Pak J. Statist.
2004 Vol. 20(1) pp 173-192

SIMPLE ACCURATE UNIQUE: THE METHODS OF MODERN LIKELIHOOD THEORY

D.A.S. FRASER
Department of Statistics, University of Toronto
100 St. George Street, Toronto, Canada M5S 3G3.
dfraser@utstat.toronto.edu
A. WONG
School of Analytic Studies and Information Technology, York University
4700 Keele Street, Toronto, Canada M3J 1P3.
august@yorku.ca
J. WU
Department of Biostatistics, St. Jude Children's Research Hospital
332 North Lauderdale Street, Memphis, TN 38105, USA.
Jianrong.Wu@stjude.org

ABSTRACT

Recent likelihood theory has given us a deconstruction of the statistical model that produces simple, highly accurate and unique methods for the analysis of data from the model. We describe these methods briefly and illustrate them with a succession of progressively more substantial examples. Computer implementation is available but presently needs automation.

Some key words.

Conditioning, Large sample theory, Likelihood, p-values, Signed likelihood ratio, Simulations, Sufficiency, Third order asymptotics.

1. INTRODUCTION

A central objective in statistics is to determine the information available concerning a parameter in the context of some given model and corresponding data. The traditional use of suffficiency and sometimes conditionality have not produced incisive and widely applicable methods. The Bayesian approach has sought to fill this gap and does yield a wealth of results: the approach does use the observed likelihood function and has thus partially compensated for traditional theory where the observed likelihood function is typically not used in a substantive way. The new likelihood theory is now widely applicable for continuous response models and is being extended to the discrete case covering categorical and contingency type models. For the discrete case, however, the high third order accuracy drops to second order as a direct consequence of the discreteness.

In Section 2 we briefly outline the recent likelihood theory and then in Section 3 give a succession of examples that illustrate for the continuous case the progressive extension from a simple exponential model to a general continuous model. The discrete case now being developed will soon be reported on.

2. BACKGROUND: RECENT LIKELIHOOD THEORY

The recent highly accurate likelihood methods need and use two functions of the parameter obtained from the model and observed data $\ell(\theta), \varphi(\theta)$ where $\ell(\theta)$ is the observed log-likelihood function

$$\ell(\theta) = \log f(y^0; \theta), \tag{1}$$

and $\varphi(\theta)$ is the gradient of likelihood at the data point in directions tangent to an exact or approximate full ancillary

$$\varphi'(\theta) = \ell_{;V}(\theta) = \frac{d}{dV}\ell(\theta; y)\big|_{y^0} \tag{2}$$

where $V = (v_1, \ldots, v_p)$ is an array of p linearly independent tangent vectors recording tangent directions at the observed data y^0. We refer to φ as the canonical parameter as it can be presented as the canonical parameter of a best fitting exponential model now to be described. Some details for calculating the vectors V are recorded in the Appendix.

The two functions $\ell(\theta), \varphi(\theta)$ define an exponential model that approximates the conditional model given an intrinsic ancillary in the neighbourhood of the data point; the approximate model is

$$f(s; \varphi) = (2\pi)^{-p/2} e^k \exp\{\ell^0(\varphi) - \ell^0(\hat{\varphi}^0) + s'(\varphi - \hat{\varphi}^0)\} |\hat{\jmath}_{\varphi\varphi}|^{-1/2}, \tag{3}$$

where k is constant of order $O(n^{-1})$, $\ell(\theta)$ has been reexpressed as $\ell\{\theta(\varphi)\}$ in terms of the new parameter φ, and s is the score variable with observed value $s^0 = 0$. For some background details see for example Fraser & Reid (1995). This tangent model provides third order inference at the data point.

An equivalent pair of functions is given by $\ell(\theta), \beta(\theta)$ where $\beta(\theta)$ is the location parameter of a location model that gives a location equivalent of the exponential model in the preceding paragraph. A formula for $\beta(\theta)$ in the scalar parameter case is given in Cakmak et al (1995); an existence result for the vector case is given in Cakmak et al (1994) and Fraser & Yi (2003).

We will see that a statistical model with data yields by very simple calculations the pair of functions $\ell(\theta), \varphi(\theta)$ and that this pair of functions leads to highly accurate p-values and likelihood function for any scalar parameter say $\psi(\theta)$ of interest. The alternate pair of functions $\ell(\theta), \beta(\theta)$ provides equivalent information but in a different form. The first pair is relatively easy to calculate and the second has advantages for transparency and ease of interpretation.

Now suppose that we are interested in some scalar parameter $\psi(\theta)$. The direct likelihood function information is then usually presented and used in the form of a signed likelihood ratio $r(\psi)$ which can be recorded as

$$(4) \qquad r = \operatorname{sgn}(\hat{\psi} - \psi)\left[2\{\ell(\hat{\theta}) - \ell(\hat{\theta}_{\psi})\}\right]^{1/2}$$

where $\hat{\theta}$ is the full maximum likelihood value and $\hat{\theta}_{\psi}$ is the maximum likelihood value subject to the constraint $\psi(\theta) = \psi$. The likelihood gradient or canonical φ function information is usually used in the form of a Wald type quantity $q(\psi)$ that includes a factor corresponding to the elimination of the related nuisance parameter; this special Wald type quantity is recorded in equation (9) with supplemental details in the Appendix. The third order p-value function for assessing ψ can then be calculated by using the Barndorff-Nielsen (1986) formula

$$p(\psi) = \Phi(r^*) = \Phi\{r - r^{-1}\log(r/q)\} \,, \tag{5}$$

where Φ is the standard normal distribution function, or by using the earlier Lugannani & Rice (1980) formula. It follows trivially that the survivor function for ψ using a flat prior for the location model just mentioned agrees to third order with the just described p-value; for some background see Fraser & Reid (2002).

Thus p-values and survivor values are available to third order for scalar parameters under moderate regularity and continuity; the examples below indicate the ease, flexibility and accuracy of the implementation.

3. SCALAR VARIABLE AND PARAMETER

Consider a simple exponential model with scalar variable and scalar parameter:

$$f(y;\theta) = \exp\{y\theta - \kappa(\theta)\}h(y)$$

where $\kappa(t)$ is the cumulant generating function for y when $\theta = 0$; a location shift for θ may be needed in some examples. A simple Wald type statistic based on the canonical parameter θ is

$$q = (\hat{\theta} - \theta)\hat{\jmath}^{1/2} \tag{6}$$

where $\hat{\theta}$ is the maximum likelihood value for θ and $\hat{\jmath}$ is the corresponding observed information

$$\hat{\jmath} = -\frac{\partial}{\partial\theta}\frac{\partial}{\partial\theta}\ell(\theta; y)\bigg|_{\hat{\theta}} \tag{7}$$

at the observed data. Then the signed likelihood ratio r and the given q produce third order p-values at the observed data using the Barndorff-Nielsen or Lugannani & Rice expressions discussed in the preceding section. This calculation is equivalent to the saddlepoint approximation (Daniels, 1954, 1987) as extended to distribution functions by Lugannani & Rice (1980). Now in an example we compare this third order approximation with the exact value and with the familiar first order p-value approximation given by $\Phi(r)$ based on the signed likelihood ratio.

Example 1. Simple exponential life model. Consider the simple exponential life model $f(y;\theta) = \theta\exp(y\theta)$ with $y < 0$ and $\theta > 0$. The log-likelihood function is

$$\ell(\theta; y) = y\theta + \log\theta$$

and the canonical parameter is just θ. The p-value records probability left of the observed maximum likelihood value and here this is just the probability left of an observed data. Without loss of generality, consider an observed value $y^0 = -1$. Table 1 records the p-values obtained by the likelihood ratio method, the third order method, and the exact method for the following values of the parameter θ: .01, .1, 10, 100. Note the very high accuracy of the third order method even for the extreme θ values.

Table 1: Left tail p-values for $\theta = .01, .1, 1$ and 100 and $y = -1$ calculated using the likelihood ratio (lr), third order (3rd) and the exact (exact) methods.

θ	.01	.1	10	100
lr	.996416	.953020	$.0^{3}1262$	$.0^{42}2921$
3rd	.989759	.903889	$.0^{4}4697$	$.0^{43}3971$
Exact	.990050	.904837	$.0^{4}4940$	$.0^{43}3921$

Now consider an asymptotic model for y with parameter θ, both scalar; this can arise with an accumulation of data affecting a scalar variable, and indeed can arise embedded in some very general models. The likelihood function is immediately available; the canonical parameter needs, as mentioned earlier, a tangent direction at the data point; but for the scalar variable, this corresponds to just increasing the variable. Thus the derivative in (2) became just an ordinary derivative and we obtain just the ordinary derivative of the log-likelihood at the data point with respect to y:

$$\varphi(\theta) = \ell_{;y}(\theta; y)|_{y^0} = \frac{\partial}{\partial y}\ell(\theta; y)\bigg|_{y^0} . \tag{8}$$

This gives third order accurate p-values for θ. For some numerical examples see Fraser (1990) and Fraser, Reid & Wu (1999). The interesting thing is that it suffices to get, beyond the observed likelihood, just the locally defined canonical parameter, $\varphi(\theta)$, the canonical parameter of some locally best fitting exponential model; and this retains the third order accuracy for the p-value function.

4. dim p VARIABLE AND PARAMETER

Consider a simple exponential model with vector variable and vector parameter having the same dimension p:

$$f(y; \theta) = \exp\{y'\theta - \kappa(\theta)\}h(y)$$

where $\kappa(t)$ is the cumulant generating function for the vector y when $\theta = 0$. Suppose that statistical interest rests on a scalar parameter component $\psi(\theta)$ and that λ is a complementing nuisance parameter. The signed likelihood ratio is given by (4) above. The Wald type statistic (6) is modified in an easily described way (see for example Fraser & Reid, 1995) giving

$$q(\psi) = \text{sgn}(\hat{\psi}^0 - \psi)|\chi(\hat{\theta}) - \chi(\hat{\theta}_\psi)| \left\{ \frac{|\hat{\jmath}_{\varphi\varphi}|}{\jmath_{(\lambda\lambda)}(\hat{\theta}_\psi)} \right\}^{1/2} ; \tag{9}$$

where $\chi(\theta)$ is a rotated coordinate of $\varphi(\theta)$ that agrees with $\psi(\theta)$ at $\hat{\theta}_\psi$ and acts as a surrogate for $\psi(\theta)$ at $\hat{\theta}_\psi$, and the full and nuisance informations are recalibrated in the φ parameterization, as indicated by the use of parentheses around $\lambda\lambda$. Formulas for these are recorded in the Appendix. This has the form of a Wald statistic (6) calculated for the modified interest parameter $\chi(\theta)$ multiplied by $|\hat{\jmath}_{(\lambda\lambda)}|^{1/2}/|j_{(\lambda\lambda)}(\hat{\theta}_\psi)|^{1/2}$ which is the root of a ratio of recalibrated nuisance informations and takes account of the elimination of the nuisance parameter. If ψ and λ are canonical parameter coordinates of θ then χ can be taken equal to ψ and the recalibrations can be overlooked.

Example 2. The Weibull model. Consider a sample from the Weibull model

$$f(y;\beta,\eta) = \frac{\beta}{\eta}\left(\frac{y}{\eta}\right)^{\beta-1}\exp\left\{-\left(\frac{y}{\eta}\right)\right\}$$

on the positive axis with shape and scale parameters β and η, respectively. The typical scalar interest parameters are: β, η, and the α th percentile which takes the form $w_\alpha = \log\{\eta[-\log(1-\alpha)]^{1/\beta}\}$.

For our empirical illustration, we use a sample of the strike duration data given in Kennan (1985). The data used by Kennan is from the Bureau of Labor Statistics and reflects strike durations in days involving at least 1,000 workers for the US manufacturing industries for the 1968 to 1976 period. The complete data set is in Kiefer (1988).

We restrict our sample to strikes beginning in June of each year for a total number of 62 observations. Table 2 records the 95% confidence intervals for the 3 parameters of interest calculated by the likelihood ratio and by the third order likelihood method. The confidence intervals by the two calculations are strikingly different for each of the parameters examined. Theory and simulations put our trust firmly on the third order results.

Table 2: 95% confidence intervals for β, η, and median using the likelihood ratio (lr) and the third order method (3rd) for the Kennan data.

	95% CI for β	95% CI for η	95% CI for median
lr	$(0.6959, 1.1141)$	$(30.6670, 58.2784)$	$(19.4879, 38.5315)$
3rd	$(0.6794, 1.0944)$	$(30.7924, 59.2684)$	$(19.4097, 38.8668)$

To illustrate the accuracy of the third order method in the small sample setting, we consider a sample size $n = 10$ from the Weibull model and simulate with $N = 10000$ repetitions. The true parameter values are taken to be $\beta = 5$ and $\eta = 10$. Table 3 records the proportion of samples whose p-values lies below $0.005, 0.025, 0.05$ and above $0.95, 0.975, 0.995$. The nominal or target values then are $0.005, 0.025, 0.05, 0.05, 0.025$ and 0.005 respectively. The third order likelihood method gives excellent coverage, very close to the nominal or target!

Table 3: Simulation proportions for the p-values lying in the lower and upper $.5\%, 2.5\%, 5\%$ ranges.

Parameter	Method	lower 0.5%	lower 2.5%	lower 5%
β	lr	0.0174	0.0619	0.1097
	3rd	0.0048	0.0264	0.0521
η	lr	0.0109	0.0430	0.0787
	3rd	0.0050	0.0261	0.0539
median	lr	0.0094	0.0356	0.0663
	3rd	0.0053	0.0262	0.0534

Parameter	Method	upper 5%	upper 2.5%	upper 0.5%
β	lr	0.0223	0.0112	0.0016
	3rd	0.0475	0.0236	0.0056
η	lr	0.0580	0.0313	0.0068
	3rd	0.0514	0.0245	0.0051
median	lr	0.0666	0.0357	0.0084
	3rd	0.0483	0.0251	0.0049

Now consider an asymptotic model for y with corresponding parameter θ, both vectors of the same dimension; this can arise with an accumulation of data affecting a vector variable, and indeed often arise embedded in quite general models, The likelihood function is immediately available. The canonical parameter needs the existence of tangent directions at the data point, but on the space of dimension p these correspond just to changes in the variables. Thus

$$\varphi(\theta) = \ell_{;y}(\theta; y^0) = \partial\ell(\theta; y)/\partial y|_{y^0}$$

which is just the gradient of the log-likelihood at the data point, that is, the vector derivative with respect to y. This gives third order accurate p-values for scalar component parameters say ψ. For some numerical examples and simulations see Fraser, Wong & Wu (1999). The interesting thing is that it suffices to get, beyond the observed likelihood, just the locally defined canonical parameter $\varphi(\theta)$ which is the canonical parameter of some locally best fitting exponential model, and yet this retains the high third order accuracy for p-value function for any choice of the scalar interest parameter $\psi(\theta)$

5. VARIABLE WITH DIMENSION GREATER THAN THE DIMENSION OF THE PARAMETER

The solution for the general case with dimension of the variable larger than that of the parameter has been resolved by showing that there exists an approximate ancillary of dimension complementing the dimension of the parameter, and that only the corresponding tangent directions at the data point are needed for third order inference; for details see Fraser & Reid (2001). The tangent directions V can in fact be obtained from a full dimensional pivotal quantity, say $z(y;\theta)$, of dimension n. With independent coordinates the pivotal quantity is naturally just the vector of distribution functions, with say $F_i(y_i;\theta)$ for the ith coordinate. The pivotal quantity shows how a change in θ affects a particular coordinate and the corresponding flow of probability generates the ancillary (Fraser & Reid, 2001). This gives

$$V = -z_y^{-1}(y^0;\hat{\theta}^0)z_{;\theta}(y^0;\hat{\theta}^0). \tag{10}$$

where the subscripts denote partial differentiation. The canonical parameter is then obtained by differentiating the log-likelihood function at the data point y^0; for $p = 2$, we would then have

$$\varphi'(\theta) = \{\varphi_1(\theta), \varphi_2(\theta)\} = \frac{d}{dV}\ell(\theta; y)\Big|_{y=y^0}$$
$$= \left\{\frac{d}{dv_1}\ell(\theta; y)\,|_{y^0}, \frac{d}{dv_2}\ell(\theta; y)\right\}\Big|_{y=y^0}.$$

where the derivatives are directional derivatives in the directions v_1, v_2.

Example 3. Nonnormal regression. The standard analysis of the regression model is based on normally distributed errors. It is however widely acknowledged that error quite generally has longer tails and something like the Student distribution on six degrees of freedom provides a more realistic error structure. The recent likelihood analysis provides a definitive analysis for this more realistic condition of the regression model. A detailed discussion may be found in Fraser, Wong & Wu (1999).

Consider the linear regression model

$$y = X\beta + \sigma e$$

where y and e are n vectors, X is the $n \times r$ design matrix and the components of e have a known distribution $f(e_i) = \exp\{\ell(\theta)\}$ which has been centered so that the slope $s(e) = d\ell(e)/de$ is zero at the origin: $s(0) = \ell'(0) = 0$. For illustration we will consider the normal case with $l(e) = -e^2/2$ and the Student (6) case with

$$l(e) = -\frac{7}{2}\log\left(1+\frac{e^2}{6}\right) + \log\frac{\Gamma(7/2)}{\Gamma(1/2)\Gamma(3)} - \frac{1}{2}\log 6.$$

The parameters are the regression coefficients β and the error scaling σ. The log likelihood has the form

$$\ell(\theta) = -n\log\sigma + \sum_{i=1}^{r} \ell_i \left\{\frac{y_i - X_i\beta}{\sigma}\right\}$$

where X_i is the i th row of X.

There is a natural pivotal $e = (y - X\beta)/\sigma$ which has full dimension n. The tangent vectors $V = (v_1, \ldots, v_{r+1})$ are obtained from (10) and have the natural form $V = \{\hat{e}, X\hat{\beta}\}$ where $\hat{e}$ is the standardized residual vector, $\hat{e} = (y - X\hat{\beta})/\hat{\sigma}$ and $X\hat{\beta}$ is the tangent plane at the observed data. From this we obtain the canonical reparameterization

$$\varphi'(\theta) = \frac{1}{\sigma}\sum_{i=1}^{n} s_i \left\{\frac{y_i - X_i\beta}{\sigma}\right\}\left\{\hat{e}_i, X_i\hat{\beta}\right\}$$

which is written as an $(r+1)$-dimension row vector. The special $r+1$ functions $\ell(\theta), \varphi_1(\theta), \ldots, \varphi_r(\theta)$ can be input to the computer program and numerical p-values can be obtained for any component parameter of interest, all following the procedures described in Section 2.

For some real data examples, see Fraser, Wong & Wu (1999). For normal error, we have no need for the present calculations of course, so for simulations, we consider the Student(6) error distribution. We have $N = 100,000$ repetitions from a regression model with a very small $n = 2$ data vector with $X = 1$ which gives just the location model $y_1 = \mu + \sigma e_1, y_2 = \mu + \sigma e_2$ and with the chosen values $\mu = 0, \sigma = 1$; we then tested the distribution of the p-value $p(\mu)$ for the null hypothesis $\mu = 0$. The observed proportions are recorded in Table 4.

Table 4: Observed proportions for $p(0)$
in the lower 0.5%, the next 2%, the next 47.5%,
in the next 47.5%, the next 2%, and the upper 0.5%

Method	Lower 0.5%	Next 2%	Next 47.5%
lr	1.975%	3.909%	44.325%
3rd	0.580%	1.567%	47.777%
Target	0.5%	2%	47.5%

Method	Next 47.5%	Next 2%	Upper 0.5%
lr	43.959%	3.613%	2.222%
3rd	47.722%	1.749%	0.605%
Target	47.5%	2%	0.5%

This records a simulation of a very extreme case with nonnormal error and with a sample of size $n = 2$. It shows the clear superiority of the higher order likelihood ratio method which in turn tends to be much more stable than the score or maximum likelihood departure methods. We do note that the values for the lower 0.5%, and next 2% and the corresponding upper 0.5% and previous 2% are much closer to the target value by the third order method than the likelihood ratio method; however, it is still not within the 95% confidence range of the proportion of coverage.

6 DISCUSSION

We have recorded the formulas for going from likelihood and canonical parameter to the p-value function for an arbitrary scalar parameter and illustrated the calculations with a succession of three examples involving models of progressively more comprehensive structure. The accuracy obtained far surpasses that available from the likelihood ratio calculations which in turn far exceed that available from Wald and Rao type statistics.

The application of these methods do require the calculations of a canonical reparameterization which has the role of a canonical parameter in an exponential model. The calculations involved are discussed in the three examples.

Our theme is that highly accurate essentially unique p-values are available by simple and direct procedures working from likelihood at and near an observed data point.

7. APPENDIX

The full information determinant calculated in the new parameterization is available as

$$|\jmath_{(\lambda\lambda)}| = |\jmath_{\theta\theta}(\hat{\theta})| \varphi_{\theta}(\hat{\theta})|^{-2}$$

using the Jacobian $\varphi_{\theta}(\theta) = \partial\phi(\theta)/\partial\theta'$. The nuisance information determinant in a somewhat similar way takes the form

$$|\jmath_{(\lambda\lambda)}(\hat{\theta}_{\psi})| = |j_{\lambda\lambda}(\hat{\theta}_{\psi})| \cdot |\varphi_{\lambda'}(\hat{\theta}_{\psi})|^{-2} = |j_{\lambda\lambda}(\hat{\theta}_{\psi})| \cdot |X'X|^{-1}$$

where the right hand determinant uses $X = \phi_{\lambda'}(\hat{\theta}_{\psi})$ and in the regression context records the volume on the regression surface as a proportion of the corresponding volume for regression coefficients; in the preceding formula this changes the scaling for the nuisance parameter to that derived from the φ parameterization. The expressions above are for the case where θ' is given as (ψ, λ') with an explicit nuisance parameterization; the more general version is recorded in Fraser, Reid & Wu (1999). The rotated coordinate $\chi(\theta)$ in the φ parameterization is obtained from the gradient vector of $\psi(\theta)$ at $\hat{\theta}_{\psi}$ and has the form

$$\chi(\theta) = \frac{\psi_{\varphi'}(\hat{\theta}_{\psi})}{|\psi_{\varphi'}(\hat{\theta}_{\psi})|} \cdot \varphi(\theta) ,$$

where the first factor is a row vector which is the unit vector corresponding to the gradient $\psi'_{\varphi}(\hat{\theta}_{\psi})$ and is obtained from

$$\psi_{\varphi'}(\theta) = \partial\psi(\theta)/\partial\varphi' = (\partial\psi(\theta)/\partial\theta') \cdot (\partial\varphi(\theta)/\partial\theta')^{-1} = \psi_{\theta'}(\theta)\varphi_{\theta'}^{-1}(\theta);$$

in this we take $\psi_{\varphi'}$ to be the Jacobian of the column vector ψ with respect to the row vector φ' and for example we would have $(\psi_{\varphi'})' = \psi'_{\varphi}$ for the transpose of the first Jacobian.

ACKNOWLEDGEMENT

The Natural Sciences and Engineering Research Council of Canada has provided support for this research.

REFERENCES

Abebe, F., Cakmak, S., Cheah, P.H., Fraser, D.A.S., Kuhn, J., McDunnough, P., Reid, N. & Tapia, A. (1995). Third order asymptotic model: exponential and location type approximations. *Parisankhyan Samikkha* **2**, 25-33.

Barndorff-Nielsen, O.E. (1986). Inference on full or partial parameters based on the standardized signed log-likelihood ratio. *Biometrika* **73**, 307-322.

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

Cakmak, S., Fraser, D.A.S., McDunnough, P., Reid, N. & Yuan, X. (1995). Likelihood centered asymptotic model: exponential and location model versions. *Int. J. math and Statist. Sci.* **4**, 1-12.

Daniels, H.E. (1954). Saddlepoint approximations in Statistics. *Ann.*

Math. Statist. **15**, 631-650.

Daniels, H.E. (1987). Tail probability approximations. *Int. Statist. Rev.* **55**, 37-48.

Fraser, D.A.S. (1990). Tail probabilities from observed likelihood. *Biometrika* **77**, 65-76.

Fraser, D.A.S. (2003). Ancillaries and conditional inference. *Journal of Statistical Planning and Inference.* **103**, 263-285.

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

Fraser, D.A.S. & Reid, N. (2001). Ancillary information for statistical inference. *Empirical Bayes and Likelihood Inference.* Ed. S.E. Ahmed and N. Reid, pp. 185-207. New York: Springer.

Fraser, D.A.S. & Reid, N. (2002). Strong matching of frequentist and Bayesian inference. *J. Statist. Plan. Infer.* **103**, 263-285.

Fraser, D.A.S., Reid, N. & Wu, J. (1999). A simple general formula for tail probabilities for frequentist and Bayesian inference. *Biometrika* **86**, 249-64.

Fraser, D.A.S., Wong, A. & Wu, J. (1999). Regression analysis, nonlinear or nonnormal: simple and accurate p values from likelihood analysis. *J. Amer. Statist. Assoc.* **94**, 1286-1295.

Fraser, D.A.S. & Yi, G.Y. (2003). Location reparametrization and default priors in statistical analysis. *J. Iranian Statist. Soc.* **1**, 55-78.

Kennan, J. (1985). The duration of contract strikes in U.S. manufacturing. *J. Econometrics* **28**, 5-28.

Kiefer, N. (1988). Economic duration data and hazard functions. *J. Econ. Lit.* **26**, 646-679.

Lugannani, R. & Rice, S. (1980). Saddlepoint approximation for the distribution function of the sum of independent variables. *Adv. Appl. Prob.* **12**, 475-490.