Exponential Models: Approximations for Probabilities

D. A. S. Fraser[1,2]**, A. Naderi**[3]**, Kexin Ji**[1]**, Wei Lin**[1]**, Jie Su**[1]

[1]Department of Statistics, University of Toronto, Ontario, Canada.
[2]Department of Statistical and Actuarial Sciences, University of Western Ontario, Ontario, Canada.
[3]Department of Mathematical Sciences, Isfahan University of Technology, Isfahan, Iran.

Abstract. Welch & Peers (1963) used a root-information prior to obtain posterior probabilities for a scalar parameter exponential model and showed that these Bayes probabilities had the confidence property to second order asymptotically. An important undercurrent of this indicates that the constant information reparameterization provides location model structure, for which the confidence property was and is well known. This paper examines the role of the scalar-parameter exponential model for obtaining approximate probabilities and approximate confidence levels, and then addresses the extension for the vector-parameter exponential model.

D.A.S. Fraser (✉)(dfraser@utstat.toronto.edu), A. Naderi(amir@cc.iut.ac.ir), Kexin Ji(kexinj@gmail.com), Wei Lin(becky@utstat.toronto.edu), Jie Su(sujie2000@hotmail.com)

1   Introduction

For assessing the value of a scalar parameter θ the departure of an estimate say $\bar{y}$ from the parameter value of interest can often be standardized $t = (\bar{y} - \theta)/s_{\bar{y}}$ and then examined against the standard Normal approximation suggested by the Central Limit Theorem. This then gives an approximation for the %-age position of the data relative to the parameter value, otherwise known as the p-value. We thus obtain the approximation $p(\theta) = \Phi(t)$ where Φ is the standard Normal distribution function. Such an observed p-value function $p(\theta)$ makes available of course all confidence intervals that respect direction on the range for the parameter, thus making the p-value function $p(\theta)$ the central expression of statistical inference. For this, the Normal distribution is providing a very useful approximation, an approximation for the distribution of a sum or average or other asymptotic variable, and as indicated then leads to approximate p-values, the key summary in statistical inference. But the Normal however as a tool has just two free parameters to facilitate the appropriate fitting to approximate a distribution.

So now we consider an exponential model with a scalar parameter θ which can be written as

$$g(s; \varphi) = \exp\{\varphi s - \kappa(\varphi)\} h(s) \tag{1}$$

where $\varphi = \varphi(\theta)$ is called the canonical parameterization, and often an antecedent data variable y with $s = s(y)$ is part of the background with the factor $h(s)$ frequently not easily available. By contrast to the Normal however this exponential has much greater flexibility, with perhaps eight free mathematical parameters hidden in the functions $h(s)$ or $\kappa(\varphi)$, as examined to third order $O(n^{-3/2})$. Thus we might reasonably have hope that it can make available more accurate approximations and thus more accurate p-values. But this raises two concerns: How to fit the exponential model; And how to obtain the needed distribution function for that exponential model. Some answers to these questions are now evolving from saddlepoint analysis and higher-order likelihood analysis. We briefly review these directions and then address extensions for the related Welch-Peers large sample theory for the exponential model.

2   Approximating the Distribution Function of an Exponential Model

Let's start with the second concern mentioned in the Introduction: how to obtain the distribution function say $H(s^0; \varphi)$ for a scalar exponential model at an observed data value $s^0 = s(y^0)$. If the variable s is stochastically increasing in φ the observed distribution function $H(s^0; \varphi)$ is the p-value: the $p(\varphi)$ of focal interest for assessing a parameter value φ and the primitive for a wealth of one sided and two-sided modifications that might be considered in various contexts. Stochastically increasing means that if a parameter value is increased then the distribution shifts to the right, meaning probability left of a data value as given by the observed distribution function decreases. In such cases we can think loosely of an observed data value as estimating the parameter, and correspondingly think of a distribution function value as giving the statistical or %-age position of the data value here s relative to the parameter value φ on the scale of that parameter. And for convenience here we assume that $\varphi(\theta)$ is increasing in θ; this avoids rethinking direction when considering how data position relates to a reparameterized value θ, and thus what %-age position means when we speak of left of a data value. The highly accurate distribution function approximation is

$$p(\theta) = \Phi\{r - r^{-1}\log(r/q)\} = \Phi\{r^*\} \tag{2}$$

where r is a signed likelihood root departure, q is a maximum likelihood departure calculated from the observed log-likelihood function $\ell(\varphi) = \log f(s^0; \varphi)$ and r^* is implicitly defined. For this let $\hat{\varphi}$ be the value that maximizes $\ell(\varphi)$ and let $\hat{\jmath}$ be the corresponding curvature of $\ell(\varphi)$ as given by the negative second derivative $\hat{\jmath}_{\varphi\varphi} = -\ell_{\varphi\varphi}(\hat{\varphi}) = -(\partial^2/\partial\varphi\partial\varphi)\ell(\varphi)|_{\hat{\varphi}}$ at that maximum value; these can be viewed as being very fundamental characteristics of an available likelihood function. We then have the definitions

$$\begin{aligned} r &= r(\varphi; s^0) = \text{sign}(\hat{\varphi} - \varphi)[2\{\ell(\hat{\varphi}) - \ell(\varphi)\}]^{1/2}, \\ q &= q(\varphi; s^0) = \text{sign}(\hat{\varphi} - \varphi)\hat{\jmath}_{\varphi\varphi}^{1/2}|\hat{\varphi} - \varphi|. \end{aligned} \tag{3}$$

which in themselves provide two ways of measuring departure of observed data from the parameter value φ.

Thus if we have an exponential model and want its distribution function value for given data at some parameter value of interest we can in effect just directly calculate it using (2); this formula (2) can give incredible accuracy. These results come from a density function approximation

(Daniels, 1954), a distribution function approximation (Lugannani & Rice, 1980), a modification of the preceding (Barndorff-Nielsen, 1986), and an extension in Fraser & Reid (1993), all with an overview in Bédard, Fraser & Wong (2008). A minor computational anomaly can occur with (2) close to the maximum likelihood value where r and q are equal to 0; but such a p-value at the center of the statistical range for a parameter is rarely of direct interest, unless some resulting computationally-bad p-values are ignored in related simulations.

3 Deriving the Distribution Function Approximation

The distribution function approximation can be obtained of course by integrating the density function (1), but this means that the ingredient $h(s)$ must be available and accessible; and often it isn't, when an underlying variable y is the primary observational variable. To access $h(s)$ we employ a third-order equivalent for the marginal density for $s(y)$ which is available by the saddlepoint of Daniels (1954), or by the p^* of Barndorff-Nielsen (1980) which is discussed in the next section. In several forms, all subject to a normalizing constant k/n, these provide

$$
\begin{aligned}
g(s;\varphi)ds &= \exp(k/n)(2\pi)^{-1/2}\exp\{\ell(\varphi;s)-\ell(\hat{\varphi};s)\}\hat{\jmath}_{\varphi\varphi}^{-1/2}ds \\
&= \exp(k/n)(2\pi)^{-1/2}\exp\{-r^2/2\}\hat{\jmath}_{\varphi\varphi}^{-1/2}ds \\
&= \exp(k/n)\phi(r)\hat{\jmath}_{\varphi\varphi}^{-1/2}ds;
\end{aligned}
\tag{4}
$$

where $\phi(r)$ is the standard Normal density function and $-r^2/2$ is implicitly defined and corresponds r in (3). Of course this approximation in comparison with (1) clearly has the correct likelihood $L(\varphi;s) = \exp\{\ell(\varphi;s)\} = \exp\{\varphi(\theta)s-\kappa(\varphi)\}$ at each data point s; and the remainder of the approximation involving $h(s)$ is clarified in the next section. First we take the differential of $r^2/2$ relative to s noting that $\hat{\varphi}$ itself is also function of s:

$$
\begin{aligned}
dr^2/2 = rdr &= d\{\ell(\hat{\varphi};s)-\ell(\varphi;s)\} \\
&= \ell_\varphi(\hat{\varphi};s)d\hat{\varphi} + \{\ell_{;s}(\hat{\varphi};s)-\ell_{;s}(\varphi;s)\}ds \\
&= 0d\hat{\varphi} + (\hat{\varphi}-\varphi)ds
\end{aligned}
$$

where the subscript $;s$ denotes differentiation with respect to the second variable s gives φ, and the differentiation with respect to φ at the maximum likelihood $\hat{\varphi}$ gives 0. We can then write $ds = \{r/(\hat{\varphi}-\varphi)\}dr$ and

then combine $(\hat{\varphi} - \varphi)$ and $\hat{\jmath}^{1/2}$ in (4) to obtain q defined at (3), giving

$$\tilde{g}(s; \varphi)ds = \exp(k/n)\phi(r)(\frac{r}{q})dr. \tag{5}$$

We then take r/q to the exponent of the Normal density as $\log(r/q)$ and force a completion of the square in the exponent:

$$-r^2/2 + \log(r/q) = -\{r - r^{-1}\log(r/q)\}^2/2 = -(r^*)^2/2,$$

where it can be shown that $r/q = 1$ to first order and $\{r^{-1}\log(r/q)\}^2$ is constant c/n to second order. With minor details this shows that r^* is standard Normal to third order and verifies the p-value at (2).

A simple example. Consider a very simple model $g(s; \theta) = \theta^{-1}\exp(-s/\theta)$ on $(0, \infty)$ with data $s^0 = 1$. The parameter θ is the expectation parameter of the model, which is the mean life $E(s; \theta) = \theta$ and the variable s is easily seen to be stochastically increasing in θ. The corresponding p-value $p(\theta)$ thus becomes the observed value of the distribution function, and is easily calculated exactly: $p_{Ex}(\theta) = F(1; \theta) = \int_0^1 \theta^{-1}\exp(-s/\theta)ds = 1 - \exp(-1/\theta)$.

The approximate value as described before the example is expressed in terms of φ, which for this example can conveniently be taken as the rate parameter $\varphi(\theta) = \theta^{-1}$, provided we view it to be the coefficient of $-s$. This does involve a reversal of direction of parameter change, as this choice of φ is monotone decreasing in θ; accordingly we use the sign of $\hat{\theta} - \theta$ in the formulas of r and q. We then have $L(\theta) = \theta^{-1}\exp\{-\theta^{-1}\}$ and $\ell(\theta) = \log\varphi - \varphi$ for the observed $s^0 = 1$ with $\hat{\theta}^0 = 1$, $\hat{\varphi}^0 = 1$, $\hat{\jmath}^0 = 1$. This leads to

$$\begin{aligned} r^2/2 &= \ell(\hat{\theta}) - \ell(\theta) = \log\hat{\varphi} - \hat{\varphi} - \log\varphi + \varphi = \varphi - 1 - \log\varphi \\ r &= \text{sign}(1 - \theta)[2\{\ell(\hat{\theta}) - \ell(\theta)\}]^{1/2} = \text{sign}(1 - \theta)\{2(\varphi - 1 - \log\varphi)\}^{1/2} \\ q &= \text{sign}(1 - \theta)\hat{\jmath}^{1/2}|1 - \varphi| = \varphi - 1. \end{aligned}$$

We record the exact p-value p_{Ex} and the corresponding r, q, r^* with the resulting approximate p-value $p_{App}(\theta) = \Phi(r^*)$ for a few values of the interest parameter θ.

θ	2	5	10	20
$\varphi(\theta)$	0.5	0.2	0.1	0.05
$p_{Ex}(\theta)$	39.35%	18.13%	9.52%	4.88%
r	-0.6215	-1.2724	-1.6749	-2.0227
q	-0.50	-0.80	-0.90	-0.95
r^*	-0.2715	-0.9077	-1.3041	-1.6491
$p_{App}(\theta) = \Phi(r^*)$	39.30%	18.20%	9.61%	4.96%

The third order approximate *p*-values are very close to the exact values; this happens in quite wide generality. And just to see what the values would be without the use of the exponential model as the primary approximation tool we record the values from the Normal approximation applied to the likelihood ratio *r* and the maximum likelihood value *q*:

θ	2	5	10	20
$\Phi(r)$	26.71%	10.16%	4.70%	2.16%
$\Phi(q)$	30.85%	21.19%	18.41%	17.11%
$p_{Ex}(\theta)$	39.35%	18.13%	9.52%	4.88%
$p_{App}(\theta) = \Phi(r^*)$	39.30%	18.20%	9.61%	4.96%

Clearly the familiar Normal approximations can be extremely aberrant relative to the exact, and the exponential model approximation can be very close to the exact.

We have examined the case of a scalar variable and scalar full parameter of interest. For the more general case of a scalar interest parameter ψ with *p*-dimensional full parameter (ψ, λ) and *n*-dimensional variable *y* preceding the canonical variable $s(y)$ some minor extras are needed such as the Laplace elimination of nuisance parameter effect and a directional derivative on the $\{y\}$-space to obtain the canonical parameter $\varphi(\theta)$; for a recent survey see Bédard, Fraser & Wong (2008).

4   Saddlepoint and p^* Approximate Densities

Now consider the first concern: How to find the exponential model that approximates some given model. Consider the statistical model $f(y; \theta)$ and an observed data y^0 value. We want to approximate this model near observed data y^0 by making use of recent asymptotic theory. The method is simple and is a matter of just determining the change in the log-likelihood under small change in the data at the observed data y^0; in other words it is to use the derivative of the log-likelihood at the observed data point y^0. It is as if one examined the log-likelihood at the

data and then at a nearby point and just took the difference, in itself a trivial calculation, just a subtraction! This immediately gives us the canonical parameter $\varphi(\theta)$ of the approximating exponential model; and this then with the observed log-likelihood presents the full approximating exponential model.

For the exponential model (1) with an antecedent variable y traditional statistics would appeal to sufficiency and focus for statistical inference on just the marginal model for s as recorded (1). Of course sufficiency is an "easy-out" for a Normal model or for an exponential. But it only works for such specialized models (for example, Fraser, 2004) while directional conditioning now being described works trivially in such preceding nice cases and but also works widely and routinely. What is needed (Fraser, Fraser & Staicu, 2010) is to write the full model in quantile form $y = y(\theta; u)$ where u is the randomness defined by the coordinate distribution functions, and then determine the directions $V = (v_1, ..., v_p) = \partial y / \partial \theta |_{y^0; \hat{\theta}^0}$ in which θ affects y at the observed data y^0 and corresponding maximum likelihood value $\hat{\theta}^0$; for this p as used here is just the dimension of the parameter.

Then, if you want accuracy going far beyond the first-order Normal approximation, this extra difference or derivative step is relatively minor, painless, and fruitful, and does provide incredible results relative to say Bayes, Bootstrap and McMC (Bédard, Fraser & 2008). The third order approximating exponential model (Reid & Fraser, 2010) is obtained by differentiating the log-model in the directions V,

$$\varphi(\theta) = \frac{\partial}{\partial s} \ell(\theta; y^0 + Vs)|_{s=0}, \tag{6}$$

and then by using this as the canonical parameter $\varphi(\theta)$ of the exponential model (1). This works for vector as well as scalar parameters; see Fraser, Wong & Sun (2009) for some interesting examples.

The approximating model can then be written to third order accuracy in the saddlepoint form

$$\tilde{g}(s; \varphi)ds = \frac{\exp(k/n)}{(2\pi)^{p/2}} \exp\{\ell(\varphi) + \varphi' s\} |\hat{j}_{\varphi\varphi}(s)|^{-1/2} ds, \tag{7}$$

where $\hat{j}_{\varphi\varphi}(s)$ is the information matrix calculated from the log-likelihood $\ell(\varphi) + \varphi' s$ and $s = s^0 = 0$ is the observed data value. This was developed in Fraser & Reid (1993, 1995) and Reid & Fraser (2010), and called the Tangent Exponential Model.

A related saddlepoint-type approximation called p^* for general models $f(y; \theta)$ with asymptotic properties was developed by Barndorff-Nielsen

(1980) and has the parametrization invariant form

$$p^*(\hat{\theta};\theta)d\hat{\theta} = \frac{e^{k/n}}{(2\pi)^{p/2}} \exp\{\ell(\theta;y) - \ell(\hat{\theta};y)\}|\hat{\jmath}_{\theta\theta}(y)|^{1/2}d\hat{\theta} \qquad (8)$$

along a contour of an ancillary, which is a statistic with distribution free of the parameter θ. This approximation can be verified by Laplace integration. Take a general data point say y_0 on the particular ancillary contour and do Laplace integration with respect to $\hat{\theta}$ using the parameter value $\theta = \hat{\theta}(y_0) = \hat{\theta}_0$. This determines the density function at the maximum giving $|-\ell_{\hat{\theta}\hat{\theta}}(y_0)|^{1/2}$ rather than the quoted root-information in (8). But the two root informations differ by a factor $1 + k/n$, as established for the location parameter choice in Cakmak, Fraser & Reid (1994) and Cakmak, Fraser, McDunnough, Reid & Yuan (1998). This can then be repeated for other data and corresponding maximum likelihood values, which thus verifies (8) to third order.

For the special exponential model (7) case there is a minor change of variable relative to expression (8). With the exponential model, the log-density can be written $\ell(\varphi) + \varphi's$ where $\ell(\varphi)$ is the log-density at some particular data value and φ and s are then taken to be departures from reference values provided by that particular data value. In this notation the maximum likelihood value $\hat{\varphi}$ is the solution of the score equation $\ell_\varphi(\hat{\varphi}) + s = 0$. Thus differentiating with respect to s gives $\ell_{\varphi\varphi}(\hat{\varphi})d\hat{\varphi} + ds = 0$ or $d\hat{\varphi} = \hat{\jmath}_{\varphi\varphi}^{-1}ds$. This gives the change from (8) to (7) when the parameter φ is used in the information.

5 Welch-Peers Analysis

Welch-Peers (1963) used a root-information prior with an exponential model (1) and showed that resulting posterior intervals were approximate confidence intervals to second order. This was a fundamental contribution then to the Bayes-frequentist dialogue, and has provided clear downstream influences in some current procedures for determining default priors (Fraser, Reid, Marras & Yi, 2010). Perhaps, however, there is yet deeper significance in the Welch-Peers derivation, that it effectively shows that a scalar exponential model is a location model to second order. At the time of the Welch-Peers result it was well known that location models with default Bayes priors gave intervals with the confidence property. Indeed Lindley (1958) had shown this as part of a critique of the confidence procedure; but as a deeply committed Bayesian

he had somehow treated the result as a criticism of confidence, despite wide recognition that confidence had full repetition reliability (Neyman, 1937).

Consider the scalar parameter exponential model (1) with maximum likelihood $\hat{\varphi}$ and information $\hat{j}_{\varphi\varphi}^{1/2}$, both from data s. We define what is called the constant information parameterization $\beta(\varphi)$:

$$\beta = \beta(\varphi) = \int_{\varphi_0}^{\varphi} \hat{j}_{\varphi\varphi}^{1/2}(\varphi)d\varphi \tag{9}$$

where φ_0 is some convenient origin or reference value, and we use the argument φ in place of the usual $\hat{\varphi}$. We then use $d\varphi = \hat{j}_{\varphi\varphi}^{-1/2}(\varphi)d\beta$ to calculate the information function in the new parameterization and obtain

$$\hat{j}_{\beta\beta} = -\ell_{\beta\beta}(\hat{\varphi}) = \hat{j}_{\varphi\varphi}^{-1/2}\{-\ell_{\varphi\varphi}(\hat{\varphi})\}\hat{j}_{\varphi\varphi}^{-1/2} = \hat{j}_{\varphi\varphi}^{-1/2}\hat{j}_{\varphi\varphi}\hat{j}_{\varphi\varphi}^{-1/2} = 1$$

with some technicalities. Thus β has constant information.

Now consider a scalar parameter regular statistical model $f(y;\theta)$ with asymptotic properties; we can approximate the log-model in a neighborhood of $(y_0, \hat{\theta}_0) = (y_0, \hat{\theta}(y_0))$ by a Taylor series expansion. Cakmak, Fraser, McDunnough, Reid & Yuan (1998) examined this expansion using departures standardized by observed information and found that it can be expressed to second order equally in location model form or in exponential model form, and also to the third order but with a technicality that for statistical inference is inconsequential; this provides theoretical background for the third order inference results mentioned in Section 3. But for such a location model the constant information reparameterization (9) is precisely the location parameterization; thus we have a direct connection between exponential model form and location model form and the constant reparameterization (9) provides the direct link; there is a similar linkage for the vector parameter case as initiated in Cakmak, Fraser & Reid (1994), with further aspects to be discussed in the concluding Section 6.

For the scalar parameter case the linkage says that $z = \beta(\hat{\varphi}) - \beta(\varphi)$ has a fixed distribution to second order, and thus the expansion gives

$$g(z) = (2\pi)^{-\frac{1}{2}}\exp\{-\frac{1}{2}z^2 - \frac{a_3}{12n^{1/2}}z^3\},$$

where the third derivative $a_3/n^{1/2}$ is written with dependence on the nominal sample size n made explicit. The total differential relationship $dz = d\hat{\beta} - d\beta$ thus gives the transfer of probability $g(z)dz$ to the

variable y as probability $g(z)d\hat{\beta}$ for given β, and the transfer of probability $g(z)dz$ to β as the confidence $g(z)d\beta$, which makes explicit the confidence calculation implicit in the Welch-Peers analysis.

6 The Exponential-Location Connection by Taylor

The Welch-Peers analysis provides a simple link between a scalar exponential model and a scalar location model. Towards extending this to vector parameter models we first exhibit the link in terms of the Taylor series expansions used at the end of the preceding section. For this we follow Cakmak, Fraser, McDunnough, Reid & Yuan (1998) but drop terms of order $O(n^{-1})$ and thus work just to second order. A scalar asymptotic log-model can be centered at a point $(y_0, \hat{\theta}_0)$ using departures standardized by observed information and can then have the variable and parameter reexpressed giving the second order asymptotic exponential form

$$f_E(s; \varphi) = \frac{1}{(2\pi)^{1/2}} \exp\{-\frac{1}{2}(s - \varphi)^2 - \frac{\alpha}{6n^{1/2}}\varphi^3 + \frac{\alpha}{6n^{1/2}}s^3 - \frac{\alpha}{2n^{1/2}}s\}, (10)$$

where the Normal quadratic comes from the limiting Normal distribution of the standardized departure, the coefficients of the two cubic terms are equal because of density determination, and other cubic terms are absent due to the reexpression in the exponential model form. In a related way the scalar asymptotic log-model can be centered at a point $(y_0, \hat{\theta}_0)$ using departures standardized by observed information and can then have the variable and parameter reexpressed giving the second order asymptotic location form

$$f_L(\tilde{y}; \tilde{\theta}) = \frac{1}{(2\pi)^{1/2}} \exp\{-\frac{1}{2}(\tilde{y} - \tilde{\theta})^2 + \frac{a}{6n^{1/2}}(\tilde{y} - \tilde{\theta})^3\}. \tag{11}$$

If the two standardized models, $f_E(s; \varphi)$ and $f_L(\tilde{y}; \tilde{\theta})$ represent the same underlying statistical model as indicated above then the two variables are different and the two parameterizations are different, although of course agreeing to first derivative at the expansion point. If we examine the cross terms in (11) and identify them with the same in (10) we obtain

$$\tilde{y}\tilde{\theta} - \frac{a}{2n^{1/2}}\tilde{y}^2\tilde{\theta} + \frac{a}{2n^{1/2}}\tilde{y}\tilde{\theta}^2 = \varphi s,$$

and can then solve for φ and s to second order:

$$\varphi = \tilde{\theta} + a\tilde{\theta}^2/2n^{1/2} \qquad\qquad \tilde{\theta} = \varphi - a\varphi^2/2n^{1/2} \qquad (12)$$
$$s = \tilde{y} - a\tilde{y}^2/2n^{1/2} \qquad\qquad \tilde{y} = s + as^2/2n^{1/2}.$$

Also we can look at the log-likelihood at the expansion point in (11) after substituting $\tilde{\theta} = \varphi - a\varphi^2/2n^{1/2}$ and obtain

$$-\frac{1}{2}\tilde{\theta}^2 - \frac{a}{6n^{1/2}}\tilde{\theta}^3 = -\frac{1}{2}\varphi^2 + \frac{a}{2n^{1/2}}\varphi^3 - \frac{a}{6n^{1/2}}\varphi^3 = -\frac{1}{2}\varphi^2 + \frac{2a}{6n^{1/2}}\varphi^3$$

which implies that $\alpha = -2a$, which is in agreement with Cakmak et al (1998). We can also calculate the observed information from (10) and obtain $\jmath_{\varphi\varphi}(s) = 1 + \alpha s/n^{1/2}$ giving $\jmath_{\varphi\varphi}^{1/2}(s) = 1 + \alpha s/2n^{1/2} = 1 - as/n^{1/2}$, and thus the Welch-Peers probability differential is $\jmath_{\varphi\varphi}^{1/2}d\varphi = (1 - a\varphi/n^{1/2})d\varphi$. We can then integrate $\jmath_{\varphi\varphi}^{1/2}d\varphi = (1 - a\varphi/n^{1/2})d\varphi$ and obtain $\varphi - a\varphi^2/2n^{1/2}$, in agreement with the location parameter $\tilde{\theta}$ in (12).

7 Discussion

We have examined the exponential model as a rich tool for approximating probabilities in a wide segment of inference contexts and we have discussed the Welch-Peers root information adjustment as a devise to obtain confidence intervals calculated by a Bayesian procedure. These uses of the exponential model can begin with a multivariate exponential but for the actual calculation are focussed on scalar exponential models. Our prime concern is to develop methods and techniques for the multivariate and multiparameter context. Accordingly our development here has been to survey and assemble the methods and theory in the scalar case to provide a basis for the broader examination of the vector case.

Acknowledgement

X. Li, W. Lin, J. Su, W. Xin, K. Xue, W. Zeng. And very special thanks to a careful and insightful referee for many positive recommendations.

References

Barndorff-Nielsen, O. E. (1980), Conditionality resolutions. Journal of Royal Statistical Society, Series. B, **67**, 293-310.

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

Bédard, M., Fraser, D. A. S., and Wong, A. (2008), Higher accuracy for Bayesian and frequentist inference: Large sample theory for small sample likelihood. Statistical Science, **22**, 301-321.

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.

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

Daniels, H. E. (1954), Saddlepoint approximations in statistics. Annals Math. Statist., **25**, 631-50.

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

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., Reid, N., Marras, E., and Yi, G. Y. (2010), Default priors for Bayesian and frequentist inference. Journal of Royal Statistical Society, Series. B, **75**, 631-654.

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

Fraser, D. A. S., Wong, A. and Sun, Ye (2009), Three enigmatic examples and inference from likelihood. Canadian Journal of Statistics, **37**, 161-181.

Lindley, D. V. (1958), Fiducial distributions and Bayes' theorem. J. Roy. Statist. Soc., B, **20**, 102-107.

Lugannani, R. and Rice, S. (1980), Saddle point approximation for the distribution of the sum of independent variables. Adv. Appl. Prob., **12**, 475-90.

Neyman, J. (1937), Outline of a theory of statistical estimation based on the classical theory of probability. Phil. Trans. Roy. Soc., A, **237**, 333-380.

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

Welch, B. L. and Peers, H. W. (1963), On formulae for confidence points based on intervals of weighted likelihoods. J. Roy. Statist. Soc., B, **25**, 318–329.