Biometrika (2008), **94**, 2, *pp*. 1–10

Advance Access publication on 31 July 2008
Printed in Great Britain

Testing canonical parameters in exponential models: Conditional analysis or an inadvertent saddlepoint

BY D.A.S. FRASER

Department of Statistics, University of Toronto, Toronto, Ontario, Canada M5S 3G3
Department of Statistical and Actuarial Sciences, University of Western Ontario, London, Ontario, Canada N6A 5B7
dfraser@utstat.toronto.edu

SUMMARY

Testing a scalar canonical parameter of an exponential model is widely based on the conditional distribution given the canonical variable for the nuisance parameter; this has the support of being uniformly most powerful similar and uniformly most powerful unbiased for given test size. Such testing optimality in turn gives support for the exact test with the 2×2 table, and only occasionally is caution expressed that the nuisance canonical variable may contain information relevant to the interest parameter. Second and third order approximate likelihood analysis examines this conditional procedure and shows generally with non-normality that: (i) $O(n^{-1/2})$ bias can be present in the conditional distribution for the interest parameter; (ii) the nuisance parameter can be separated from the interest parameter to third order making available a definitive marginal assessment of the interest parameter; (iii) $O(n^{-3/2})$ accurate p-values for the interest parameter then come from this marginal approach; and (iv) this marginal approach is not restricted to exponential model form. These somewhat surprising results are obtained when, quite properly, similarity and unbiasedness are also addressed approximately. And an inadvertent choice of saddlepoint for the conditional analysis actually gives the marginal analysis.

Some key words: Asymptotic likelihood; Likelihood bias; Exponential model; Score conditioning; Score conditioning; UMP similar; UMP unbiased.

1. INTRODUCTION

Consider a full exponential model with canonical parameter $\varphi = (\psi, \lambda)$ where ψ is a scalar interest parameter and λ is a possibly vector-valued nuisance parameter; the density is

$$f(s, t; \psi, \lambda) = \exp\{\psi s + \lambda t - \kappa(\psi, \lambda)\}h(s, t) \quad (1)$$

relative to canonical variables s and t. An antecedent variable y may of course be part of the background but here we examine just the reduced model for (s, t), whether available directly by sufficiency or indirectly by integration or saddlepoint approximation.

The model has the attractive property that the conditional model for $s|t$ has scalar variable form and depends only on the scalar interest ψ:

$$f(s|t; \psi) = \exp\{\psi s - \kappa^t(\psi)\}h_t(s),$$

where $h_t(s)$ is the t-section of $h(s, t)$ and $\kappa^t(\psi)$ provides the normalization. This conditional distribution is widely viewed as highly appropriate for statistical inference; and it is directly available when the marginal density for (s, t) is accessible and is available approximately when that density is obtained by saddlepoint analysis. An assessment of $\psi = \psi_0$ or of $\psi \le \psi_0$ is known to have a uniformly most powerful similar test or a uniformly most powerful unbiased test against $\psi > \psi_0$, at any given level α; see for example Lehman & Romano (2005, p 121). Also this test is directly derived within the conditional model $f(s|t;\psi)$ and its properties are then justified by the familiar completeness of the conditioning statistic t. The test has substantial advantages coming from its ease of computation, but we will also see that this conditional approach can seriously neglect ψ-information in the conditioning distribution.

By contrast recent likelihood-based methods, Daniels (1954) and Barndorff-Nielsen & Cox (1979) for asymptotic exponential models and Fraser & Reid (2001) and Fraser, Reid & Wu (1999) for asymptotic models more generally, give a marginal variable and model for assessing the interest parameter ψ, and show that the model and analysis is third-order free of the nuisance parameter. We compare these conditional and marginal procedures for assessing a canonical parameter of an exponential model.

In Section 2 we develop coordinates, standardized in moderate deviations about an observed data point or some suitable reference value, and restrict attention to the case with a model information characteristic $\jmath_{\lambda\lambda\psi}$, say a, which describes how nuisance parameter information varies with the interest parameter ψ. We then derive a quick third-order Taylor expansion (4) for the model and verify that it agrees with the saddlepoint approximation that seems less directly accessible.

Then in Section 3 we derive a third order expansion (5) for the model for $s|t$ and find to second order that it has the simplified form

$$\begin{aligned} f(s|t;\psi) &= \phi(s-\psi)\{1 + as(t^2-1)/2n^{1/2} - a\psi(t^2-1)/2n^{1/2}\} \\ &= \phi\{s - a(t^2-1)/2n^{1/2} - \psi\}, \end{aligned} \tag{2}$$

where $\phi(z)$ is the standard Normal density function; this says that the conditional distribution of $s|t$ is standard Normal located at $\psi + a(t^2-1)/2n^{1/2}$. Thus the canonical variable has an apparent location bias of order $O(n^{-1/2})$ for assessing ψ; and this bias varies quadratically with respect to the conditioning canonical variable.

In Section 4 we find that a modified variable S is Normal $(\psi, 1)$ to third order, thus providing highly accurate likelihood inference for ψ free of the nuisance parameter λ. And in Section 5 we obtain the joint distribution of (S, t) and see that the related conditional distribution of $(t|S)$ is Normal $(\lambda, 1)$ but with a first-order shape modification $-\jmath_{\lambda\lambda\psi}\psi(\lambda^2 - t^2 + 1)/2n^{1/2}$. Thus the first order bias in the conditional distribution of $s|t$ for inference concerning ψ is removed when we switch to S for inference and the bias is transfered to the distribution of $t|S$, where it does not affect inference concerning ψ.

Section 6 discusses the usual use of the saddlepoint to approximate the conditional distribution for $s|t$ and shows that this usual use inadvertently gives a saddlepoint approximation for a different distribution, the distribution for S that is deemed here to be the appropriate third order inference distribution concerning the interest ψ. And Section 7 examines an $n = 1$ example with some concluding discussion in Section 8.

2. APPROXIMATIONS: SADDLEPOINT AND TAYLOR

Consider the exponential model $f(s, t; \psi, \lambda)$ with some reference data point $(s, t) = (s^0, t^0)$ and assume regularity and asymptotic properties as some nominal sample size n becomes large. We derive a third-order approximate model that is applicable in moderate deviations relative to data. As a first step we center the variable and parameter about the data and corresponding maximum likelihood value and for convenience rewrite $(s - s^0, t - t^0)$ and $(\psi - \hat{\psi}^0, \lambda - \hat{\lambda}^0)$ as just (s, t) and (ψ, λ). We then examine the resulting standardized model $f(s, t; \psi, \lambda)$ and find that the general likelihood function has the simple form $\ell(\psi, \lambda; s, t) = \ell(\psi, \lambda) + s\psi + t\lambda$ where $\ell(\psi, \lambda)$ is likelihood at the given reference data point $(s, t) = (s^0, t^0)$ which is just $(0, 0)$ in the modified coordinates.

We can apply the likelihood-based saddlepoint approximation to obtain the third order reexpression of the model:

$$f(s, t; \psi, \lambda) ds dt = \frac{e^{k/n}}{2\pi} \exp\{\ell - \hat{\ell}\} |\hat{\jmath}_{\varphi\varphi}(s, t)|^{-1/2} ds dt \tag{3}$$

where

$$\hat{\ell} - \ell = \ell(\hat{\psi}, \hat{\lambda}; s, t) - \ell(\psi, \lambda; s, t) = \ell(\hat{\psi}, \hat{\lambda}) - \ell(\psi, \lambda) + s(\hat{\psi} - \psi) + t(\hat{\lambda} - \lambda)$$

is the log-likelihood ratio $r^2(\psi, \lambda)/2$ at (s, t) and $\hat{\jmath}_{\varphi\varphi}(s, t) = -\ell_{\varphi\varphi}(\hat{\psi}, \hat{\lambda})$ is the related information array with subscripts to ℓ denoting derivatives with respect to $\varphi = (\psi, \lambda)$.

Or more directly we can Taylor expand the log-model $\log f(s, t; \psi, \lambda)$ about $(s, t) = (0, 0)$ using variable and parameter departures that are standardized by observed information in such a way that the new ψ is a monotone equivalent of the initial ψ. For this let $\jmath^{1/2}$ be the positive lower-triangular right root of the observed information matrix, $\jmath_{\varphi\varphi} = (\jmath^{1/2})^T(\jmath^{1/2})$ giving the rescaled departures

$$\jmath^{1/2} \begin{pmatrix} \psi - \hat{\psi}^0 \\ \lambda - \hat{\lambda}^0 \end{pmatrix},$$

which again for convenience we designate as just $(\psi, \lambda)^T$. The corresponding transformation of the score variable $(s, t)^T$ that retains the form of the exponential tilt term $s\psi + t\lambda$ uses the transposed inverse of $\jmath^{1/2}$. We first expand r^2 to first-order and obtain $\chi^2 = (s - \psi)^2 + (t - \lambda)^2$, which is the quadratic in the usual Normal approximation.

We next expand to the third order and because of exponential model form find that the new terms for φ come from $\ell(\varphi)$ and the new terms for (s, t) come solely from parameter-free terms that are typically unavailable but here are trivially accessible from model normalization. The first-order terms in φ involve third derivatives of $-\ell(\varphi)$; there are four such derivatives but we examine just one that is of crucial interest, $\jmath_{\lambda\lambda\psi} = -\ell_{\lambda\lambda\psi} = a$, and use the letter a for convenience; this is the ψ-derivative of the nuisance information $\jmath_{\lambda\lambda}(\psi, \lambda)$, and from later discussion is seen to be the clear source for the phenomena of interest here. Each possible third derivative term when brought down from the exponent has a compensating data term so we lose no generality in examining just the single term of special interest here. Accordingly we examine to third order the standardized model

$$\begin{aligned} f(s, t; \psi, \lambda) &= \frac{1}{2\pi} \exp\{-\chi^2/2 - a\psi\lambda^2/2n^{1/2}\} h(s, t) \\ &= \phi(s - \psi)\phi(t - \lambda)\{1 - a\psi\lambda^2/2n^{1/2} + a^2\psi^2\lambda^4/8n\} h(s, t) \\ &= \phi(s - \psi)\phi(t - \lambda)\{1 - a\psi\lambda^2/2n^{1/2} + a^2\psi^2\lambda^4/8n\} \\ &\quad \{1 + as(t^2 - 1)/2n^{1/2} + a^2(s^2 - 1)(t^4 - 6t^2 + 3)/8n\} \end{aligned} \tag{4}$$

where the second equality uses the third-order expansion $\exp(c/n^{1/2}) = 1 + c/n^{1/2} + c^2/2n$ and the third equality uses $(1 - c/2n^{1/2} + c^2/8n)^{-1} = (1 + c/2n^{1/2} + c^2/8n)$ and uses the expectations $E(y^2 - 1) = \theta^2$ and $E(y^4 - 6y^2 + 3) = \theta^4$ for the Normal $(\theta, 1)$.

The preceding uniqueness of a data term corresponding to a particular parameter term is the present Taylor series equivalent of the uniqueness of a Fourier inverse; and a reverse correspondence is the equivalent of the uniqueness of the Fourier transformation itself. For the implications of this with approximation models see Cakmak et al (1998) and Andrews et al (2005).

The expression (4) for the modified model $f(s,t;\psi,\lambda)$ can be derived directly from the saddlepont expression (3) but with additional work; this derivation uses the expansion of $\ell(\varphi)$ and the substitution of $\hat{\varphi}$ given by

$$\hat{\psi} = s - at^2/2n^{1/2} + a^2st^2/n \qquad \hat{\lambda} = t - ast/n^{1/2} + a^2s^2t/n + a^2t^3/2n$$

leading to $\hat{\ell} = -ast^2/2n^{1/2} + a^2t^4/8n + a^2s^2t^2/2n$ and $|\hat{\jmath}| = 1 + as/n^{1/2} - 3a^2t^2/2n$. After combining the resulting factors, the earlier expression (4) is obtained but also the original Laplace constant $k/n = 3a^2/8n$ is inferred from a comparison of (3) with (4).

3. CONDITIONAL MODEL CONCERNING ψ

Consider the exponential model (1) for (s,t) with the special $\jmath_{\lambda\lambda\psi}$ characteristic as used in the standardized third order version (4). As indicated in the Introduction the conditional distribution of s given the nuisance parameter score t is the prominent theoretical recommendation for powerful unbiased and similar tests. We obtain this conditional distribution for our specialized model (4) by taking the t-section of the joint model, and then renormalizing:

$$\begin{aligned} f(s|t;\psi) &= c\phi(s-\psi)\{1 + as(t^2-1)/2n^{1/2} + a^2(s^2-1)(t^4-6t^2+3)/8n)\} \\ &= \phi(s-\psi)\{1 + as(t^2-1)/2n^{1/2} + a^2(s^2-1)(t^4-6t^2+3)/8n)\} \\ &\qquad \{1 + a\psi(t^2-1)/2n^{1/2} + a^2\psi^2(t^4-6t^2+3)/8n)\}^{-1} \\ &= \phi(s-\psi)\{1 + as(t^2-1)/2n^{1/2} + a^2(s^2-1)(t^4-6t^2+3)/8n\} \\ &\qquad \{1 - a\psi(t^2-1)/2n^{1/2} + a^2\psi^2(t^4+2t^2-1)/8n)\} \\ &= \phi(s-\psi)\{1 + as(t^2-1)/2n^{1/2} + a^2(s^2-1)(t^4-6t^2+3)/8n\} \\ &\qquad \exp\{-a\psi(t^2-1)/2n^{1/2} + a^2\psi^2(4t^2-2)/8n\}, \end{aligned} \tag{5}$$

where the second equality comes from evaluating c as the reciprocal of an integral, the third from calculating the reciprocal and the fourth from taking the constant to the exponent.

If we now take the log-likelihood for the original model (4) and subtract the log-likelihod for the conditional model (5) we obtain

$$-(t-\lambda)^2/2 - a\psi(\lambda^2 - t^2 + 1)/2n^{1/2} - a^2\psi^2(4t^2-2)/8n, \tag{6}$$

which is of course the log-likelihood for the marginal model for t. And we find that it depends $O(n^{-1/2})$ on the interest parameter ψ through $\lambda^2 - t^2 + 1$, which does have mean 0. Now suppose we think of ψ as being fixed and address the likelihood for λ: from (5) or (6) we then see that the marginal model for t does not have the full model likelihood for λ; this will be of particular interest for us in Section 5.

Now suppose we examine this conditional distribution (5) further, but just to second order; we then have the easier expression

$$\begin{aligned} f(s|t;\psi) &= \phi(s-\psi)\{1 + as(t^2-1)/2n^{1/2} - a\psi(t^2-1)/2n^{1/2}\} \\ &= \phi\{s - a(t^2-1)/2n^{1/2} - \psi\}. \end{aligned} \tag{7}$$

This says the conditional distribution of $s|t$ is standard Normal located at $\psi + a(t^2-1)/2n^{1/2}$; it thus has a location bias of order $O(n^{-1/2})$ which varies quadratically with the conditioning variable. This bias has no real effect on the likelihood or on the p-value function; indeed the conditional model is standard Normal to the second order in terms of the modified variable $\tilde{s} = s - a(t^2 - 1)/2n^{1/2}$ with second order model $\phi(\tilde{s} - \psi)$.

4. Marginal model concerning ψ

With a fixed value for an interest parameter, asymptotic theory provides contours that are second order ancillary in the presence of a nuisance parameter (Fraser & Reid, 2001); this makes available the third order tests for scalar interest parameters (Fraser, Reid & Wu, 1999) and builds on Barndorff-Nielsen (1986); and the analysis then in Fraser, Fraser & Staicu (2010) shows that to second order the contours are quadratic in the canonical variable of the actual or approximating exponential model. This quadratic property does not define the ancillary contour to third order which could be of interest for calculations here; but the related likelihood theory has the existence of third order contours and then derives the marginal distribution of the needed statistic from local properties near the observed data. Accordingly we pursue this latter route using the saddlepoint method.

Let $S(s,t)$ be a third order ancillary with respect to the nuisance parameter λ for some given ψ value. Following Fraser & Reid (1995), Fraser (2003) and Fraser & Rousseau (2008) we examine the ancillary S where its contours intersect the observed profile curve for the interest parameter ψ; specifically the curve is the line $L = \{(s,t) : \hat{\lambda}_\psi = \hat{\lambda}_\psi^0\}$. As the nuisance score equation for calculating $\hat{\lambda}_\psi$ from the exponential model (1) is $t - \kappa_\lambda(s,t) = 0$, we have that the profile curve is the line $L = \{t = t^0 = 0\}$ in our data standardized notation. Accordingly we have that S is a function of s along the line L, and by reexpressing $S(s,0)$ can write $S(s,t) = s$ for points on the line L. Thus along the profile line L we have $S = s$ and $dS = ds$.

Consider the particular version (4) of the exponential model (1) and a point $(s,0)$ on the profile line L. The saddlepoint approximation for the full model $f(s,t)$ at $(s,0)$ is given by (3) with $t = 0$.

Consider further the version (4) of the exponential model but now with the interest parameter ψ fixed at the value of interest. The available parameter is then λ and it can be viewed as operating on the conditional distribution given the ancillary S. Accordingly the saddlepoint approximation can be applied to this conditional and gives the density $f(t|S;\lambda)$ at the interest point $t = 0$ on the line L. This makes use of the full likelihood ℓ but with ψ fixed and takes the form

$$f(t|S)dt = \frac{e^{k/n}}{(2\pi)^{1/2}} \exp\{\ell - \hat{\ell}_{(\psi)}\} |\hat{\jmath}_{\lambda\lambda}(s,0)|^{-1/2} dt, \tag{8}$$

where

$$\hat{\ell}_{(\psi)} - \ell = \ell(\psi, \hat{\lambda}_\psi; s, 0) - \ell(\psi, \lambda; s, 0) = \ell(\psi, \hat{\lambda}_\psi) - \ell(\psi, \lambda) + 0(\hat{\lambda}_\psi - \lambda) \tag{9}$$

is the log-likelihood ratio for λ with ψ fixed and $\hat{\psi}, \hat{\lambda}, \hat{\lambda}_\psi$ and $\hat{\jmath}_{\psi\psi}$ are evaluated at $(s,0)$ on L.

The quotient of a full by a conditional gives a marginal density, and is not a common calculation. Accordingly we now divide the joint model (3) at $(S,0)$ by the conditional model (8) at $(S,0)$ and obtain the saddlepoint-based approximation for the marginal model for S, which here describes the probability projected to the profile line L along the ancillary contours for fixed S:

$$\begin{aligned} f(S;\psi)dS &= \frac{e^{k/n}}{(2\pi)^{1/2}} \exp\{\hat{\ell}_{(\psi)} - \hat{\ell}\} |\hat{\jmath}^{\psi\psi}|^{1/2} dS \\ &= (2\pi)^{-1/2} \exp\{\ell(\hat{\varphi}_\psi) - \ell(\hat{\varphi}) + s(\psi - \hat{\psi})\} |\jmath^{\psi\psi}(s,0)|^{1/2} dS, \end{aligned} \tag{10}$$

where

$$\hat{\ell} - \hat{\ell}_{(\psi)} = \ell(\hat{\psi}, \hat{\lambda}) - \ell(\psi, \hat{\lambda}_\psi) + s(\hat{\psi} - \psi)$$

is the profile log-likelihood ratio for ψ and $(\jmath^{\psi\psi})^{-1}k = |\jmath_{\varphi\varphi}|/\jmath_{\lambda\lambda}$ is the profile information for ψ; the dependence on λ cancels in the quotient.

The approximation (10) for the density concerning ψ on the line L can be evaluated by using $\hat{\psi} = s$ and $\hat{\lambda} = 0$ obtained from (4) and recorded at the end of Section 2, and by using $\hat{\lambda}_\psi = -a\psi\lambda/n^{1/2}$ obtained from the nuisance score equation, and by then substituting into the saddlepoint ingredients to obtain $\ell(\hat{\varphi}_\psi) = -(s-\psi)^2/2$, $\ell(\hat{\varphi}) = 0$ and $\jmath^{\psi\psi}(s, 0) = 1 + as/n^{1/2}$. Then simplifying the saddlepoint expression (10), we obtain

$$f(S;\psi) = \phi(S - \psi). \tag{11}$$

Thus S is Normal $(\psi; 1)$ to third order and is equivalent to the corresponding r^* quantity. But we have not yet defined the modified variable S other than along the profile contour L.

The present development with minor adjustments is the present context simplification of the third order inference development for curved parameters $\psi(\varphi)$ in exponential models (Barndorff-Nielsen, 1986) and for curved parameters in general asymptotic models (Fraser, Reid and Wu, 1999; Fraser, Wong Wu, 1999). These in turn provide third-order p-values by r^*-integration.

5. JOINT DISTRIBUTION FOR (S, t)

In Section 3 we obtained the density for the widely recommended conditional distribution (5) of $s|t$. And in Section 4 we obtained the marginal distribution (11) of S calculated by projecting to L along ancillary contours for the nuisance parameter, in the same way that the Student distribution concerning μ is obtained in the Normal (μ, σ) context. Both cases lead to a distribution that is free of the nuisance parameter λ. We first seek an explicit expression for $S(s, t)$ generally, not just along the profile line L.

For this we have from Fraser, Fraser & Staicu (2010) that the ancillary to second order for given ψ is quadratic in the canonical variables. Accordingly, we try a plausible quadratically defined statistic $S = s - a(t^2-1)/2n^{1/2}$ with an intuited coefficient a, and then derive the joint density $f(S, t)$ from the available density (4) for (s, t).

First we have that the Jacobian from (s, t) to (S, t) is unity. Next we examine the substitution $s = S + a(t^2-1)/2n^{1/2}$ in the Normal density factor $\phi(s - \psi)$ of (4):

$$\phi(S-\psi)[1 - (S-\psi)a(t^2-1)/2n^{1/2} + \{(S-\psi)^2 - 1\}a^2(t^2-1)^2/8n].$$

And then we make the same substitution in the data factor of (4):

$$\{1 + aS(t^2-1)/2n^{1/2} + a^2(t^2-1)^2/4n + a^2(S^2-1)(t^4 - 6t^2 + 3)/8n\}.$$

And finally after collecting and simplifying we obtain the joint density $f(S, t)$:

$$\phi(S-\psi)\phi(t-\lambda)\exp\{-a\psi(\lambda^2 - t^2 + 1)/2n^{1/2}\}\{1 - a^2(S^2-1)(4t^2-2)/8n\} \tag{12}$$

Now suppose we temporarily examine the joint distribution (12) but just to second order; we then have the easier expression

$$f(S, t) = \phi(S-\psi)\phi(t-\lambda)\exp\{-a\psi(\lambda^2 - t^2 + 1)/2n^{1/2}\} \tag{13}$$

This says that the variable $S = s - a(t^2-1)/2n^{1/2}$ has a standard Normal distribution located at ψ with no nuisance parameter effect, and thus the variable S gives ancillary contours for any fixed value of ψ relative to the nuisance parameter λ. It also says that the variable t is statistically independent with a Normal distribution about λ but with first order shape modification $-\jmath_{\lambda\lambda\psi}\psi(\lambda^2 - t^2 + 1)/2n^{1/2}$. This independent variable can reasonably be viewed as irrelevant for inference concerning ψ.

6. Conditional inference and the inadvertent saddlepoint

Consider further the conditional model $f(s|t)$ at (5) for inference concerning ψ; it has Normal density form $\phi(s - \psi)$ with two modifying factors. If we take these factors to the exponent, combine the linear terms with the quadratic of the Normal density, and then collect the remaining terms which are $O(n^{-1})$, we obtain

$$f(s|t) = \phi\{s - \psi - a(t^2 - 1)/2n^{1/2}\} \exp\{a^2(4t^2 - 2)(\psi^2 - s^2 + 1)/8n\}. \tag{14}$$

This says that the conditional model is Normal to second order as noted at (7) but for third order has an additional factor factor $\exp\{a^2(4t^2 - 2)(\psi^2 - s^2 + 1)/8n\}$ which is quadratic in the operative variable s but is also quadratic in the conditioning variable t and in the interest parameter ψ. Thus inference using the conditional model (14) for $s|t$ can differ substantially from inference using the marginal model (11) for S. To second order we can of course view S as a bias corrected version of the conditional variable $s|t$. But to third order we have that the S remains standard Normal located at ψ while $s|t$ has a contraction or expansion in the centre and the opposite in the tails. In fact the distribution of $s|t$ can be viewed as having the expansion or contraction property for its basic shape and this property averages out under variation of the nuisance score variable to give the marginal version based on S. This seems to argue in favour of S for inference concerning ψ, that is, to use the result obtained from the higher-order likelihood route and the saddlepoint.

The conditional distribution, however, of $s|t$ as mentioned in Section 1 is the widely recommended reference for statistical inference concerning ψ, and we have seen now that it contains a bias of order $O(n^{-1/2})$, the same order as with the use of the score, maximum likelihood departure and signed likelihood ratio; and we have also seen that it leads to inference that oscillates about that provided by the marginal model for S. However in the typical implementation using $s|t$, for example Davison (1988) and Fraser, Reid & Wong (1991), a saddlepoint approach is used, and it seeks the conditional distribution as a ratio joint-over-marginal, in contrast with Section 4 where a ratio of joint-over-conditional was used.

As part of the joint-over-marginal calculation, the likelihood attributed to t is the full likelihood function but with ψ-fixed and with λ as the operating parameter. But from the second last paragraph of Section 3 we have that a canonical variable does not capture all the likelihood for the related parameter. Thus by associating the full likelihood function with the canonical variable t we have inadvertently included likelihood properly associated with the remaining variable $s|t$. But this by inadvertence happens to be the appropriate likelihood from another viewpoint and indeed corresponds to the third order distribution $f(S;\psi)$. It does not lead to the conditional model for $s|t$, but it does to the marginal model for S. In effect the variable S was targeted inadvertently while seeking the biased variable $s|t$.

7. AN $n = 1$ EXAMPLE

Consider two independent scalar variables y_1, y_2 with simple exponential life models and rate parameters φ_1, φ_2; the statistical model is

$$f(y_1, y_2; \varphi_1, \varphi_2) = \varphi_1 \varphi_2 \exp(-y_1\varphi_1 - y_2\varphi_2)$$

with y_1 and y_2 greater than zero. This has simplicity like the location Normal but also has the minor anomaly that coordinates as written are stochastically decreasing in their parameters; accordingly the underlying variables can conveniently be viewed as reciprocals y_1^{-1}, y_2^{-1} or negatives; the sample space is the positive quadrant. With minor loss of generality we now consider the observed data point $y^0 = (1,1)'$ which conveniently has observed information $\hat{\jmath} = I$, the identity matrix.

As a simple interest parameter consider $\varphi_1 + \varphi_2$, the total-rate parameter; but for convenience we use the scaled modification ψ with the related nuisance parameter λ given by a rotation from (φ_1, φ_2):

$$\begin{pmatrix} \psi \\ \lambda \end{pmatrix} = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{pmatrix} \begin{pmatrix} \varphi_1 \\ \varphi_2 \end{pmatrix}$$

The same transformation gives the canonical variable (s,t) from the given (y_1, y_2); the reexpressed model is then

$$f(s,t;\psi,\lambda) = \frac{1}{2}(\psi^2 - \lambda^2)\exp(-\psi s - \lambda t)$$

on the quadrant opening symmetrically to the right from the origin $(s,t) = (0,0)$; the observed data point is $(s^0, t^0) = (\sqrt{2}, 0)$ with maximum likelihood value $(\sqrt{2}, 0)$ and observed information I. From the observed log-likelihood function: $\ell(\psi,\lambda) = \log(\psi^2 - \lambda^2) - \psi\sqrt{2}$ we find that the third order likelihood characteristic $\jmath_{\lambda\lambda\psi}(\psi,\lambda)$ has value $a = -\sqrt{2}$ at the maximum likelihood value $(\sqrt{2}, 0)$.

The p-value from $s|t$ for assessing ψ is given as probability left of the data relative to ψ; as $s|t$ is stochastically decreasing relative to ψ with density $\psi\exp(-\psi s)$ we thus obtain the p-value $p(\psi) = pr(s > s^0|t) = \exp(-\sqrt{2}\psi)$. The third order p-value from S calculated using Barndorff-Nielsen's $r^* = r - r^{-1}\log(r/q)$ uses the signed log-likelihood ratio $r = \text{sign}(\hat{\psi} - \psi)[\{2(\hat{\ell} - \hat{\ell}_\psi)\}]^{1/2}$ and the information adjusted maximum-likelihood departure $q = (\hat{\psi} - \psi)|\hat{\jmath}_{\varphi\varphi}|^{1/2}/\jmath_{\lambda\lambda}^{1/2}(\hat{\varphi}_\psi)$; these with observed data values substituted give

$$r = \text{sign}(\hat{\psi} - \psi)\{2(\sqrt{2}\psi + \log 2 - 2\log\psi - 2)\}^{1/2}, \qquad q = \frac{(\sqrt{2} - \psi)\psi}{\sqrt{2}}.$$

For four moderate values of the total failure rate $\varphi_1 + \varphi_2$ we now record the p-values from the variable $s|t$ and from the variable S. Among these, the value $\psi = 1.4142$ is the maximum likelihood value $\sqrt{2}$ and has both r and q equal to zero making r/q in the Barndorff-Nielsen formula somewhat problematic; this is a general concern for third order calculations when made at or near the maximum likelihood value; routes around are available (for example Fraser et al, 2003) but simple limits here show that $r^* = 4/3\sqrt{2}$ giving the value -0.9428.

$\varphi_1 + \varphi_2$	1	1.5	2	2.5
ψ	0.7071	1.0607	1.4142	1.7678
$p_{s\|t}(\psi)$	36.8%	22.3%	13.5%	8.2%
r	0.8790	0.3882	”0”	-0.3279
q	0.3536	0.2652	”0”	-0.4419
r^*	-0.1571	-0.5938	-0.9428	-1.2397
$p_s(\psi)$	43.8%	27.6%	17.3%	10.8%

The two p-value calculations clearly give substantially different numbers. The $s|t$ calculation is of course an analytic answer. The S calculation is a third order approximation; such third-order calculations have an extensive literature demonstrating high accuracy, as indicated in the citations in preceding sections and in the references from those citations. However, the $p_{s|t}(\psi)$-value calculation here is for $t = 0$ which labels the line through the vertex $(0, 0)$ of the range for (s, t), and can be viewed as extreme. The variation of such $p_{s|t}(\psi)$ values over repetitions on the conditioning variable t represents noise in an evaluation of “where the data is relative to the parameter” and such variation is averaged out in the determination of the $p_s(\psi)$-value. We thus have a situation somewhat parallel to that with an ancillary, where a conditional p-value is viewed as more sensitive to information concerning accuracy; but here the conditioning variable t actually has a distribution with dependence on the interest parameter and thus is not ancillary. If there is question concerning conditioning on an ancillary one could say there is bigger question when the ancillary morphs into interest-parameter dependence. Here our primary concern is just that the conditional p-value is not obtained by the usual saddlepoint calculation.

8. DISCUSSION

We have examined higher order p-values in the specialized context of an exponential model with a canonical interest parameter ψ, where classical methods are available in the form of most powerful unbiased tests. The saddlepoint-based distribution (10) for the higher order statistic S has a root information $|j^{\psi\psi}(s,0)|^{-1/2} = |j_{\varphi\varphi}(s,0)|^{1/2}|j_{\lambda\lambda}(s,0)|^{-1/2}$ that distinguishes the distribution from the original joint distribution (3) on the section designated L; as a consequence a nuisance information factor $j_{\lambda\lambda}^{1/2}(\hat{\varphi}_\psi)$ appears as an adjustment to the maximum likelihood departure q cited in the Example. This adjustment arises in the third order derivation of the corresponding r^* as part of the Laplace integration elimination of nuisance parameter effects.

The Neyman-Pearson theory for assessing a scalar interest parameter in the presence of nuisance parameters tends to focus on unbiased or similar tests (see Lehman & Romano, 2005) and are found typically in the exponential model context discussed in this paper. Such tests are essentially conditional tests given the nuisance score or canonical variable. A frequent approach to such conditional tests is by a saddlepoint approximation. We have shown that this saddlepoint approximation actually gives the third order likelihood p-value that is available quite generally with moderate regularity. Thus an inadvertent choice of saddlepoint leads to a marginal distribution rather than to a conditional distribution, and it is the marginal distribution approach that is available widely.

This also raises a question as to how finely to condition in statistical inference and in particular whether to condition at all in the exponential model context where the interest parameter affects the conditioning variable; there seems to be strong argument in favor of not conditioning in the present exponential model context. A strong advantage of the marginal approach using $f(S)$ is that it is applicable generally with just moderate regularity.

ACKNOWLEDGEMENT

The author acknowledges the support of the Natural Sciences and Engineering Research Council of Canada. Very special thanks to Kexin Ji for extensive support in the development of the present material and and for the write up; an extended version of the calculations by Kexin is available as http://www.utstat.toronto.edu/dfraser/documents/260ext.pdf

REFERENCES

Andrews, D.F., Fraser, D.A.S. and Wong, A. (2005). Computation of distribution functions from likelihood information near observed data. *Journal Statist Plann. Inference*, **134**, 180-193.

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

Barndorff-Nielsen O.E. (1986 Inference on full or partial parameters based on the standardized, signed log likelihood ratio. *Biometrika* **73** 307-22.

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

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

Davison, A.C. (1988). Approximate conditional inference in generalized linear models. *J. R. Statist. Soc., B***50**. 445-61.

Fraser, D.A.S. (2003). Likelihood for component parameters. *Biometrika* **90**, 327-339.

Fraser, D.A.S. (2004). Ancillaries and conditional inference. *Statistical Science* **19**, 333-369.

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

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

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

Fraser, D.A.S. and Reid, N. (2001). Ancillary information for statistical inference. In S.E. Ahmed and N. Reid (Eds), *Empirical Bayes and Likelihood Inference*, 185-209. New York: Springer-Verlag.

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

Fraser, D.A.S., Reid, N. and Wong, A. (1991). Exponential linear models: a two pass procedure for saddle point approximation. *J. Royal Statist. Soc.* **B53**, 483-492.

Fraser, D.A.S., Reid, N., Li, R., and Wong, A.(2003). p-value formulas from likelihood asymptotics: Bridging the singularities. *J. Statist. Research.* **37**, 1-15.

Fraser, D.A.S., Reid, N. and Wu, J. (1999). A simple general formula for tail probabilities for frequentist and Bayesian inference.

Fraser, D.A.S., Wong, A. and Wu, J. (1999). Regression Analysis, Nonlinear or Nonnormal: Simple and accurate p-values from Likelihood Analysis. *Journal American Statistical Assocociation* **94**, 1286-1295.

Lehman, E.L. and Romano, J.P. (2005) *Testing Statistical Hypotheses*, New York: Springer.

[*Received January* 2008. *Revised March* 2009]