215:2001mar21

ON BRIDGING THE SINGULARITIES OF p-VALUE FORMULAS FROM LIKELIHOOD ANALYSIS

By

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

and

A. Wong
Department of Mathematics and Statistics, York University,
Toronto, Canada M3J 1P3
august@mathstat.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 certain singularities, at the maximum likelihood value and at extremes of the range of the parameter being assessed. The singularity at the maximum likelihood value is a downstream version of one 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, for 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 which handles cases both without and with nuisance parameters. The combining formulas can also produce p-values outside the allowable $[0, 1]$ range. Several alternative combining formulas are developed that avoid these improper p-values. Simulations examine reliability and accuracy.

1. INTRODUCTION

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) used the saddlepoint method to develop a distribution function approximation as an alternative to numerical integration of the approximate density function. The Lugannani & Rice (1980) 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. Barndorff-Nielsen (1986) developed an alternative distribution function approximation as part of extending results beyond the exponential model context.

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 (1.3). The second is some appropriately defined maximum likelihood departure q; the search for the appropriate q has been the recent focus for obtaining progressively more general p-values in likelihood asymptotics.

These two inputs are combined using either of the following two formulas to give a third order p-value for testing a scalar parameter value:

$$\Phi_{\mathrm{LR}}(r,q) = \Phi(r) + \phi(r)\left(\frac{1}{r} - \frac{1}{q}\right) , \tag{1.1}$$

$$\Phi_{\mathrm{BN}}(r,q) = \Phi\{r - r^{-1}\log(r/q)\} , \tag{1.2}$$

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

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

$$r(\psi, y) = \mathrm{sgn}(\hat{\psi} - \psi)\left[2\{\ell(\hat{\theta}; y) - \ell(\hat{\theta}_{\psi}; y)\}\right]^{1/2} \tag{1.3}$$

where $\ell(\theta; y)$ is 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 typically depends on more than just observed likelihood; several expressions are recorded below at (2.2), (2.11), (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) and (1.2) 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 statistics, but only the tangents to the approximate ancillary are needed 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 $z = z(y;\theta)$; related details are recorded in Sections 2 and 3.

The r and q in (1.1) and (1.2) 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. Also, for extreme values of ψ, the second term in (1.1) can overwhelm the tail probability from the first term and give a value outside the acceptable range $[0,1]$ for a p-value; see Figure 2 for an example where a range of values less than zero are recorded for large values of the parameter.

In Section 2 we examine an asymptotic model with scalar variable and parameter and define two measures of how the asymptotic density of the signed likelihood ratio departs from the standard normal; asymptotic expressions are obtained for the measures and they are seen to be essentially equivalent.

In Section 3 we examine the scalar parameter case and use the two measures of departure to develop a third order bridge for the singularity at the maximum likelihood value.

In Section 4 we examine the case of vector full parameter θ with scalar interest parameter $\psi = \psi(\theta)$. The techniques from Section 2 are then used to develop a second order bridge at the maximum likelihood value.

Then in Section 5 we use functional properties of the departure measures to develop simple third order graphical procedures for bridging at the maximum value, for both the scalar and vector full parameter cases.

In Section 6 we consider alternatives to the combining formulas (1.1) and (1.2) to avoid the possible singularities at the extremes of the parameters being tested.

2. DEPARTURES FROM STANDARD NORMALITY: SCALAR CASE

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. 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 likelihood quantity q above 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 is the canonical parameter of the tangent exponential model at the data point y^0 of interest (Fraser, 1990). This is used to form the standardized measure of departure

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

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 standardization to that for φ. The likelihood based third order approximation to the density of the likelihood root $r(\theta;y)$ is then given by

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

where k 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 investigate how the third order likelihood based distribution for r differs from the nominal standard normal of first order theory; and we can do this in terms of the density function (2.3) or the distribution function versions (1.1) and (1.2).

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 a 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 then examine how r/q exceeds 1 taken relative to r:

$$d_1(r) = \frac{1}{r}\left(\frac{r}{q} - 1\right) = \frac{1}{q} - \frac{1}{r} . \tag{2.4a}$$

We could also examine the distribution function (1.1) and how it falls short of the nominal $\Phi(r)$, taken relative to the density $\phi(r)$,

$$d_1(r) = \frac{\Phi(r) - \Phi_1(r,q)}{\varphi(r)} = \frac{1}{q} - \frac{1}{r} ; \tag{2.4b}$$

this gives the same measure.

For a second measure we examine the argument of the distribution function approximation (1.2); the argument is usually designated r^*. We consider how it falls short of the nominal normal deviate r:

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

We now determine 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 relative to a particular data point, say y^0. In Appendix A these results are used to obtain 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 obtain the tangent exponential model approximation relative to the data point y^0. Explicit expressions for $\varphi(\theta)$ and $x(y)$ are recorded in Andrews, Fraser, and Wong (2001).

In a parallel way Abebe et al (1995) determined the local form of the model relative to a particular parameter value, say θ_0. In Appendix A this is used to obtain asymptotic expressions for d_1 and d_2 for fixed θs and varying y:

$$\begin{aligned} d_1 &= -\frac{a_3}{6n^{1/2}} - \frac{3a_4 + 4a_3^2 + 6c}{24n} r \\ d_2 &= -\frac{a_3}{6n^{1/2}} - \frac{9a_4 + 13\alpha_3^2 i + 18c}{72n} r \ . \end{aligned} \tag{2.7}$$

In (2.7) 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 local reexpressions of θ and y that are used to obtain the tangent exponential model relative to the 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 $a_3 = \alpha_3 + O(n^{-1/2})$, $a_4 = \alpha_4 - 3\alpha_3^2 - 6c + O(n^{-1/2})$. Both 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 measure is then describing how r and q differ beyond the first derivative.

We could have started with a point (θ_0, y_0) with some particular value for $r = r(\theta_0, y_0)$ and then used (2.5) to examine change in r for fixed y_0 or (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 at www.....

Example 2.1. Cauchy location model.

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

$$r = \text{sgn}(y-\theta)[2\log\{1+(y-\theta)^2\}]^{1/2} ,$$

$$q = \sqrt{2}(y-\theta)/\{1+(y-\theta)^2\} .$$

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 $d_1 = d_2 = \frac{3}{8}r$. The lack of skewness removes differences between the two versions of the departure measure.

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

$$v = -z_{y'}^{-1} z_{;\theta'}\big|_{(y^0,\hat{\theta}^0)} \tag{2.9}$$

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, and then defining q as the standardized maximum likelihood departure in this parametrization ensures that (1.1) and (1.2) are third order approximations to the p-value conditional on an approximately ancillary statistic. Thus 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 derivative of ℓ in the directions (2.9) for the ancillary at the data point. Using v, the reparameterization φ in (2.1) is generalized to

$$\varphi(\theta; y^0) = \frac{d}{dv}\ell(\theta; y)\big|_{y^0} = \ell_{;y'}(\theta; y^0)v , \tag{2.10}$$

and the expanded expression for q in (2.2) uses derivatives in the direction v rather than with respect to the original scalar y:

$$q(\theta; y) = \{\ell_{;v}(\hat{\theta}; y) - \ell_{;v}(\theta; y)\}\hat{\jmath}_{\theta\theta}^{1/2}\ell_{\theta;v}^{-1}(\hat{\theta}; y). \tag{2.11}$$

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 available, as they would need model information along the contour of the observed approximate ancillary.

3. BRIDGING THE SINGULARITY: SCALAR CASE

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 (2.4b) and (1.1) we obtain

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

and from (2.5) and (1.2) we obtain

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

These can be viewed as Bartlett type corrections to the likelihood ratio but are derived from observed likelihood.

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

$$\begin{aligned} p_1(\theta) &= \Phi(r) - \frac{3}{8}r\varphi(r) \\ p_2(\theta) &= \Phi\left(\frac{5}{8}r\right) \ . \end{aligned}$$

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 close to $r = 0$ we check numerically; at the point say $r = 0.1$ we have

$$p_1 = .524942$$

$$p_2 = .524918$$

$$p = .522499$$

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 done a full Taylor series expansions in r but, as in many similar asymptotic calculations, there are advantages to retaining the $\varphi(r)$ 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,

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

with data $y = 10$. The significance function $p(\theta)$ is plotted in Figure 5.1. Note the computational irregularities near the maximum likelihood value $\hat{\theta} = 10.495838$. Simple calculations give $\alpha_3 = -0.315901$ and $\alpha_4 = 0.199422$ from which we obtain

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

giving the bridge

$$p_B(\theta) = \Phi(0.9473498r - 0.0526502) .$$

The likelihood approximations $\Phi_{\text{LR}}(\theta)$, $\Phi_{\text{BN}}(\theta)$ are plotted in Figure 3.1 together with the bridge $p_B(\theta)$ and the exact $p_X(\theta)$. Clearly a simple algorithm can choose between the approximation and the bridge to give a close construction for the exact.

4. BRIDGING THE SINGULARITY: SCALAR INTEREST

Now consider 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 say $\theta' = (\lambda', \psi)$. Again there is an approximate ancillary with 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 $n \times p$ array 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 d/dV gives a row vector of directional derivatives; for some discussion of examples see Fraser, Wong & Wu (1999).

For testing $\psi(\theta) = \psi)$ using (1.1) or (1.2) the r is given by (1.3) and the q by the following extension (Fraser, Reid & Wu, 1999) of (2.2) and (2.11):

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

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.

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 a $p = 2$ dimensional parameter.

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 $\psi(\varphi) = \varphi_1 + c\varphi_2^2/2n^{1/2}$ where c is a second derivative measuring the curvature of $\psi = \hat{\psi}$ at $\varphi = 0$.

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

$$r = -\varphi_1 - \frac{\alpha^{111}\varphi_1^2}{6n^{1/2}} \ . \tag{4.3}$$

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, however, $\hat{\chi} - \hat{\chi}_\psi = -\varphi_1$ to second order based on the cosine of an $O(n^{-1/2})$ angle. The nuisance information is

$$\hat{\jmath}_{(\lambda\lambda)}(\hat{\theta}_\psi) = 1 + \frac{\alpha^{122}\varphi_1}{n^{1/2}} - \frac{c\varphi_1}{n^{1/2}} \ .$$

It follows that

$$q = -\varphi_1\left(1 - \frac{\alpha^{122}\varphi_1}{2n^{1/2}} + \frac{c\varphi_1}{2n^{1/2}}\right) \ . \tag{4.4}$$

Combining (4.4) and (4.3) we obtain

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

from which it follows that

$$d = -(\alpha^{111} + 3\alpha^{122} - 3c)/6n^{1/2} \ . \tag{4.4}$$

The bridging p-value formula is then

$$p(\psi) = \Phi(r) - d\phi(r) = \Phi(r - d)$$

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 from any 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 standardizations of the signed likelihood ratio that gives a third order standard normal variable.

Now consider a particular assessment of a paramater ψ with given data together with possible instability in the p-value formulas (1.1) and (1.2). We propose plotting d_1 and 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 propose fitting a line for d_1 or d_2 plotted against r, excluding the middle possibly unstable values and the extreme 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,

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

and data from Fraser, Reid and Wong (1997). For testing the parameter μ we record the approximations $\Phi_{\text{LR}}(\mu)$ and $\Phi_{BN}(\mu)$ in Figure 5.1. Note the aberrant behaviour near the maximum likelihood value $\hat{\mu}^0 =$. 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 using (3.2) with the marked segment of the straight line fit for d_2 is then recorded in Figure 5.1.

6. DISCONTINUITY AT THE EXTREMES

The familiar combining formulas (1.1), (1.2) typically give values very close to each other and the Lugannani & Rice version (1.1), is often closer to the exact; see for example Pierce & Peters (1991) and Fraser, Wong & Wu (1999). The first formula however has a disadvantage in that it can produce values outside the acceptable $[0, 1]$ range for p-values. The mechanics of this can be seen in the scalar parameter case with say large values of r. We consider this from a distribution function viewpoint (fixed θ) rather than the p-value viewpoint (fixed y).

For large values of r the first formula can be viewed using (2.3) as the integral of a normal density $\varphi(r)$ together with an adjustment factor r/q. The first correction term to $\Phi(r)$ in (1.1) is $\varphi(r)/r$ which provides the Mills ratio evaluation of the right tail of the normal. As the Mills ratio for the normal is typically on the large side, this first correction can produce an approximate value greater than 1. The second correction is r/q times the Mills ratio and provides an adjusted Mills ratio appropriate to the scaled density (2.3). If the right tail is very thin and r/q is small, then this compensating adjustment may not be enough to bring the value below 1. A reasonable objective is a modified formula that generally tracks the Lugannani & Rice (1.1) but avoids the singularity just described.

Formula (1.1) can be written

$$\Phi_1(r, d) = \Phi(r) + d\varphi(r)$$

using the nonnormality measure d_1 from Section 2 which takes the form (2.7) in the fixed θ context. We can consider this for any fixed r and then examine convergence as d goes to zero:

$$\Phi(r;0) = \Phi(r) \qquad\qquad O(n^{-3/2})$$

$$\Phi_d(r;0) = \varphi(r) \qquad\qquad O(n^{-1}) \tag{6.1}$$

$$\Phi_{dd}(r;0) = 0 \qquad\qquad O(n^{-1/2}) \ ,$$

where the subscripts denote differentiation. The second formula (1.2) can be written

$$\Phi_2(r;d) = \Phi\big(r - r^{-1}\log(1 - rd)\big)$$
$$= \Phi\big(r + d + rd^2/2\big)$$

which of course satisfies (4.1) as is easily checked by differentiation or expansion.

For simulations the tail singularity with the Lugannani & Rice (1.1) formula can be avoided by compressing towards the Barndorff-Nielsen (1.2) formula; for the right tail of the distribution function use

$$\Phi_C(r,q) = \min\{\Phi_1(r,q), .50\Phi_2(r,q) + .50\} \tag{6.2a}$$

and for the left tail use

$$\Phi_C(r,q) = \max\{\Phi_1(r,q), .50\Phi_2(r,q)\} \ . \tag{6.2b}$$

This retains the third order asymptotic property (6.1) but limits the value to being at most half way from $\Phi_2(r,q)$ to the particular bound. Alternative proportions even proportions dependent on r can replace the .50 above.

Another way to avoid the tail singularity is to back reference towards a model with asymptotic properties but also an exact p-value answer. Consider a sample from the normal (μ, σ^2). Then r and q for testing μ are

$$r = \pm\left\{n\log\left(1 + \frac{t^2}{n-1}\right)\right\}^{1/2}$$
$$q = \pm\left(\frac{n}{n-1}\right)^{1/2}\frac{t}{1 + t^2/(n-1)} \tag{6.4}$$

and the exact p-value is $H_{n-1}(t)$ where $t = n^{1/2}(\bar{y} - \mu)/s_y$ and $H_f(\cdot)$ is the Student distribution function with f degrees of freedom. It is of interest that the same r and q

arise if we make reference to the simple location model $H_{n-1}(t-\mu)$. The Student has thick tails, that is $r/q > 1$. Thus for a general problem with $r/q > 1$ we can use

$$\Phi_S(r, q) = H_f(t) \tag{6.5}$$

where

$$\begin{aligned} r &= \pm\{(f+1)\log(1+t^2/f)\}^{1/2} \\ q &= \pm\sqrt{(f+1)/f}t(1+t^2/f)^{-1} \ . \end{aligned} \tag{6.6}$$

An alternative approach is to use the gamma model

$$f(y;\theta) = \Gamma^{-1}(p)\theta^p y^{p-1} e^{-\theta y} \ , \tag{6.7}$$

which has a thick tail on the left and a thin tail on the right, although this may be somewhat less than obvious. In terms of $\hat{\theta}$ these properties are reversed. The exact p-value is

$$G_p(z) = \int_z^\infty \Gamma^{-1}(p) z^{p-1} e^{-z} dz \tag{6.8}$$

using $z = \theta y$ and allowing for the inverse relationship between $\hat{\theta}$ and y. The r and q are

$$\begin{aligned} r &= \{2(z + \log p - p - \log z)\}^{1/2} \\ q &= p^{-1/2}(p - z) \ ; \end{aligned}$$

from which z and p are easily obtained numerically.

The gamma combining formula can be described as follows. If $r > 0$ then

$$\begin{aligned} \Phi_G(r, q) &= G_p(z) && \text{if } r/q < 1 \\ &= 1 - G_p(z) && \qquad > 1; \end{aligned}$$

and if $r < 0$ then

$$\begin{aligned} \Phi_G(r, q) &= G_p(z) && \text{if } r/q > 1 \\ &= 1 - G_p(z) && \qquad < 1. \end{aligned}$$

The Student and Gamma approximations also inherit the third order properties described by (6.1).

In normal distribution sampling the Student modifications Φs will give an exact p-value for testing the location μ. Then, as suggested by Prof. John Tukey, this approximation would be preferable for location scale analysis where the error distribution could be at or near the normal form; for some related discussion see Fraser, Wong and Wu (1999).

Example 6.1. Consider the noncentral chi-square distribution with noncentral $\rho^2 = 1$ and degrees of freedom equal to $f = 5$. With ρ as parameter we use the asymptotic approximations (1.1) and (1.2) to approximation the distribution function for the chi squared variable r^2 when $\rho^2 = 1$; see Figure 6.1. The modification Φ_C from (6.2b) is also plotted there; it does avoid the long range of negative p-values for small values of r^2 but still falls short of the exact. We do note that the left end of the distribution corresponds to a singularity in the asymptotic model.

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

Derivation of (2.6), (2.7)

In the case that y and θ are both scalars we can examine the asymptotic form of $d(r,q)$ as a function of r and q by expanding the log density $\ell(\theta; y) = \log f(y;\theta)$ about a reference point in terms of standardized 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 in 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

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

and

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

The only nonzero mixed derivative terms to this order are $y\theta$ and $cy^2\theta^2/4n$. The constant $a = -\log(2\pi)/2$. The α_3 and α_4 are the standardized cumulants of the null density, and c is a measure of nonexponentiality; these are intrinsic parameters describing shape characteristics of the model. For some recent details see Andrews, Fraser & Wong (2000). 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

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

and thus gives an expression for the nonnormality measure with y fixed,

$$d_1 = -\frac{\alpha_3}{6n^{1/2}} + \frac{\alpha_4 - \alpha_3^2}{24n}r . \tag{A.3}$$

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

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

and

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

together with 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

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

and thus gives an expression for the nonnormality measure with θ held fixed:

$$d_2 = -\frac{a_3}{6n^{1/2}} - \frac{3a_4 + 4a_3^2 + 6c}{24n}r \ . \qquad (A.6)$$

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

$$d_2 = -\frac{\alpha_3}{6n^{1/2}} - \frac{3\alpha_4 - 5\alpha_3^2 - 12c}{24n}r \ . \qquad (A.7)$$

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

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.

Andrews, D.F., Fraser, D.A.S. and Wong, A. (2000). Higher order Laplace integration and the hyperaccuracy of recent likelihood methods. in revision for *J. Am. Statist. Assoc.*

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

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

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., 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`.

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

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

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

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

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.

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.

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.

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.*

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

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.

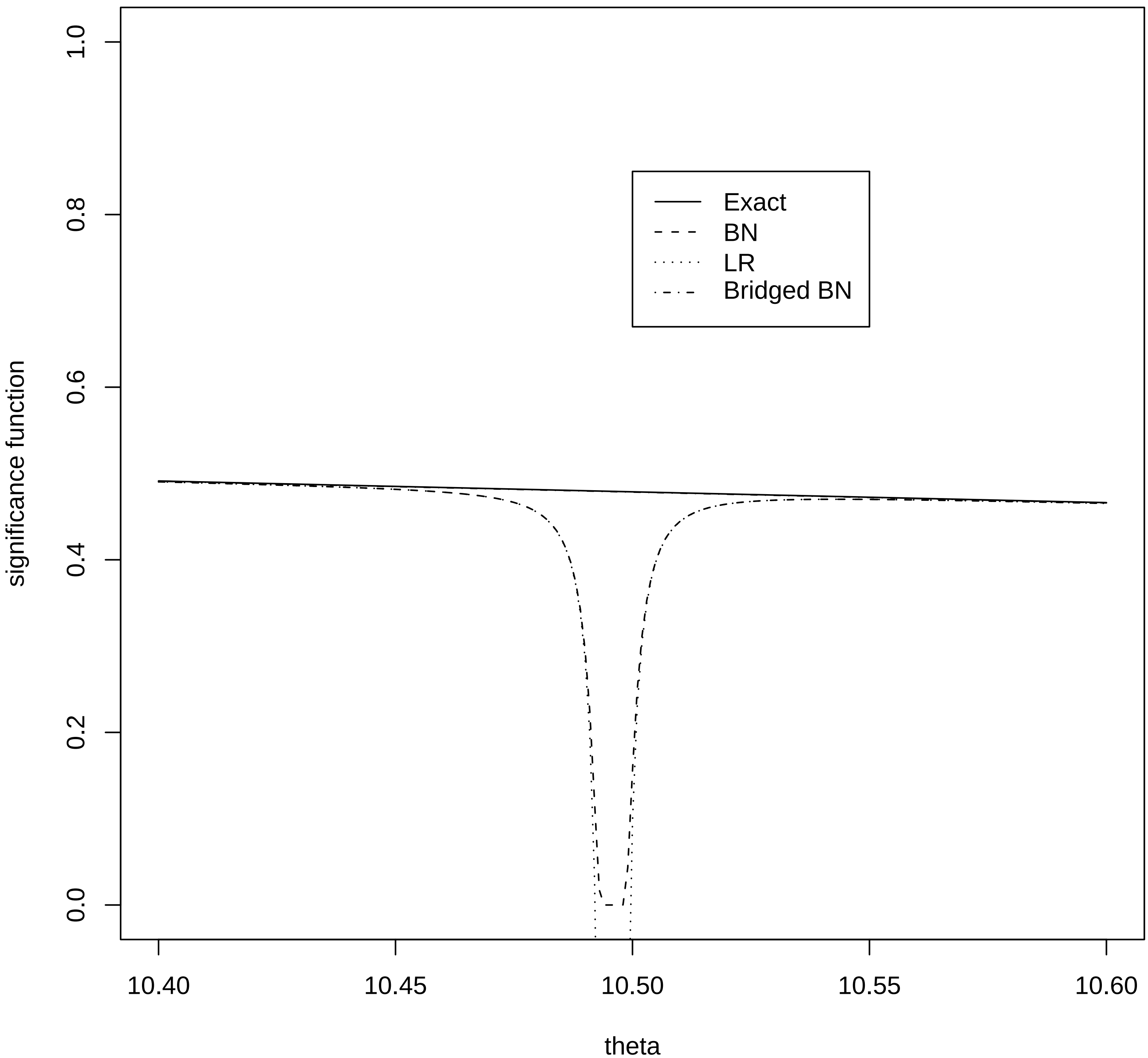
LEGENDS

Figure 3.1. The gamma model $\Gamma^{-1}(\theta)y^{\theta-1}$ with $y^0 = 10$. The asymptotic approximations $\Phi_{\text{LR}}(\theta)$ and $\Phi_{\text{BN}}(\theta)$ for the p-value function $p_X(\theta)$ for testing θ are plotted against θ. The bridge $p_B(\theta)$ at the Maximum likelihood value is superimposed on the exact $p_X(\theta)$.

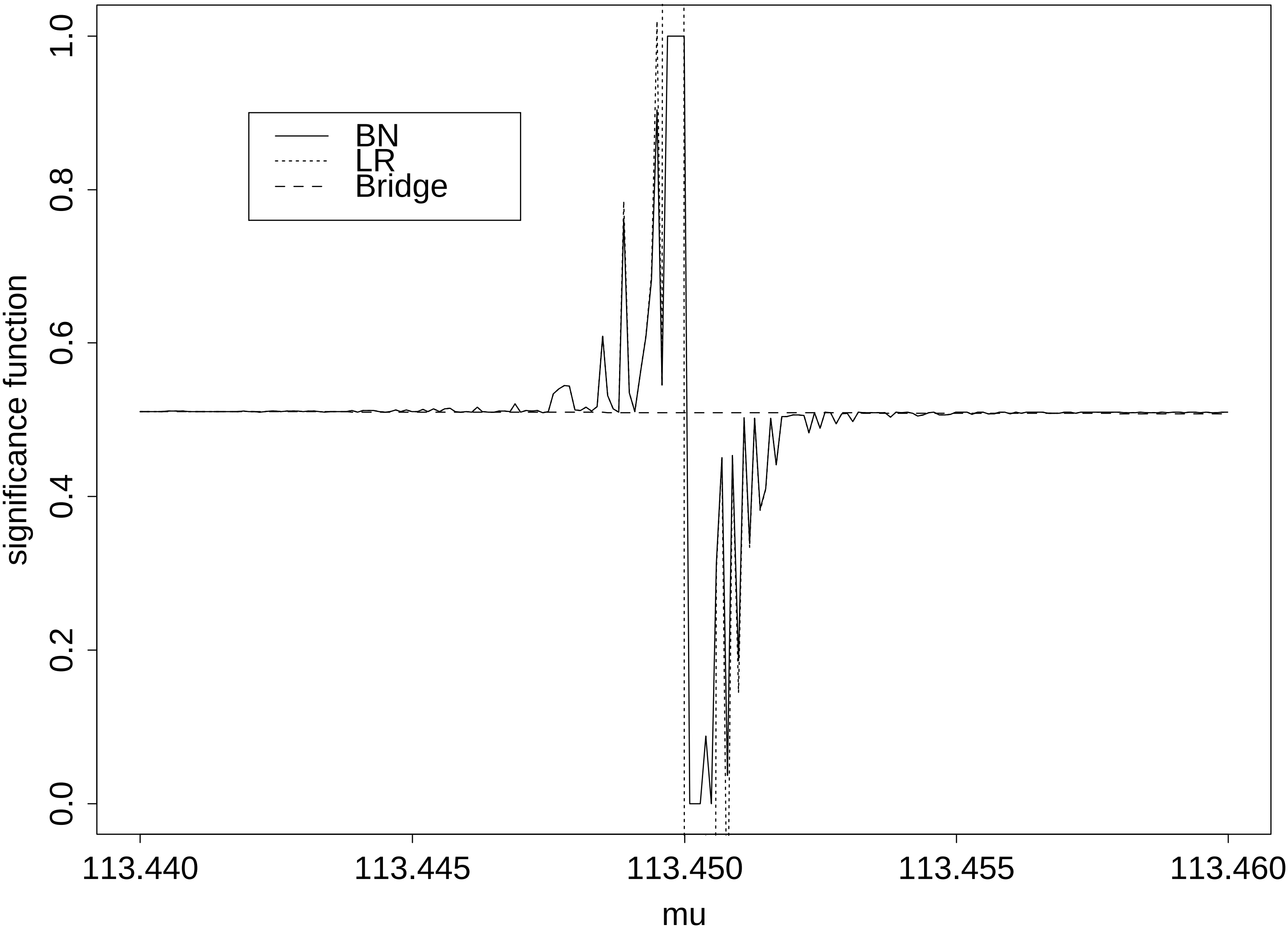
Figure 5.1. For the gamma model with mean μ and shape β the p-value approximations $\Phi_{\text{LR}}(\mu)$ and $\Phi_{\text{BN}}(\mu)$ for testing μ are plotted for a sample of 20. The aberrant behavior at the maximum likelihood value is succesfully bridged using (3.2) together with a graphical d_2 determined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

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.

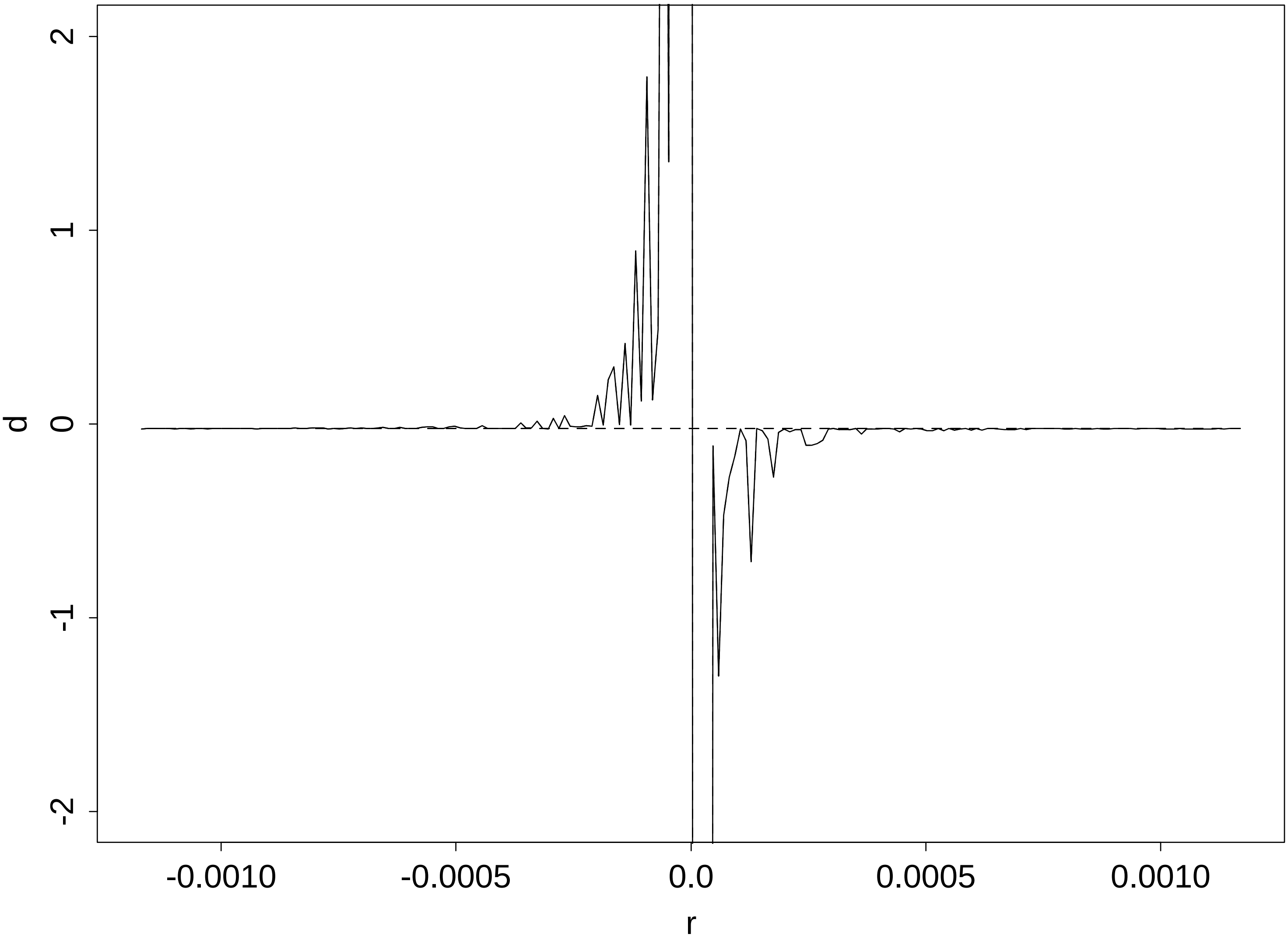
Figure 6.1. The approximations Φ_{LR} and Φ_{BN} for the distribution function of the noncentral chi-squared distribution with degrees of freedom 5 and noncentral $e^2 = 1$. The compression modification Φ_C avoids the negative values found with the Φ_{LR} approximation.



Exact
BN
LR
Bridged BN
significance function
theta
0.0
0.2
0.4
0.6
0.8
1.0
10.40
10.45
10.50
10.55
10.60



BN
LR
Bridge
significance function
0.0
0.2
0.4
0.6
0.8
1.0
113.440
113.445
113.450
113.455
113.460
mu



2
1
0
-1
-2
d
-0.0010
-0.0005
0.0
0.0005
0.0010
r