ASSESSING A VECTOR PARAMETER

By D.A.S. Fraser and N. Reid
Department of Statistics, University of Toronto
100 St. George Street, Toronto, Canada M5S 3G3
dfraser@utstat.toronto.edu

SUMMARY

Some key words. Ancillary; Conditional likelihood; Composite hypothesis; Conditioning; Conical test; Directional test; Vector interest parameter; Likelihood; Marginal likelihood; Profile likelihood; p-value.

1 Introduction

The assessment of a vector parameter is central to statistical theory. The analysis of variance with tests and confidence regions for treatment effects is well established and the related distribution theory is conveniently quite straightforward, particularly in the normal error case. In more general contexts such as generalized linear models, the assessment is usually based on the likelihood ratio or maximum likelihood departure measures and the related distribution theory is first order. As such, and confirmed by experience, both the type of departure and the distributional accuracy can be strongly misleading.

Some improvements in the departure measure and the distributional accuracy have been obtained with directional tests (Fraser & Massam, 1985; Skovgaard, 1988; Cheah et al, 1994) but these typically require an exponential model.

For greater generality we now examine a continuous statistical model with parameter θ of dimension p and interest parameter $\psi(\theta)$ of dimension d. We examine an observed data point y^0 recorded on a score variable space which has suitable linearity, and we consider the directed departure of the data point from expectation under a hypothesized value ψ for $\psi(\theta)$. The directed p-value is obtained as the observed conditional distribution function for the distance of the score variable from its expected value, given the direction of the departure. We view this as arguably the appropriate statistical assessment of the data point with respect to the parameter value ψ; for some background see Fraser & Massam (1985).

This distribution function value can be approximated by

$$p(\psi) = G_d(\chi^0) + b_d\{G_{d+1}(\chi^0) - G_d(\chi^0)\}\delta(\chi^0) \tag{1.1}$$

where $G_d(\chi)$ is the ordinary chi distribution function

$$G_d(\chi) = \int_0^{\chi} c_d t^{d-1} e^{-t^2/2} dt \tag{1.2}$$

with norming constant $c_d = 1/\Gamma(d/2)2^{d/2-1}$, and χ^0 is the observed value obtained from the score-based departure measure described in §3. The constant $b_d = c_d/c_{d+1} = 2^{1/2}\Gamma\{(d+1)/2\}/\Gamma(d/2)$ is the ratio of norming constants from adjacent χ density functions. The adjustment factor $\delta(\chi)$, obtained from (3.4) and (3.5), provides a measure of how much the underlying distribution differs from normality.

The preceding p-value has second order accuracy, by which we mean the error of the approximation to the exact distribution is $O(n^{-1})$. The usual chi-squared approximations have relative error $O(n^{-1/2})$. The approximation is derived in §3 after a brief summary of needed likelihood results in §2; it has second order accuracy and has components that are accessible and easily interpreted. By comparison, the Skovgaard (1988) directional test is also accurate to second order but requires full exponential form and cumulant type calculations, and the Cheah et al (1994) directional test is third order but also requires full exponential model form and more detailed model characteristics. The proposed test combines the likelihood ratio and score test statistics in a manner not unrelated to that found with the Lugannani & Rice (1980) approximation formula. Examples are discussed in §4.

The Bartlett (1955) correction of the likelihood ratio test can also be used for assessing vector parameters. This has the attractive property that the distribution ascribed to the related pivot has $O(n^{-2})$ distributional accuracy and the procedure can then be described as accurate to fourth order. Our viewpoint is that the assessment should take account of where the data point is relative to a parameter value, in much the same way as recent third order p-values do for a scalar parameter of interest, by recording the percentile position of the data point relative to this parameter. From this viewpoint we argue in §5 that the Bartlett corrected likelihood ratio assessment is misleading if viewed as being of other than first order. Skovgaard (2001) derives an adjustment to the likelihood ratio statistic somewhat analogous to that derived here, although from a different point of view. The relationship between the two approaches deserves further study.

2 Likelihood Background

Suppose we have a statistical model with density $f(y;\theta)$, where $y = (y_1,\ldots,y_n)'$ and θ is p-dimensional. Recent third order theory for assessing scalar or vector parameters can be based on just two ingredients, $\ell(\theta)$ and $\varphi(\theta)$, where $\ell(\theta) = \log f(y^0;\theta)$ is the observed log-likelihood function and $\varphi(\theta) = (d/dV)\ell(\theta;y)|_{y^0}$ is the observed log-likelihood gradient, calculated in directions $(v_1,\ldots,v_p) = V$ tangent to an exact or approximate ancillary. The parameter $\varphi(\theta)$ is for convenience written as a row vector, and more explicitly we have

$$\begin{aligned}\varphi(\theta) &= \left\{\frac{d}{dv_1}\ell(\theta;y^0),\ldots,\frac{d}{dv_p}\ell(\theta;y^0)\right\}\\ &= \left\{\frac{d}{dt}\ell(\theta;y^0+tv_1)\bigg|_{t=0},\ldots,\frac{d}{dt}\ell(\theta;y^0+tv_p)\bigg|_{t=0}\right\}\end{aligned}$$

where the $n\times 1$ vectors v_j make up the columns of the $n\times p$ matrix V.

With an appropriate n-dimensional pivotal quantity $z(y,\theta)$ we can obtain V from the expression

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

where the subscripts denote partial differentiation with respect to the row vectors y' and θ'. The use of the gradient implements conditioning on an approximate ancillary statistic (Fraser & Reid, 2001), and several examples of pivotal quantities z are illustrated in Fraser, Reid & Wu (1999), Fraser, Wong & Wu (1999) and Reid (2003).

Third order p-values for a scalar parameter $\psi(\theta)$ can be obtained from just $\{\ell(\theta),\varphi(\theta)\}$ using the likelihood ratio r and a special maximum likelihood departure q; see, for example, equations (1.5) and (2.11) in Fraser, Reid & Wu, or equations (3.13) and (3.16) in Reid (2003).

The two ingredients $\ell(\theta)$ and $\varphi(\theta)$ also yield an approximation to the original model that produces equivalently the third order p-values and likelihoods just described. The approximation is obtained as a first derivative calculation on the sample space, using the derivative $\varphi(\theta)$; the approximate model can be written as

$$\hat f(s;\theta)ds = (2\pi)^{-p/2}e^{k/n}\exp[\ell(\theta)-\ell(\hat\theta^0)+\{\varphi(\theta)-\varphi(\hat\theta^0)\}s]|\hat\jmath_{\varphi\varphi}|^{-1/2}ds \qquad (2.1),$$

where $s = s(y) = -\ell_\varphi\{\hat\theta(y^0);y\}$ is the score variable determined at the maximum likelihood estimate corresponding to y^0, and $s^0 = s(y^0) = 0$. The approximate model is called the tangent exponential model and provides a full second order approximation to the original model in the moderate deviation range and a third order approximation to first derivative at the data point. It implements conditioning on an approximate ancillary which corresponds to the vectors V.

In §3 we will derive a p-value for testing the full p-dimensional parameter vector θ. For testing a sub-parameter ψ of length d, we can apply the same approximation, using the adjusted log-likelihood derived in Fraser (2003). Third order log-likelihood for a scalar or vector parameter $\psi(\theta)$ can also be obtained from $\ell(\theta)$ and $\varphi(\theta)$ and has the form

$$\ell_{adj}(\psi) = \ell_{\mathrm{P}}(\psi) + \frac{1}{2}\log|\jmath_{[\lambda\lambda]}(\hat{\theta}_\psi)| \tag{2.2}$$

where λ is a complementing nuisance parameter, $\hat{\theta}_\psi = (\psi, \hat{\lambda}_\psi)$ is the constrained maximum likelihood value given $\psi(\theta) = \psi$, $\ell_{\mathrm{P}}(\psi) = \ell(\hat{\theta}_\psi) = \ell(\psi, \hat{\lambda}_\psi)$ is the profile log-likelihood, and

$$|\jmath_{[\lambda\lambda]}(\hat{\theta}_\psi)| = |\jmath_{\lambda\lambda}(\hat{\theta}_\psi)||\varphi_\lambda(\hat{\theta}_\psi)\hat{\jmath}_{\varphi\varphi}\varphi'_{\lambda'}(\hat{\theta}_\psi)|^{-1} \tag{2.3}$$

is the nuisance information determinant calculated relative to a symmetrized version of λ. Use of this log-likelihood to construct a conditional p-value is illustrated in the second example in §4.

3 A conditional p-value for inference about θ

We follow the approach developed in Fraser & Massam (1985), Skovgaard (1988) and Cheah et al. (1994), which considers the magnitude of the departure from a fixed value θ_0, conditional on the direction of the departure. One approach for example might be to use the distribution of $|\hat{\theta} - \theta_0|$ conditional on $(\hat{\theta} - \theta_0)/|\hat{\theta} - \theta_0|$, but this is not parameterization invariant. In Cheah et al. (1994), where the model is a full exponential family in the canonical parametrization, the departure from θ_0 is measured by the score variable or sufficient statistic s, and an approximation is derived for the distribution of $|s - s_0|$, given $(s - s_0)/|s - s_0|$, where s_0 is the value of the score variable with maximum likelihood value θ_0; it is parametrization invariant.

Here we work directly with the tangent exponential model approximation (2.1). We denote by χ^2 the score based measure of departure from θ_0 using the tangent exponential model:

$$\chi^2 = \chi^2(\theta_0; s) = (s - s_0)'\jmath_{\varphi\varphi}^{-1}(\theta_0)(s - s_0), \tag{3.1}$$

where $s = -\ell_\varphi(\hat{\theta}; y^0)$ is locally defined at y^0 as a function of y, satisfying $s^0 = 0$, and $s_0 = -\ell_\varphi(\theta_0; y^0)$. Note that χ is proportional to $|s - s_0|$. The dependence of χ on the tested value θ_0 is suppressed here. The tangent exponential model (2.1), with $\theta = \theta_0$ can be expressed in terms of χ^2 as

$$\hat{f}(s; \theta_0) ds = (2\pi)^{-p/2} \exp\{-\frac{\chi^2}{2}\} A(s) |j_{\varphi\varphi}(\theta_0)|^{-1/2} ds$$

where

$$A(s) = e^{k/n} \exp\{\frac{\chi^2}{2} - \frac{r^2}{2}\}\{\frac{|j_{\varphi\varphi}(\theta_0)|}{|\hat{j}_{\varphi\varphi}|}\}^{1/2}$$

and r^2 is the likelihood ratio statistic for testing $\theta = \theta_0$:

$$r^2 = r^2(\theta_0; s) = 2[\ell(\hat{\theta}) - \ell(\theta_0) + \{\varphi(\hat{\theta}) - \varphi(\theta_0)\}s]. \qquad (3.2)$$

When r^2 is evaluated at the observed data point s^0, it coincides with the likelihood ratio statistic from the original model $\ell(\theta; y^0)$. Both r^2 and χ^2 have a limiting χ_p^2 distribution under the model $f(y; \theta_0)$, and both are parametrization invariant.

We use the density approximation to compute the directional test of θ_0 conditioning on the score direction $e = (s^0 - s_0)/|s^0 - s_0| = -s_0/|s_0|$, where $s^0 = 0$ by the definition of the maximum likelihood estimate. The density of s is essentially a multivariate normal with an adjustment factor. Using this to compute the conditional distribution introduces a Jacobian effect s^{p-1} which combines with the normal kernel to give a chi density. The p-value based on this conditional test has the form

$$p(\theta_0) = \int_0^{\chi^0} g_d(\chi) \frac{a(\chi)}{c} d\chi$$

where c is the norming constant, $g_d(\chi)$ is the chi density used in (1.2), and

$$a(\chi) = A(\theta_0 + \chi e) \qquad (3.3)$$

gives the adjustment factor along the line in the direction e.

It is shown in the Appendix that for functions $a(\chi)$ having an asymptotic expansion in χ, this p-value can be approximated to $O(n^{-1})$ by (1.1) where $\delta(\chi) = \chi^{-1}\{a(\chi)/a(0)\} - \chi^{-1}$ which in this case is

$$\delta(\chi) = \chi^{-1}[\exp\{\frac{\chi^2}{2} - \frac{r^2}{2}\}\{\frac{|j_{\varphi\varphi}(\theta_0)|}{|\hat{j}_{\varphi\varphi}|}\}^{1/2} - 1] \qquad (3.4)$$

Formula (3.4) has third order accuracy but its use in (1.1) yields just second order accuracy. The calculations in Andrews et al (2004) indicate that third order accuracy for the directional

p-value is not available from the usual asymptotic information $\{\ell(\theta), \varphi(\theta)\}$ and thus is not available without additional model information such as exponential or other specified model form (Cheah et al, 1994).

With just second order accuracy available from (1.1), a reasonable option is to replace (3.4) by a further second order approximation:

$$\delta(\chi) = [\chi^2 - r^2 + \log\{\frac{|j_{\varphi\varphi}(\theta_0)|}{|\hat{j}_{\varphi\varphi}|}\}]/2\chi. \tag{3.5}$$

The likelihood and p-value expressions described here are easily seen to be invariant to the choice of parametrization, and invariant under re-expression of the response variable.

When (3.1), (3.2) and (3.5) are used in (1.1), they are evaluated at $s = s^0$.

The p-value has been developed in the context of inference for the full parameter vector θ. If we are interested in a sub-parameter vector $\psi(\theta)$ of dimension d, then the pair $\{\ell(\theta), \varphi(\theta)\}$ is replaced by the adjusted log-likelihood (2.2) and the corresponding likelihood gradient $\varphi(\psi)$. For scalar ψ this gradient is $\varphi(\psi) = \nu(\hat{\theta}) - \nu(\hat{\theta}_\psi)$ with ν calculated using a locally orthogonalized version of $\varphi(\theta)$. The p-value for assessing $\psi(\theta) = \psi$ depends only on the set $\Omega_\psi = \{\theta; \psi(\theta) = \psi\}$ and not on the structure of the parametrization $\psi(\theta)$ near $\psi(\theta) = \psi$. For vector ψ the gradient is determined component by component, using a version of ψ that has been orthogonalized at the observed data point.

4 Some examples

Example 1. Location normal. Consider $(y_1, ..., y_p)'$ normally distributed with mean $(\theta_1, ..., \theta_p)'$ and identity covariance matrix, and suppose the full parameter $\theta = (\theta_1, \ldots, \theta_p)'$ is of interest. The log-likelihood for θ is

$$\ell(\theta; y) = -\frac{1}{2}\Sigma_1^p (y_i - \theta_i)^2,$$

both obviously and by the third order determination (Fraser,2003). The related third order symmetrical canonical parametrization is $\varphi(\theta) = \theta$; and the related information is $j_{\varphi\varphi}(\theta_0) = I$. The likelihood ratio quantity (4.1) for assessing say θ_0 is

$$r^2 = \Sigma_1^p (s_i - s_{i0})^2 = \Sigma_1^p (y_i - \theta_{i0})^2$$

and the score based departure (3.1)

$$\chi^2 = (s - s_0)' I^{-1} (s - s_0) = \Sigma_1^p (y_i - \theta_{i0})^2.$$

It follows that $A(s) = 1$ and $\delta(\chi) = 0$, to second order. Thus from (1.1) we obtain

$$p(\theta_0) = G_p(\chi^0)$$

which is just the ordinary χ departure p-value but derived conditionally.

Example 2. Normal regression. Consider the normal regression model $y = X\beta + \sigma e$ where e is a sample from the standard normal. We assume that β has dimension r and thus $p = r + 1$, and we suppose that the parameter of interest is $\beta_{(1)} = (\beta_1, ..., \beta_d)'$. The directional results are parametrization and sample space invariant, so here it suffices to take a canonical design matrix $X = (I \quad O)'$. Let $s^2 = \Sigma_{i=1}^n (y_i - \hat{y}_i)^2 = \Sigma_{i=r+1}^n y_i^2$ be the sum of squared residuals, $t = (\hat{\beta}_{(1)} - \beta_{(1)})/s$ be the Student statistic for the parameter of interest, and $T^2 = |t|^2 = \Sigma_1^d (y_i - \beta_i)^2/s^2$ be the ratio of the usual sums of squares for inference about $\beta_{(1)}$. The third order likelihood $\beta_{(1)}$ is derived in Fraser (2003),

$$L(\beta_{(1)}) = (1 + T^2)^{-(n-f)/2} \tag{4.1}$$

where $f = r - d$ is the number of nuisance regression coefficients. This corresponds to the multivariate Student density for the usual multivariate quantity t for assessing $\beta_{(1)}$.

Location scale invariance shows (Fraser, 2003) that much of the analysis can without loss of generality be developed for a particular data point, with say, $\hat{\beta}_{(1)} = 0, s^2 = n$. A differential change in the data point with respect to the first d coordinates gives a corresponding translation of the likelihood in terms of the coordinates of $\beta_{(1)}$. From Cakmak, Fraser, & Reid (1994) we have that the asymptotic model determined at and to first derivative at a data point can to second order be approximated by a location model, and that the location parameter determination is unique. It follows that (4.1) records the relative density for that model and that the score variable is t and the score parameter is $\beta_{(1)}$. For this the directional assessment of $\beta_{(1)}$ is given by $|t|$ with a $\{dF/(n-p)\}^{1/2}$ distribution where F has the F distribution with d and $n - f$ degrees of freedom. The expression for $p(\beta_{(1)})$ from (1.1) then gives a second order approximation to that directional test.

5 Discussion

We have developed a simple directed departure measure (1.1) with (3.1) and (3.5) that provides a statistical measure of where the observed data point is relative to a vector parameter value. The measure combines the familiar likelihood ratio departure measure and the score departure measure and attains second order accuracy. The calculations are based on the log-likelihood and log-likelihood gradient quantities that are available in wide generality under moderate regularity conditions and a nominal assumption of asymptotic properties. Higher accuracy than the secord order seems unavailable without additional information concerning the model, such as say explicit exponential model form. Some background on related likelihood theory may be found in §6 of Fraser (2003).

An alternative approximation for testing a vector parameter can be obtained by Bartlett correction of the likelihood ratio quantity (3.2). This is a scale correction, and the resulting approximate p-value is accurate to $O(n^{-2})$ (Barndorff-Nielsen & Hall, 1988). This higher accuracy comes from the restriction to likelihood ratio as the departure measure and is attained by sacrificing information specific to the direction of departure.

This is most obvious in the case of a scalar parameter of interest, where a Bartlett-corrected version is inherently two-sided, and when there is an underlying asymmetry can provide a very poor approximation to the p-value in the direction indicated by the data. As a simple scalar case where skewness is involved we take y to have the extreme value distribution with location parameter θ: $y = \theta + z$ and

$$f(z) = \exp\{-z - e^{-z}\}, \qquad F(z) = \exp\{-e^{-z}\}.$$

The distribution is quite asymmetric; it has mode at 0 but the median is 0.366. The likelihood ratio quantity is

$$r^2 = 2\{\ell(\hat{\theta}) - \ell(\theta)\} = 2(z + e^{-z} - 1) \tag{5.1}$$

which has mean value 2γ, where $\gamma = 0.5772...$ is Euler's constant. Bartlett correction of the likelihood ratio leads to the confidence interval $(y^0 - 3.175, y^0 + 1.565)$; the exact confidence in the left tail is 0.85%, and in the right is 4.15%. The interval with exact confidence 2.5% in each tail is $(y^0 - 3.676, y^0 + 1.305)$. Both intervals are asymmetric around the point y^0, but the equi-tailed interval seems more appropriate as a summary of the data. The approximation $\Phi(r^*)$ where $r^* = r + (1/r)\log(q/r)$, r is defined in (5.1), and $q = 1 - \exp(-z)$ provides very accurate inference in each tail, leading in this example to the third order 95% confidence interval $(y^0 - 3.699, y^0 + 1.309)$, with exact confidence 2.46% in the left tail and 2.44% in the right tail.

The anomaly is likely to be more pronounced in the case of vector parameters. As an illustration suppose y_1 and y_2 are independent observations from the extreme value distribution with location parameters θ_1 and θ_2, respectively. The log-likelihood is

$$\ell(\theta; y) = -y_1 + \theta_1 - y_2 + \theta_2 - e^{(y_1-\theta_1)} - e^{(y_2-\theta_2)}$$

and

$$\varphi'(\theta) = \{\ell_{;y_1}(\theta; y^0), \ell_{;y_2}(\theta; y^0\} = \{-1 + e^{(\theta_1-y_1^0)}, -1 + e^{(\theta_2-y_2^0)}\}$$

giving

$$\ell_\varphi(\theta; y^0) = \{1 - e^{-(y_1^0-\theta_1)}, 1 - e^{-(y_2^0-\theta_2)}\}'$$

and thus

$$\begin{aligned} \chi^2(\theta_0, s^0), &= \{e^{(\theta_{10}-y_1^0)} - 1\}^2 + \{e^{(\theta_{20}-y_2^0)} - 1\}^2 \\ r^2(\theta_0, y^0) &= 2\{-2 + y_1^0 - \theta_{10} + y_2^0 - \theta_{20} + e^{(y_1^0-\theta_{10})} + e^{(y_2^0-\theta_{20})}\}. \end{aligned}$$

Figure 1 shows the contours in the parameter space using the first order χ^2_2 approximation to the distribution of r^2 and the first order χ^2_2 approximation to the distribution of χ^2, the Bartlett corrected version of r^2, and the second order approximation using (1.1) and (3.5), all at the observed data point $(y_1^0, y_2^0) = (0, 0)$.

The Bartlett adjustment is a uniform rescaling of the likelihood ratio in all regions of the parameter space, whereas the adjustment to the score statistic is conditional on the direction. The second-order approximation here uses the standardized score function as the primary measure of departure, with an adjustment that is a function of the difference between that and the likelihood ratio. It should also be possible to start with the likelihood ratio and adjust it by some function related to the difference from the score. This is the approach taken in Skovgaard (2001), although Skovgaard approximates the needed sample space derivatives by expected values of various likelihood quantities; we compute the sample space derivatives using the tangent exponential model. Work on comparing the two methods in more complex examples is in progress.

6 Appendix

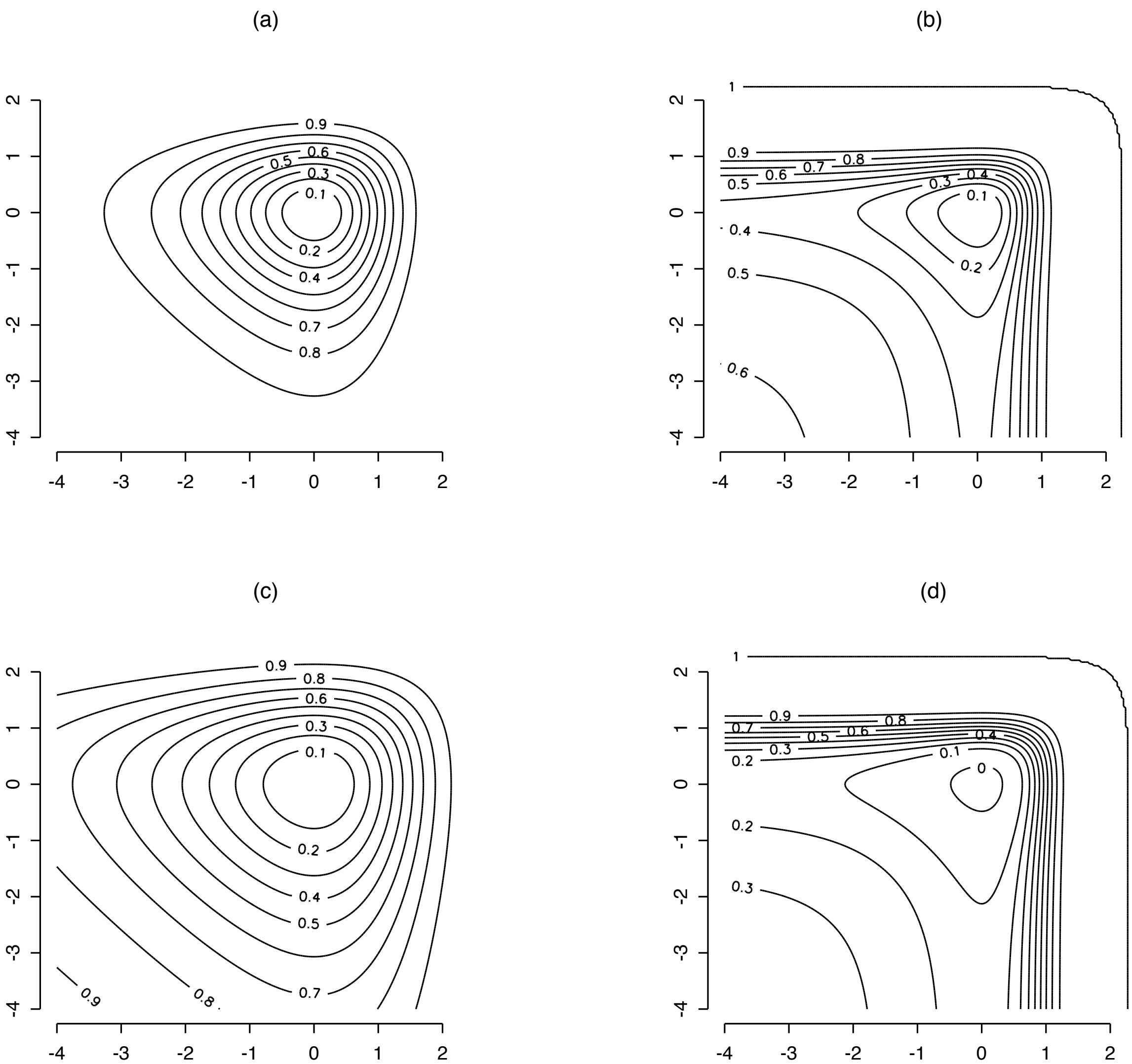
We want to evaluate

$$p(\theta_0) = \int_0^{\chi^0} g_d(\chi) \frac{a(\chi)}{c} d\chi$$

where c is the norming constant, $g_d(\chi)$ is the chi density used in (1.2), and

$$a(\chi)/a(0) = 1 + a_1\chi/n^{1/2} + a_2\chi^2/2n + O(n^{-3/2}).$$

Figure 1: Contours based on: (a), the first order approximation to the likelihood ratio statistic; (b), the first order approximation to the score statistic; (c), the Bartlett-corrected likelihood ratio statistic; and (d), the second order approximation given by (1.1) using δ computed from (3.5).



Using $c = a(0)$ as a temporary norm to give say $\bar{p}(\theta_0)$, separating off the chi density and rewriting the discrepancy as

$$\delta(\chi) = \chi^{-1}\frac{a(\chi)}{a(0)} - \chi^{-1},$$

we obtain

$$\bar{p}(\theta_0) = G_d(\chi^0) + \int_0^{\chi^0} b_d g_{d+1}(\chi)\delta(\chi)d\chi,$$

where b_d is the ratio of χ norming constants recorded after (1.2). Since $\delta(\chi)$ is constant to $O(n^{-1})$ we obtain

$$\bar{p}(\theta_0) = G_d(\chi^0) + b_d G_{d+1}(\chi^0)\delta(\chi^0)$$

by integration by parts. Then adjusting for the temporary norm we obtain the second order p-value (1.1).

ACKNOWLEDGEMENT The Natural Sciences and Engineering Research Council of Canada has provided support for this research.

REFERENCES

Andrews, D.F., Fraser, D.A.S., Li, R. & Wong, A. (2004). Distribution functions from local likelihood information. *J. Statist. Plan. Inf.*, to appear.

Barndorff-Nielsen, O.E. & Hall, P. (1988). On the level error after Bartlett adjustment of the likleihood ratio statistic. *Biometrika* **75**, 374–378.

Bartlett, M.S. (1955). Approximate confidence intervals III. A bias correction. *Biometrika* **42**, 201-4.

Cakmak, S., Fraser, D.A.S., & Reid, N. (1994) Multivariate asymptotic model: exponential and location approximations. *Utilitas Mathematica* **46**, 21-31.

Cheah, P.K., Fraser, D.A.S. & Reid, N. (1994) Multiparameter testing in exponential models: third order approximations from likelihood. *Biometrika* **81**, 27 1-8.

Fraser, D.A.S. (2003) Likelihood for component parameters. *Biometrika* **90**, to appear.

Fraser, D.A.S., & Massam, H. (1985) Conical tests: Observed levels of significance and confidence regions. *Statistische Hefte* **26**, 1-18.

Fraser, D.A.S. & Reid, N. (1993) Third order asymptotic models: Likelihood functions leading to accurate approximations for distribution functions. *Statist. Sinica* **3**, 67 -82.

Fraser, D.A.S. & Reid, N. (1995). Ancillaries and third-order significance. *Utilitas Math.* **47**, 33-53.

Fraser, D.A.S. & Reid, N. (2000). Ancillary information for statistical inference. *Empirical Bayes and Likelihood Inference*. Ed. S.E. Ahmed and N. Reid, 185-207 . New York: Springer.

Fraser, D.A.S., Reid, N. & Wu, J. (1999). A simple general formula for tail probabilities for frequentist and Bayesian inference. *Biometrika* **86**, 249-64.

Lugannani, R. & Rice, S. (1980). Saddlepoint approximation for the distribution function of the sum of independent variables. *Adv. Appl. Prob.* **12**, 47 5-90.

Reid, N. (2003). Asymptotics and the theory of inference. *Ann. Statist.* **31**, 1695–1731.

Skovgaard, I.M., (1988). Saddlepoint expansions for directional test probabilities. *J. R. Statist. Soc. B* **50** 269-80.

Skovgaard, I.M. (2001). Likelihood asymptotics. *Scand. J. Statist.* **28** 3–32.