*Journal of Statistical Research* *ISSN 0256 - 422 X*
*2016, Vol. 48-50, No. 1, pp. 47-59*

# DEFINITIVE TESTING OF AN INTEREST PARAMETER: USING PARAMETER CONTINUITY

D. A. S. FRASER
*Department of Statistical Sciences, University of Toronto, Toronto, Canada M5S 3G3, Canada*
*Email: dfraser@utstat.toronto.edu*

SUMMARY

For a scalar or vector parameter of interest with a regular statistical model, we determine the definitive null density for testing a particular value of the interest parameter: continuity gives uniqueness without reference to sufficiency but the use of full available information is presumed. We start with an exponential family model, that may be either the original model or an approximation to it obtained by ancillary conditioning. If the parameter of interest is linear in the canonical parameter, then the null density is third order equivalent to the conditional density given the nuisance parameter score; and when the parameter of interest is also scalar then this conditional density is the familiar density used to construct unbiased tests. More generally but with scalar parameter of interest, linear or curved, this null density has distribution function that is third order equivalent to the familiar higher-order p-value $\Phi(r^*)$. Connections to the bootstrap are described: the continuity-based ancillary of the null density is the natural invariant of the bootstrap procedure. Also ancillarity provides a widely available general replacement for the sufficiency reduction. Illustrative examples are recorded and various further examples are available in Davison et al. (2014) and Fraser et al. (2016).

*Keywords and phrases:* Ancillary; Exponential model; Information; Likelihood asymptotics; Nuisance parameter; $p$-value; Profile likelihood; Score conditioning; Similar test

## 1 Introduction

We consider the problem of testing a value for a $d$-dimensional parameter of interest $\psi$ in the presence of a $(p-d)$-dimensional nuisance parameter $\lambda$, in the context of a statistical model $f(y;\psi;\lambda)$ on $\mathbb{R}^n$ that we assume has the usual regularity conditions for deriving higher order approximations. We show that continuity and ancillarity directly determine a density that is free of the nuisance parameters, a density that can be viewed as providing measurement of the parameter of interest. The saddlepoint approximation then gives an expression for this density with error of $O(n^{-3/2})$. If the parameter of interest is scalar, inference based on this null density leads immediately to the familiar $r^*$ approximation (Barndorff-Nielsen, 1991; Fraser, 1990; Brazzale et al., 2007). An associated average $p$-value can also be approximated to the same order by a parametric bootstrap, as initiated in

Lee and Young (2005), Fraser and Rousseau (2008) and DiCiccio and Young (2008); computation time and ease of use can however differ dramatically.

In §2 we present the model, in §3 develop the null density 3.4 for testing the interest parameter $\psi$, and then in §4 specialize this to the linear interest parameter case obtaining the the null density 4.3; this is then shown to be equivalent to the familiar conditional distribution 4.5, which in the scalar interest case is widely used to derive unbiased or similar tests. In §5, for $\psi$ a scalar parameter, we relate the null density to the higher-order likelihood based p-values obtained from the familiar $r^*$ approximation. For a vector $\psi$ we propose the use of directional p-values, which can be obtained by one-dimensional integration. Numerical examples of the latter application are given in Davison et al. (2014) and Fraser et al. (2016). Intrinsic connections with the parametric bootstrap are addressed in §6.

## 2 Exponential model

Suppose we have a statistical model $f(y;\theta)$ for a response $y \in \mathbb{R}^n$ with parameter $\theta \in \mathbb{R}^p$ that takes the exponential form

$$f(y;\theta) = \exp\left[\varphi(\theta)^\tau \upsilon(y) - \kappa\{\varphi(\theta)\}\right] h(y), \tag{2.1}$$

where the canonical $\varphi(\theta)$ in $\mathbb{R}^p$ is one-to-one equivalent to $\theta$, and the canonical $\upsilon(y)$ in $\mathbb{R}^p$ is the usual variable directly affected by the parameter. The assumption of exponential form is more general than it may appear, as this form arises widely with regular statistical models as the tangent exponential approximation, tangent at the observed value $y^0$ with tangent vectors $V$. The construction of the tangent exponential model is briefly outlined in Appendix A, together with references to the literature.

Two key simplifications offered by 2.1 are that the distribution of $\upsilon$ provides all the information about $\varphi$ and that the density of $\upsilon$ can be approximated by the saddlepoint method. Thus our model for inference can be written

$$\begin{aligned} g(\upsilon;\varphi) &= \exp\{\varphi^\tau \upsilon - \kappa(\varphi)\} g(\upsilon) & (2.2)\\ &= \frac{e^{k/n}}{(2\pi)^{p/2}} \exp\{\ell(\varphi;\upsilon) - \ell(\hat\varphi;\upsilon)\} |\jmath_{\varphi\varphi}(\hat\varphi)|^{-1/2}\{1 + O(n^{-3/2})\}, & (2.3) \end{aligned}$$

where $\ell(\varphi;\upsilon) = \varphi^\tau\upsilon - \kappa(\varphi)$ is the log-likelihood function, $\hat\varphi = \hat\varphi(\upsilon)$ is the maximum likelihood estimator, $\jmath_{\varphi\varphi}(\hat\varphi) = -\partial\ell/\partial_\varphi\partial_\varphi^T|_{\hat\varphi}$ is the observed information array in the canonical parameterization, and $k/n$ is a generic normalizing constant (Daniels, 1954; Barndorff-Nielsen and Cox, 1979). From some original regular model this approximation needs only the observed log-likelihood function $\ell^0(\theta)$ from $y^0$ and the observed gradient $\varphi(\theta)$ of the log-likelihood in the directions $V$, and then effectively implements the integration for the original model or its approximation 2.1 to produce the marginal density $g(\upsilon;\varphi)$ to third order from that of $y$.

## 3 Curved interest and exponential model

In 2.2 and 2.3 we suppressed the dependence of $\varphi$ on $\theta$ for convenience; and we now assume that our parameter of interest is $\psi(\varphi) \in \mathbb{R}^d$, and use 2.3 to obtain the density 3.4 for testing $\psi(\varphi) = \psi_0$, eliminating the nuisance parameter $\lambda$. Thus, we consider $\psi(\varphi)$ to be fixed at $\psi_0$ in 2.3, so the model has a $p$-dimensional variable $\upsilon$, and a $(p-d)$-dimensional unknown parameter $\lambda$. With $\psi(\varphi)$ fixed at $\psi_0$, there is an approximate ancillary statistic $S$ for $\lambda$, a function of $\upsilon$ with a marginal distribution free of $\lambda$ (Fraser et al., 2010), and the ancillary density is uniquely determined to $O(n^{-3/2})$. Thus the reference marginal density for inference about a value $\psi$ based on this function of $\upsilon$ is also unique.

To describe this density we define a plane $L^0$ in the sample space by fixing the constrained maximum likelihood estimator of $\lambda$ at its observed value:

$$L^0 = \{\upsilon \in \mathbb{R}^p : \hat{\lambda}_{\psi_0} = \hat{\lambda}^0_{\psi_0}\}$$

where $\hat{\lambda}_{\psi_0}(\upsilon)$ is obtained as the solution of the score equation $\partial\ell(\varphi;\upsilon)/\partial\lambda = 0$ with notation $\hat{\lambda}_{\psi_0}(\upsilon^0) = \hat{\lambda}^0_{\psi_0} = \bar{\lambda}^0$. The constrained estimate of the full parameter $\varphi$ at $(s,t^0)$ is $\bar{\varphi}^0$. In some generality the interest parameter $\psi$ can be non-linear; in that case we define a new parameter $\chi = \chi(\varphi)$ linear in $\varphi$ that is tangent to $\psi(\varphi)$ at $\bar{\varphi}^0$; the right hand panel of Figure 1 shows the curve with $\psi$ fixed, the constrained maximum likelihood estimate $\bar{\varphi}^0$, and the linear approximation

$$\chi(\varphi) = \psi(\bar{\varphi}^0) + \bar{\psi}^0_{\varphi}(\varphi - \bar{\varphi}^0), \tag{3.1}$$

as well as the overall maximum likelihood estimate $\hat{\varphi}^0$; here $\bar{\psi}^0_{\varphi} = (\partial\psi/\partial\varphi)|_{\bar{\varphi}^0}$ is the needed Jacobian. The complementing parameter $\lambda$ in the full parameter space is shown in Figure 1 as orthogonal to $\chi$, for convenience. The left panel of Figure 1 shows the sample space, using corresponding rotated canonical variables s and t: in particular the profile plane $L^0$ on the sample space corresponds to a $p-d$ dimensional variable $t$, fixed at its observed value $t^0$. The $d$-dimensional variable $s$ on $L^0$ indexes the ancillary contours where they intersect $L^0$. In effect $(s,t)$ plays the role of the full canonical variable in an approximating exponential model, and $\chi$ is linear in the canonical parameter.

On $L^0$ the saddlepoint approximation to the joint density is, from 2.3

$$g(s,t^0) = \frac{e^{k/n}}{(2\pi)^{p/2}} \exp\{\ell(\bar{\varphi};s,t^0) - \ell(\hat{\varphi};s,t^0)\}|\jmath_{\varphi\varphi}(\hat{\varphi})|^{-1/2}, \tag{3.2}$$

where $\hat{\varphi} = \hat{\varphi}(s,t^0)$. The conditional density of $t$ given the ancillary labelled by $S = s$ has a $p^\star$ approximation at its maximum which when evaluated on $L^0$ at $\bar{\varphi}$ simplifies to

$$\frac{e^{k/n}}{(2\pi)^{(p-d)/2}}|\jmath_{(\lambda\lambda)}(\bar{\varphi})|^{-1/2}. \tag{3.3}$$

The marginal density for the ancillary variable $S$ as indexed by $s$ on the observed $L^0$ is then obtained by dividing the joint density 3.2 at $(s;t_0)$ by the conditional density 3.3 of $t$ given the ancillary $S$, with both evaluated at $(s, t^0)$ on $L^0$:

$$g_m(s; \psi_0) = \frac{e^{k/n}}{(2\pi)^{d/2}} \exp\{\ell(\bar{\varphi}; s, t^0) - \ell(\hat{\varphi}; s, t^0)\} |\jmath_{\varphi\varphi}(\hat{\varphi})|^{-1/2} |\jmath_{(\lambda\lambda)}(\bar{\varphi}|^{1/2} ds, \qquad (3.4)$$

to third order. In 3.4, the exponent $\ell(\bar{\varphi}) - \ell(\hat{\varphi})$ is the log-likelihood ratio statistic at $(s, t^0)$ for the tested value $\psi_0$, and the nuisance information determinant in the exponential parameterization $(\lambda)$ can be obtained from that in terms of $\lambda$ by applying the Jacobian $\varphi_\lambda$,

$$|\jmath_{(\lambda\lambda)}(\psi_0, \hat{\lambda}^0_{\psi_0})| = |\jmath_{\lambda\lambda}(\bar{\varphi}^0)|\, |\varphi_\lambda^T(\bar{\varphi}^0)\varphi_\lambda(\bar{\varphi}^0)|^{-1}, \qquad (3.5)$$

as described in Fraser and Reid (1993), Brazzale et al. (2007) or Davison et al. (2014). In the left panel of Figure 1 we show the curve $\psi(\varphi) = \psi_0$, and two different lines $L^0$ and $L^{00}$ corresponding to two different points $u^0$ and $u^{00}$ on an ancillary contour for the particular $\psi_0$ value.

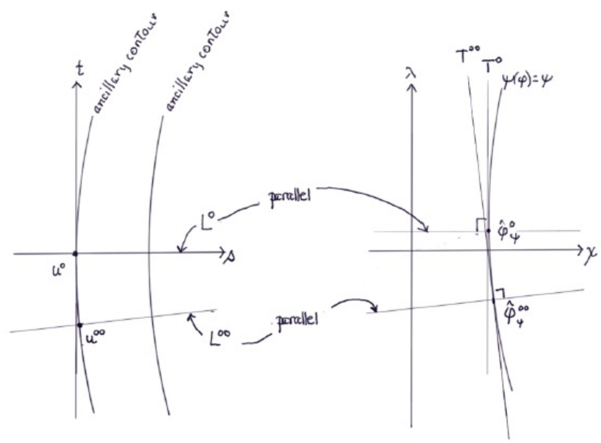


Figure 1: Score space on left; canonical parameter space on right; ancillary contours through observed $u^0$ and through a nearby point $u^{00}$ on same ancillary contour for $\psi$.

The ancillary distribution 3.4 for testing is recorded in terms of $s$ on $L^0$ but represents the result of integrating along the ancillary contours relative to $\psi(\varphi) = \psi_0$, not by integrating for fixed $s$; accordingly the distribution appears to depend on $t^0$, but this is an artifact of its presentation using coordinates that do depend on $t^0$ (Fraser and Reid, 1995; Fraser and Rousseau, 2008); see Example 4.1 in the next section. The ancillary distribution is developed above within an exponential model, either the given model or a tangent approximation to it as described in Appendix A. The development for a regular model from the point of view of approximate studentization is available in Fraser and Rousseau (2008); the distribution has third order uniqueness even though the third order ancillary itself is not unique.

# 4 Linear interest and exponential model

Consider a special case of the exponential model 2.1 where the interest parameter $\psi = \chi$ is linear and the full canonical parameter $\varphi$ is just $(\varphi, \lambda)$:

$$g(\upsilon;\theta) = \exp\{\upsilon_1^T\psi + \upsilon_2^T\lambda - \kappa(\psi,\lambda)\}h(\upsilon). \tag{4.1}$$

It is helpful to centre $\upsilon$ at the observed value: letting $s = \upsilon_1 - \upsilon_1^0$ and $t = \upsilon_2 - \upsilon_2^0$ gives

$$g(s,t;\theta) = \exp\{s^T\psi + t^T\lambda + \ell^0(\psi,\lambda)\}h(s,t). \tag{4.2}$$

where $\ell^0(\psi,\lambda)$ is the negative cumulant generating function for the latent density $h(s,t)$. The marginal density 3.4 then simplifies to

$$g_m(s;\psi) = \frac{e^{k/n}}{(2\pi)^{d/2}}\exp\{\ell(\hat\theta_\psi) - \ell(\hat\theta)\}|\jmath_{\theta\theta}(\hat\theta)|^{-1/2}|\jmath_{\lambda\lambda}(\hat\theta_\psi)|^{1/2}ds, \tag{4.3}$$

where $\hat\theta = \hat\theta(s,t^0)$ and $\hat\theta_\psi = (\psi, \hat\lambda_\psi^0)$, all to third order.

The conditional density of $s$ given $t$ is more conventionally used for inference about $\psi$ in this linear setting. From 4.2 we have

$$g_c(s|t;\psi) = \exp\{s^T\psi - \kappa_t(\psi)\}h_t(s), \tag{4.4}$$

and its saddlepoint approximation is

$$g_c(s|t;\psi) = \frac{e^{k/n}}{(2\pi)^{d/2}}\exp\{\ell_P(\psi) - \ell_P(\hat\psi)\}|\jmath_P(\hat\psi)|^{-1/2}\left\{\frac{|\jmath_{\lambda\lambda}(\psi,\hat\lambda_\psi^0)|}{|\jmath_{\lambda\lambda}(\hat\psi,\hat\lambda)|}\right\}^{1/2}, \tag{4.5}$$

where $\ell_P(\psi)$ is the profile log-likelihood function $\ell(\psi,\hat\lambda_\psi)$, and $\jmath_P(\psi) = -\partial^2\ell_P(\psi)/\partial\psi\partial\psi^T$ is the associated information function (Davison, 1988; Barndorff-Nielsen and Cox, 1979). The two densities 4.3 and 4.5 are identical, as $|\jmath_{\theta\theta}(\hat\theta)| = |\jmath_P(\hat\psi)||\jmath_{\lambda\lambda}(\hat\theta)|$.

In the above we have not distinguished the tested value $\psi_0$, because in the exponential model with canonical interest parameter $\psi$, the planes $L^0$ for different tested values of $\psi$ are parallel, and the distributions as recorded on each plane are equivalent, so the resulting marginal density 4.3 can be used as the pivotal quantity to test any value of $\psi$ and thereby provide confidence intervals or regions.

**Example 4.1.** We illustrate this with a simple exponential model, for which the detailed calculations are readily obtained. Take $p = 2$ and suppose that the joint density of $s, t$ is of the form

$$g(s,t;\psi,\lambda) = \phi(s-\psi)\phi(t-\psi)\exp\{-a\psi\lambda^2/(2n^{1/2})\}h(s,t), \tag{4.6}$$

where $\phi(.)$ is the standard normal density. The function $h(s,t)$ can be explicitly obtained as

$$h(s,t) = 1 + \frac{1}{2}as(t^2-1)n^{-1/2} + \frac{1}{8}a^2(s^2-1)(t^4-6t^2+3)n^{-1} + O(n^{-3/2}), \tag{4.7}$$

and we can re-write the density as

$$g(s,t;\psi,\lambda) = \{1 - a\psi\lambda^2/(2n^{1/2}) + a^2\psi^2\lambda^4/(8n)\}\phi(s-\psi)\phi(t-\psi) \\ \times \{1 + as(t^2-1)/(2n^{1/2}) + a^2(s^2-1)(t^4-6t^2+3)/(8n) + O(n^{-3/2})\}. \quad (4.8)$$

The related marginal density is obtained by taking all terms in 4.8 to the exponent and completing the square; this shows that, ignoring terms of $O(n^{-3/2})$, there is a pivotal function $Z_\psi$ which follows a standard normal distribution to third order:

$$Z_\psi = s\{1 + a^2(2t^2-1)/4n\} - \psi\{1 - a^2(2t^2-1)/4n\} - a(t^2-1)/2n^{1/2}. \quad (4.9)$$

From this we see that s has conditional bias $a(t^2-1)/2n^{1/2} + O(n^{-1})$, but this bias in the measurement of $\psi$ is of no consequence for inference, as it is removed as part of forming the pivot $Z_\psi$. If we ignore terms of $O(n^{-1})$ then $s - a(t^2-1)/(2n^{1/2})$ is standard normal to $O(n^{-1})$, *i.e.* to this order only a location adjustment is needed to obtain an approximately standard normal pivotal quantity.

# 5 Inference for $\psi$ from the reference density

The base density $g_m(s;\psi)$ on $\mathbb{R}^d$ given at 3.4 is to third order the unique density for inference about $\psi$, in the sense that it is a direct consequence of requiring model continuity to be retained in the elimination of the nuisance parameter (Fraser et al., 2010). The density can be computed from the distribution 2.2 or 2.3 for the canonical variable $u$ or from the observed log-likelihood from the original model $\ell(\varphi;y^0) = \log\{f(y^0;\varphi)\}$ together with the observed log-likelihood gradient $\varphi(\theta) = \ell_{;V}(\theta;y^0)$ in directions $V$; see Appendix A.

If $d=1$, the one-dimensional density can be integrated numerically. It can also be shown to be third-order equivalent to a standard normal density for the familiar pivot $r^* = r^*(\psi;y^0)$, defined by

$$r^*(\psi;y^0) = r - r^{-1}\log\frac{r}{Q}, \quad (5.1)$$

$$r = \pm\left(2[\ell\{\varphi(\hat\theta);y^0\} - \ell\{\varphi(\hat\theta_\psi);y^0\}]\right)^{1/2}, \quad (5.2)$$

$$Q = \pm|\chi(\hat\theta) - \chi(\hat\theta_\psi)|/\hat\sigma_\chi, \quad (5.3)$$

where $\hat\sigma^2_\chi = |\jmath_{(\lambda\lambda)}\{\varphi(\hat\theta_\psi)\}|/|\jmath_{\varphi\varphi}\{\varphi(\hat\theta)\}|$ is a particular estimate of the variance of the numerator of $Q$, and $\pm$ designates the sign of $\hat\psi^0 - \psi$. From the definition 3.1 of $\chi$ as tangent to $\psi$ at $\varphi(\hat\theta_\psi)$, we obtain an alternate expression for $Q$,

$$Q = \frac{|\varphi(\hat\theta) - \varphi(\hat\theta_\psi)\ \varphi_\lambda(\hat\theta_\psi)|}{|\varphi_\theta(\hat\theta)|}\frac{|\jmath_{\theta\theta}(\hat\theta)|^{1/2}}{|\jmath_{\lambda\lambda}(\hat\theta_\psi)|^{1/2}} \quad (5.4)$$

which can be more convenient for computation. Several examples of the use of $r^*(\psi;y^0)$ as a standard normal pivotal for inference about a scalar parameter of interest are given in Fraser et al. (1999) and Brazzale et al. (2007). For Example 4.1, straightforward calculations verify that $r^*(\psi;y^0) = Z_\psi$ of 5.4 to $O(n^{-3/2})$.

**Example 5.1.** As an illustration of exact and approximate p-value contours, consider two exponential life variables $y_1, y_2$ with failure rates $\varphi_1, \varphi_2$ and with interest parameter chosen as the total failure rate $\psi = \varphi_1 + \varphi_2$; the model is $\varphi_1\varphi_2 \exp\{-\varphi_1 y_1 - \varphi_2 y_2\}$ with $0 < y_1, y_2 < \infty$. A rotation of variable and of parameter through $\pi/4$ gives new variables $s = (y_1 + y_2)/2^{1/2}, t = (-y_1 + y_2)/2^{1/2}$ and new parameters $\chi = (\varphi_1 + \varphi_2)/2^{1/2}, \lambda = (-\varphi_1 + \varphi_2)/2^{1/2}$ on equivalent rotated quadrants; the model then becomes

$$f(s,t) = (\chi^2/2 - \lambda^2/2)\exp(-\chi s - \lambda|t|),$$

with $s > |t| > 0, \chi > |\lambda| > 0$ and parameter of interest $\psi = 2^{1/2}\chi$. The exact conditional density of $s$, given $t$, is $f(s|t;\chi) = \chi \exp\{-\chi(s - |t|)\}$, for $s > |t|$, i.e. the pivotal quantity $Z_\chi = \chi(s - |t|)$ follows an exponential distribution with rate 1. The approximation 4.3 is an $O(n^{-3/2})$ approximation to this, equivalent to a standard normal distribution for the adjusted log-likelihood root $r^*_\chi$.

In Figure 5 we illustrate three quantile contours, at levels 25%; 50%, 75%, for the exact conditional distribution and for the normal approximation to the distribution of $r^*$, for testing the value of $\psi = 0.6931$ or equivalently $\chi = 0.4901$. The contours of the exact conditional density are line segments, and the contours of the approximate normal distribution for $r^*_\chi$ are smooth curves. The conditional and marginal approaches are identical to third order: the difference that appears in Figure 5 is due entirely to the approximation to the marginal density. From one point of view the normal approximation to $r^*_\chi$ replaces exact similarity of the test with similarity to $O(n^{-3/2})$, and the smoothed version is somewhat less sensitive to the exact value of $t$. Third-order similarity of tests based on $r^*$ is established in Jensen (1992).

**Example 5.2.** As an illustration of an exponential model with a curved interest parameter suppose in Example 5.1 that the parameter of interest is now taken to be $\psi = \varphi_1\varphi_2$; we let $\lambda = \varphi_2$ be an initial nuisance parameter. Then

$$\hat\psi = 1/(y_1 y_2), \quad \hat\lambda = 1/y_1, \quad \hat\lambda^2_\psi = \psi_1/y_2 = \psi\hat\lambda^2/\hat\psi.$$

The linear parameter $\chi(\varphi)$ is

$$\chi(\varphi) = \chi(\bar\varphi) + \psi_\varphi(\bar\varphi)(\varphi - \bar\varphi) = \chi(\bar\varphi) - 2\bar\varphi_1\bar\varphi_2 + \bar\varphi_2\varphi_1 + \bar\varphi_1\varphi_2$$

and $s$ is the corresponding linear combination of $y_1$ and $y_2$. The information determinants $|\jmath_{\varphi\varphi}(\hat\varphi)|$ and $|\jmath_{(\lambda\lambda)}(\bar\varphi)|$ are $\hat\psi^{-2}$ and $2\psi\hat\lambda_\psi\hat\lambda/(\hat\psi(\psi^2+1))$, respectively; the latter is obtained using $\varphi_\lambda = (-\psi/\lambda^2, 1)$. The marginal density of $s$ is approximated by 3.4, from which it follows that $r^*_\psi$ is a pivotal quantity following a standard normal distribution to $O(n^{-3/2})$. The quantile curves will be similar in shape to those in Example 4.1, but there is no exact conditional density for comparison.

When $\psi$ is a vector, the marginal density on $L^0$ does not immediately lead to a single $p$-value function. A directional approach is available following Fraser and Massam (1985), Skovgaard (1988), Fraser and Reid (2006) and references therein. On $L^0$ the mean value under the fitted null

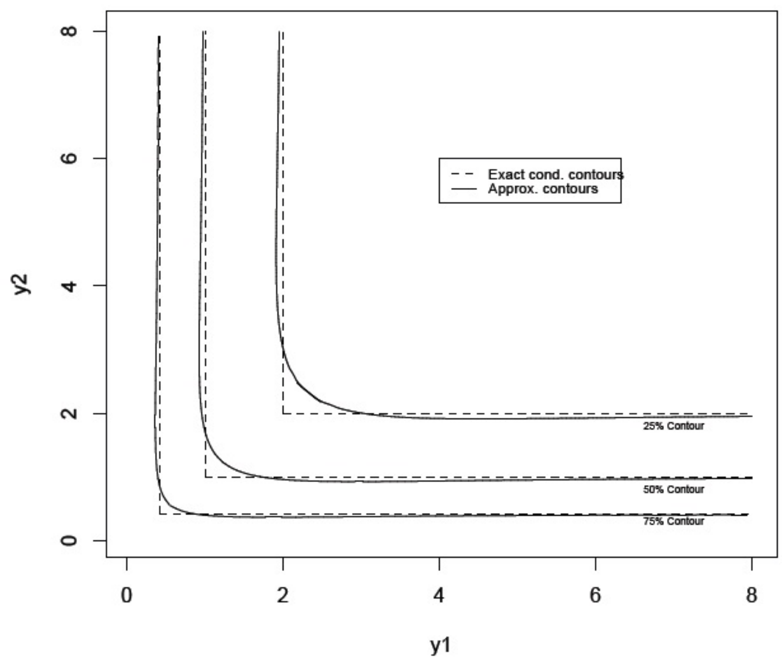


Figure 2: The exact conditional contours and the third order approximate contours at quantile levels 25%, 50% and 75% for testing $\psi = 0.6931$ in the simple exponential life model.

parameter value $\hat{\varphi}^0_\psi$ is $s_\psi = -\ell^0_\psi(\psi, \hat{\lambda}^0_\psi)$ with corresponding data $s = 0$ in the standardized coordinates. Then conditioning on the direction from expected $s_\psi$ to observed $s = 0$ gives the directional $p$-value

$$p(\psi) = \frac{\int_1^\infty g_m\{s_\psi + t(0 - s_\psi)\}t^{d-1}dt}{\int_0^\infty g_m\{s_\psi + t(0 - s_\psi)\}t^{d-1}dt};$$

which can easily be evaluated numerically. A number of examples based on familiar exponential models where calculations are particularly accessible are presented in Davison et al. (2014) and Fraser et al. (2016).

# 6 Bootstrap and higher order likelihood

For the exponential model 2.1, improved $p$-values for testing a scalar interest parameter $\psi = \psi_0$ can also be obtained using bootstrap sampling from the estimated model $f(y; \hat{\theta}^0_\psi)$. In particular, this bootstrap applied to the signed log-likelihood root $r_\psi = \pm[2\{\ell(\hat{\theta}) - \ell(\hat{\theta}_\psi)\}]^{1/2}$ produces $p$-values that are uniformly distributed with accuracy $O(n^{-3/2})$, and are asymptotically equivalent to this order to $p$-values obtained from the normal approximation to $r^*_\psi$. If however the estimated sampling model is taken to be the traditional $f(y; \hat{\theta}^0)$ then the relative error drops to $O(n^{-1})$ (DiCiccio et al (2001); Lee & Young (2005)).

The bootstrap has an intrinsic connection with the ancillary distribution 3.3 for the marginal variable $S$, as recorded in terms of $s$ on the observed $L^0$. Indeed, the bootstrap distribution $f(y; \hat{\theta}^0_\psi)$ directly produces the preceding null distribution for the ancillary $S$; this follows by noting that the distribution of $S$ is free of $\lambda$ and thus a particular choice $\lambda = \hat{\lambda}^0_\psi$ in the re-sampling model just generates the same marginal null distribution. Accordingly the distribution 3.3 can be viewed as an invariant of the bootstrap procedure. It also follows that the bootstrap distribution of any statistic that is a function of the ancillary $S$ is also an invariant of the bootstrap distribution to third order.

More generally with an asymptotic model having full parameter dimension $p$, and null parameter dimension $p-d$, the moderate deviations region can be presented as a product space with coordinates $(S, \hat{\lambda}_\psi)$ and a bootstrap step can be viewed as a projection along contours of the ancillary variable $S$ such that dependence on the conditional $\hat{\lambda}_\psi$ is reduced by a factor $O(n^{-1/2})$ (Fraser and Rousseau, 2008).

Meanwhile for the higher-order likelihood approach, the standard normal approximation to the usual $r^*_\psi$ is accurate to $O(n^{-3/2})$. This, with the preceding bootstrap result, shows that the higher-order $r^*_\psi$ approximation can be implemented directly by bootstrap resampling of $r^*_\psi$ or equivalently the bootstrap resampling of $r_\psi$ which is known to be affinely equivalent to $r^*_\psi$ to third order, using of course the estimated null model $f(y; \hat{\theta}^0_\psi)$; computation times however can be significantly different: for a recent example calculations used 20 hours for the bootstrap calculation to achieve the same accuracy as the higher order likelihood calculation achieved in 0.09 seconds.

For an exponential model with scalar linear interest parameter, DiCiccio & Young (2008) show that the null model bootstrap distribution of $r^*_\psi$ directly approximates the conditional distribution of $r^*_\psi$ even though the bootstrap is an unconditional simulation; this follows from 4.3 and 4.5 by noting that the marginal and conditional distributions are identical to third order.

More generally with a regular model and conditioning based on tangent vectors $V$ a bootstrap step provides an average over the conditioning indexed by the vectors $V$ and thus does not record the precision information that is routinely available by the higher order approach and even certain higher order Bayesian calculations. Thus we can say that the parametric bootstrap based on the observed maximum likelihood estimate under the null reproduces an average of the higher order $r^*_\psi$ evaluations rather than the individual precision-tuned $p$-values coming from the higher-order method.

# 7 Conclusion

For general regular models with scalar or vector, linear or curved interest parameters, we have determined the score space distribution that has nuisance parameter effects removed and has the full information to provide $r^*$ tests for scalar parameters and directed tests for vector parameters. We have thus extended available distribution theory for statistical inference, and integrated the directional methodology with the higher order distribution theory. In particular for the vector parameter case this can fine-tune the Bartlett-corrected l-dimensional numerical integration (Davison et al., 2014).

## Acknowledgment

This research is part of a continuing joint research project on statistical inference using methods of likelihood asymptotics, joint with Nancy Reid, Department of Statistical Sciences, University of Toronto, and received support from the Natural Sciences and Engineering Research Council of Canada, the Department of Statistical Sciences at the University of Toronto, and the Senior Scholars Funding of York University. Very special thanks to Mahbub Latif for support with type setting, and to Wei Lin and Uyen Hoang for manuscript development.

# Appendix A

In the context of a regular statistical model with continuity suppose that the variable $y$ has $n$ independent coordinates, with $n > p$, the dimension of the parameter $\theta$. Let $p = p(y; \theta)$ be the vector with $i$th coordinate $p_i = F_i(y_i; \theta)$, where $F_i(.; \theta)$ is the distribution function for the $i$th component of $y$. By inverting $p = p(y; \theta)$ to solve for $y$ we obtain the generalized quantile function $y = y(p; \theta)$. This quantile function links change in the parameter with change in the variable $y$; the assumed model continuity provides the inverse. The local effect of the continuity at the observed data $y^0$ can then described by the $n \times p$ matrix of gradient vectors, called ancillary directions,

$$V = \left. \frac{\partial y(p; \theta)}{\partial \theta} \right|_{y^0, \hat{\theta}^0} \tag{.1}$$

which link change in the coordinates of $\theta$, at $\hat{\theta}^0$, to change in the response, at $y^0$, for fixed p. A number of examples of the matrix $V$ are given in Fraser et al. (2010, §3) and Brazzale et al. (2007, §8.4). The column vectors of $V$ are tangent to the flow of probability under $\theta$-change near $\hat{\theta}^0$; this flow defines the continuity-based ancillary contours concerning $\theta$. Fraser et al. (2010) show that these vectors define a surface in the sample space that is ancillary to $O(n^{-1})$.

Our continuity assumption, which we view as intrinsic to a general approach to inference using approximate ancillarity, rules out unusual pivots as in the inverted Cauchy introduced in McCullagh (1992); see Fraser et al. (2010, Example 5 ). A different, although related, approach is needed for discrete responses $y$; see Davison et al. (2006) for discussion.

For vector parameters the approach builds on the presence of the quantile function presentation of the model and with independent vector coordinates may leave arbitrariness that can be addressed in other ways.

Given this matrix of ancillary directions $V$, a tangent exponential model with canonical parameter

$$\varphi(\theta) = \ell_{;V}(\theta; y^0),$$

can be constructed, where

$$\ell_{;V}(\theta; y^0) = \left.\frac{\partial \ell(\theta; y)}{\partial V}\right|_{\hat{\theta}^0, y^0}$$

is shorthand for the set of directional derivatives of $\ell(\theta; y)$ in the sample space, each direction determined by a row of $V$. The tangent exponential model is

$$f_{TEM}(s; \theta) = \exp\{\varphi(\theta)^T s + \ell(\theta; y^0)\} h(s), \qquad (.2)$$

where $s \in \mathbb{R}^p$ has the same dimension as the parameter $\varphi$, and can be thought of as the score variable. The tangent exponential model was introduced in Fraser (1990); see also Reid and Fraser (2010, §2) and the references therein. The model was introduced mainly as a tool to obtain an $r^*$ approximation for inference about a scalar component parameter, without the need to explicitly compute an ancillary density. Here we are using the tangent model as a descriptive device for obtaining a conditional density for inference about a scalar or vector parameter of interest, via saddlepoint approximations.

**Example .1.** Suppose $y_i$ are independent observations from the curved exponential family $N(\psi; c^2\psi^2)$, where $c$ is fixed. The $i$th component of the quantile vector $p$ is $(y_i - \psi)/(c\psi)$, and the $i$th entry of the $n \times 1$ vector $V$ is $y_i^0/\hat{\psi}^0$. Using this to define $\varphi(\theta)$ we have

$$\varphi(\theta) = \ell_{;V}(\theta; y^0) = \sum_{i=1}^{n} \frac{\partial \ell(\theta; y^0)}{\partial y_i} V_i = \frac{1}{c^2\psi} \sum (y_i^0/\hat{\psi}^0) - \frac{1}{c^2\psi^2} \sum \{(y_i^0)^2/\hat{\psi}^0\},$$

a linear combination of $1/\psi$ and $1/\psi^2$. In terms of the sufficient statistic $(\sum y_i, \sum y_i^2)$ an exact ancillary is $\sum y_i^2/(\sum y_i)^2$. The ancillary based on the $V_i$ is consistent with this as both $\{y_i^0\}$ and $c\{y_i^0\}$ on the linear space $\mathcal{L}V$ give the same value to $\sum y_i^2/(\sum y_i)^2$.

# Appendix B

*(i) From likelihood to density by Taylor expansion.* Example 4.1 is motivated by the usual Taylor series expansion of the log-likelihood function for a regular $p$-dimensional statistical model: the leading term is the log-likelihood for a normal distribution, with higher order terms that drop off as $n^{1/2}; n^{-1}; n^{-3/2}$; see for example, DiCiccio et al. (1990) and Cakmak et al. (1998). To simplify the calculations we introduce just one third derivative term: $a\lambda^2\chi/(2n^{1/2})$, where $a = \partial^3\ell/\partial\lambda^2\partial\chi$, evaluated at the expansion point. The resulting likelihood function can be inverted to provide an expression for the latent density $h(s, t)$, to $O(n^{-3/2})$, verifying 4.7:

$$\begin{aligned} g(s, t; \chi, \lambda) =& \frac{1}{2\pi} \exp\{-(s-\chi)^2/2 - (t-\lambda)^2/2 - a\chi\lambda^2/2n^{1/2}\} h(s, t) \qquad (.3) \\ =& \phi(s-\chi)\phi(t-\lambda)\{1 - a\chi\lambda^2/2n^{1/2} + a^2\chi^2\lambda^4/8n\} h(s, t) + O(n^{-3/2}) \\ =& \phi(s-\chi)\phi(t-\lambda)\{1 - a\chi\lambda^2/2n^{1/2} + a^2\chi^2\lambda^4/8n\} \\ & \times \{1 + as(t^2-1)/2n^{1/2} + a^2(s^2-1)(t^4 - 6t^2 + 3)/8n\} + O(n^{-3/2}); \end{aligned}$$

the second equality uses $\exp(c/n^{1/2}) = 1 + c/n^{1/2} + c^2/2n + O(n^{-3/2})$, and the third equality uses $(1 - c/2n^{1/2} + c^2/8n)^{-1} = 1 + c/2n^{1/2} + c^2/8n + O(n^{-3/2})$ together with $E(x^2 - 1) = \theta^2, E(x^4 - 6x^2 + 3) = \theta^4$ when $x$ follows a $N(\theta, 1)$ distribution.

*(ii) From density to likelihood by Taylor expansion.* The conditional model for $s$ given $t$ is available as the $t$-section of the density 4.6 and gives 4.8, up to a normalizing constant as:

$$\begin{aligned} g(s|t;\chi) =& c\phi(s-\chi)\{1 + as(t^2-1)/2n^{1/2} + a^2(s^2-1)(t^4-6t^2+3)/8n\} \\ =& \phi(s-\chi)\{1 + as(t^2-1)/2n^{1/2} + a^2(s^2-1)(t^4-6t^2+3)/8n\} \\ & \{1 + a\chi(t^2-1)/2n^{1/2} + a^2\chi^2(t^4-6t^2+3)/8n\}^{-1} \\ =& \phi(s-\chi)\{1 + as(t^2-1)/2n^{1/2} + a^2(s^2-1)(t^4-6t^2+3)/8n\} \\ & \{1 - a\chi(t^2-1)/2n^{1/2} + a^2\chi^2(t^4+2t^2-1)/8n\} \\ =& \phi(s-\chi)\{1 + as(t^2-1)/2n^{1/2} + a^2(s^2-1)(t^4-6t^2+3)/8n\} \\ & \exp\{-a\chi(t^2-1)/2n^{1/2} + a^2\chi^2(4t^2-2)/8n\} \end{aligned} \quad (.4)$$

The second equality comes by evaluating the constant c as the reciprocal of an integral with respect to $s$ and uses $E(x) = \theta$ and $E(x^2-1) = \theta^2$ when $x$ follows a $N(\theta, 1)$ distribution; the third equality comes from calculating the reciprocal of the factor coming from the preceding integration; and the fourth comes by taking the preceding to the exponent.

# References


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

[2] Barndorff-Nielsen, O. E. and D. R. Cox (1979). Edgeworth and saddlepoint approximations with statistical applications. *J. R. Statist. Soc. B* **41**, 187-220.

[3] Brazzale, A. R., A. C. Davison, and N. Reid (2007). *Applied Asymptotics. Cambridge: Cambridge University Press.*

[4] Cakmak, S., D. A. S. Fraser, P. McDunnough, N. Reid, and X. Yuan (1998). Likelihood centered asymptotic model: exponential and location model versions. *J. Statist. Planning and Inference* **66**, 211-222.

[5] Daniels, H. E. (1954). Saddlepoint approximations in statistics. *Annals Math. Statist.* **46**, 21-31.

[6] Davison, A. C. (1988). Approximate conditional inference in generalized linear models. *J. R. Statist. Soc., B* **50**, 445-461.

[7] Davison, A. C., D. A. S. Fraser, and N. Reid (2006). Improved likelihood inference for discrete data. *J. R. Statist. Soc., B* **68**, 495-508.

[8] Davison, A. C., D. A. S. Fraser, N. Reid, and N. Sartori (2014). Accurate directional inference for vector parameters in linear exponential models. *J. Amer. Statist. Assoc.* **109**, 302–314.

[9] DiCiccio, T. J., C. A. Field, and D. A. S. Fraser (1990). Approximations of marginal tail probabilities and inference for scalar parameters. *Biometrika* **77**, 77-95.

[10] DiCiccio, T. J. and G. A. Young (2008). Conditional properties of unconditional parametric bootstrap procedures for inference in exponential families. *Biometrika* **95**, 497-504.

[11] Fraser, A. M., D. A. S. Fraser, and A. M. Staicu (2010). Second order ancillary: A differential view with continuity. *Bernoulli* **16**, 1208-1223.

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

[13] Fraser, D. A. S. and H. Massam (1985). Conical tests: Observed levels of significance and confidence regions. *Statist. Hefte* **26**, 1-17.

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

[15] Fraser, D. A. S. and N. Reid (1995). Ancillaries and third order significance. *Utilitas Mathematica* **47**, 33-53.

[16] Fraser, D. A. S. and N. Reid (2006). Assessing a vector parameter. *Student* **5**, 247-256.

[17] Fraser, D. A. S., N. Reid, and N. Sartori (2016). Accurate directional inference for vector parameters. *Submitted*.

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

[19] Fraser, D. A. S. and J. Rousseau (2008). Studentization and deriving accurate p-values. *Biometrika* **95**, 1-16.

[20] Jensen, J. (1992). The modified signed likelihood statistic and saddlepoint approximations. *Biometrika* **79**, 693-703.

[21] Lee, S. and G. Young (2005). Parametric bootstrapping with nuisance parameters. *Statist. Prob. Lett.* **71**, 143-153.

[22] McCullagh, P. (1992). Conditional inference and Cauchy models. *Biometrika* **79**, 247-259.

[23] Reid, N. and D. A. S. Fraser (2010). Mean likelihood and higher order approximations. *Biometrika* **97**, 159-170.

[24] Skovgaard, I. (1988). Saddlepoint expansions for directional test probabilities. *J.R. Statist. Soc. B* **50**, 269-280.