Journal of Statistical Research *ISSN 0256 - 422 X*
2008, Vol. 42, No. 2, pp. xx-xx
Bangladesh

IS r^* LINEAR IN r?

D.A.S. Fraser

Department of Statistics
University of Toronto, Toronto, Canada M5S 3G3
Email: dfraser@utstat.toronto.edu

L. Faye, Y. Cai, K. Ji, M. Mallo, and R. Thinniyam

Department of Statistics
University of Toronto, Toronto, Canada M5S 3G3
Email:

SUMMARY

The quantity $r^*(\psi, y)$ was introduced by Barndorff-Nielsen, in part as a distributional refinement of the signed likelihood root $r(\psi, y)$, a refinement that can also be approximated by a mean and standard deviation adjustment of the root $r(\psi, y)$. We clarify: that r^* is not linear in r; that r^* achieves large distributional improvement on $r(\psi, y)$; and that r^* provides the definitive separation of inference information concerning scalar component parameters of a statistical model. These distributional and inference properties deserve broader awareness.

Keywords and phrases: higher order likelihood, meta-parameter, p-value, parameter assessing, signed likelihood root, survivor function, tail probability

AMS Classification: 62E20 62F99

1 Introduction

Is r^* linear in r? Well, Yes and No! If the variable is held fixed then to third order r^* is linear in r. And if the parameter is held fixed then to third order r^* is linear in r. But collectively r^* is not linear in r. More specifically we have linearity to the third order, if and only if the Bayes posterior has a confidence interpretation, that is, if and only if the variable is location related to the parameter.

The signed likelihood root $r = r(\psi, y)$ for a scalar parameter $\psi(\theta)$ is a normal-quantile re-expression of the rise of log likelihood from the value ψ to the maximum likelihood value $\hat{\psi}$; thus if we set the rise $\ell(\hat{\theta}; y) - \ell(\hat{\theta}_\psi; y)$ equal to $r^2/2$ and solve for r we obtain

$$r = \text{sign}(\hat{\psi} - \psi)[2\{l(\hat{\theta}; y) - l(\hat{\theta}_\psi; y)\}]^{1/2},$$

where $\hat{\theta}$ and $\hat{\theta}_\psi$ are the maximizing values generally and subject to the constraint $\psi(\theta) = \psi$. Central limit theorem conditions then show that r is standard normal to first order; and thus show that $p(\psi, y) = \Phi\{r(\psi, y)\}$ is Uniform $(0, 1)$, where $\Phi(z)$ is the standard normal distribution function. Accordingly, if the value $y = y^0$ is substituted in $p(\psi; y)$ we obtain to first order the observed p-value function $p^0(\psi)$ that records the percentage position of the data relative to a value of interest ψ. The quantity $r(\psi, y)$ can arguably be viewed as the intrinsic likelihood based measure of departure of data from a scalar parameter, and the observed $p^0(\psi)$ can be viewed as the intrinsic first order interpretation of where the data point is with respect to the parameter. So why r^*?

The modified likelihood ratio quantity $r^*(\psi, y)$ uses additional likelihood information in the form of an appropriate maximum likelihood departure $q = q(\psi, y)$; r^* then has the form

$$r^* = r - r^{-1}\log(r/q) \tag{1.1}$$

and was introduced by Barndorff-Nielson (1991) as a refinement of r. Under moderate regularity r^* is standard normal to the third order and is thus pivotal to the third order; it leads to a corresponding third order p-value $p(\psi, y) = \Phi\{r^*(\psi, y^0)\}$ that provides a highly accurate statistical position of the location of data y^0 relative to an interest value ψ.

Central to this however is the definition of the maximum likelihood departure $q(\psi, y)$, and it uses a re-calibration of the parameter somewhat paralleling the use of suitable default prior in Bayes analysis. If the model is exponential with scalar canonical parameter $\varphi(\theta)$ then $q(\theta, y)$ can be written

$$q(\theta, y) = (\varphi(\hat{\theta}) - \varphi)\hat{j}_{\varphi\varphi}^{1/2}, \tag{1.2}$$

as the maximum likelihood departure of $\varphi(\hat{\theta})$ from the canonical parameter value $\varphi = \varphi(\theta)$ but standardized by observed information $\hat{j}_{\varphi\varphi} = (-\partial^2/\partial\varphi^2)\ell(\theta)|_{\hat{\theta}}$; this follows from Daniels (1987), but the same $q(\theta, y)$ also leads (Lugannani & Rice, 1980) to an asymptotically equivalent version of r^*:

$$r^* = \Phi^{-1}\left\{\Phi(r) + \varphi(r)\left(\frac{1}{r} - \frac{1}{q}\right)\right\}.$$

We examine the connection between r^* and r for the scalar full parameter case. The more general connection between r^* and r can then be obtained as an adjustment to the scalar case using Cheah et al (1995). Thus, if the model is a p-dimensional full exponential model and $\theta^{'} = (\psi, \lambda')$ with canonical parameter $\varphi(\theta)$ then an appropriate $q(\psi, y)$ giving third order for the p-value is available (Barndorff-Nielson, 1986; Fraser & Reid, 1993). And if the model is general with dim y=n and with appropriate pivots for independent coordinates then a data dependent canonical $\varphi(\theta, y^0)$ with related $q(\psi, y)$ similarly gives third order accuracy (Fraser & Reid, 1995). Also if the model involves discrete data such as with contingency tables then again an appropriate $\varphi(\theta, y^0)$ is available (Davison et al, 2006), but this however yields only second order accuracy due to the discreteness.

Accordingly, for comparing r^* with r it suffices to consider a statistical model $f(y; \theta)$ with scalar variable and scalar full parameter, and with asymptotic properties inherited from some antecedent context with a sample-size parameter n. It is well known that r can be mean and variance corrected,

$$r^* \approx \frac{r - E(r)}{SD(r)}, \tag{1.3}$$

to give a variable that is third order standard normal and third order equivalent to r^*. This suggests that r^* is linear or more correctly affine, as a function of r. Details however show that the mean correction depends on the parameter, so strict linearity is not present in general. And also we note that while versions (1.1) and (1.3) are both third order, the first (1.1) which can be viewed as a functional expression is found typically to have numerical accuracy far exceeding that of the other (1.3). While this appears to be linear, it does need more than the term involving $r(\theta, y)$, it requires the extra term involving $r(\hat{\theta}^0, y)$ which corresponds to the Taylor series reference value ($\hat{\theta}^0$) that underlay the definition of the meta-parameters.

In Guanajuato on March 22, 2006, a conference was held to honour Barndorff-Nielson's many major contributions to statistical inference. As part of the discussion during a likelihood theory session, it was suggested by Donald Pierce that in regards to the inferential basis for r^* it may be difficult, perhaps impossible, to go beyond the points that: (a) to second order it gives the same ordering of datasets against the hypothesis as does r, and hence that, in obvious notation, (b) $\Phi(r^*) = \text{pr}(R < r; \psi, \lambda) + O_p(n^{-1})$ (Pierce & Bellio, 2006; DiCiccio, Martin & Stern, 2001). Accordingly, to clarify the focus, we examine the linearity just mentioned and argue that r^* was a major advance on r as input to statistical inference. Our principal tool involves the Taylor expansion of the logarithm $\log f(y; \theta)$ of the statistical model to third order in terms of appropriate y and θ.

For fixed data value $y = y^0$, we find in Section 2 that $r^*(\theta, y^0)$ can be expressed in terms of $r(\theta, y^0)$ to third order as

$$r^*(\theta, y^0) = \frac{\alpha_3}{6n^{1/2}} + \left(1 - \frac{3\alpha_4 - 4\alpha_3^2}{72n}\right) r(\theta, y^0) \tag{1.4}$$

where α_3 and α_4 are mathematical or meta parameters of the model of exponential type, as expanded relative to the data value y^0, and the effect of the sample size parameter n on the coefficients is made explicit; in particular α_3 and α_4 are specially standardized third and fourth derivatives of observed likelihood. This describes the r^* to r relationship on the y^0-section of the sample-cross-parameter space.

Then for fixed parameter value $\theta = \theta_0$ we find in Section 3 that $r^*(\theta_0, y)$ can be expressed in terms of $r(\theta_0, y)$ to third order as

$$r^*(\theta_0, y) = \frac{\tilde{\alpha}_3}{6n^{1/2}} + \left(1 + \frac{9\tilde{\alpha}_4 + 13\tilde{\alpha}_3^2 + 18\tilde{\gamma}}{72n}\right) r(\theta_0, y) \tag{1.5}$$

where $\tilde{\alpha}_3, \tilde{\alpha}_4, \tilde{\gamma}$ are exponential type meta parameters relative to the θ_0-section on the sample-cross-parameter space; in particular $\tilde{\alpha}_3$ and $\tilde{\alpha}_4$ are special standardized derivatives of log-density and $\tilde{\gamma}$ is a similarly standardized measure of nonexponentiality. Then also in

Section 3 we use a connection between the two types of meta parameters with θ_0 set equal to $\hat{\theta}^0$ to show that (1.5) can be rewritten as

$$r^*(\hat{\theta}^0, y) = \frac{\alpha_3}{6n^{1/2}} + \left(1 - \frac{3\alpha_4 - 4\alpha_3^2}{72n}\right) r(\hat{\theta}^0, y) - \frac{1}{12n}(3\alpha_3^2 - 2\alpha_4 + 6\gamma) r(\hat{\theta}^0, y). \tag{1.6}$$

Comparing (1.4) and (1.6) we see clearly that r^* is not in general linear in r.

In Section 4 we reexpand about a modified data point and find that for third order accuracy free of both y and θ in moderate deviations an additional term can be appended to give:

$$r^*(\theta, y) = \frac{\alpha_3}{6n^{1/2}} + \left(1 - \frac{3\alpha_4 - 4\alpha_3^2}{72n}\right) r(\theta, y) - \frac{1}{12n}(3\alpha_3^2 - 2\alpha_4 + 6\gamma) r(\hat{\theta}^0, y); \tag{1.7}$$

while this appears to be linear it does need not only the term $r(\theta, y)$ but also the term $r(\hat{\theta}^0, y)$ corresponding to the expansion point that led to the definition of the meta-parameters.

2 Taylor Expansion at the Observed Data Point

Taylor series are an incredibly powerful tool in statistics and seem essential for obtaining distribution theory results concerning likelihood methods. We go beyond the usual expansion of the log-model in terms of the parameter, and expand in terms of both variable and parameter, and for this we can draw on extensive calculations that are available in the literature. The form of an expansion in the two variables can depend heavily on the expansion point say (y_0, θ_0) as we will see. If our focus is on the model near a data point y^0, we would rather naturally use $(y_0, \theta_0) = (y^0, \hat{\theta}^0)$ where $\hat{\theta}^0 = \hat{\theta}(y^0)$ is the observed maximum likelihood value, and the expansion then would use observable characteristics of the model with data; we address this in the present section. Then in the next section we focus on distributional properties when θ has some value θ_0 of interest; we then would rather naturally use $(y_0, \theta_0) = (\hat{y}_0, \theta_0)$ where $\hat{y}_0 = \hat{y}(\theta_0)$ is the maximum density point of the θ_0 distribution.

For the expansion of $\log f(y; \theta)$ about $(y^0, \hat{\theta}^0)$ we use results from the extensive calculations in Cakmak et al (1998). First, the variable is centered; thus $y - y^0$ is taken algorithmically to be the new y. Then the parameter is centered and scaled by the observed information $\hat{j} = -\partial^2/\partial\theta^2 \log f(y^0; \theta)|_{\hat{\theta}^0}$; thus $\hat{j}^{1/2}(\theta - \hat{\theta}^0)$ is taken to be the new parameter θ. And then, the variable is rescaled so that the cross derivative $(\partial/\partial\theta)(\partial/\partial y)\log f(y; \theta)|_{(y^0, \hat{\theta}^0)} = 1$. So far this can be viewed as an attempt to fit a standard normal, and in the new coordinates the model would have an expansion beginning as

$$\log f(y; \theta) = a - \frac{1}{2}\theta^2 + \theta y,$$

to first derivative in y and second derivative in θ.

But we can do more than this to obtain a simple form for the Taylor expansion. As with Taylor series in general the form of the expansion depends heavily on the variables used for the expansion and we are of course free to choose those variables to obtain an easily understood model form. But then what target form would be helpful for our understanding? Two types come quickly to mind, the location model and the exponential model.

Following the first pattern in Cakmak et al (1998) we examine an expansion that is targetted on exponential model form. Our data point now is $y^0 = 0$ and $\hat{\theta}^0 = 0, \hat{j}^0 = 1$ and the cross derivative $\hat{\ell}^0_{\theta;y} = (\partial/\partial\theta)(\partial/\partial y)\ell(\theta;y)|_{(y^0,\hat{\theta}^0)} = 1$ where $\ell(\theta;y) = \log f(y;\theta)$. With an exponential model in canonical form the only nonzero cross derivative is $\hat{\ell}^0_{\theta;y}$ which here is equal to 1. Accordingly we reexpress the parameter to get $(\partial^2/\partial\theta^2)(\partial/\partial y)\ell = 0$, $(\partial^3/\partial\theta^3)(\partial/\partial y)\ell = 0$ at the expansion point; and also we similarly reexpress the variable taking account of Jacobian effects to get $(\partial/\partial\theta)(\partial^2/\partial y^2)\ell = 0$, $(\partial/\partial\theta)(\partial^3/\partial y^3)\ell = 0$ at the expansion point. The details for this are extensive and have been verified by computer algebra; and the steps themselves are recorded in Fraser & Wong(2002). For our purposes here the existence of the steps and the reality of the final model are the primary interest. As a result with the final choice of standardized variable and standardized parameter we have a model expressed as $\log f(y;\theta) = \sum_{i+j\le 4} a_{ij}\theta^i y^i/i!j!$ to order $O(n^{-3/2})$. The terms drop off with increasing order as powers of $n^{-1/2}$ and only terms that are $O(1)$, $O(n^{-1/2})$, $O(n^{-1})$ are recorded with terms of order $O(n^{-3/2})$ omitted and marked as -; this conforms to recent likelihood analysis. The drop off is made explicit by showing the dependence on n. Thus we have

$$(a_{ij}) = \begin{pmatrix} a + k/n & k_{01}/n^{1/2} & -\{1 + k_{02}/n\} & \alpha_3/n^{1/2} & (\alpha_4 - 3\alpha_3^2 - 6\gamma)/n \\ 0 & 1 & 0 & 0 & - \\ -1 & 0 & \gamma/n & - & - \\ -\alpha_3/n^{1/2} & 0 & - & - & - \\ -\alpha_4/n & - & - & - & - \end{pmatrix}$$

to order $O(n^{-3/2})$ where $a = -(1/2)\log(2\pi)$, $k = (3\alpha_4 - 5\alpha_3^2 - 12\gamma)/24$, $k_{01} = -\alpha_3/2$, $k_{02} = (\alpha_4 - 2\alpha_3^2 - 5\gamma)/2$, and α_3, α_4, and γ are critical model characteristics that we call mathematical or meta parameters. An intriguing property is that α_3, α_4, and γ uniquely determine the elements in the first row to the given order; this derives directly from the assumed norming property $\int f(y;\theta)dy \equiv 1$. The model to first order is standard location normal $f(y;\theta) = \phi(y-\theta)$; and to second order is exponential

$$f(y;\theta) = \phi(y-\theta)\exp\{-\alpha_3\theta^3/6n^{1/2} + \alpha_3 y^3/6n^{1/2} - \alpha_3 y/2n^{1/2}\};$$

and to third order is an expanded exponential with an added cross term $\gamma\theta^2y^2/4n$ and related compensating extras as shown in the first row.

How can we use this canonical model? An observed likelihood function is of course available immediately from data, $\ell^0(\theta) = a + \log f(y^0;\theta)$. But also the special defined canonical parameter of the above exponential type model can be directly extracted by calculating a first derivative of log-likelihood at the data; thus if designating the canonical parameter by $\varphi(\theta)$ we have

$$\varphi(\theta) = \frac{\partial}{\partial y}\log f(y;\theta)|_{y^0};$$

however, to be exactly equal to a given canonical parameter, we would need to differentiate with respect to the corresponding score variable; here differentiating with respect to an original data variable gives the same result save for a constant factor. Thus the magic of the sample space derivative of likelihood, or at least part of the magic! We can thus write our canonical model as

$$f(y;\theta) = \frac{e^{k/n}}{(2\pi)^{1/2}}\exp\{\ell^0(\theta) + (\varphi(\theta) - \hat{\varphi}^0)s\}h(s)$$

where s is a score type variable with observed value $s^0 = 0$ and $\varphi(\theta)$ is the just defined likelihood gradient. The standard saddlepoint expansion (Daniels, 1954) then gives

$$f(y;\theta)dy = \frac{e^{k/n}}{(2\pi)^{1/2}}\exp\{-\frac{r^2(\theta,y)}{2}\}\left|\hat{j}_{\varphi\varphi}(s)\right|^{-1/2} ds$$

as the third order version of the approximated model at the data point $s^0 = 0$; this uses observed information calculated relative to the reexpressed parameter and it ignores the form of the original variable other than near the data. The reexpression has the role of a third order Fourier inversion of likelihood at the data point.

At the data point we can calculate $q(\theta;y^0) = -\theta$ and also calculate $r = r(\theta;y^0)$ and $r^*(\theta;y^0)$; for details see Cakmak et al (1998). From these we obtain formula (1.4) immediately which expresses r^* in terms of r when $y = y^0$. Our interest however is in the more general dependence of r^* on r.

3 Taylor expansion at a parameter value θ_0

Now in order to explore r^* and r under change in the data variable we first expand $\log f(y;\theta)$ about a parameter value θ_0 of interest, together with a natural value y_0 which we take to be $\hat{y}(\theta_0)$, the value with maximum density when the parameter value is θ_0. For this we now use the extensive calculations in Abebe et al (1995). First the parameter is centered; thus $\theta - \theta_0$ is taken for notational convenience to be a new θ. Then the variable is centered at $\hat{y}(\theta_0)$ and scaled by the negative second derivative of the log-density under $\theta = \theta_0$ following the related pattern in the preceding section. And then the variable is rescaled so that the cross derivative $(\partial/\partial\theta)(\partial/\partial y)\log f(y;\theta)|_{\hat{y}(\theta_0),\theta_0} = 1$. Again this can be viewed as a standardization towards standard normal form, yielding the first terms

$$\log f(y;\theta) = a - 1/2y^2 + y\theta,$$

to first derivative in θ and second derivative in y.

Now as in the preceding section we further reexpress y and θ so as to target on exponential model form but work here with the first row having higher coefficients $\tilde{\alpha}_3/n^{1/2}$, $\tilde{\alpha}_4/n$; we obtain

$$(a_{ij}) = \begin{pmatrix} a+k/n & 0 & -1 & \tilde{\alpha}_3/n^{1/2} & \tilde{\alpha}_4/n \\ k_{10}/n^{1/2} & 1 & 0 & 0 & - \\ -\{1+k_{20}/n\} & 0 & \tilde{\gamma}/n & - & - \\ -\tilde{\alpha}_3/n^{1/2} & 0 & - & - & - \\ -(\tilde{\alpha}_4+3\tilde{\alpha}_3^2+6\tilde{\gamma})/n & - & - & - & - \end{pmatrix}$$

to $O(n^{-3/2})$, where $a = -(1/2)\log(2\pi)$, $k = -(3\tilde{\alpha}_4 + 5\tilde{\alpha}_3^2)/24$, $k_{10} = -\tilde{\alpha}_3/2$, $k_{20} = (\tilde{\alpha}_4 + 2\tilde{\alpha}_3^2 + \tilde{\gamma})/2$, and $\tilde{\alpha}_3$, $\tilde{\alpha}_4$, and $\tilde{\gamma}$ are meta parameters corresponding to this parameter-based expansion pattern; the details follow from Abebe et al (1995). Also we have here that the coefficients in the first column are determined by the meta parameters $\tilde{\alpha}_3$, $\tilde{\alpha}_4$, and $\tilde{\gamma}$ using the norming property. The coefficient $\tilde{\gamma}$ correspondingly is a measure of departure from exponential model form.

With this modified exponential type expansion focussed on a parameter value θ_0 we can calculate $q(\theta_0; y)$ and $r(\theta_0; y)$ and $r^*(\theta_0; y)$; for details see Abebe et al (1995). From these we directly obtain formula (1.5) expressing r^* in terms of r, when the parameter value is held fixed at θ_0. And we noted in Section 1 that r^* was not linear in r. Our objective of course is to obtain a general formula under variation in both y and θ.

If we compare the Taylor expansion in this section with the Taylor expansion in the preceding section we can note a lot of similarity. We could for example examine the preceding expansion with θ_0 set equal to $\hat{\theta}^0$. We would then have that the earlier expansion would correspond to $y_0 = y^0$ and the present expansion would correspond to $y_0 = \hat{y}(\theta_0) = \hat{y}(\hat{\theta}^0)$ here. From results in Section 4 these can be seen to differ just by order $O(n^{-1/2})$. Also we can then see by substitution that this has no effect $O(n^{-3/2})$ on the meta parameters. Thus letting $\tilde{\alpha}_3$, $\tilde{\alpha}_4$, $\tilde{\gamma}$ designate the present meta parameters we can record the connection to the earlier meta parameters:

$$\begin{aligned} \tilde{\alpha}_3 &= \alpha_3 & \qquad \alpha_3 &= \tilde{\alpha}_3 \\ \tilde{\alpha}_4 &= \alpha_4 - 3\alpha_3^2 - 6\gamma & \qquad \alpha_4 &= \tilde{\alpha}_4 + 3\tilde{\alpha}_3^2 + 6\tilde{\gamma} \\ \tilde{\gamma} &= \gamma & \qquad \gamma &= \tilde{\gamma} \end{aligned}$$

to the third order.

If we then make the relevant substitutions in formula (1.5) we obtain formula (1.6), which demonstrates that different linearity between r^* and r applies under y change for fixed θ_0 and under θ change for fixed y_0. We next quantify this difference in the linearity pattern.

4 Meta-Parameters and the Data Point

In Section 2 the Taylor expansion of the log-model used coordinates standardized to exponential model form relative to an observed data point. The resulting mathematical parameters α_3, α_4 and γ could then reasonably be expected to depend on that data point. And we did mention this in the Introductory Section 1. Thus a direct approach to determining the r^* to r relationship would be to evaluate the dependence of the meta parameters on the expansion point. We pursue this now.

For the Taylor expansion in Section 2 using the array (a_{ij}) we take a possible new expansion point with $y_0 = a$ and derive the corresponding parameters say $\bar{\alpha}_3$, $\bar{\alpha}_4$, $\bar{\gamma}$. In principle this is simple. We take $(y_0, \theta_0) = (a, \hat{\theta}(a))$ and carry through the various steps used to derive the original array (a_{ij}): centre data variable at a; centre parameter at $\hat{\theta}(a)$; scale parameter by new observed information; scale data by cross derivative; reexpress parameter towards exponential form; reexpress variable towards exponential form. The actual algebra follows Cakmak et al (1998) and Fraser & Wong (2002) and has been independently derived by two of the authors together with follow up checks. From this we obtain the following meta parameters corresponding to the exponential expansion based on a data point $y = a$

$$
\begin{aligned}
\bar{\alpha}_3 &= \alpha_3 - (3\alpha_3^2 - 24\alpha_4 + 6\gamma)a/2n^{1/2} \\
\bar{\alpha}_4 &= \alpha_4 \\
\bar{\gamma} &= \gamma.
\end{aligned}
$$

When we change the expansion point from the nominal $y = 0$ to the new $y = a$ we see that only one meta parameter changes, the α_3, and it changes linearly in a. We also note that the α_3 in the expansion is used as $-\alpha_3\theta^3/6n^{1/2}$. Thus if we were to consider an expansion point that differs by $O(n^{-1/2})$, say $y = a = b/n^{1/2}$, then the net effect would be zero to order $O(n^{-3/2})$. This then justifies the remark in Section 1 that the versions (1.4) and (1.5) are directly comparable if θ_0 for (1.5) is taken equal to the $\hat{\theta}(y^0)$ obtained for (1.4), with the result that the new y_0 value differs by just $O(n^{-1/2})$.

But now let us use the new meta parameters corresponding to the expansion relative to $y_0 = a$ to determine the general relationship between $r^*(\theta, y)$ and $r(\theta, y)$. Formula (1.4) gives that connection when y is held fixed at say y_0 and the formula uses the meta parameters corresponding to the y_0 expansion. But now we have the change in the meta parameters that comes from the change in expansion point from $y = 0$ to $y = a$. Thus we have

$$
\begin{aligned}
r^*(\theta, a) &= \frac{\bar{\alpha}_3}{6n^{1/2}} + \left(1 - \frac{3\bar{\alpha}_4 - 4\bar{\alpha}_3^2}{72n}\right) r(\theta, a) \\
&= \frac{\alpha_3}{6n^{1/2}} + \left(1 - \frac{3\alpha_4 - 4\alpha_3^2}{72n}\right) r(\theta, a) - \frac{1}{12n}(3\alpha_3^2 - 2\alpha_4 + 6\gamma)a
\end{aligned}
$$

where as usual terms of order $O(n^{-3/2})$ are ignored. In this expression we see that a appears only in the last term and appears with a factor n^{-1}. For this we can rewrite a as $r(a, \hat{\theta}^0)$ to order $O(n^{-1/2})$; in this alternate form it is using an invariant reexpression. If we then substitute this in the above formula we obtain formula (1.7) which gives the general expression for $r^*(\theta, y)$ in terms of $r(\theta, y)$ in moderate deviations about an expansion point with $(y_0, \theta_0) = (y^0, \hat{\theta}^0)$.

Our general formula (1.7) shows that there is linearity in a restricted sense but that a constant correction is needed relative to the expansion point and the resulting coefficients.

5 Cross Validation of the r^* to r Relationship

As discussed at the beginning of Section 2 we can examine Taylor expansion about two rather natural points $(y^0, \hat{\theta}^0)$ or $(\hat{y}(\theta_0), \theta_0)$ and then use a reexpressed variable and reexpressed parameter that are targetted on exponential model form or on location model form. In the preceding section we worked entirely with the exponential model form relative to the two natural expansion points. We used extensive calculations from the literature and derived the general third order relationship between r^* and r. To provide independent validation of this approach we now outline the steps using the targetted location model form. For the targetted location form we also have the option of either natural expansion point.

First consider the expansion about the data oriented point $(y^0, \hat{\theta}^0)$. As before we center the variable at y^0; then center the parameter at $\hat{\theta}^0$ with scaling from the observed information $\hat{j}^0$; and then scale the variable so the cross derivative $\ell^0_{\theta;y}$ becomes equal to 1. We then reexpress the parameter to location form and reexpress the variable to location form, giving the coefficient array

$$(a_{ij}) = \begin{pmatrix} a + k/n & 0 & -(1 + k_{02}/n) & a_3/n^{1/2} & (-a_4 - 6c)n \\ 0 & -1 & -a_3/n^{1/2} & a_4/n & - \\ -1 & a_3/n^{1/2} & (-a_4 + c)/n & - & - \\ -a_3/n^{1/2} & a_4/n & - & - & - \\ -a_4/n & - & - & - & - \end{pmatrix}$$

where the constants k and k_{02} are different from those used for the exponential form and are available in Section 4 of Cakmak et al (1998); and the current meta parameters designated as a_3, a_4, c are oriented towards the location model form; the connection to the earlier, exponential parameters is also avaiable from Cakmak et al (1998) and is given as

$$\begin{aligned} a_3 &= -\alpha_3/2 \\ a_4 &= (-4\alpha_4 + 9\alpha_3^2)/12 \\ c &= \gamma + (-2\alpha_4 + 3\alpha_3^2)/6 \end{aligned} \tag{5.1}$$

We now return to the r^* to r relationship and record the version for fixed $y = y^0$ which is available from Cakmak et al (1998):

$$r^*(\theta; y^0) = -\frac{a_3}{3n^{1/2}} + \left(1 + \frac{9a_4 - 11a_3^2}{72n}\right) r(\theta; y^0). \tag{5.2}$$

If we are then interested in the r^* to r relationship for another data expansion point say a we can follow the pattern in the preceding section and determine the new meta parameters. This is easier than in the exponential context due to the present location structure and we write the new meta parameters as $\bar{a}_3$, $\bar{a}_4$ and $\bar{c}$ and obtain

$$\begin{aligned}
\bar{a}_3 &= a_3 + 3ca/2n^{1/2} \\
\bar{a}_4 &= a_4 \\
\bar{c} &= c
\end{aligned}$$

If we then substitute in the r^* to r relationship we obtain

$$r^*(\theta, a) = -\frac{a_3}{3n^{1/2}} + \left(1 + \frac{9a_4 - 11a_3^2}{72n}\right) r(\theta, a) - ca/2n.$$

which can then be made coordinate invariant as before

$$r^*(\theta, a) = -\frac{a_3}{2n^{1/2}} + \left(1 + \frac{9a_4 - 11a_3^2}{72n}\right) r(\theta, a) - \frac{c}{2n} r(y; \hat{\theta}^0). \tag{5.3}$$

This provides an alternative version of (1.7) using the present location meta parameters rather than the exponential meta parameters in (1.7). To check we substitute the conversion formulas (5.1) to return to the exponential parameters; the substitutions give the formula (1.7), thus establishing equivalence.

Abebe et al (1995) also do a location model Taylor series expansion based on a chosen parameter value θ_0. It yields the relationship

$$r^*(\theta_0, y) = -\frac{\tilde{a}_3}{3n^{1/2}} + \left(1 - \frac{9\tilde{a}_4 + 11\tilde{a}_3^2 - 18\tilde{c}}{72n}\right) r(\theta_0; y). \tag{5.4}$$

The meta parameters of the two location expansions are related if we take $\theta_0 = \hat{\theta}^0$

$$\begin{aligned}
\tilde{a}_3 &= a_3 && (5.5) \\
\tilde{a}_4 &= -a_4 + 6c && (5.6) \\
\bar{c} &= c && (5.7)
\end{aligned}$$

If we then substitute (5.5), (5.6) (5.7) in (5.4) we obtain

$$r^*(\theta, a) = -\frac{a_3}{3n^{1/2}} + \left(1 + \frac{9a_4 - 11a_3^2}{72n}\right) r(\hat{\theta}^0, y) - \frac{c}{2n} r(\hat{\theta}^0, y). \tag{5.8}$$

which agrees with (5.3).

Other cross validations have been checked. Perhaps the most useful of the r^* to r formulas are those recorded as (1.7) in the Introduction and as (5.3) and (5.8) above.

6 Discussion

The quantity $r^*(\psi, y)$ is a statistical refinement of the familiar signed likelihood root $r(\psi, y)$, a refinement of several magnitudes with respect to having a standard normal approximation, with respect to its ability to precisely assess values ψ of the interest paraqmeter, and with its substantial freedom from the effects of nuisance parameters λ when $\theta = (\psi, \lambda)$. Our analysis in preceding sections has focussed on a technical aspect of the relationship of r^* with r, and whether and to what degree there is linearity. And we have found that it is linear to third order if one or the other argument for the quantities is free but not if they are both free.

This restricted linearity has important implications for general statistical inference. From the results in Section 5 we have that r^* is fully linear in r to third order if and only if the model has location structure to the third order. This curiously has close ties to current research that shows that default Bayes has a confidence interpretation if and only if that model has the third order location structure. Departures from location structure can be calibrated and defined as a particular form of model curvature. Thus in summary if the the model has appropriate linearity, then the signed likelihood root suffices for much of statistical inference. And in the same circumstances the Bayesian use of a default prior can give accurate results, accurate in the sense that statements concerning the parameter in relationship to a posterior quantile can conform to the stated percentage level for the quantile.

While our emphasis has been on the relationship of r^* with r we do find it appropriate to give some indication of the wide range of application of r^* in statistical inference. Some examples of the high accuracy of the standard normal approximation for r^* may be found in Barndorff-Nielsen (1986) and Fraser (1990). A general formula for the use of r^* may be found in Fraser, Reid & Wu (1999) and a detailed discussion of its use with regression under normality or nonnormality and linearity or nonlinearity may be found in Fraser, Wong & Wu (1999). In general contexts there is a need for approximate ancillarity as developed for example in Fraser & Reid (2001). An extension that avoids the direct determination of ancillary characteristics but is restricted to second order accuracy may be found in Skovgaard (1996), and an extension to the case of discrete variables in Davison et al (2006).

Acknowledgements

Many many thanks to Donald Pierce for dicussions, contributions, and enthusiasm for modern likelihood, particularly higher order likelihood and inference. The authors wish to thank the many participants in a continuing workshop in likelihood theory and methodology at the University of Toronto who contributed extensive time and energy on various aspects of the r^* to r theory and calculations, in particular, C. Poon, T. Ponnampalam, and Y. Sun .

This research was supported in part by the National Science and Engineering Research Council of Canada.

References

[1] Abebe, F., Cakmak, S., Cheah, P. K., Fraser, D.A.S., Kuhn, J., McDunnough, P., Reid, N., and Tapia, A. (1995). Third order asymptotic model: Exponential and location type approximations. *Pari–sankhyan Samikkha.*, **2**, 25-33.

[2] Barndorff-Nielson, O.E. (1986). Inference on full or partial parameters based on the standardized signed log likelihood ratio. *Biometrika*, **78**, 557-563.

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

[4] 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.

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

[6] Daniels, H.E. (1954). Saddlepoint approximation in statistics. *Ann. Math. Statist.*, **25**, 631-650.

[7] Daniels, H.E. (1987). Tail probability approximations. *Int. Statist. Rev.*, **54**, 37-48.

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

[9] DiCiccio, T. J., Martin, M. A. and Stern, S. E. (2001). Simple and accurate one-sided inference from signed roots of likelihood ratios. Canad. J. Statistics **29**, 67-76.

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

[11] 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.

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

[13] 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.

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

[15] Fraser, D.A.S. and Wong, A. (2002). Algebraic extraction of the canonical asymptotic model: scalar case. *Journal of Statistical Studies*, D **1**, 29-49.

[16] 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 Association*, **94**, 1286-1295.

[17] Lugannani, R. and Rice, S. (1980) Saddlepoint approximation for the distribution of the sum of independent random variables. *Adv. in Appl. Probab.*, **12**, 475-490.

[18] Pierce, D.A. and Bellio, R. (2006). Effects of the reference set on frequentist inferences. Biometrika **93**, 425-438.

[19] Skovgaard, I.M. (1996). An explicit large-deviation approximation to one-parameter tests. *Bernoulli*, **2**(2), 145-165.