reid, and Wong [14].

Now consider a real parameter ψ in an exponential linear model

$$\exp\{\psi t + \lambda' t_1 - K(\lambda, \psi)\} h(y) .$$

The observed likelihood function from the conditional distribution of t given t_1 has as its saddle-point approximation

$$l_c^0(\psi) \doteq l(\psi, \hat\lambda_\psi^0; y^0) + \frac{1}{2} \log |j_{\lambda\lambda}(\psi, \hat\lambda_\psi^0; y^0)| \tag{2.5}$$

where $\hat\lambda_\psi^0$ is the observed maximum likelihood estimate of λ for given ψ and $j_{\lambda\lambda}$ is the observed information for λ again for given ψ. This extended approximation is also accurate to $O(n^{-3/2})$, Fraser and Reid [12].

The numerical procedure applied to $l_c^0(\psi)$ then gives tests and confidence intervals for ψ. This is called a sequential saddlepoint in Fraser, Reid, and Wong [14] as opposed to the double saddlepoint procedure discussed in Davison [7].

As a further extension consider a parameter ψ that is the ratio of two canonical parameters:

$$\exp\{\psi\lambda t + \lambda t_1 - K(\lambda, \psi\lambda)\}h(y_1, y_2)\ ; \tag{2.6}$$

additional nuisance parameters are possible by combining the discussion here with the conditional analysis just described. For testing a value ψ_0 the model can be expressed in the form

$$\exp\{\gamma t + \lambda(t_1 + \psi_0 t) - \bar{K}(\gamma, \lambda)\}h(y_1, y_2) \tag{2.7}$$

where $\gamma \sqsubseteq \lambda(\psi - \psi_0)$; the conditional distribution of t given $t_1 + \psi_0 t$ can then be examined for a null hypothesis test of $\psi \sqsubseteq \psi_0$ or $\gamma \sqsubseteq 0$. The numerical procedure directly gives the observed level of significance, and is also accurate to $O(n^{-3/2})$. For further details see Fraser, Reid and Wong [13], where confidence intervals are obtained for the gamma mean; also see Jensen [15].

3. GENERALIZED TAIL PROBABILITY FORMULAS

Barndorff-Nielsen's [1] approximation to the density of the maximum likelihood estimator $\hat\theta$ for a k-dimensional parameter has the form

$$h(\hat\theta|a;\theta)d\hat\theta \approx (2\pi)^{-k/2}\frac{L(\theta;\hat\theta,a)}{L(\hat\theta;\hat\theta,a)}|j(\hat\theta,a)|^{1/2}d\hat\theta \tag{3.1}$$

where a is some exact or approximate ancillary. This can be viewed (Fraser, [9]) as a saddlepoint approximation to a tangent exponential model.

For a real parameter θ the Lugannani and Rice formula can be applied (Fraser, [10]) to the tangent exponential model giving

$$F(\hat\theta;\theta) \approx \Phi(r) + \phi(r)\left\{\frac{1}{r} - \frac{1}{q}\right\} \tag{3.2}$$

where r as before is the signed square root of the likelihood ratio statistic,

$$r = \operatorname{sgn}(\hat{\theta} - \theta) \cdot [2\{l(\hat{\theta}; y) - l(\theta; y)\}]^{1/2} , \tag{3.3}$$

q is a standardized estimate of a linearized parameter,

$$q = \{l_y(\hat{\theta}; y) - l_y(\theta; y)\}|j(\hat{\theta})|^{1/2} k^{-1}(y) , \tag{3.4}$$

with $l_y(\theta; y) = (\partial/\partial y)l(\theta; y)$, $k(y) = (\partial^2/\partial\theta\partial y)l(\theta; y)|_{\hat{\theta}}$. For the case of a vector variable y the derivative $\partial/\partial y$ needs to be interpreted as a directional derivative given the exact or approximate ancillary a. For the likelihood function $l(\theta; y)$ it suffices to use any fixed representative such as $\ln\{f(y; \theta)/f(y; \theta_0)\}$ or $\ln\{f(y; \theta)/f(y; \hat{\theta})\}$.

For exponential models the formula coincides with the Lugannani and Rice formula but does not require θ to be the canonical parameter; thus it is a parametrization invariant version of that formula, and is accurate to $O(n^{-3/2})$. An approximation reported in Barndorff-Nielson [2,4] as $O(n^{-1})$ for exponential families is similar to the present approximation in certain contexts.

To examine the accuracy of the approximation more generally, location models were chosen as being in some sense orthogonal to the exponential model. For a model $f(x - \theta)$ we have

$$\begin{aligned} r &= \operatorname{sgn}(x - \theta)\{2[\ln f(0) - \ln f(x - \theta)]\}^{1/2} \\ q &= \{-f'(x - \theta)/f(x - \theta)\}|j(0)|^{-1/2} \end{aligned} \tag{3.5}$$

where it is assumed that f has been centred at the origin. For a sample of size n from such models the formula has been shown to be accurate to $O(n^{-3/2})$ in DiCiccio, Field and Fraser [8]. Of various numerical examples considered, the Cauchy was the most extreme and for testing $\theta = 0$ gave the values F_a in Table 1, expressed in percent.

TABLE 1. Cauchy distribution approximations, in percent.

$\hat{\theta}$	-30	-10	-2
$F_{LR}(\hat{\theta};\theta)$	.01	.12	3.64
$F_{CLR}(\hat{\theta};\theta)$	.07	.54	8.81
$F_a(\hat{\theta};\theta)$	.94	2.81	13.30
$F_e(\hat{\theta};\theta)$	1.06	3.17	14.76

The usual asymptotic approximation is $F_{LR}(\hat{\theta};\theta) = \Phi(r)$; the corrected likelihood ratio value (Barndorff-Nielsen, [1]) is recorded as $F_{CLR}(\hat{\theta};\theta)$; the exact value is recorded as $F_e(\hat{\theta};\theta)$. For continuous unimodal distributions the approximation was quite generally good. It should be noted that the numerical example corresponds to the $n = 1$ case.

Somewhat more generally, consider the regression model $y = X\beta + \sigma e$ where e has some nonnormal density $f(e)$ on R^n and X is full rank $n \times r$. Standard conditional methods of analysis lead to the $r+1$ dimensional conditional density

$$g(t, v|d) = cf\{(Xt + d)e^v\}e^{nv} \tag{3.6}$$

where $t = (\hat{\beta} - \beta)/\hat{\sigma}$, $v = \log(\hat{\sigma}/\sigma)$, and $d = (y - X\hat{\beta})/\hat{\sigma}$. Mild assumptions on f show that $\ln g(t, v)$ is $O_p(n)$. For testing a real parameter component say $\beta_r = 0$ or $\sigma = \sigma_0$ the *marginal* density of the corresponding t_r or v is needed from the conditional distribution (3.6). Two recent approaches lead to tail probability approximations and observed levels of significance.

In Fraser, Lee and Reid [11] the needed marginal density, for say t_r, obtained by integrating out the remaining components in (3.6) is approximated by a constructed conditional distribution, adapted to mimic the behaviour of $\int g(t, v|d)dt_1 \cdots dt_{r-1}dv$ near its maximum. The resulting approximation is a one-dimensional density, and two options are available: to numerically integrate the one dimensional distribution; or to apply the tail probability approximations (3.2) with (3.5). For the numerical examples considered, the two approximations were very close, although the numerical integration was slightly better. The approximating density was also used to develop a Monte Carlo importance sampling estimate of the needed marginal observed level of significance and this was used to confirm the accuracy of the two approximate conditional methods.

In DiCiccio, Field, Fraser [8] a detailed asymptotic analysis is given for a density of the form (3.6). This leads directly to a tail probability approximation for the marginal distribution of the component t_r or v which is accurate to $O_p(n^{-3/2})$ as $n \to \infty$. The result coincides with that mentioned above based on the adjusted conditional distribution. For quite general continuous statistical models, the tail probability formula (3.2) with (3.5) is shown in Fraser and Reid [12] to be accurate to $O(n^{-1})$.

4. FURTHER GENERALIZATIONS

Close connections exist between density, likelihood and cumulant generating functions, particularly in the exponential model context. Let $f(y;\theta)$ be the exponential extension of $f(y)$ with cumulant generating function $K(t)$, and let $c(t;\theta)$ be the corresponding cumulant generating function. Then

$$f(y;\theta) \;=\; \exp\{\theta' y - K(\theta)\} f(y) \tag{4.1}$$

and

$$c(t;\theta) \;=\; K(t+\theta) - K(\theta) \;=\; \Delta_t K(\theta)\,. \tag{4.2}$$

The likelihood function from a value y is

$$l(\theta; y) \;=\; \theta' y - K(\theta)\,. \tag{4.3}$$

The cumulant generating function for a θ_2 distribution can be obtained from that for a θ_1 distribution:

$$\begin{aligned} c(t;\theta_2) &= c(t+\theta_2-\theta_1;\theta_1) - c(\theta_2-\theta_1;\theta_1) \\ &= \Delta_t c(\theta_2-\theta_1;\theta_1) \end{aligned} \tag{4.4}$$

The y_2 likelihood function can be obtained from a y_1 likelihood function:

$$l(\theta; y_2) \doteq \theta'(y_2 - y_1) + l(\theta; y_1) \tag{4.5}$$

We can move from a θ_1 generating function to a y_1 likelihood function

$$l(\theta; y_1) \doteq (\theta - \theta_1)' y_1 - c(\theta - \theta_1; \theta_1) \,. \tag{4.6}$$

The return from likelihood to generating function is more complicated. From (4.1) we see that $l(\hat\theta(y_1); y_1) - l(t + \hat\theta(y_1); y_1)$ is the cumulant generating function of $y - y_1$ with parameter $\theta \doteq \hat\theta(y_1)$. It follows that the cumulant generating function of y is

$$c\{t; \hat\theta(y_1)\} \doteq t' y_1 + l(\hat\theta(y_1); y_1) - l(t + \hat\theta(y_1); y_1)$$

for $\theta \doteq \hat\theta(y_1)$, and in general is

$$\begin{aligned} c(t; \theta) &\doteq \Delta_t c\{\theta - \hat\theta(y_1); \hat\theta(y_1)\} \\ &\doteq t' y_1 + l(\theta; y_1) - l(t + \theta; y_1) \end{aligned}$$

These close connections allow an asymptotic analysis of density, cumulant generating functions and likelihood functions and lead to generalizations of the tail probability formulas in Fraser, Lee, Reid [11] and DiCiccio, Field, Fraser [8]. The details will be reported in Fraser and Reid [12].

REFERENCES

[1] Barndorff-Nielsen, O.E. Inference on full or partial parameters based on the standardized signed log likelihood ratio. *Biometrika* 73 (1986) 307-322.

[2] Barndorff-Nielsen, O.E. Comment on saddlepoint methods and statistical inference. *Statist. Sc.* 3(1988) 228-229.

[3] Barndorff-Nielsen, O.E. Approximate interval probabilities. *J. Royal Statist. Soc. B*, to appear(1990).

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

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

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

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

[8] DiCiccio, T.J., Field, C.A. and Fraser, D.A.S. Marginal tail probabilities and inference for real parameters. *Biometrika* 77 (1990),to appear

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

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

[11] Fraser, D.A.S., Lee, H.S. and Reid, N. Nonnormal linear regression; an example of significance levels in high dimensions. submitted to *Biometrika* (1989).

[12] Fraser, D.A.S. and Reid, N. From multiparameter likelihood to tail probabilities for a scalar parameter. manuscript (1989).

[13] Fraser, D.A.S., Reid, N. and Wong, A. Inference for the gamma mean with unknown shape or for the ratio of exponential parameters. submitted to *Scand. J. Statist.* (1989).

[14] Fraser, D.A.S., Reid, N. and Wong, A. Exponential linear models: a two-pass procedure for saddlepoint approximation. submitted to *J. Royal Statist. Soc. B.* (1989).

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

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

[17] Reid, N. Saddlepoint methods and statistical inference. *Statist. Sc.* 3 (1988) 213-227.