Journal of Statistical Research *ISSN 0256 - 422 X*
2013, Vol. 47, No. 1, pp. 1-10

SIGNED LIKELIHOOD ROOT WITH A SIMPLE SKEWNESS CORRECTION: REGULAR MODELS, SECOND ORDER

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

WEI LIN
Department of Statistics, University of Toronto, Toronto, Ontario, Canada M5S 3G3
Email: becky@utstat.toronto.edu

ANGELA LEI WANG
Department of Statistics, University of Toronto, Toronto, Ontario, Canada M5S 3G3
Email: lei@utstat.toronto.edu

SUMMARY

A standardized maximum likelihood departure, a standardized score departure, the signed likelihood root: these are familiar inference outputs from statistical packages, with the signed likelihood root often viewed as the most reliable. A third-order adjusted signed likelihood root called r^* is available from likelihood theory, but the formulas and development methods are not always easily implemented. We use a log-model Taylor expansion to develop a simple second order adjustment to the signed likelihood root, an adjustment that is easy to calculate and easy to explain, and easy to motivate. The theory is developed, simulations are recorded to indicate repetition accuracy, real data are analyzed, and connections to alternatives are discussed.

Keywords and phrases: Exponential model, Likelihood asymptotics, Maximum likelihood departure, Score departure, Second-order, Signed likelihood root, Taylor expansion, Third-order.

1 Introduction

Exponential models are widely used in applications and widely used as building blocks for more general statistical models; they also provide theoretical background for many results from modern likelihood theory. An exponential model has the form

$$f(y;\theta) = \exp\{s'(y)\varphi(\theta) - k(\theta)\}h(y),$$

where here y has dimension n, and $s(y)$ and $\varphi(\theta)$ have dimension p. Many familiar models such as the Normal, the Poisson, the Binomial, the Gamma, the Exponential life and various generalized linear models have this form. With appropriate linear independence of the coordinates of $s(y)$ and of $\varphi(\theta)$, the model can be re-parameterized in terms of φ itself called the canonical parameter and re-expressed using sufficiency in terms of $s(y)$ called the canonical variable:

$$g(s;\varphi) = \exp\{s'\varphi - \kappa(\varphi)\}g(s), \tag{1.1}$$

where $g(s)$ can be chosen as a density function. The exponential factor provides what is called an exponential tilt of the "underlying" density $g(s)$, and $\kappa(\varphi)$ is the cumulant generating function or log moment generating function of the density $g(s)$:

$$m(t) = e^{\kappa(t)} = \int \exp(s't)g(s)\mathrm{d}s.$$

Example 1, The Exponential life model. The Exponential life model is a very simple example; it can be written in the form:

$$f(x;\varphi) = \varphi\exp(-x\varphi) = \exp[-x(\varphi-1) + \log\varphi]e^{-x}, \tag{1.2}$$

with lifetime x and failure rate φ. We can see that $\varphi = 0$ is an excluded boundary point and that the modified parameter $\varphi - 1$ set equal to the value 0 gives the simple exponential distribution $f(x) = e^{-x}$ which has the moment generating function

$$m(t) = e^{\kappa(t)} = \int_0^\infty e^{xt}e^{-x}\mathrm{d}x = (1-t)^{-1} = \exp(-\log(1-t)), \tag{1.3}$$

and cumulant generating function κ(t) = $-\log(1-t)$. The offset in the failure rate φ allows the new $\varphi - 1 = 0$ to be an interior point and correspond to the basic e^{-x} distribution.

The use of exponential models together with simple asymptotic analysis relative to increasing sample data size n allows Normal approximations to be upgraded to the highly accurate exponential-model or saddle-point approximations. For this a key result is a distribution function approximation for the density $g(s;\varphi)$: When the canonical variable is stochastically increasing in φ, the distribution function $G(s;\varphi)$ has the approximation

$$\bar{G}(s;\varphi) = \Phi(r - r^{-1}\log\frac{r}{q}) = \Phi(r^*)$$

where Φ is the standard Normal distribution function, and r and q are the signed likelihood root (SLR) and the standardized maximum-likelihood departure for data s and parameter value φ, and r^* is the third order approximation;

$$\begin{aligned} r &= r(\varphi;s) = \mathrm{sign}(\hat{\varphi}-\varphi)[2\{l(\hat{\varphi};s) - l(\varphi;s)\}]^{1/2} \\ q &= q(\varphi;s) = \mathrm{sign}(\hat{\varphi}-\varphi)(\hat{\varphi}-\varphi)\jmath_{\varphi\varphi}^{1/2}; \end{aligned}$$

for this $l(\varphi;s) = \log g(s;\varphi)$ is the log density, and $\hat{\varphi} = \hat{\varphi}(s)$ is the parameter value that maximizes $l(\varphi;s)$. These departures are often available as output from computer packages provided the parameter is chosen as the canonical parameter φ; such output can also often include a score departure $z = l_\varphi(\varphi;s)\jmath_{\varphi\varphi}^{-1/2}$.

Example 1 continued. Consider the example with data $x^0 = 17$ for the exponential life model (2): The log-likelihood function $l(\varphi; x) = -\varphi x + \log \varphi$ can be differentiated giving

$$l_\varphi(\varphi; x) = -x + \frac{1}{\varphi},$$

and the the equation $l_\varphi = 0$ gives the maximum-likelihood value $\hat{\varphi}(x) = x^{-1}$. Differentiating further gives

$$l_{\varphi\varphi}(\varphi; x) = -\frac{1}{\varphi^2}, \;\; l_{\varphi\varphi\varphi}(\varphi; x) = \frac{2}{\varphi^3},$$

and thus gives the information $\hat{\jmath}_{\varphi\varphi} = \jmath_{\varphi\varphi}(\hat{\varphi}; x) = -l_{\varphi\varphi}(\hat{\varphi}; x) = x^2$ from the second derivative and gives a skewness coefficient $\gamma = -l_{\varphi\varphi\varphi}/(-l_{\varphi\varphi})^{3/2} = -(2/\varphi^3)/(1/\varphi^2)^{3/2} = -2$ as the third standardized derivative.

For the data x^0 = 17 we have $\hat{\varphi}^0 = 1/17 = .0588$ and $\hat{\jmath}^0_{\varphi\varphi} = 17^2$; the observed log-likelihood function is plotted in Figure 1. The observed information is calculated as the negative second derivative of log-likelihood at the observed maximum-likelihood $\hat{\varphi}^0$, with respect to the canonical parameter φ; other estimates of information are possible but the just mentioned choice has various advantages. From this we have:

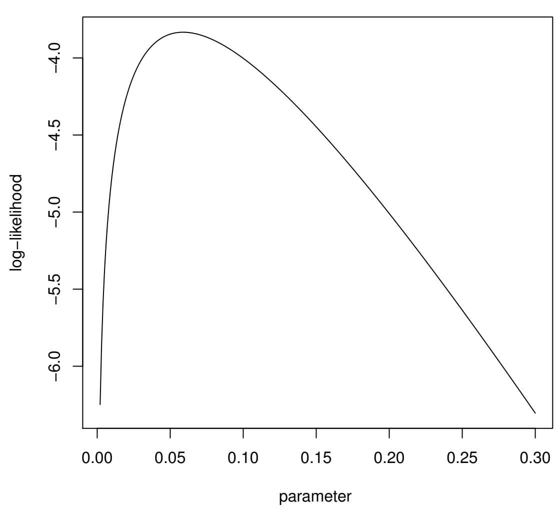

Figure 1: The observed log-likelihood function with data $x^o = 17$ from the exponential life model

$$r(\varphi; x^0) = \text{sign}(\hat{\varphi}^0 - \varphi)[2\{l(\hat{\varphi}^0; x^0) - l(\varphi; x^0)\}]^{1/2} = [2(-1 + 17\varphi - \log(17\varphi))]^{1/2}$$
$$q(\varphi; x^0) = \text{sign}(\hat{\varphi}^0 - \varphi)(\hat{\varphi}^0 - \varphi)\hat{\jmath}_{\varphi\varphi}^{1/2} = (\frac{1}{17} - \varphi)17 = 1 - 17\varphi$$
$$z(\varphi; x^0) = l_\varphi(\varphi; x^0)\hat{\jmath}_{\varphi\varphi}^{-1/2} = (-17 + \frac{1}{\varphi})\frac{1}{17} = -1 + \frac{1}{17\varphi}.$$

If the familiar Normal approximation is applied to these standardized departure measures we obtain the following first order p-values for the data above:

$$\Phi(r) = \Phi[2\{-1 + 17\varphi - \log(17\varphi)\}^{1/2}] = \Phi(1.372557) = 0.9150549$$
$$\Phi(q) = \Phi(1 - 17\varphi) = \Phi(0.83) = 0.7967306$$
$$\Phi(z) = \Phi(-1 + \frac{1}{17\varphi}) = \Phi(4.882353) = 0.9999995,$$

where the right hand numerical values are for testing the Null hypothesis $\varphi = 0.01$. Also here, the exact p-value is easily calculated

$$p(.01) = \int_{x^0}^{\infty} 0.01e^{-x(0.01)}\text{d}x = \exp(-0.01x^0) = 0.8436648.$$

This is probability left of $\hat{\varphi}^0$ on the $\hat{\varphi}$ scale and right of x^0 on the x scale, a consequence of the model being stochastically decreasing in φ.

In the next two sections we derive the simple skewness corrected $r^\dagger = r + \gamma/6$. Then for the example we have $r^\dagger = 1.3726 - 1/3 = 1.0392$, where $\gamma = -2$ is the standardized third derivative of the observed log-likelihood, as obtained above. The approximation $\Phi(r^\dagger) = .8506497$ is much closer to the exact $p(.01) = .8437$ than the first order p-values recorded earlier, thus indicating the improved accuracy of $r^\dagger$.

2 The standardized scalar exponential model to second order

Consider the scalar exponential model (1) with canonical variable s, canonical parameter φ, and dim s = dim $\varphi = 1$. We use asymptotics to examine the Taylor expansion of the log model

$$\log g(s; \varphi) = s\varphi - \kappa(\varphi) + \log h(s)$$

in the neighbourhood of observed data $s = s^0$ and the related maximum-likelihood value $\varphi = \hat{\varphi}^0$. For this we assume that the log-model has asymptotic properties as some background data size n becomes large. How can this background condition arise? The particular log-model can be a conditional distribution given some condition that restricts data variability to the dimension of the parameter, here $p = 1$. As such the log density will be a sum of components from an original sum of n components and correspondingly would grow as O(n); the normalization will affect this but the additive properties of the log-model are retained. In order to examine this asymptotically, we first center and scale-standardized the parameter φ; for this we examine the parameter departure of φ from $\hat{\varphi}^0$ using the scaling $\hat{\jmath}_{\varphi\varphi}^{1/2}$ indicated by the observed information, thus using the scale-standardize parameter (φ - $\hat{\varphi}^0$)$\hat{\jmath}_{\varphi\varphi}^{1/2}$. And then to avoid notation growth we temporarily designate this as just φ, a simple data based rewrite of the original canonical parameter. The resulting observed log-likelihood, then has the form

$$l(\varphi) = a - \varphi^2/2 - \gamma\varphi^3/6n^{1/2} + \mathrm{O}(\mathrm{n}^{-1}). \tag{2.1}$$

The cubic term is of order $\mathrm{O}(\mathrm{n}^{-1/2})$ as follows from $\jmath_{\varphi\varphi}$ and l_φ both being $\mathrm{O}(\mathrm{n})$. For some background on these expansions see Chakmak et al. (1998) and Fraser and Wong (2002). Also note that $\gamma/n^{1/2}$ is the third derivative of negative log-likelihood provided the parameter is scaled to have second derivative at the maximum equal to -1.

Now consider the canonical variable in a similar manner, obtaining $s - s^0$ as the departure, and then rescaling to retain the form of the tilt term $s\varphi$ with coefficient equal to 1. It follows then that to first order $\mathrm{O}(\mathrm{n}^{-1/2})$ the log model has the form

$$\log g(s;\varphi) = a - \frac{1}{2}(s-\varphi)^2 + \mathrm{O}(\mathrm{n}^{-1/2}).$$

This is in agreement with the Central Limit Theorem result but comes directly from the large sample form of the log density.

The inclusion of fractional powers such as $n^{-1/2}$ can be purely formal, to indicate asymptotic magnitude and should not be confused with an actual sample size in an example. Thus the second derivative $l^{''}(\varphi)$ at $\varphi = 0$ is 1 and the third derivative $l^{'''}(\varphi)$ at $\varphi = 0$ is $-\gamma/n^{1/2}$ with the $n^{1/2}$ typically having a purely symbolic role to remind us of the asymptotic size; the standardized third derivative is $\gamma/\sqrt{n}$.

Now for second order analysis corresponding to the case with $\gamma \neq 0$ as in (4), we first determine the terms in s that are required so that $\log(g;\varphi)$ represents a density function that integrates to unity. If we expand the parameter factor $\exp(-\gamma\varphi^3/6n^{1/2})$ to second order we obtain $1 - \gamma\varphi^3/6n^{1/2}$. A data term that integrates to give φ^3 with the Normal $(\varphi, 1)$ distribution is $s^3 - 3s$; thus the factor $(1 + \gamma s^3/6n^{1/2} - 3s\gamma/6n^{1/2})$ compensates the factor $1 - \gamma\varphi^3/6n^{1/2}$ to order $\mathrm{O}(\mathrm{n}^{-1})$. This gives the second order density expression

$$g(s;\varphi) = (2\pi)^{-1/2}\exp\{-\frac{1}{2}(s-\varphi)^2 - \gamma\varphi^3/6n^{1/2} + \gamma s^3/6n^{1/2}\}(1 - \gamma s/2n^{1/2}) + \mathrm{O}(\mathrm{n}^{-1}),$$

where we have chosen not to put the linear term in s in the exponent. This agrees with the Taylor coefficient array (5) in Chakmak et al. (1998) which records results to third order; see also Fraser and wong (2002).

We will next see that the exponent in the above expression is just $-r^2/2$ to second order.

3 Corrected SLR: second order.

Consider the standardized version of the scalar exponential model as determined to second order. We calculate the likelihood and score,

$$l(\varphi;s) = -\frac{1}{2}(s-\varphi)^2 - \gamma\varphi^3/6n^{1/2}, \;\; l_\varphi(\varphi;s) = (s-\varphi) - \gamma\varphi^2/2n^{1/2},$$

which lead to

$$\hat{\varphi} = \hat{\varphi}(s) = s - \gamma s^2/2n^{1/2} + \mathrm{O}(\mathrm{n}^{-1}).$$

We then have

$$\hat{l} = l(\hat{\varphi}; s) = -\frac{1}{2}(\gamma s^2/2n^{1/2})^2 - \gamma s^3/6n^{1/2} = -\gamma s^3/6n^{1/2}$$
$$\hat{l} - l = \frac{1}{2}(s-\varphi)^2 - \gamma(s^3 - \varphi^3)/6n^{1/2} = \frac{1}{2}(s-\varphi)^2 - \gamma(s-\varphi)(s^2 + s\varphi + \varphi^2)/6n^{1/2};$$

which is of course just $r^2/2$. We then calculate the signed likelihood root,

$$r = s - \varphi - \gamma(s^2 + s\varphi + \varphi^2)/6n^{1/2}.$$

We next find the second order distribution of the signed likelihood root r. From the definition above we have to second-order

$$\frac{\mathrm{d}r}{\mathrm{d}s} = 1 - \gamma(2s+\varphi)/6n^{1/2}, \;\; \mathrm{d}s = \{1 + \gamma(2s+\varphi)/6n^{1/2}\}\mathrm{d}r.$$

Then substituting in the expression $g(s;\varphi)ds$ we obtain to second order;

$$\begin{aligned} f(r;\varphi)dr &= \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-r^2/2}(1 - \gamma s/2n^{1/2})(1 + \gamma(2s+\varphi)/6n^{1/2})\mathrm{d}r \\ &= \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-r^2/2}(1 - \gamma(s-\varphi)/6n^{1/2})\mathrm{d}r \\ &= \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-r^2/2}(1 - \gamma r/6n^{1/2})\mathrm{d}r \\ &= \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-r^2/2 - \gamma r/6n^{1/2}}\mathrm{d}r \\ &= \phi(r + \gamma/6n^{1/2})\mathrm{d}r, \end{aligned}$$

where ϕ is the standard Normal density. It follows that

$$r^\dagger = r^\dagger(s;\varphi) = r + (\gamma/n^{1/2})/6$$

is standard Normal using the additive correction which is 1/6 of the scale-standardized third derivative $\gamma/n^{1/2}$ of likelihood.

For the example at the end of Section 1 we now record simulations to compare the Normal distribution accuracy of r and $r^\dagger$.

Example 1 Continues. Consider the extreme case of a single observation, y^0 from the Exponential life model example in Section 1. The quantities for r and $r^\dagger$ are given as

$$r = \mathrm{sign}(1/y^0 - \varphi)\sqrt{2[-\log(\varphi y^0) + \varphi y^0 - 1]}, \;\; r^\dagger = r + \gamma/6 = r - 1/3,$$

where $\gamma = -l_{\varphi\varphi\varphi}(\hat{\varphi}, y^0)/[-l_{\varphi\varphi}(\hat{\varphi}, y^0)]^{3/2}$ is the third scale-standardized derivative of the negative log-likelihood. Based on the calculations in Section 1, we have $\gamma = -2$ for this case.

To examine the Normal approximations for r and $r^\dagger$, we simulated 10,000 repetitions of the of size 1 sample from the Exponential life model with rate $\varphi = 1$ and compared the sample distribution function values with the Normal (0,1) distribution function values. For the 10,000 samples, we obtain the $P - P$ plot in Figure 2 (a); we also record the simulations for the less extreme case with sample size $n = 3$ plotted as Figure 2 (b). The diagonal line corresponds to the standard uniform distribution; The solid and dotted lines are the p-value for r and $r^\dagger$ respectively. Note that the $P - P$ plot of $r^\dagger$ closely approximates the diagonal line and thus $r^\dagger$ closely follows the standard Normal distribution.

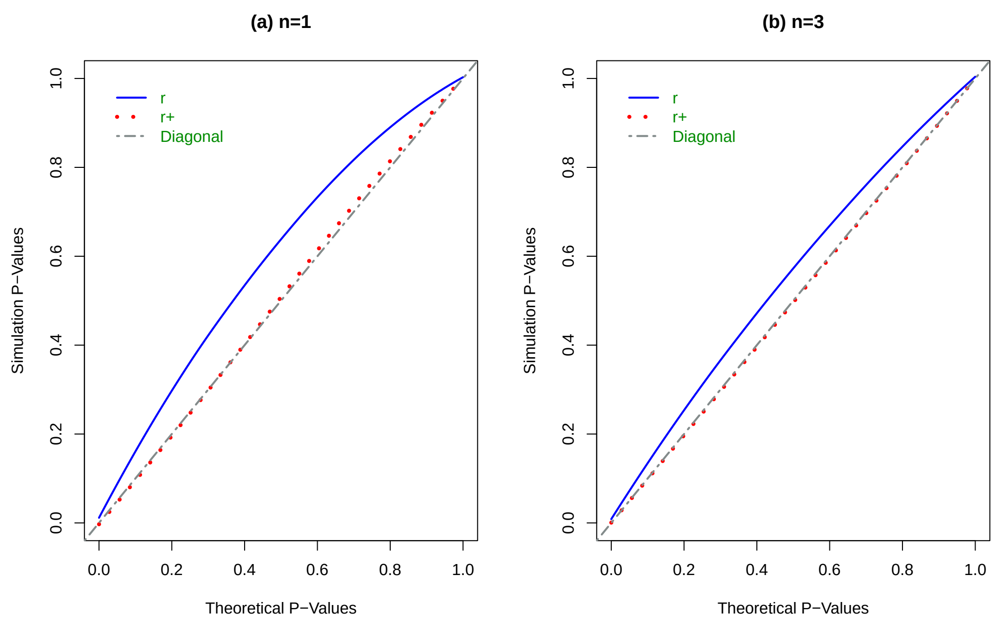

Figure 2: The $P - P$ plot, for the simulation of r and $r^\dagger$ from the Exponential 1; (a) $n = 1$, (b) $n = 3$, with $N = 10,000$

4 Some further examples

With the scalar exponential model $f(y;\theta) = \exp\{s'(y)\varphi(\theta) - k(\theta)\}h(y)$, we have various familiar examples. In each of these the skewness is proportional to $n^{-1/2}$ so it suffices to do the calculation for the $n = 1$, and then adjust accordingly for other n values. We also record simulations but do these just for $n = 1$ case. With larger values of n the discrepancies fall off quickly as indicated by the example in the preceding section.

Example 2. The Gamma(α_0, β) model. The gamma model is widely used in application to describe variables that take positive values. The Gamma(α_0, β) model is

$$f(y; \alpha_0, \beta) = \frac{\beta_0^{\alpha} y^{\alpha_0 - 1}}{\Gamma(\alpha_0)} \exp\left(-\beta y\right)$$

with

$$l(\alpha_0, \beta; y) = \alpha_0 s_1 - \beta s_2 + \alpha_0 \log \beta - \log \Gamma(\alpha_0),$$

where $s_1 = \log(y), s_2 = y$; The signed likelihood root for a sample of n is

$$r = \text{sign}(\hat{\varphi} - \varphi)\sqrt{2n\alpha_0 \log\left(\hat{\varphi}/\varphi\right) - 2S_2(\hat{\varphi} - \varphi)}$$

where $\varphi = \beta$, $S_2 = \sum_i y_i$ and $\hat{\varphi} = n\alpha_0/S_2$

Taking second and third derivatives of the log-likelihood with respect to β we obtain

$$\gamma = -2\alpha_0^{-1/2}$$

Consider $N = 10,000$ simulations. With $n = 1, \alpha = \alpha_0 = 1$, we have plots for Figure 3 (a) for testing the hypothesis $H_0 : \beta = 2$. We see that γ makes a substantial correction.

Examples 3. The Gamma(α, β_0) model. This model has a fixed scale parameter $\beta = \beta_0$ and varying shape parameter α; the model is

$$f(y; \alpha, \beta_0) = \frac{\beta_0^{\alpha} y^{\alpha - 1}}{\Gamma(\alpha)} \exp\left(-\beta_0 y\right)$$

with

$$l(\alpha, \beta_0; y) = \alpha s_1 - \beta_0 s_2 + \alpha \log \beta_0 - \log \Gamma(\alpha)$$

where s_1 and s_2 are as in the preceding example. We use derivatives of the log gamma function; the first derivative $D(x)$ is called the digamma function, and higher derivatives are readily available:

$$D(x) = \frac{d}{dx} \log \Gamma(x); \quad D^{(i)}(x) = \frac{d^i}{dx^i} \log \Gamma(x).$$

The skewness γ is then just

$$\gamma = D^{(2)}(\alpha)/\{D^{(1)}(\alpha)\}^{3/2}.$$

We evaluate the null distribution of $r, r^\dagger$ using $N = 10,000$ simulations with $n = 1, \beta = \beta_0 = 2$ giving the $P - P$ plots in Figure 3 (b) for testing the hypothesis $H_0 : \alpha = 1$. We see that the skewness γ makes a substantial improvement.

Examples 4. The Inverse Gaussian Model with fixed mean. This Inverse Gaussian model has the following form

$$f(y; \mu, \lambda) = \sqrt{\lambda/2\pi y^3} \exp\{-\lambda(y - \mu)^2/2\mu^2 y\}.$$

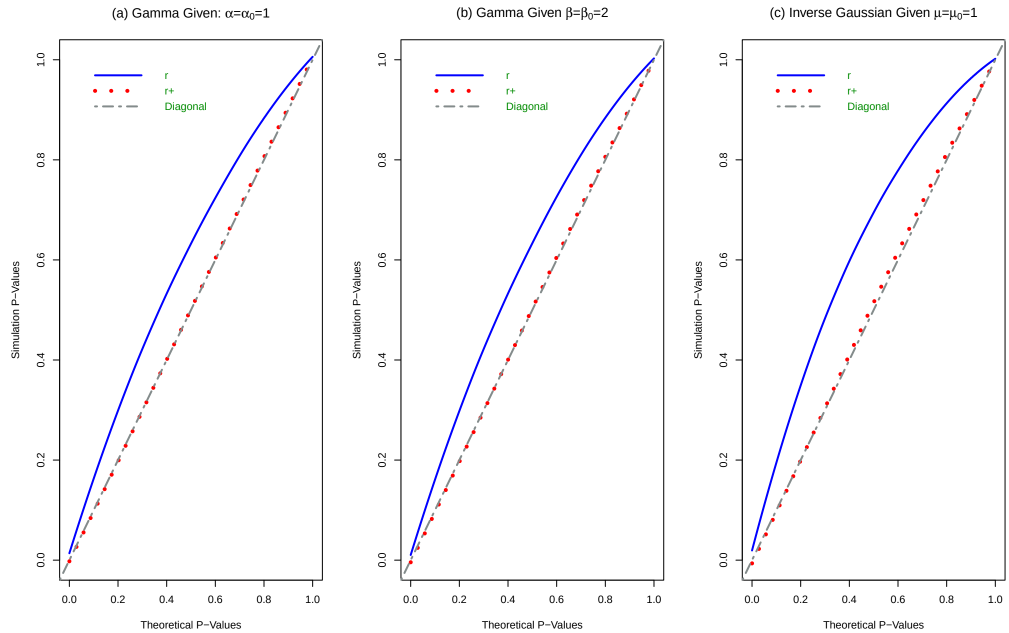

Figure 3: For these models we have simulations with $N = 10,000$ to compare $r, r^\dagger$. For the $n = 1$ case: (a) Gamma model parameter with $\alpha = \alpha_0 = 1$, testing $H_0 : \beta = 2$; (b) Gamma model parameter given $\beta = \beta_0 = 2$, testing $H_0 : \alpha = 1$; (c) Inverse Gaussian model given $\mu = \mu_0 = 1$, testing $H_0 : \lambda = 1$

We examine this model with $\mu = \mu_0$ fixed and test the scale parameter λ with hypothesis $H_0 : \varphi = \lambda$. This gives

$$l(\varphi; y_1, \ldots, y_n) = \frac{n}{2} \log \varphi - \varphi \sum_{i=1}^{n} \frac{(y_i - \mu_0)^2}{2\mu_0^2 y_i},$$

and the skewness is $\gamma = -2^{3/2}$ for $n = 1$, giving the corrected

$$r^\dagger = r - 2^{1/2}/3n^{1/2}.$$

The normality of the null distribution of $r, r^\dagger$ was evaluated with $N = 10,000$ simulations using the extreme $n = 1$; the $P - P$ plots recorded in Figure 3 (c) With $\mu = \mu_0 = 1$, and testing $H_0 : \varphi = \lambda = 1$, we see that γ makes a substantial improvement.

5 Discussion

We have examined the scalar exponential model expressed in terms of its canonical parameter. By taking second and third derivatives of the log-likelihood with respect to the canonical parameter we obtain a standardized third derivative of log-likelihood. Subtracting one sixth of this from the signed likelihood root (SLR) gives a remarkably accurate corrected SLR; substituting in the Normal(0,1) then gives highly accurate p-value, we plan to extend this to the scalar interest parameter in higher dimensional exponential models.

Acknowledgements

This research has been partly supported by a grant from the Natural Science and Engineering Research Council of Canada. The development of the material began with a graduate course on inference and progressed through a following research seminar. Several participants in that seminar contributed to the initial development of this paper: Zheng Li, Ali Zaidi, Huong Le. They deserve special thanks for their contributions; course work and other pressures prevented them from continuing with the project.

References

[1] 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.

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