Bayes Posteriors for Scalar Interest Parameters

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

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

SUMMARY

The posterior distribution for an interest parameter is usually obtained by integrating the nuisance parameters out of a full posterior distribution. For this, considerable attention has focussed on finding a prior that is noninformative for the interest parameter and yet allows the effective elimination of the nuisance parameters; of course this needs model structure beyond the likelihood function. As a paradigm for generalizing this, we take a location model for the interest parameter and a separate model for the nuisance parameter. Then for a general asymptotic model we use recent third order asymptotics to determine location model structure relative to a scalar interest parameter; we thus obtain a posterior for the interest parameter based on a flat or diffuse prior for only the interest parameter.

1. INTRODUCTION

Consider a full parameter $\theta = (\lambda, \psi)$ that is reexpressed in terms of a nuisance parameter λ and a scalar interest parameter ψ.

As a simple paradigm suppose that y_2 measures ψ with location model $g(y_2 - \psi)$ and separately y_1 measures λ with model $h(y_1 - \lambda)$. A standard flat or reference prior for ψ then gives the posterior $g(y_2 - \psi)$ for the interest parameter ψ.

Recent asymptotic theory leads to p-values for a scalar interest parameter that have third order accuracy. The theory can be extended to give third order likelihoods and then local location structure relative to the interest parameter. This thus produces a posterior for the scalar interest parameter using only a flat or diffuse prior for that interest parameter.

2. BACKGROUND

Consider a statistical model that is asymptotic (for example, DiCiccio, Fields & Fraser, 1990, Fraser & Reid, 1993) as some parameter $n \to \infty$, and suppose that $\theta = (\lambda, \psi)$ where ψ is a scalar interest parameter and λ is the nuisance parameter. Recent third-order likelihood-based asymptotics produces a significance function

$$p(\psi) = P(\hat{\psi} \leq \hat{\psi}^{\circ}; \psi) = \Phi(R, Q) + O(n^{-3/2}) \tag{1}$$

where R and Q are an observed signed likelihood ratio and an appropriately standardized maximum likelihood departure, and the function $\Phi(R, Q)$ can be given the Lugannani & Rice (1980) form

$$\Phi(R, Q) = \Phi(R) + \varphi(R)(R^{-1} - Q^{-1}) \tag{2}$$

or the Barndorff-Nielsen (1991) form

$$\Phi(R, Q) = \Phi\{R - R^{-1} \log(R/Q)\} \tag{3}$$

using the standard normal distribution and density functions $\Phi(z)$ and $\varphi(z)$.

For the case (Fraser & Reid, 1990, 1993) of a scalar full parameter $\theta = \psi$ the signed likelihood ratio R has the form

$$r = \text{sgn}(\hat{\psi} - \psi) \cdot [2\{\ell(\hat{\psi}; y) - \ell(\psi; y)\}]^{1/2} \tag{4}$$

using the log likelihood $\ell(\psi; y) = \log f(y; \psi)$ and the maximum likelihood departure Q has the form

$$q = (\hat{\varphi} - \varphi) \cdot \hat{\jmath}_{\varphi}^{1/2} \tag{5}$$

where φ is a notional parameter recording the gradient of the likelihood in a direction v at the observed data y°,

$$\varphi = \ell_{;v}(\psi; y^\circ) = \frac{d}{dv}\ell(\psi; y)\Big|_{y^\circ} , \tag{6}$$

and v is tangent at the data to an exact first derivative (at $\psi = \hat{\psi}^\circ$) location ancillary (Fraser, 1964; Fraser & Reid, 1990, 1993); the observed information for φ is available by recalibrating the observed information for ψ,

$$\hat{\jmath}_\varphi = -\ell_{\psi\psi}(\hat{\psi}^\circ; y^\circ)(\partial\varphi/\partial\psi\big|_{\hat{\psi}^\circ})^{-2} = \hat{\jmath}_\psi(\partial\varphi/\partial\psi\big|_{\hat{\psi}^\circ})^{-2} \tag{7}$$

with $\ell_{\psi\psi}(\psi; y) = (\partial^2/\partial\psi^2)\ell(\psi; y)$. If the model is exponential (Lugannani & Rice, 1980) with canonical parameter ψ, then φ as given in (6) is affine in ψ and it suffices to use ψ in place of φ in (5), (7). If the model is location (Fraser, 1990a; DiCiccio, Field & Fraser, 1990) with canonical parameter ψ, then φ as given in (6) becomes the negative score function $-\ell_\psi(\psi; y^\circ)$ and q becomes the standardized score $\ell_\psi(\psi; y^\circ)\hat{\jmath}_\psi^{-1/2}$. For the general scalar parameter case, the location ancillary direction (Fraser, 1964; Fraser & Reid, 1993) can also be used with an alternative to (5) given by Barndorff-Nielsen (1991).

For the general case (Fraser & Reid, 1990, 1993, 1994) with p dimensional parameter $\theta = (\lambda, \psi)$ and scalar interest parameter ψ, the likelihood ratio R is calculated from (4) using the profile likelihood $\ell_p(\psi) = \ell(\hat{\lambda}_\psi, \psi; y)$ where $\hat{\theta}_\psi = (\hat{\lambda}_\psi, \psi)$ is the constrained maximum likelihood estimate with ψ fixed. The maximum likelihood departure Q has the form

$$q = (\hat{\varphi}_{(p)} - \varphi_{(p)})\{|\hat{\jmath}_\varphi|/|j_{(\lambda)}(\hat{\theta}_\psi)\}|^{1/2} . \tag{8}$$

For this, φ is a notional full reparameterization obtained as the gradient of the likelihood in directions $V = (v_1, \ldots, v_p)$ at the data y°,

$$\varphi = \ell_{;V}(\theta; y^\circ) = \left\{\frac{d}{dv_1}\ell(\theta; y), \ldots, \frac{d}{dv_p}\ell(\theta; y)\right\}\Big|_{y^\circ} , \tag{9}$$

with $\mathcal{L}(V)$ tangent at y° to a location ancillary (Fraser, 1964; Fraser & Reid, 1990, 1993, 1994) for the full parameter; and $\varphi_{(p)} = \varphi u_\psi$ is a scalar reparameterization which is calculated (Fraser & Reid, 1990, 1993) using a unit vector u_ψ perpendicular to the ψ fixed curve at $\varphi(\hat{\theta}^\circ_\psi)$ on the φ space and pointing in the ψ increasing direction; and further $\hat{\jmath}_\varphi$ is the full information matrix for θ recalibrated in terms of φ at the overall maximum likelihood value $\hat{\theta}^\circ$ and $\jmath_{(\lambda)}(\hat{\theta}_\psi)$ is the nuisance information matrix recalibrated in terms of φ at the constrained maximum likelihood value.

In certain contexts we can encounter a nominal likelihood $\ell^\circ_p(\psi) = \ell^\circ_p(\psi; y^\circ)$ (often a profile likelihood) with signed likelihood ratio r given by (4) and standardized maximum likelihood departure q given by (5) relative to some reparameterization φ with corresponding information $\tilde{\jmath}_\varphi$ obtained from a φ tilt of the profile. The corresponding exponential model in saddlepoint form (Barndorff-Nielsen & Cox, 1979; Fraser & Reid, 1993) is given by the following with $a(\hat{\varphi}, \varphi) = 1$:

$$\begin{aligned} &\frac{c}{(2\pi)^{1/2}} \exp\{\ell^\circ(\psi) - \ell^\circ(\hat{\psi}) + s\varphi\} a(\hat{\varphi}, \varphi) \hat{\jmath}^{1/2} d\hat{\varphi} \\ &= \frac{c}{(2\pi)^{1/2}} \exp\{\ell^\circ(\psi) - \ell^\circ(\hat{\psi}) + s\varphi\} a(\hat{\varphi}, \varphi) \frac{r}{q} dr \end{aligned} \tag{10}$$

where $c = 1 + O(n^{-1})$ is constant to order $O(n^{-3/2})$. And further with $c(\hat{\varphi}, \varphi) = 1$ the left tail probability at the data point is given by (1) and $R = r$ and $Q = q$.

More generally we may have an adjustment factor $a(\hat{\varphi}, \varphi) = 1 + O(n^{-1/2})$ where

$$a(\hat{\varphi}, \varphi) = 1 + a_1(\hat{\varphi} - \varphi)/n^{1/2} + a_2(\hat{\varphi} - \varphi)^2/2n \tag{11}$$

in standardized coordinates where a constant of order $O(n^{-1})$ may be combined with the c in (10); see Cheah, Fraser & Reid (1994). Various forms for the left tail probability are then possible depending on combining the adjustment $a(\hat{\varphi}, \varphi)$ in various ways with the r and q. The most direct form corresponds to the double saddlepoint (Skovgaard, 1987) uses (1) with

$$R = r \quad , \quad Q = q/a \ . \tag{12}$$

An alternative corresponds to the sequential saddlepoint (Fraser, Reid & Wong, 1991) and calculates the signed likelihood ration R and standardized maximum likelihood Q from (4) and (6) using the modified likelihood $\tilde{\ell}_p(\psi) = \ell_p(\psi) + \log a(\hat{\varphi}, \varphi)$. Some comparisons in the exponential model context may be found in Pierce & Peters (1992) and Cheah, Fraser & Reid (1994). In particular this allows the ψ dependence appearing in the information factor of (8) to be moved to the likelihood factor.

3. LARGE SAMPLE MODEL STRUCTURE

Consider a general asymptotic model $f(y;\theta) = \exp\{\ell(\theta, y)\}$ as described at the beginning of Section 2 and in the paragraph preceding formula (8). The limiting form of the model involves an approximate third order ancillary with tangent vectors V at the data y° obtained by parameter forcing (Fraser, 1964; Fraser & Reid, 1994). The conditional model given the ancillary can be described to third order accuracy near the data y° by the p-dimensional tangent exponential model

$$\frac{c}{(2\pi)^{p/2}} \exp\{\ell^\circ(\theta) - \ell^\circ(\hat{\theta}^\circ) + (\varphi - \hat{\varphi}^\circ)s\} |\tilde{\jmath}_\varphi|^{-1/2} \, ds \tag{13}$$

with φ the notional parameter given by (9) and $s = \ell^\circ_{\varphi;}(\theta; y)|_{\hat{\theta}^\circ}$, the corresponding score variable. For testing a parameter value ψ, the unique (to third order) pivotal distribution is determined on the line $\hat{\theta}_\psi = \hat{\theta}^\circ_\psi$ in the space conditional on the ancillary; it has density

$$\frac{c'}{(2\pi)^{1/2}} \exp\{\ell^\circ_p(\varphi_{(p)}) + \varphi_{(p)} s_{(p)}\} |\tilde{\jmath}_\varphi|^{-1/2} |\tilde{\jmath}_{(\lambda)}(\hat{\theta}_\psi)|^{1/2} \, ds_{(p)} \tag{14}$$

where ℓ°_p is the profile relative to the scalar parameter $\varphi(p)$ and other components are described after (9). This is an adjusted exponential model (10) along a line in score space through the data y° and parallel to the unit vector u_ψ. The left tail probability is available from (1) using R and Q in the paragraph involving (8).

Formula (14) gives an observed log-likelihood for ψ,

$$\ell_a(\psi) = \ell(\hat{\lambda}_\psi, \psi; y^\circ) + \frac{1}{2} \log |\jmath_{(\lambda)}(\hat{\theta}_\psi)| \ , \tag{15}$$

relative to a score increment $ds_{(p)}$ in the direction u_ψ at the data point; the direction u_ψ can vary $O(n^{-1/2})$ with ψ. This is an example common with marginal likelihoods where the support volume measure has parameter dependence. If we replace V in $\mathcal{L}(V)$ by VB where B is $p \times p$, then we effectively replace φ by $V\varphi$ and this does not affect the R and Q for (1) but does affect $O(n^{-1/2})$ the expression (15).

We now calculate a new parameterization φ by a linear transformation as just described so as to obtain an identity observed information $\hat{\jmath}_\varphi = I$. The score density corresponding the maximum likelihood parameter value then has maximum density curvature (Cakmak, Fraser & Reid, 1994) equal to $I + O(n^{-1})$, and the curvature along the direction u_ψ then is equal to a constant to order $O(n^{-3/2})$. This removes the parameter dependence from the maximum likelihood density in the conditional model (14) and presents the model as an adjusted exponential (10) with the parameter dependence removed from the score variable.

4. POSTERIOR FOR THE INTEREST PARAMETER

Consider the general asymptotic model as described in Section 3 and let φ be the notional parameter with identity observed information. The likelihood for ψ is given by (15) as calculated from this parameter. The exponential model parameterization (Cakmak, Cheah, Fraser, Reid, & Tapia, 1993) for ψ in (14) is $\varphi_{(p)} = \varphi u_\psi = \alpha(\psi)$. The corresponding location model parameterization (ibid) is

$$\chi = \chi(\psi) = \int_0^\psi \frac{z(\alpha)}{q(\alpha)} d\alpha(\psi)$$

where $z(\alpha) = \ell_\alpha(\psi)\hat{\jmath}_\varphi^{-1/2}$ and $q(\alpha) = (\hat{\alpha} - \alpha)\hat{\jmath}_\varphi^{1/2}$ (with $\hat{\jmath}_\varphi = 1$) are the standardized score and maximum likelihood departure as based on the exponential parameterization.

Now let $\ell_*(\chi)$ be the likelihood (15) reexpressed in the parameterization $\chi = \chi(\psi)$ and $\pi(\psi)$ be a prior density for ψ. Then the posterior for ψ is

$$p(\psi|y^\circ)d\psi \propto \exp\{\ell_*(\chi)\}\pi\{\psi^{-1}(\chi)\}d\chi$$

and with a flat prior in terms of the location parameterization is

$$p(\psi|y^\circ)d\psi \propto \exp\{\ell_*(\chi)\}d\chi \ .$$

We have used recent asymptotics to determine a third order likelihood for a scalar interest component parameter, and have obtained a posterior for the interest parameter using a location reparameterization.

REFERENCES

Barndorff-Nielsen, O.E. (1991). Modified signed log likelihood ratio. *Biometrika* **78**, 557–563.

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

Cakmak, S., Cheah, P.K., Fraser, D.A.S., Reid, N. and Tapia, A. (1993). Third order asymptotic model: exponential and location type approximations, *Parisankhyan Samikkha*, to appear.

Cakmak, S., Fraser, D.A.S., and Reid, N. (1994). Multivariate asymptotic model: exponential and location approximations, *Utilitas Mathematica*, to appear.

Cheah, P.K., Fraser, D.A.S., and Reid, N. (1994). Adjustments to likelihoods and densities, submitted *Canadian J. Statist.*

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

Fraser, D.A.S. (1964). Local conditional sufficiency. *J.R. Statist. Soc. B* **26**, 52–62.

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

Fraser, D.A.S. and Reid, N. (1990). From multiparameter likelihood to tail probabilities for a scalar parameter. University of Toronto Technical Report, Department of Statistics.

Fraser, D.A.S. and Reid, N. (1993). Simple asymptotic connections between densities and cumulant generating function leading to accurate approximations for distribution functions. *Statist. Sinica* 3, 67-82.

Fraser, D.A.S. and Reid, N. (1994). Ancillaries and third order significance, *Utilitas Mathematica*, to appear.

Fraser, D.A.S., Reid, N. and Wong, A. (1991). From observed likelihood to tail area: a two pass procedure. *J. R. Statist. Soc. B* **52**, 483–492.

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

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

Pierce, D.A. and Peters, D. (1992). Practical use of higher order asymptotics for multiparameter exponential families, *J. Royal Statist. Soc. B* **54**, 701-737.

Skovgaard, I.M. (1987). Saddlepoint expansions for conditional distributions, *J. App. Prob.* **24**, 875-87.