Journal of Statistical Research *ISSN 0256 - 422 X*
2003, Vol. 37, No. 1, pp. 1-15
Bangladesh

p-VALUE FORMULAS FROM LIKELIHOOD ASYMPTOTICS BRIDGING THE SINGULARITIES:

D.A.S. Fraser, N. Reid, R. Li
Department of Statistics, University of Toronto,
Toronto, Canada M5S 3G3
Email: dfraser@utstat.toronto.edu, reid@utstat.toronto.edu, rongcai@utstat.toronto.edu

A. Wong
Department of Mathematics and Statistics
York University, Toronto, Canada M3J 1P3
Email: august@yorku.ca

SUMMARY

Recent likelihood asymptotics has produced highly accurate p-values for many very general contexts. The terminal formulas for producing these p-values can however have serious singularities in the neighbourhood of the maximum likelihood value. The singularities near the maximum likelihood value are downstream versions of those addressed by Daniels (1987) for the scalar saddlepoint context; he provided an approximate value at the singularity which involved a standardized third order cumulant. For a general statistical context we develop a third order bridge for the p-value formula at the maximum likelihood singularity in the case with no nuisance parameters, and a second order bridge at the singularity for the case with nuisance parameters. We also develop a third order graphical procedure for bridging the singularities; it handles both the cases without and with nuisance parameters.

Keywords: Place keywords here

AMS Classification: Place Classification here.

1 Introduction

Recent likelihood methods lead to highly accurate p-values for scalar parameters without and with nuisance parameters. These p-value formulas however can have serious computational singularities in the neighbourhood of the maximum likelihood value. We develop a third order bridge (3.1) or (3.2) that bypasses these singularities for the case without nuisance parameters and a second order bridge (4.5) for the case with nuisance parameters; the latter can however be difficult to compute. As an alternative, we develop a third order graphical procedure for bypassing the singularity. This allows a full presentation of a p-value graph even though in some applications the central portion of the graph may be just of secondary interest. For simulations however it is essential to be able to correctly handle central values; Also, the central portion is of prime interest if we want a point estimate based on the median p-value.

The saddlepoint method introduced to statistics by Daniels (1954) and Barndorff-Nielsen & Cox (1979) gives a highly accurate approximation for a density function with known cumulant generating function. Lugannani & Rice (1980) developed a corresponding distribution function approximation as an alternative to the numerical integration of the approximate density function. This distribution function approximation has a singularity at the saddlepoint, which can be replaced (Daniels 1987) by its limiting value, a multiple of a third order standardized cumulant. An alternative distribution function approximation was developed by Barndorff-Nielsen (1986) as part of extending results beyond the exponential model context; this also has a related singularity.

These distribution function approximations quite generally use two rather different inputs of information from likelihood. The first is almost always the signed square root r of the log likelihood ratio given below at (1.2).

The second is an appropriately defined maximum likelihood departure q; the search for the appropriate q has led to the progressively more general p-values now available from likelihood asymptotics. The two inputs are combined using one or other of the following two formulas and give a third order p-value for testing a scalar parameter value:

$$\Phi_{\mathrm{LR}}(r,q) = \Phi(r) + \phi(r)\left(\frac{1}{r} - \frac{1}{q}\right) \ , \ \Phi_{\mathrm{BN}}(r,q) = \Phi\{r - r^{-1}\log(r/q)\} \ . \qquad (1.1)$$

These are due to Lugannani & Rice (1980) and Barndorff-Nielsen (1986) respectively and were developed for specific model contexts; $\phi(r)$ and $\Phi(r)$ are the standard normal density and distribution functions.

In the cases we consider, r is the signed square root of the log likelihood ratio statistic,

$$r(\psi, y) = \mathrm{sgn}(\hat{\psi} - \psi)\Big[2\{\ell(\hat{\theta}; y) - \ell(\hat{\theta}_{\psi}; y)\}\Big]^{1/2} \qquad (1.2)$$

where $\ell(\theta; y)$ is the log likelihood, $\psi(\theta)$ is the scalar interest parameter with tested value ψ, and $\hat{\theta}$ and $\hat{\theta}_{\psi}$ are the maximum likelihood values without and with the constraint $\psi(\theta) = \psi$. The definition of q is less straightforward as it generally depends on more than just observed likelihood; several expressions are recorded below at (2.2), (2.10), and (4.1). Both r and q are standard normal to order $O(n^{-1/2})$, but with the appropriately defined q the p-values given by (1.1) are distributed under the model as uniform $(0, 1)$ to order $O(n^{-3/2})$.

We are assuming that we have a continuous statistical model $f(y; \theta)$ with dimensions n and p for y and θ, and that $\psi(\theta)$ is a scalar parameter of interest. The reduction in dimension from n to p is achieved in principle by conditioning on an approximate ancillary statistic, but in an application we need only the tangents to the approximate ancillary at the data point of interest. It is shown in Fraser & Reid (1995, 2000) that these tangents can be obtained from a full dimension pivotal quantity $z = z(y; \theta)$; needed background is recorded in Sections 2 and 3. The r and q in (1.1) are functions of ψ and y and are both close to zero when ψ is near $\hat{\psi}(y)$. This poses obvious numerical difficulties for the evaluation of $r^{-1} - q^{-1}$ and r/q; see Figure 3.1 for an example of numerical perturbations near the maximum likelihood value.

In Section 2 we record background on the asymptotic model with scalar variable and parameter; we define two measures (2.4) and (2.5) of how the asymptotic density of the signed likelihood ratio departs from the standard normal; and we obtain asymptotic expressions for the two measures. Then in Sections 3 to 5, we develop the bridging formulas and methods.

2 Background: Departures from Standard Normality

Consider first the case of an asymptotic model with scalar variable and scalar parameter. Many properties for the more general context can be derived from this case. In this section we record needed properties of the asymptotic model; the bridging formulas are then derived in Section 3.

We assume the model $f(y;\theta)$ leads to a log density $\ell(\theta;y) = \log f(y;\theta)$ that has the usual asymptotic properties, such as $\ell(\theta;Y) = O_p(n)$, var $\{\ell'(\theta;Y)\} = O(n)$ and so on. For testing a value θ, the needed likelihood quantity q is intrinsically based on a locally defined reparameterization,

$$\varphi(\theta;y^0) = \frac{\partial}{\partial y}\ell(\theta;y)\big|_{y^0} = \ell_{;y}(\theta;y^0) \ , \tag{2.1}$$

which gives the canonical parameter of the tangent exponential model at the data point y^0 of interest (Fraser, 1990). This is then used to form the maximum likelihood departure measure

$$\begin{aligned} q(\theta;y) &= (\hat{\varphi} - \varphi)\hat{\jmath}_{\varphi\varphi}^{1/2} \\ &= \{\ell_{;y}(\hat{\theta};y) - \ell_{;y}(\theta;y)\}\hat{\jmath}_{\theta\theta}^{1/2}\ell_{\theta;y}^{-1}(\hat{\theta};y) \end{aligned} \tag{2.2}$$

where $\hat{\jmath}_{\theta\theta} = -\ell''(\hat{\theta};y)$ is the observed information and $\ell_{\theta;y} = \partial^2\ell(\theta;y)/\partial\theta\partial y = \partial\varphi/\partial\theta$ evaluated at the maximum likelihood value $\hat{\theta}(y)$ adjusts the information from the θ scale to the φ scale. The likelihood based third order approximation to the density of the likelihood root $r(\theta;y)$ is then given as

$$\exp(k/n)\phi(r)(r/q)\,dr \tag{2.3}$$

where k/n is a constant to third order. This can be obtained by change of variable from y to r starting from the p^* formula (Barndorff-Nielsen, 1983) or starting from the saddlepoint approximation to the tangent exponential model (Fraser & Reid, 1995, 2000).

We examine how the third order likelihood based distribution for r differs from the nominal standard normal of first order theory and we can do this using either the density function (2.3) or the distribution function (1.1).

In terms of its functional form the expression (2.3) has the factor r/q attached to the basic standard density $\phi(r)$. The factor r/q is greater (or less) than 1 according as the tail of the density for r is thicker (or thinner) than that of the standard normal. As a measure of departure from standard normal we can examine how r/q exceeds 1 taken relative to r, or we can examine how the distribution function (1.1) falls short of the nominal $\Phi(r)$ as taken relative to the density $\phi(r)$. In either case, we obtain

$$d_1(r) = \frac{1}{r}\left(\frac{r}{q} - 1\right) = \frac{\Phi(r) - \Phi_{LR}(r,q)}{\phi(r)} = \frac{1}{q} - \frac{1}{r} \ . \tag{2.4}$$

For an alternative measure we examine the argument of the second distribution function approximation (1.1); the argument is usually recorded as r^*. We consider how it falls short of the nominal normal deviate r:

$$d_2(r) = r - r^* = r^{-1}\log(r/q) \ . \tag{2.5}$$

Now consider the asymptotic form of these departure measures as a means to bridge the singularity at the maximum likelihood point. Taylor series expansion methods were used in Cakmak et al (1998) to determine the local form of a statistical model for a fixed data point, say y^0. These results are used in Appendix A to give asymptotic expressions for d_1 and d_2 for fixed data $y = y^0$ and varying θ:

$$\begin{aligned} d_1 &= -\frac{\alpha_3}{6n^{1/2}} + \frac{\alpha_4 - \alpha_3^2}{24n} r \ , \\ d_2 &= -\frac{\alpha_3}{6n^{1/2}} + \frac{3\alpha_4 - 4\alpha_3^2}{72n} r \ . \end{aligned} \tag{2.6}$$

For this α_3 and α_4 are standardized third and fourth derivatives of the log density $\ell(\varphi; x)$ with respect to φ at $\{\varphi(\hat{\theta}^0), x^0\}$ and $\varphi = \varphi(\theta)$ and $x = x(y)$ are local reexpressions of θ and y that are used to record the tangent exponential model approximation at the data point y^0. Explicit expressions for $\varphi(\theta)$ and $x(y)$ are recorded in Andrews, Fraser, and Wong (2002).

In a parallel way Abebe et al (1995) determined the local form of the model relative to a particular parameter value, say θ_0. This is used in Appendix A to derive asymptotic expressions for d_1 and d_2 for fixed $\theta_0 = \theta$ and varying y:

$$\begin{aligned} d_1 &= -\frac{a_3}{6n^{1/2}} - \frac{3a_4 + 4a_3^2 + 6c}{24n} r \\ d_2 &= -\frac{a_3}{6n^{1/2}} - \frac{9a_4 + 13a_3^2 + 18c}{72n} r \ . \end{aligned} \tag{2.7}$$

The a_3 and a_4 are standardized third and fourth derivatives of $\ell(\varphi; x)$ with respect to x at $\{\varphi(\theta_0), \hat{x}(\theta_0)\}$ where $\hat{x}(\theta_0)$ is the maximum density point for θ_0, and $\varphi = \varphi(\theta)$ and $x = x(y)$

are the local reexpressions of θ and y that are used to record the tangent exponential model for fixed parameter value θ_0; the constant c is a measure of nonexponentiality and is given as $\partial^4\ell/\partial\varphi^2\partial x^2$ evaluated at $\{\varphi(\theta_0), \hat{x}(\theta_0)\}$.

If $\theta_0 = \hat{\theta}(y_0)$ or if $y_0 = \hat{y}(\theta_0)$ then the expansion coefficients are linked by the norming property which gives $a_3 = \alpha_3 + O(n^{-1})$, $a_4 = \alpha_4 - 3\alpha_3^2 - 6c + O(n^{-1/2})$. This does not give equality of (2.6) and (2.7) as the two versions describe dependence on r for fixed value of y and for fixed value of θ respectively. The expressions for the d_i are accurate to $O(n^{-3/2})$.

Some clarity on the roles for the two versions (2.6) and (2.7) arises by noting that r and q are functions of y and θ for a moderate deviations range from some initial y_0 or θ_0 of interest. Along the curve $C = \{(\theta, y) : \theta = \hat{\theta}(y)\}$ we have $r = q = 0$, and to first derivative we have $r = q$. The departure measures are then describing how r and q differ beyond the first derivative. We could have started with a point (θ_0, y_0) and some particular value for $r = r(\theta_0, y_0)$ and then used (2.6) to examine change in r for fixed y_0 or used (2.7) to examine change in r for fixed θ_0. For this we note that the α_3, α_4 would be values determined on C with the particular y_0, and a_3, a_4, c would be values determined on C with the particular θ_0. The details for this may be found in Li (2001).

Example 2.1. Cauchy location model.

Consider the location Cauchy $f(y-\theta) = \pi^{-1}\left\{(1+(y-\theta)^2\right\}^{-1}$ with $n = 1$. For this we have $\hat{\theta} = y$ and

$$\begin{aligned} r &= \operatorname{sgn}(y-\theta)[2\log\{1+(y-\theta)^2\}]^{1/2} , \\ q &= \sqrt{2}(y-\theta)/\{1+(y-\theta)^2\} . \end{aligned}$$

The exponential parameter can be standardized, $\varphi = \sqrt{2}\theta/(1+\theta^2)$, giving

$$\ell(\varphi) = -\frac{\varphi^2}{2} - \frac{3}{8}\varphi^4 ,$$

from which we obtain $\alpha_3 = 0$, $\alpha_4 = 9$, and thus $d_1 = d_2 = \frac{3}{8}r$; note that φ is redescending but this anomaly seems not to bother the calculation of p-values for this extreme application of likelihood theory (Fraser, 1990) with $n = 1$. The lack of skewness in the model leads to the equality of the two departure measure.

More generally when the parameter θ is a scalar but the observable variable has dimension n, we define a vector v by

$$v = -z_{y'}^{-1} z_{;\theta}\big|_{(y^0, \hat{\theta}^0)} \tag{2.8}$$

where $z = z(y, \theta)$ is an $n \times 1$ vector of natural pivotal quantities. As shown in Fraser & Reid (1995), this vector can be used to define a canonical parametrization φ for the original model; then defining q as the standardized maximum likelihood departure in this parametrization ensures that the expressions in (1.1) are third order approximations to the p-value conditional on an approximately ancillary statistic. The dimension reduction from n to 1 is achieved by conditioning on an approximate ancillary statistic, but this ancillary is not explicitly needed, just the gradient of ℓ in the directions (2.8) of the ancillary at the data point.

Using v, the reparameterization φ in (2.1) is generalized to

$$\varphi(\theta; y^0) = \frac{d}{dv}\ell(\theta; y)\big|_{y^0} = \ell_{;y'}(\theta; y^0)v \ , \tag{2.9}$$

and the generalized version of q becomes

$$q(\theta; y) = \{\ell_{;v}(\hat{\theta}; y) - \ell_{;v}(\theta; y)\}\hat{\jmath}_{\theta\theta}^{1/2}\ell_{\theta;v}^{-1}(\hat{\theta}; y). \tag{2.10}$$

In this more general context the expressions (2.6) for the departure measures remain available, but the versions (2.7) for varying data point are typically not accessible as they would need model information along the contour of the observed approximate ancillary.

3 Bridging the Singularity: Without Nuisance Parameter

The measures of departure developed in the preceding section provide a simple and direct means for bridging the maximum likelihood singularity in the p-value formulas. ¿From (2.4), (2.5) and (1.1) we obtain

$$p_1(\theta) = \Phi(r) - d_1\phi(r) = \Phi(r) + \left(\frac{\alpha_3}{6n^{1/2}} - \frac{\alpha_4 - \alpha_3^2}{24n}r\right)\phi(r) \ , \tag{3.1}$$

$$p_2(\theta) = \Phi(r - d_2) = \Phi\left(r + \frac{\alpha_3}{6n^{1/2}} - \frac{3\alpha_4 - 4\alpha_3^2}{72n}r\right) \ . \tag{3.2}$$

These can be viewed as Bartlett type corrections to the likelihood ratio but are derived from observed likelihood. At the maximum likelihood value where $r = 0$, we have that (3.1) reduces to

$$p_1(\theta) = \Phi(0) + \frac{\alpha_3}{6n^{1/2}}\phi(0)$$

which agrees with Daniels(1987) for the exponential model case, but is now applicable in wide generality with the extended definition of α_3.

Example 3.1. Cauchy location model.

Consider the location Cauchy model with data $y = 0$, as examined in Example 2.1. From the two bridging formulas we obtain

$$p_1(\theta) = \Phi(r) - \frac{3}{8}r\phi(r), \quad p_2(\theta) = \Phi\left(\frac{5}{8}r\right) \ .$$

The exact p-value is of course available as

$$p(\theta) = .5 + \pi^{-1}\tan^{-1}\left\{\pm\left(e^{r^2/2} - 1\right)^{1/2}\right\} \ .$$

At $r = 0$ all three are equal to 0.5. At a point close to $r = 0$ we check numerically: at $r = 0.1$ we have

$$p_1 = .524942, \quad p_2 = .524918, \quad p = .522499.$$

The rather small departure of the approximations from the exact is of course due to the almost impossibly small sample size $n = 1$ and to the sharp peak at the centre of the Cauchy model.

For the bridging formulas (3.1) and (3.2) we could have considered a full Taylor series expansions in r but there are advantages to retaining the $\phi(r)$ and $\Phi(r)$ which reflect the dominant role of the signed likelihood ratio r.

Example 3.2. Consider the simple gamma model on the positive axis,

$$f(y;\theta) = \Gamma^{-1}(\theta) y^{\theta-1} e^{-y} \ ,$$

and the data point $y = 10$. The significance function $p(\theta)$ is plotted in Figure 3.1. Note the computational irregularities near the maximum likelihood value $\hat{\theta} = 10.495838$; of course, doubling the precision would limit the domain of irregularity. Simple calculations give $\alpha_3 = -0.315901$ and $\alpha_4 = 0.199422$ from which we obtain

$$d_2 = 0.0526502 + .00276517r \ ,$$

which gives the bridge

$$p_B(\theta) = \Phi(0.9972348r - 0.0526502) \ .$$

The likelihood based p-values $\Phi_{\rm LR}(\theta)$, $\Phi_{\rm BN}(\theta)$ are plotted in Figure 3.1 together with the bridge $p_B(\theta)$ and the exact $p(\theta)$; the bridge is seen to be superimposed on the exact. A simple algorithm can choose between the approximation and the bridge.

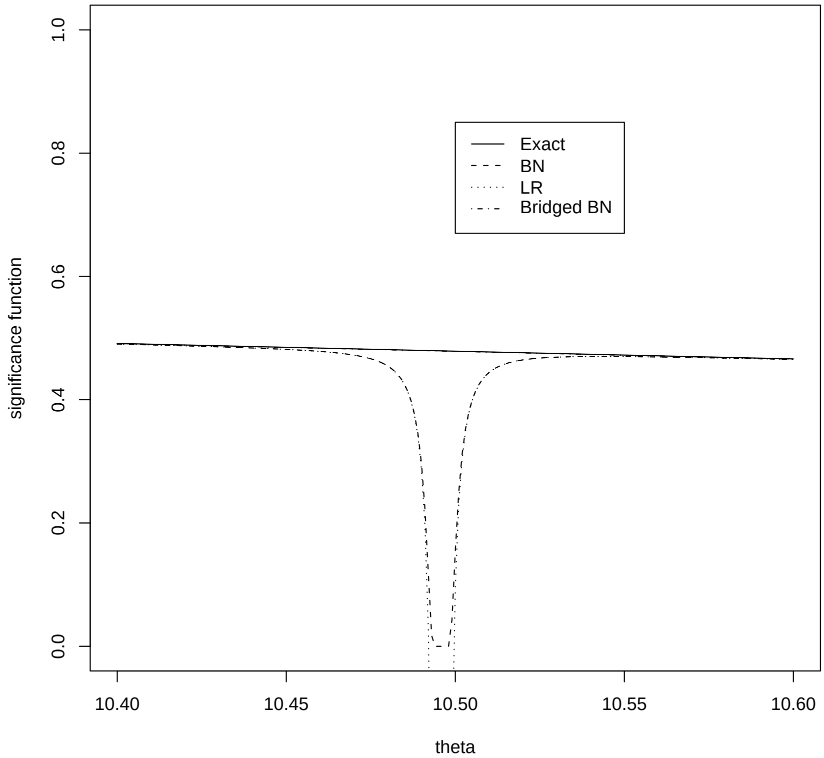

Figure 3.1. The gamma model $\Gamma^{-1}(\theta)y^{\theta-1}$ with $y^0 = 10$. The asymptotic approximations $\Phi_{\rm LR}(\theta)$ and $\Phi_{\rm BN}(\theta)$ for the p-value $p(\theta)$ for testing θ are plotted against θ. The bridge $p_B(\theta)$ at the maximum likelihood is visually identical to the exact $p(\theta)$.

4 Bridging the Singularity: With Nuisance Parameters

Consider now a continuous statistical model $f(y;\theta)$ with dimensions n and p for y and θ and let $\psi(\theta)$ be a scalar interest parameter with $\theta' = (\lambda', \psi)$. Again there is an approximate ancillary with tangent vectors $V = (v_1 \cdots v_p)$ given as

$$V = -z_{y'}^{-1} z_{;\theta'}\big|_{(y^0,\hat{\theta}^0)} \ ,$$

where $z = z(y,\theta)$ is now an $n \times 1$ vector of natural pivotal quantities. The exponential type parameter is

$$\varphi'(\theta; y^0) = \frac{d}{dV}\ell(\theta; y)\big|_{y^0} = \ell_{;y'}(\theta; y^0)V$$

and the gradient d/dV gives a row vector of directional derivatives of log likelihood function; for some discussion and examples see Fraser, Wong & Wu (1999). For testing $\psi(\theta) = \psi$ using (1.1), the r is given by (1.2) and the q by the following extension (Fraser, Reid & Wu, 1999) of (2.2) and (2.10),

$$q = \text{sgn}(\hat{\psi} - \psi)(\hat{\chi} - \hat{\chi}_\psi)\frac{|\hat{\jmath}_{\varphi\varphi}|^{1/2}}{|\jmath_{(\lambda\lambda)}(\hat{\varphi}_\psi)|^{1/2}} \ , \tag{4.1}$$

where the numerator and denominator determinants are the full and nuisance information determinants recalibrated on the φ scale and $\chi(\varphi) = u'_\psi \varphi$ is a rotated φ coordinate based on a unit vector

$$u_\psi = \frac{\psi_\varphi(\hat{\varphi}_\psi)}{|\psi_\varphi(\hat{\varphi}_\psi)|}$$

perpendicular to $\psi\{\theta(\varphi)\}$ at the constrained maximum likelihood value $\hat{\varphi}_\psi$.

For bridging the discontinuity at the maximum likelihood value $\psi(\hat{\theta}) = \hat{\psi}$, the calculations are more complex and we temporarily restrict our attention to $O(n^{-1})$ accuracy. Even with this the results may not easily be computed and we describe a relatively simple graphical procedure in the next Section 5. Let $\ell(\varphi) = \ell^0\{\theta(\varphi); y^0\}$ be the observed likelihood reexpressed in terms of φ and suppose it has been normed and recentered and rescaled so that $\hat{\varphi}^0 = 0$, $\ell(0) = \ell_\varphi(0) = 0$, and $\ell_{\varphi\varphi'} = -I$. Then in tensor summation notation we have

$$\ell(\varphi) = -\frac{1}{2}\alpha^{ij}\varphi_i\varphi_j - \frac{1}{6n^{1/2}}\alpha^{ijk}\varphi_i\varphi_j\varphi_k \tag{4.2}$$

to second order. Also for convenience we restrict attention to the case with $p = 2$. For the scalar interest parameter $\psi\{\theta(\varphi)\}$ we suppose that the φ coordinates have been rotated so that $\psi(\varphi) = \hat{\psi}$ is tangent to $\varphi_1 = 0$ and has been relocated and rescaled so that $\psi = 0$, $\partial\psi/\partial\varphi_1 = 1$, $\partial^2\psi/\partial\varphi_1^2 = 0$ at the maximum likelihood value $\varphi = 0$; then $\psi(\varphi) = \varphi_1 + a_{12}\varphi_1\varphi_2/n^{1/2} + a_{22}\varphi_2^2/(2n^{1/2})$ where the coefficients are second derivatives measuring the curvature of $\psi(\varphi)$ at $\varphi = 0$.

The signed likelihood ratio for testing ψ can be calculated to the second order giving

$$r = -\psi - \frac{\alpha^{111}\psi^2}{6n^{1/2}} \ . \tag{4.3}$$

The maximum likelihood departure (4.1) uses a unit vector u_ψ which is the first coordinate vector at $\psi = 0$ and locally can change direction by $O(n^{-1/2})$; the departure $\hat{\chi} - \hat{\chi}_\psi = -\psi$ to second order based on the cosine of an $O(n^{-1/2})$ angle. The nuisance information is $\hat{\jmath}_{(\lambda\lambda)}(\hat{\theta}_\psi) = 1 + \alpha^{122}\psi/n^{1/2} - a_{22}\psi n^{1/2}$. It follows that

$$q = -\psi\left(1 - \frac{\alpha^{122}\psi}{2n^{1/2}} + \frac{a_{22}\psi}{2n^{1/2}}\right) \ . \tag{4.4}$$

Combining (4.4) and (4.3) we obtain

$$q = r\left\{1 + (\alpha^{122} - a_{22} + \alpha^{111}/3)\frac{r}{2n^{1/2}}\right\}$$

from which it follows that $d = -(\alpha^{111} + 3\alpha^{122} - 3a_{22})/6n^{1/2}$. The bridging p-value formula is then

$$p(\psi) = \Phi(r) - d\phi(r) = \Phi(r - d) \tag{4.5}$$

to second order.

5 Graphical Bridging of the Singularity

For the case of a scalar full parameter we have seen in Section 2 that the departure measures (2.4) and (2.5) are linear in r to the third order and thus provide simple third order bridging, using (3.1) and (3.2). For the more general p-dimensional full parameter ψ we have from Section 4 that the departure measures are constant (4.4) to the second order. The development (Fraser & Reid, 1995, 2000) of the p-value formulas from tangent exponential model approximations records the p-value as a tail probability obtained from an adjusted asymptotic density; and Cheah et al (1995) show that such an adjusted density is itself an asymptotic model. Together these show that the departure measures

$$d_1 = q^{-1} - r^{-1} \ , \quad d_2 = r^{-1} \log(r/q) \tag{5.1}$$

are asymptotically linear in r to the third order under parameter change for fixed data. This is of course consistent with the familiar location-scale standardization of the signed likelihood ratio which gives a third order standard normal variable.

Now consider a particular assessment of a parameter ψ with given data together with possible instability in the computation with the p-value formulas (1.1). We propose plotting d_1 or d_2 against the signed likelihood ratio r. Any instability in the p-value formulas will show in d_1 and d_2, as $\Phi(r)$ is typically smooth. Accordingly we recommend fitting a line to d_1 or d_2 plotted against r, excluding the middle possibly unstable values. The fitted d_1 or d_2 is then used with (3.1) and (3.2) to bridge the singularity.

Example 5.1. Consider the gamma model with mean μ and shape parameter β,

$$f(y; \mu, \beta) = \Gamma^{-1}(\beta) \left(\frac{\beta}{\mu}\right)^{\beta} y^{\beta-1} e^{-\beta y/\mu} \ ,$$

and with data from Fraser, Reid and Wong (1997). For testing the parameter μ we record the approximations $\Phi_{\mathrm{LR}}(\mu)$ and $\Phi_{BN}(\mu)$ in Figure 5.1.

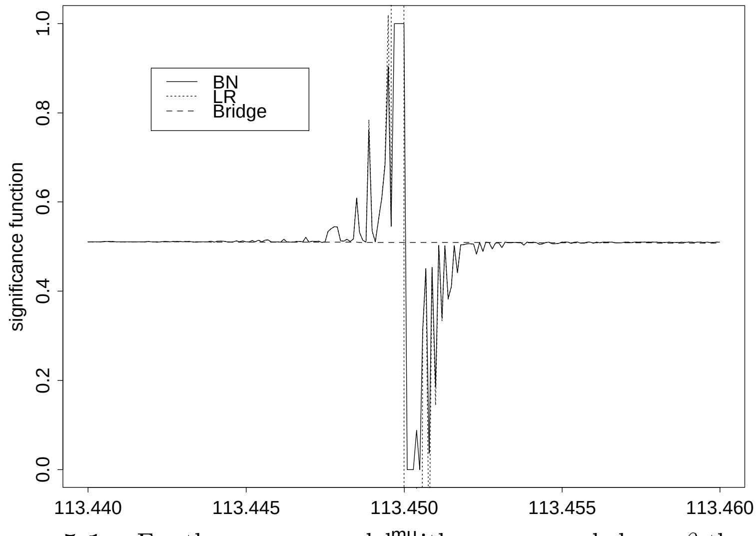

Figure 5.1. For the gamma model with mean μ and shape β the p-value approximations $\Phi_{\mathrm{LR}}(\mu)$ and $\Phi_{\mathrm{BN}}(\mu)$ for testing μ are plotted for a sample of 20. The aberrant behavior at the maximum likelihood value is successfully bridged using (3.2) together with a graphical d_2 determined from Figure 5.2.

Note the aberrant behaviour near the maximum likelihood value $\hat{\mu}^0 = 113.45$. For bridging the $\hat{\mu}^0$ value we plot d_1 and d_2 from (5.1) against the likelihood ratio r, in Figure 5.2. The bridging p-value is then obtained using (3.2) together with the straight line join obtained from Figure 5.2 and is recorded in Figure 5.1.

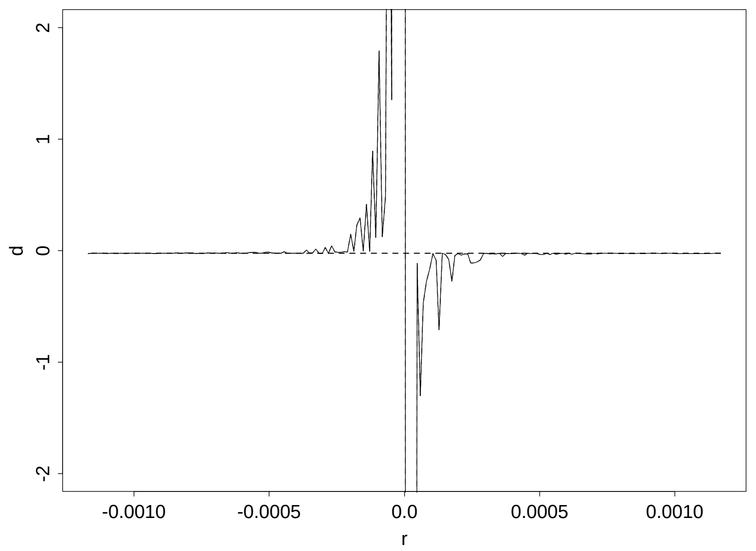

Figure 5.2. For the gamma model and data for Figure 5.1, the departure measures d_1 and d_2 are plotted against the signed likelihood ratio r and a bridging straight line is obtained graphically. As d_1 and d_2 are so close that they overlap in this figure.

Acknowledgements

This research was partially supported by the Natural Sciences and Engineering Research Council of Canada. The authors wish to acknowledge stimulating and valuable discussion with Nicola Sartori from the University of Padova while he was visiting the University of Toronto.

APPENDIX A

Derivation of (2.6) and (2.7)

In the case that both y and θ are scalars we can examine the asymptotic form of the departure $d(r,q)$ as a function of r and q by expanding the log density $\ell(\theta;y) = \log f(y;\theta)$ about a reference point (θ_0, y_0) in terms of reexpressed deviations for y and for θ. A data-oriented expansion (Cakmak et al, 1998) uses $(\theta_0, y_0) = \big(\hat{\theta}(y_0), y_0\big)$ and standardizes with respect to coefficients of θ^2 and θy. If the departures are then reexpressed towards exponential model form to the third order we obtain the log likelihood at y_0 and the log density at θ_0 given respectively by

$$\ell(\theta;y_0) = -\frac{1}{2}\theta^2 - \frac{\alpha_3}{6n^{1/2}}\theta^3 - \frac{\alpha_4}{24n}\theta^4, \tag{A.1}$$

and

$$\ell(\theta_0;y) = a + \frac{k_1}{n} - \frac{\alpha_3}{2n^{1/2}}y - \frac{1}{2}\left\{1 + \frac{\alpha_4 - 2\alpha_3^2 - 5\gamma}{2n}\right\}y^2 + \frac{\alpha_3}{6n^{1/2}}y^3 + \frac{\alpha_4 - 3\alpha_3^2 - 6\gamma}{24n}y^4. \tag{A.2}$$

The only nonzero mixed derivative terms to third order are $y\theta$ and $\gamma y^2\theta^2/4n$. The constant $a = -\log(2\pi)/2$. The α_3 and α_4 are the standardized cumulants of the null density, and γ is a measure of nonexponentiality; these are intrinsic parameters describing shape characteristics of the model. For some recent details see Andrews, Fraser & Wong (2002). From this an expansion for q in terms of r for fixed $y = y_0$ is obtained

$$q = r + \frac{\alpha_3}{6n^{1/2}}r^2 + \frac{5\alpha_3^2 - 3\alpha_4}{72n}r^3 \ ,$$

which gives

$$\frac{1}{q} = \frac{1}{r}\left\{1 - \frac{\alpha_3}{6n^{1/2}}r + \frac{\alpha_4 - \alpha_3^2}{24n}r^2\right\} \ .$$

These results give an expression for the two nonnormality measures with y fixed,

$$\begin{aligned} d_1 &= -\frac{\alpha_3}{6n^{1/2}} + \frac{\alpha_4 - \alpha_3^2}{24n}r \\ d_2 &= -\frac{\alpha_3}{6n^{1/2}} + \frac{3\alpha_4 - 4\alpha_3^2}{72n}r \ . \end{aligned}$$

$$(A.3)$$

For a parameter-oriented expansion (Abebe et al, 1995) we can use $(\theta_0, y_0) = \big(\theta_0, \hat{y}(\theta_0)\big)$ and standardize with respect to coefficients of y^2 and $y\theta$. If the departures are then reexpressed towards exponential form we obtain the log density at θ_0 and the log likelihood at y_0 given respectively by

$$a + \frac{k_2}{n} - \frac{1}{2}y^2 + \frac{a_3}{6n^{1/2}}y^3 + \frac{a_4}{24n}y^4 \ , \qquad (A.4)$$

and

$$-\frac{a_3}{2n^{1/2}}\theta - \frac{1}{2}\left\{1 + \frac{a_4 + 2a_3^2 + c}{2n}\right\}\theta^2 - \frac{a^3}{6n^{1/2}}\theta^3 - \frac{a_4 + 3a_3^2 + 6c}{24n}\theta^4 \qquad (A.5)$$

together with the mixed derivative terms $y\theta$ and $cy^2\theta^2/4n$. From this an expansion for q in terms of r for fixed $\theta = \theta_0$ is obtained,

$$q = r + \frac{a_3}{6n^{1/2}}r^2 + \frac{9a_4 + 14a_3^2 + 18c}{72n}r^3 \ ,$$

which gives

$$\frac{1}{q} = \frac{1}{r}\left\{1 - \frac{a_3}{6n^{1/2}}r - \frac{3a_4 + 4a_3^2 + 6c}{24n}r^2\right\} \ .$$

These results give an expression for the two nonnormality measures with θ held fixed:

$$\begin{aligned} d_1 &= -\frac{a_3}{6n^{1/2}} - \frac{3a_4 + 4a_3^2 + 6c}{24n}r \\ d_2 &= -\frac{a_3}{6n^{1/2}} - \frac{9a_4 + 13a_3^2 + 18c}{72n}r \ . \end{aligned}$$

$$(A.6)$$

The two expressions for d can be interrelated by taking the model (A.1,2) centered at $(\hat{\theta}^0, y^0)$ and reexpressing it in the form (A.4,5). For this we take $\theta_0 = \hat{\theta}^0$ which is zero in (A.2), and obtain $\hat{y}(\theta_0) = \alpha_3/2n^{1/2} + O(n^{-1})$; this then gives $a_3 = \alpha_3$, to order $O(n^{-1})$ and $a_4 = \alpha_4 - 3\alpha_3^2 - 6c$ to order $O(n^{-1/2})$. The constants k_1 and k_2 check under the reexpressions. We can then record (A.6) as

$$\begin{aligned} d_1 &= -\frac{\alpha_3}{6n^{1/2}} - \frac{3\alpha_4 - 5\alpha_3^2 - 12\gamma}{24n}r \\ d_2 &= -\frac{\alpha_3}{6n^{1/2}} - \frac{9\alpha_4 - 14\alpha_3^2 - 36\gamma}{72n}r \ . \end{aligned}$$

$$(A.7)$$

Formulas (A.3) and (A.7) both use standardized likelihood cumulants α_3, α_4 for a point (y_0, θ_0) with $r = 0$; they record the change in d for fixed y_0 and for fixed θ_0 respectively.

References

[1] Abebe, F., Cakmak, S., Cheah, P.K., Fraser, D.A.S., Kuhn, J., and Reid, N. (1995). Third order asymptotic model: Exponential and location type approximations. *Parisankhyan Samikhha* **2**, 25-34.

[2] Andrews, D.F., Fraser, D.A.S. and Wong, A. (2002). Computation of distribution functions from likelihood information near observed data. Technical Report, Department of Statistics, University of Toronto.

[3] Barndorff-Nielsen, O.E. (1983). On a formula for the distribution of the maximum likelihood estimate. *Biometrika* **70**, 343-365.

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

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

[6] Cakmak, S., Fraser, D.A.S., McDunnough, P., Reid, N. and Yuan, X. (1998). Likelihood centered asymptotic model: exponential and location model version. *J. Statist. Plann. Inf.* **66**, 211-222. Some corrections may be accessed through <http://www.utstat.toronto.edu/dfraser/research.html>.

[7] Cheah, P.K. Fraser, D.A.S. and Reid, N. (1995). Adjustments to likelihood and densities: calculating significance. *J. Statist. Res.* **29**, 1-13.

[8] Daniels, H.E. (1954). Saddlepoint approximations in statistics. *Ann. Math. Statist.* **25**, 631-650.

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

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

[11] Fraser, D.A.S. and Reid, N. (1995). Ancillaries and third order significance. *Utilitas Mathematica* **47**, 39-53.

[12] Fraser, D.A.S. and Reid, N. (2000). Ancillary information for statistical inference. *Empirical Bayes and Likelihood Inference*, Eds: S.E. Ahmed and N. Reid, Springer-Verlag, NY. 185-207.

[13] Fraser, D.A.S., Reid, N. and Wong, A. (1997). Simple and accurate inference for the mean of the gamma model, *Canad. Jour. Statist.* **25**, 91-99.

[14] Fraser, D.A.S., Reid, N. and Wu, J (1999). A simple general formula for tail probabilities for frequentist and Bayesian inference. *Biometrika* **86**, 249-264.

[15] Fraser, D.A.S., Wong, A. and Wu, J. (1999). Regression analysis, Nonlinear or nonnormal: simple and accurate p-values from likelihood analysis. *J. Am. Statist. Assoc.* **94**, 1286-1295.

[16] Li, R. (2001). On Asymptotic Likelihood Inference: Removing P-Value Singularities. *Ph.D. Thesis, Department of Statistics, University of Toronto.*

[17] Lugannani, R. and Rice, S. (1980). Saddlepoint approximation for the distribution function of the sum of independent variables. *Adv. Appl. Prob.* **12**, 475-490.

[18] Pierce, D.A. and Peters, D. (1994). Practical use of higher order asymptotics for multiparameter exponential models. *J. R. Statist. Soc. B* **54**, 701-738.

Journal of Statistical Research *ISSN 0256 - 422 X*
2003, Vol. 37, No. 1, pp. 17-19
Bangladesh

ON THE PROPRIETY OF POSTERIORS FOR PROPORTIONAL HAZARDS MODELS

Gauri S. Datta
University of Georgia

Malay Ghosh
University of Florida
Email: ghoshm@stat.ufl.edu

Bani K. Mallick
Texas A & M University

SUMMARY

The paper considers the propriety of posteriors for proportional hazards models with exponential baseline hazards.

Keywords: Baseline hazard, exponential, gamma prior, simple regression, uniform prior.

AMS Classification: 62C10

1 Introduction

Recent years have witnessed a surge of interest in the Bayesian analysis of Cox's (1972) proportional hazards models. A very succinct account of the same is given in Ibrahim, Chen and Sinha (2001). Although much of the Bayesian literature is devoted to semiparametric inference for these models, some authors (e.g. Dellaportas and Smith, 1993) have considered fully parametric Bayesian analysis of the same. In such cases, one common choice is a constant baseline hazard rate (Dellaportas and Smith, 1993; Ibrahim, Chen and Sinha, 2001) which amounts to an exponential baseline distribution. Typically, a conjugate gamma prior is assigned to the constant hazard rate. For the regression vector, on the other hand, an oft recommended prior is either the multivariate normal or uniform over the appropriate dimensional Euclidean space. While the former always ensures a proper posterior, the same cannot necessarily be said of the latter when the uniform distributions have infinite support. However, use of the latter is not uncommon (e.g. Dellaportas and Smith, 1993; Ibrahim, Chen and Sinha, 2001). Thus, it is of interest to find conditions under which the resulting posteriors are proper, so that inference based on them becomes meaningful. To our knowledge, this propriety issue has not been discussed elsewhere, in spite of its importance in Bayesian inference.

The paper addresses this problem in the special case of a simple linear regression, i.e. when the regression vector is one-dimensional. As stated in the preceding paragraph, the baseline distribution is assumed to be exponential. We have provided necessary and sufficient conditions for the propriety of posteriors in Section 2. Section 3 contains some concluding remarks.