Numerical approximation of the p–value for generalized linear models

Jonathan R.D. Kuhn, Purdue University North Central
1401 S. US 421, Westville, Indiana 46391–9528 (jkuhn@purduenc.edu)

August 25, 2002

Key Words: interest and nuisance parameters, ancillarity, sufficiency, generalized linear model, profile likelihood, constrained maximum likelihood estimate
MSC 2000: Primary 62F99; secondary 62F03 62E20 62H17

Abstract: A numerical procedure for an approximation of the p–value associated with the likelihood ratio of a variable associated with the observed constrained maximum likelihood estimate which has been conditioned on a ancillary statistic, and which is used to test an interest parameter in the presence of nuisance parameters for a generalized linear model (McCullagh and Nelder, 1989), is described. Both the p^* approximation formula (Barndorff–Nielsen 1983) and the local tangent ancillary statistic (Fraser and Reid 2001) are used in this paper.

1 Introduction

In generalized linear models, the parameter space can be partitioned into two types of parameters,

$$\begin{aligned}\boldsymbol{\beta} &= (\beta_1, \ldots, \beta_{r-1}, \beta_r) \\ &= (\lambda_1, \ldots, \lambda_{r-1}, \psi) \\ &= (\boldsymbol{\lambda}, \psi)\end{aligned}$$

where $\boldsymbol{\lambda}$ is a vector of nuisance parameters and ψ is a scalar interest parameter. If it is possible to factor the likelihood into two pieces where one is a function of the nuisance parameters alone and the other is a function of the interest parameter alone, then it seems reasonable to use the second function for inference about the interest parameter since the first function is independent of the interest parameter ψ. (Cox and Hinkley 1974, Lindsey 1996, Reid 2000).

Many attempts have been made to find an appropriate factorization of the likelihood which use some combination of the variously defined notions of sufficiency, ancillarity and information (Barndorff–Nielsen 1978, Lindsey 1996). In particular, it has been suggested that it would be ideal to identify from the data variable $\mathbf{y}$ of dimension n, an ancillary statistic $\mathbf{a}(\mathbf{y})$ of dimension $n-r$ for all of the parameters, $\boldsymbol{\beta} = (\boldsymbol{\lambda}, \psi)$, such that the model factorizes as

$$f(\mathbf{y}; \boldsymbol{\beta}) = f_1(\mathbf{s};) f_2(\mathbf{s}|\mathbf{a}; \boldsymbol{\beta}).$$

and so the conditional likelihood $L_2(\boldsymbol{\beta}; \mathbf{s}|\mathbf{a})$ could, in general, be used for inference purposes (Fraser, McDunnough and Reid 1987, Reid 2000). Since this conditional likelihood is not available in general and in the present generalized linear model context in particular, it has further been suggested that the constrained maximum likelihood estimator of the interest parameter in the presence of the nuisance parameters, $\hat{\boldsymbol{\beta}}_\psi(\mathbf{s}|\mathbf{a})$, which appears in the profile likelihood for ψ,

$$L_p(\psi; \mathbf{s}|\mathbf{t}) = L_p(\psi, \hat{\boldsymbol{\beta}}_\psi(\mathbf{s}|\mathbf{a}); \mathbf{s}|\mathbf{a})$$

and, more particularly, the likelihood ratio of the profile likelihood for ψ,

$$\frac{L_p(\psi; \mathbf{s}|\mathbf{a})}{L_p(\hat{\psi}; \mathbf{s}|\mathbf{a})}$$

could be used for inference on ψ because, in this case, the profile likelihoods can then be approximated very accurately by the p^* formula in an asymptotic sense (Barndorff–Nielsen 1983, 1999).

One difficulty is the ancillary statistic required in the calculation of p^* formula calculation is left unspecified, aside from requiring that the dimension of the variable space be reduced to the same dimension as the parameter space upon conditioning on this ancillary statistic and other technical considerations. In other words, many possible particular ancillary statistics could be used in general and for the generalized linear model inference model in particular when using the p^* formula. The particular ancillary statistic used in this paper which allows the use of the p^* formula is the local tangent ancillary statistic suggested by Fraser and Reid (Fraser and Reid 1995, Fraser 1964). Although this statistic is technical in nature and so does not necessarily apply in any and all practical inference problems (Cox and Hinkley 1974), it does have the advantage of always reducing the dimension of the variable space to the required dimension of the parameter space and so always allow the use of the accurate p^* formula approximation.

The test statistic developed in this paper is the variable associated with the observed constrained maximum likelihood estimate of the interest parameter in the presence of the nuisance parameters, which has been conditioned on the local tangent ancillary statistic $\mathbf{a}(\mathbf{y}) = \mathbf{y}$. This test statistic is defined by a system of differential equations and describes a one–dimensional curve in n–dimensional variable space. Since the test statistic is a one–dimensional curve, this allows for convenient integration of the associated distribution defined on this curve and so calculation of the p–value for testing purposes.

The p–value is approximated by a numerical procedure in two steps. First, the likelihood ratio of the profile likelihood for ψ,

$$\frac{L_p(\psi; \mathbf{s}|\mathbf{a})}{L_p(\hat{\psi}; \mathbf{s}|\mathbf{a})},$$

is calculated at points all along the one–dimensional test statistic curve. Second, the sum of likelihood ratios from the point on this curve specified by the null hypothesized value of the interest parameter, ψ_0, is divided by the sum of the likelihood ratios over the entire length of the test statistic curve. Further differential–geometric analytical methods in statistics can be found in, for example, Amari (1985).

The version of the generalized linear model and testing problem assumed in this paper is described. Both the system of differential equations of the test statistic and the associated likelihood ratio of the profile likelihood for ψ in the context of this generalized linear model are derived. The results of a numerical example are described, followed by some discussion.

2 Generalized Linear Model, Inference and Some Notation

After the version of the generalized linear model used for this analysis is described, some notation is specified.

Generalized Linear Model and Inference For Interest Parameter The version of the generalized linear model assumed in this paper consists of three components. The random component consists of a sequence of independent random variables $(y_1, \ldots, y_n)$,

$$y_i \sim F_i(y_i; \theta_i), \quad i = 1, \ldots, n,$$

where each distribution F_i is continuous and monotonic in the parameter θ_i. The systematic component is given by,

$$\begin{pmatrix} \eta_1 \\ \vdots \\ \eta_n \end{pmatrix} = \begin{pmatrix} x_{11} & \ldots & x_{1r} \\ \vdots & \ddots & \vdots \\ x_{n1} & \ldots & x_{nr} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_r \end{pmatrix}, \quad r \leq n,$$

or, in short, $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$. The inverse link functions are given by,

$$\theta_i = g(\eta_i), \quad i = 1, \ldots, n,$$

where g is a monotonic differentiable function.

The problem dealt with in this paper is: given an observed set of the random variables, $(y_1^0, \ldots, y_n^0)$, as specified in detail above, and in the presence of the unknown nuisance test parameters, $(\beta_1, \ldots, \beta_{r-1}) = (\lambda_1, \ldots, \lambda_{r-1})$, test an interest parameter $\beta_r = \psi$ by calculating an appropriate p–value.

Notation Capital bold letters are matrices. For example, $\mathbf{A}$ is a matrix. The matrix $\mathbf{A}_{(r-1)\times n}$ is a $(r-1)\times n$ matrix.

The notation $\mathrm{d}y_i$, $i = 1, \ldots, n$ is used to describe the differentials (one–forms) of the variables and $\mathrm{d}\theta_i$, $i = 1, \ldots, n$ are differentials for the parameters. A (column) n–vector of variable differentials is given by $\mathbf{dy} = (\mathrm{d}y_1, \ldots, \mathrm{d}y_n)^t$. A (column) n–vector of parameter differentials is given by $\mathbf{d}\boldsymbol{\beta} = (\mathrm{d}\beta_1, \ldots, \mathrm{d}\beta_n)^t$.

The notation used to describe the derivative of the log–likelihood $\tilde{\ell} = \sum_{i=1}^n \ell_i$, $ell_i = \log L_i$, with respect to the parameter $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_{r-1})$ is given by $\tilde{\ell}_{\boldsymbol{\lambda};}$. In a similar way, the derivative of the log–likelihood $\tilde{\ell}$ with respect to the variable $\mathbf{y} = (y_1, \ldots, y_n)$ is given by $\tilde{\ell}_{;\mathbf{y}}$.

Since the systematic component of the generalized linear model is given by $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$, with inverse link functions, $\theta_i = g(\eta_i)$, $i = 1, \ldots, n$, the following notation, with regard to constrained (with respect to the interest parameter ψ) maximum likelihood estimators is used,

$$
\begin{aligned}
\eta_i(\hat{\boldsymbol{\beta}}_\psi) &= \eta_i(\hat{\boldsymbol{\lambda}}_\psi, \psi), \\
&= \eta_i(\hat{\lambda}_1^\psi, \ldots, \hat{\lambda}_{r-1}^\psi, \psi) \\
&= x_{i1}\hat{\lambda}_1^\psi + \ldots + x_{i,r-1}\hat{\lambda}_{r-1}^\psi + x_{i,r}\psi \\
&= \hat{\eta}_i^\psi, \quad i = 1, \ldots, n,
\end{aligned}
$$

and,

$$\hat{\boldsymbol{\eta}}_\psi = (\hat{\eta}_1^\psi, \ldots, \hat{\eta}_n^\psi),$$

and also,

$$\hat{\boldsymbol{\theta}}_\psi = (g(\hat{\eta}_1^\psi), \ldots, g(\hat{\eta}_n^\psi)) = (\hat{\theta}_1^\psi, \ldots, \hat{\theta}_n^\psi).$$

Neither of the two functions, $\hat{\boldsymbol{\theta}}_\psi$, and $\hat{\boldsymbol{\eta}}_\psi$, are, in general, true maximum likelihood estimators, but are, in fact, only functions of the true constrained maximum likelihood estimator, $\hat{\boldsymbol{\beta}}_\psi$.

3 Test Statistic For Generalized Linear Model

There are two important pieces to the system of differential equations that describe the test statistic for the generalized linear model. One is the local tangent ancillary statistic, $\mathbf{a}(\mathbf{y}) = \mathbf{a}$, and the other is the variable associated with the observed constrained maximum likelihood estimate of the interest parameter in the presence of the nuisance parameters. The systems of differential equations associated with these two pieces are first discussed separately and then combined into a single system of differential equations for the one–dimensional test statistic curve. Simple analytical (non–numerical) examples are also given.

3.1 Local Tangent Ancillary Statistic

The local tangent ancillary statistic, $\mathbf{a}(\mathbf{y}) = \mathbf{a}$, is defined by the assumption that all of the total differentials of the distributions of the independent random variables $(y_1, \ldots, y_n)$, $F_i(y_i; \theta_i)$, are zero,

$$\mathrm{d}F_i = F^i_{y_i;}\mathrm{d}y_i + F^i_{;\theta_i}\mathrm{d}\theta_i = 0, \quad i = 1, \ldots, n.$$

Roughly speaking, this assumption imposes a restriction on the "movement" of a parameter, θ_i, and a variable, y_i with respect to one another. If the parameter changes by a $\mathrm{d}\theta_i$ amount, this causes (or forces) a change of

$$\mathrm{d}y_i = -\frac{F^i_{;\theta_i}}{F^i_{y_i;}}\mathrm{d}\theta_i \quad i = 1, \ldots, n$$

in the variable (and vis–versa), while, at the same time, the value of each $F_i(y_i; \theta_i)$ remains constant (does not change) with respect to changes in either the variable or parameter, $F_i(y_i; \theta_i) = c$, $i = 1, \ldots, n$. More than this, Fraser and Reid (1995, 2001) suggest this system of differential equations is ancillary in a local tangent asymptotic approximate sense. This type of ancillary was first proposed in Fraser (1964).

Applying the local tangent ancillary assumption to the generalized linear model (Kuhn 1994, pp. 78 and 83), the following local tangent ancillary directions system of differential equations can be derived as,

$$\mathbf{dy} = \mathbf{V}\mathbf{d}\boldsymbol{\beta}, \tag{1}$$

where $\mathbf{V} = -\mathbf{F}\mathbf{E}\mathbf{X}_{n\times r}$ and

$$\mathbf{F} = \mathrm{diag}_n\left\{\frac{F^i_{;\theta_1}}{F^i_{y_i;}}|_{\hat{\theta}_i}\right\}, \quad \mathbf{E} = \mathrm{diag}_n\left\{\frac{\partial g(\eta_i)}{\partial \eta_i}|_{\hat{\eta}_i}\right\}, \quad \mathbf{X}_{n\times r} = \begin{pmatrix} x_{11} & \ldots & x_{1r} \\ \vdots & \ddots & \vdots \\ x_{n1} & \ldots & x_{nr} \end{pmatrix},$$

and where this local set of differential equations is evaluated at the maximum likelihood estimators $\hat{\eta}_i$, $i = 1, \ldots, n$.

Exponential Example If the random component of the generalized linear model is the exponential distribution,

$$F_i(y_i; \theta_i) = 1 - \exp\{-y_i\theta_i\}, \quad y_i > 0, \quad i = 1, \ldots, 4,$$

where each F_i is monotonic with respect to each θ_i, and the systematic component is

$$\begin{pmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \\ \eta_4 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}, = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{pmatrix} \begin{pmatrix} \lambda \\ \psi \end{pmatrix},$$

and the link function is $\theta_i = g(\eta_i) = \exp(\eta_i)$, $i = 1, \ldots, n$, where g is monotonic, then

$$\frac{F_{;\theta_i}}{F_{y_i;}} = \frac{y_i e^{-y_i\theta_i}}{\theta_i e^{-y_i\theta_i}} = \frac{y_i}{\theta_i} \quad i = 1, \ldots, 4,$$

and

$$\frac{\partial g(\eta_i)}{\partial \eta_i} = \frac{\partial e^{\eta_i}}{\partial \eta_i} = e^{\eta_i} = g(\eta_i) = \theta_i, \quad i = 1, \ldots, 4.$$

and so

$$\mathbf{V} = -\mathbf{FEX}_{n\times r} = -\begin{pmatrix} y_1 & y_1 \\ y_2 & 2y_2 \\ y_2 & 3y_2 \\ y_n & 4y_n \end{pmatrix}.$$

At the variable point $\mathbf{y}_0 = (y_1^0, y_2^0, y_3^0, y_4^0) = (2, 3, 4, 5)$, for example, where $\mathbf{dy} = \mathbf{V}\mathbf{d}\boldsymbol{\beta}$ or $-\mathbf{dy} + \mathbf{V}\mathbf{d}\boldsymbol{\beta} = \mathbf{0}$ and so

$$\begin{pmatrix} -\mathbf{I}_{4\times 4} & \mathbf{V}_{4\times 2} \end{pmatrix} \begin{pmatrix} \mathrm{d}\mathbf{y} \\ \mathrm{d}\boldsymbol{\beta} \end{pmatrix} = \begin{pmatrix} -1 & 0 & 0 & 0 & 2 & 2 \\ 0 & -1 & 0 & 0 & 3 & 6 \\ 0 & 0 & -1 & 0 & 4 & 12 \\ 0 & 0 & 0 & -1 & 5 & 20 \end{pmatrix} \begin{pmatrix} \mathrm{d}y_1 \\ \mathrm{d}y_2 \\ \mathrm{d}y_3 \\ \mathrm{d}y_4 \\ \mathrm{d}\lambda \\ \mathrm{d}\psi \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

This system of differential equations maps from the six–dimensional ($n+r = 4+2 = 6$) variable–parameter space to a two–dimensional plane (the local tangent ancillary statistic), which, intentionally, is the same dimension as the parameter space alone.

In this case, the explicit analytic calculation of the local tangent ancillary statistic is relatively easy to do. In particular, the θ_i parameters cancel out during the analysis and, as a consequence, the maximum likelihood estimators, $\hat{\theta}_i$ need not be determined. However, in general, the determination of the maximum likelihood estimators and the $F_{y_i;}$, $F_{;\theta_i}$ for the local tangent ancillary statistic are difficult to determine. More than this, since the analysis is a local one, these calculations need to be repeated at every variable point $\mathbf{y}_0$ along the test statistic curve. It is because of these considerations, that the calculation of the local tangent ancillary statistic is done in a numerical fashion.

3.2 Observed Constrained Maximum Likelihood Estimator

The variable associated with the observed constrained maximum likelihood estimate of the interest parameter in the presence of the nuisance parameters is used for testing purposes in this paper essentially because the likelihood ratio of the profile likelihood associated with this statistic can be approximated very accurately by the p^* formula in an asymptotic sense (Barndorff–Nielsen 1983, 1999). For this statistic, the score of $\tilde{\ell}(\hat{\boldsymbol{\theta}};\mathbf{y}) = \sum_{i=1}^{n} \ell_i(\theta_i; y_i)$ with respect to the nuisance parameters, at the constrained maximum likelihood estimator of the parameter, $\hat{\theta}_\psi$, is zero; that is,

$$\frac{\partial\tilde{\ell}}{\partial\boldsymbol{\lambda}}|_{\hat{\theta}_\psi} = \mathbf{0}.$$

Since the score of the constrained maximum likelihood estimator, $\tilde{\ell}_{\boldsymbol{\lambda}}(\hat{\boldsymbol{\theta}};\mathbf{y})|_{\hat{\theta}_\psi}$, is zero (implying this estimator is constant with respect to the parameters in a local sense), then the total differential of this score is zero, $\mathrm{d}\tilde{\ell}_{\boldsymbol{\lambda}}(\hat{\boldsymbol{\theta}};\mathbf{y})|_{\hat{\theta}_\psi} = \mathbf{0}$, and, more than this, is influenced by the variable $\mathbf{y}$ only,

$$\begin{aligned} \mathrm{d}\tilde{\ell}_{\boldsymbol{\lambda}}(\hat{\boldsymbol{\theta}};\mathbf{y})|_{\hat{\theta}_\psi} &= \tilde{\ell}_{\lambda,\theta_\psi;}(\theta_\psi;\mathbf{y})|_{\hat{\theta}_\psi}\mathrm{d}\theta_\psi + \tilde{\ell}_{\lambda;\mathbf{y}}(\theta_\psi;\mathbf{y})|_{\hat{\theta}_\psi}\mathrm{d}\mathbf{y} \\ &= \tilde{\ell}_{\lambda;\mathbf{y}}(\theta_\psi;\mathbf{y})|_{\hat{\theta}_\psi}\mathrm{d}\mathbf{y} = \mathbf{0}. \end{aligned}$$

Whereas for the local tangent ancillary statistic above, where each $F_i(y_i;\theta_i)$ remains constant with respect to changes in either the variable or parameter, $F_i(y_i;\theta_i) = c$, $i = 1,\ldots,n$, here, for the constrained maximum likelihood estimator, $\tilde{\ell}(\hat{\boldsymbol{\theta}};\mathbf{y})$ remains constant (in a local sense) with respect to the parameters alone. In other words, each statistic exhibits a different kind of local ancillarity.

For the generalized linear model, $\mathrm{d}\tilde{\ell}_{\boldsymbol{\lambda}}(\hat{\boldsymbol{\theta}};\mathbf{y})|_{\hat{\theta}_\psi} = \mathbf{0}$ can be described by the following system of differential equations (Kuhn 1994, pp. 90–92),

$$\mathbf{A}\mathbf{d}\mathbf{y} = \mathbf{0}, \tag{2}$$

where $\mathbf{A} = \mathbf{X}^t_{n\times(r-1)}\mathbf{D}_\psi\mathbf{E}_\psi$ and where

$$\mathbf{X}^t_{n\times(r-1)} = \begin{pmatrix} x_{11} & \dots & x_{1,r-1} \\ \vdots & \ddots & \vdots \\ x_{n1} & \dots & x_{n,r-1} \end{pmatrix}^t, \quad \mathbf{D}_\psi = \mathrm{diag}_n \left\{ \ell^i_{\theta_i;y_i}|_{\hat{\theta}^\psi_i} \right\}, \quad \mathbf{E}_\psi = \mathrm{diag}_n \left\{ \frac{\partial g(\eta_i)}{\partial \eta_i}|_{\hat{\eta}^\psi_i} \right\}.$$

The constrained maximum likelihood estimator, $\hat{\boldsymbol{\beta}}_\psi$, maps from the $(n+1)$–dimensional variable–(interest parameter) space to an $(r-1)$–dimensional space. At an observed point in the variable space, $\mathbf{y}_0$, and for a hypothesized parameter value, ψ_0, this estimator becomes an estimate and takes on one value, $\hat{\boldsymbol{\beta}}_\psi(\mathbf{y}_0;\psi_0) = \hat{\boldsymbol{\beta}}^0_\psi$. In addition to the one observed variable point, $\mathbf{y}_0$, there are typically other variable points, $\mathbf{y}_1, \mathbf{y}_2, \dots$, which also have the same estimator value, $\hat{\boldsymbol{\beta}}_\psi(\mathbf{y}_1;\psi_0) = \hat{\boldsymbol{\beta}}^0_\psi$, $\hat{\boldsymbol{\beta}}_\psi(\mathbf{y}_2;\psi_0) = \hat{\boldsymbol{\beta}}^0_\psi$, ... This set of points, the statistic $\{\mathbf{y}|\hat{\boldsymbol{\beta}}_\psi(\mathbf{y};\psi_0) = \hat{\boldsymbol{\beta}}^0_\psi\}$, is what is described by $\mathbf{A}\mathbf{d}\mathbf{y} = \mathbf{0}$. The observed constrained maximum likelihood estimate remains unchanged (constant) for all values of the statistic described by $\mathbf{A}\mathbf{d}\mathbf{y} = \mathbf{0}$. It is more typical to discuss the profile likelihood associated with the constrained maximum likelihood estimator statistic (Barndorff–Nielsen 1983, Reid 2000), rather than focusing on the differential–geometric interpretation (Amari 1985) of the statistic itself , as is done here.

Exponential Example Revisited Continuing with the exponential example started above, since

$$\ell(\theta_i;y_i) = \ln\theta_i - y_i\theta_i, \quad \ell_{\theta_i;}(\theta_i;y_i) = \frac{1}{\theta_i} - y_i, \quad \ell_{\theta_i;y_i}(\theta_i;y_i) = -1,$$

and $\frac{g(\eta_i)}{\eta_i} = \theta_i$, for $i = 1, \dots, 4$,

$$\mathbf{A} \;=\; \mathbf{X}^t_{n\times(r-1)}\mathbf{D}_\psi\mathbf{E}_\psi$$

$$
\begin{aligned}
&= \begin{pmatrix} 2 \\ 3 \\ 4 \\ 5 \end{pmatrix}^t \begin{pmatrix} \ell^1_{\theta_1;y_1}|_{\hat\theta_1^\psi} \frac{\partial g(\eta_1)}{\partial \eta_1}|_{\hat\eta_1^\psi} & 0 & 0 & 0 \\ 0 & \ell^2_{\theta_2;y_2}|_{\hat\theta_2^\psi} \frac{\partial g(\eta_2)}{\partial \eta_2}|_{\hat\eta_2^\psi} & 0 & 0 \\ 0 & 0 & \ell^3_{\theta_3;y_3}|_{\hat\theta_3^\psi} \frac{\partial g(\eta_3)}{\partial \eta_3}|_{\hat\eta_3^\psi} & 0 \\ 0 & 0 & 0 & \ell^4_{\theta_4;y_4}|_{\hat\theta_4^\psi} \frac{\partial g(\eta_4)}{\partial \eta_4}|_{\hat\eta_4^\psi} \end{pmatrix} \\
&= \begin{pmatrix} -2 & -3 & -4 & -5 \end{pmatrix} \begin{pmatrix} -\hat\theta_1^\psi & 0 & 0 & 0 \\ 0 & -\hat\theta_2^\psi & 0 & 0 \\ 0 & 0 & -\hat\theta_3^\psi & 0 \\ 0 & 0 & 0 & -\hat\theta_4^\psi \end{pmatrix} \\
&= \begin{pmatrix} 2\hat\theta_1^\psi & 3\hat\theta_2^\psi & 4\hat\theta_3^\psi & 5\hat\theta_4^\psi \end{pmatrix},
\end{aligned}
$$

where $\hat\theta_i^\psi = \exp\{x_{i1}\hat\beta_1^\psi + x_{i2}\psi\}$, $i = 1, 2, 3, 4$, and so

$$
\mathbf{A}_{1\times 4}\mathrm{d}\mathbf{y} = \begin{pmatrix} 2\hat\theta_1^\psi & 3\hat\theta_2^\psi & 4\hat\theta_3^\psi & 5\hat\theta_4^\psi \end{pmatrix} \begin{pmatrix} \mathrm{d}y_1 \\ \mathrm{d}y_2 \\ \mathrm{d}y_3 \\ \mathrm{d}y_4 \end{pmatrix} = 0.
$$

This system of differential equations maps from the four–dimensional variable space to a four–dimensional hyperplane. Unlike for the local tangent ancillary statistic above, where the θ_i parameters cancel out during the analysis and, as a consequence, the maximum likelihood estimators, $\hat\theta_i$, need not be determined, here, in this case, they do not cancel out and are difficult to determine without using a numerical analysis.

3.3 Test Statistic

The test statistic is the variable associated with the observed constrained maximum likelihood estimate of the interest parameter in the presence of the nuisance parameters, $\{\mathbf{y}|\hat{\boldsymbol\beta}_\psi(\mathbf{y};\psi_0) = \hat{\boldsymbol\beta}^0_\psi\}$, which has been conditioned on the local tangent ancillary statistic $\mathbf{a}(\mathbf{y}) = \mathbf{y}$. This amounts to finding the intersection of the two systems of differential equations, equations 1 and 2, given above,

$$
\boldsymbol\phi_{(n+r-1)\times(n+r)} = \begin{pmatrix} \mathbf{A}_{(r-1)\times n} & \mathbf{0}_{(r-1)\times r} \\ -\mathbf{I}_{n\times n} & \mathbf{V}_{n\times r} \end{pmatrix} \begin{pmatrix} \mathrm{d}\mathbf{y} \\ \mathrm{d}\boldsymbol\beta \end{pmatrix} = \mathbf{0}_{n+r-1}
$$

where $\mathbf{A}\mathbf{dy} = \mathbf{0}$ has been rewritten as $\mathbf{A}\mathbf{dy} + \mathbf{0}\mathbf{d}\boldsymbol\beta = \mathbf{0}$. The test statistic system of differential equations is an example of a Pfaffian system of one–forms and the conditions under which this system is integrable are given in, for example, Abraham, Marsden and Ratiu (1983).

Exponential Example Revisited At the variable point $\mathbf{y}_0 = (y_1^0, y_2^0, y_3^0, y_4^0) = (2, 3, 4, 5)$,

$$\begin{pmatrix} \mathbf{A}_{1\times 4} & \mathbf{0}_{1\times 2} \\ -\mathbf{I}_{4\times 4} & \mathbf{V}_{4\times 2} \end{pmatrix} \begin{pmatrix} \mathrm{d}\mathbf{y} \\ \mathrm{d}\boldsymbol{\beta} \end{pmatrix} = \begin{pmatrix} 2\hat{\theta}_1^{\psi} & 3\hat{\theta}_2^{\psi} & 4\hat{\theta}_3^{\psi} & 5\hat{\theta}_4^{\psi} & 0 & 0 \\ -1 & 0 & 0 & 0 & 2 & 2 \\ 0 & -1 & 0 & 0 & 3 & 6 \\ 0 & 0 & -1 & 0 & 4 & 12 \\ 0 & 0 & 0 & -1 & 5 & 20 \end{pmatrix} \begin{pmatrix} \mathrm{d}y_1 \\ \mathrm{d}y_2 \\ \mathrm{d}y_3 \\ \mathrm{d}y_4 \\ \mathrm{d}\lambda \\ \mathrm{d}\psi \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

This system of differential equations maps from the six–dimensional variable–parameter space to a one–dimensional line (the test statistic).

Normal Example Consider the problem where,

$$y_i \sim F(y_i; \theta_i) = N(y_i; \theta_i, \sigma_i), \quad i = 1, 2, 3$$

where, $\sigma_1 = \sigma_2 = \sigma_3 = 1$ and $\theta_i = \eta_i$, $i = 1, 2, 3$, where,

$$\begin{pmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \end{pmatrix} = \begin{pmatrix} 1 & -3 \\ -2 & 1 \\ -3 & -2 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}.$$

Locally, at the observed variable point $\mathbf{y}_0 = (1, 1, 1)$ and for hypothesized interest parameter value, $\psi_0 = 2$, it can be shown the test statistic system of differential equations is given by,

$$\begin{pmatrix} \mathbf{A}_{1\times 3} & \mathbf{0}_{1\times 2} \\ -\mathbf{I}_{3\times 3} & \mathbf{V}_{3\times 2} \end{pmatrix} \begin{pmatrix} \mathrm{d}\mathbf{y} \\ \mathrm{d}\boldsymbol{\beta} \end{pmatrix} = \begin{pmatrix} 1 & -2 & -3 & 0 & 0 \\ -1 & 0 & 0 & 1 & -3 \\ 0 & -1 & 0 & -2 & 1 \\ 0 & 0 & -1 & -3 & -2 \end{pmatrix} \begin{pmatrix} \mathrm{d}y_1 \\ \mathrm{d}y_2 \\ \mathrm{d}y_3 \\ \mathrm{d}\lambda \\ \mathrm{d}\psi \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix},$$

where, because $\boldsymbol{\phi}_{4\times 5}$ is (locally) constant with respect to $(\mathbf{y}, \boldsymbol{\beta}) = (y_1, y_2, y_3, \lambda, \psi)$, can be rewritten as the following (easier to solve) system of linear *equations* (rather than *differential equations*),

$$\begin{pmatrix} 1 & -2 & -3 & 0 & 0 \\ -1 & 0 & 0 & 1 & -3 \\ 0 & -1 & 0 & -2 & 1 \\ 0 & 0 & -1 & -3 & -2 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \lambda \\ \psi \end{pmatrix} = \begin{pmatrix} c \\ c \\ c \\ c \end{pmatrix},$$

where, if the constant of integration is zero, $c = 0$, has a parametric solution, of $(-\frac{43}{14}b, \frac{8}{7}b, -\frac{25}{14}b, -\frac{1}{14}b, b)$, where the parameter is b, or, using basis vectors $\mathbf{e}_1 = (1,0,0,0,0)^t$, ..., $\mathbf{e}_5 = (0,0,0,0,1)^t$, of

$$-\frac{43}{14}b\mathbf{e}_1 + \frac{8}{7}b\mathbf{e}_2 - \frac{25}{14}b\mathbf{e}_3 - \frac{1}{14}b\mathbf{e}_4 + b\mathbf{e}_5$$

or, with respect to the variables (y_1, y_2, y_3) alone, of

$$-\frac{43}{14}b\begin{pmatrix}1\\0\\0\end{pmatrix} + \frac{8}{7}b\begin{pmatrix}0\\1\\0\end{pmatrix} - \frac{25}{14}b\begin{pmatrix}0\\0\\1\end{pmatrix}$$

which is the one–dimensional test statistic curve. The given observed variable point $\mathbf{y}_0 = (1,1,1)$ is defined to be (must be) on this curve. In a numerical procedure, where the parametric value is, say, $b = 14 \times 0.1$, the next variable point $\mathbf{y}_1$ of this test statistic curve is

$$\mathbf{y}_1 = \mathbf{y}_0 + \mathbf{dy} = \begin{pmatrix}1\\1\\1\end{pmatrix} + \begin{pmatrix}-\frac{43}{14} \times 14 \times 0.01\\ \frac{8}{7} \times 14 \times 0.01\\ -\frac{25}{14} \times 14 \times 0.01\end{pmatrix} = \begin{pmatrix}0.57\\1.16\\0.75\end{pmatrix}$$

The test statistic is, in this case, a line, although, in general, it need not be (it could be curved).

4 Test Statistic Distribution For Generalized Linear Model

Applying a version of the p^* formula with an appropriate Jacobian to allow for invariant reparameterization (Barndorff–Nielsen and Cox 1994) to both the numerator and denominator of the likelihood ratio of the profile likelihood for interest parameter ψ,

$$\frac{L_p(\psi; \mathbf{s}|\mathbf{a})}{L_p(\hat{\psi}; \mathbf{s}|\mathbf{a})},$$

a local asymptotic approximate distribution for the test statistic is given by (Kuhn 1994),

$$\frac{c}{(2\pi)^{\frac{1}{2}}}\exp\{\tilde{\ell}(\hat{\theta}_\psi; \mathbf{y}) - \tilde{\ell}(\hat{\theta}; \mathbf{y})\}\frac{|\tilde{\ell}_{\beta;\mathbf{V}}(\hat{\theta}; \mathbf{y})||\jmath_{\boldsymbol{\lambda}}|^{1/2}}{|\tilde{\ell}_{\boldsymbol{\lambda};\mathbf{V}}(\hat{\theta}_\psi; \mathbf{y})||\jmath|^{1/2}}$$

where,

$$\begin{aligned}
\tilde{\ell}_{\boldsymbol{\beta};\mathbf{V}}(\hat{\theta};\mathbf{y}) &= \mathbf{X}^t_{n\times r}\mathbf{DEV}, \\
\tilde{\ell}_{\boldsymbol{\lambda};\mathbf{V}}(\hat{\theta}_\psi;\mathbf{y}) &= \mathbf{X}^t_{n\times(r-1)}\mathbf{D}_\psi\mathbf{E}_\psi\mathbf{V}, \\
\hat{\jmath} &= -\mathbf{X}^t_{n\times r}(\mathbf{E}^t\mathbf{JE}+\mathbf{CL})\mathbf{X}_{n\times r}, \\
\hat{\jmath}_{\boldsymbol{\lambda}} &= -\mathbf{X}^t_{n\times(r-1)}(\mathbf{E}_\psi\mathbf{J}_\psi\mathbf{E}^t_\psi+\mathbf{C}_\psi\mathbf{L}_\psi)\mathbf{X}_{n\times(r-1)},
\end{aligned}$$

and, in addition to the previously defined matrices, there is also,

$$\mathbf{C}=\mathrm{diag}_n\left\{\frac{\partial g(\eta_i)}{\partial\eta_i\partial\eta_i}|_{\hat{\eta}_i}\right\},\quad \mathbf{C}_\psi=\mathrm{diag}_n\left\{\frac{\partial g(\eta_i)}{\partial\eta_i\partial\eta_i}|_{\hat{\eta}_i^\psi}\right\},\quad \mathbf{D}_\psi=\mathrm{diag}_n\left\{\ell^i_{\theta_i;y_i}|_{\hat{\theta}^i_\psi}\right\},$$

$$\mathbf{L}=\mathrm{diag}_n\left\{\ell^i_{\theta_i;}(\hat{\theta};\mathbf{y})\right\},\quad \mathbf{L}_\psi=\mathrm{diag}_n\left\{\ell^i_{\theta_i;}(\hat{\theta}_\psi;\mathbf{y})\right\},$$

$$\mathbf{J}=\mathrm{diag}_n\left\{\ell^i_{\theta_i\theta_i;}(\hat{\theta};\mathbf{y})\right\},\quad \mathbf{J}_\psi=\mathrm{diag}_n\left\{\ell^i_{\theta_i\theta_i;}(\hat{\theta}_\psi;\mathbf{y})\right\}.$$

This local asymptotic approximate distribution for the test statistic is calculated at points all along the one–dimensional test statistic curve, ultimately to be used to determine the p–value.

5 Examples of Numerical Calculation of P–Value

In general, the p–value is approximated by the numerical procedure in two steps. First, the local asymptotic approximation of the likelihood ratio of the profile likelihood for ψ,

$$\frac{L_p(\psi;\mathbf{s}|\mathbf{a})}{L_p(\hat{\psi};\mathbf{s}|\mathbf{a})},$$

is calculated at points plotted all along the one–dimensional test statistic curve, $\{\mathbf{y}|\hat{\boldsymbol{\beta}}_\psi(\mathbf{y};\psi_0)=\hat{\boldsymbol{\beta}}^0_\psi\}$. Second, the sum of these values starting from the observed point, $\mathbf{y}_0$, on this curve and also specified by the null hypothesized value of the interest parameter, ψ_0, is divided by the sum of all values over the entire length of the test statistic curve. A normal example is described in detail; other examples are also given.

5.1 Normal Example

Let $n = 2$ and $r = 2$. The underlying distribution, F_i, $i = 1, 2$, is chosen to be the non-standard normal with interest distribution mean θ_i and nuisance distribution standard deviation, σ_i, $i = 1, 2$, arbitrarily set equal to 2. The initial point on the test statistic curve is the observed variable value, assumed to be $\mathbf{y}_0 = (y_1^0, y_2^0) = (1, 1)$. The interest test parameter is assumed to be $\beta_2^\psi = \psi = 3$ and the one nuisance test parameter is given by $\beta_1^\psi = \lambda$. The systematic component is given by,

$$\mathbf{X}_{2\times 2} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix},$$

and let the inverse link function be the identity, where $\theta_i = \eta_i$, $i = 1, 2$. The p–value is calculated to test the interest parameter, $\psi = 3$ at the observed point $\mathbf{y}_0 = (y_1^0, y_2^0) = (1, 1)$.

The first couple of iterations of the numerical method are given in some detail. Then, the remaining iterations are summarized into a table which gives a sequence of variable values of the test statistic curve and also the associated local asymptotic approximation values at these variable points.

Iteration 1: Density Approximation at $\mathbf{y}_0$*.* The un–normed density approximation at $\mathbf{y}_0$ given by,

$$\begin{aligned} f_0 &\approx \frac{c}{2\pi^{\frac{1}{2}}} \exp\left\{\tilde{\ell}_0(\hat{\theta}_\psi^0; \mathbf{y}) - \tilde{\ell}_0(\hat{\theta}_0; \mathbf{y})\right\} \frac{|\tilde{\ell}^0_{\beta;\mathbf{V}}(\hat{\theta}_0; \mathbf{y})||\hat{\jmath}^0_\lambda|^{1/2}}{|\tilde{\ell}^0_{\lambda;\mathbf{V}}(\hat{\theta}^0_\psi; \mathbf{y})||\hat{\jmath}_0|^{1/2}} \\ &= \frac{c}{2\pi^{\frac{1}{2}}} \exp\{-8.238 + 1.838\} \frac{(8.994)(2.236)}{(40.975)(3)}, \\ &\approx 7.7 \times 10^{-5}. \end{aligned}$$

Iteration 2: Test Statistic Calculation, $\mathbf{y}_1^+ = \mathbf{y}_0 + \mathbf{dy}_0^+$ *and* $\mathbf{y}_1^- = \mathbf{y}_0 + \mathbf{dy}_0^-$*.* The first variable value point along the test statistic curve in the positive direction from the observed point, $\mathbf{y}_1^+ = \mathbf{y}_0 + \mathbf{dy}_0^+$, is determined by first calculating $\mathbf{dy}_0^+$ in the test statistic system of equations, where $\mathbf{A}_0 = \mathbf{A}_0^+$ and $\mathbf{V}_0 = \mathbf{V}_0^+$,

$$\begin{pmatrix} \mathbf{A}^{0,+} & \mathbf{0} \\ -\mathbf{I} & \mathbf{V}^{0,+} \end{pmatrix} \begin{pmatrix} \mathbf{dy}_0^+ \\ \mathbf{d}\beta_0^+ \end{pmatrix} = \begin{pmatrix} 0.5 & 0.25 & 0 & 0 \\ -1 & 0 & 2 & 1 \\ 0 & -1 & 1 & 2 \end{pmatrix} \begin{pmatrix} \mathrm{d}y_1^+ \\ \mathrm{d}y_2^+ \\ \mathrm{d}\lambda_1^+ \\ \mathrm{d}\psi^+ \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

and so,

$$\mathbf{dy}_0^+ = \begin{pmatrix} -0.075 \\ 0.150 \end{pmatrix};$$

thus,

$$\mathbf{y}_1^+ = \mathbf{y}_0 + \mathbf{dy}_0^+ = \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \begin{pmatrix} -0.075 \\ 0.150 \end{pmatrix} = \begin{pmatrix} 0.925 \\ 1.150 \end{pmatrix}.$$

In a similar way, the first variable value point along the test statistic curve in the negative direction from the observed point is given by,

$$\mathbf{y}_1^- = \mathbf{y}_0 + \mathbf{dy}_0^- = \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \begin{pmatrix} 0.075 \\ -0.150 \end{pmatrix} = \begin{pmatrix} 1.075 \\ 0.850 \end{pmatrix}.$$

Iteration 2: Density Approximation at $\mathbf{y}_1^+$ and $\mathbf{y}_1^-$. The un–normed density approximations at $\mathbf{y}_1^+$ and $\mathbf{y}_1^-$ are given by, respectively,

$$f_1^+ \approx 6.7 \times 10^{-4}, \quad f_1^- \approx 5.6 \times 10^{-6}.$$

Table of Test Statistic and Approximate Distribution Values. The numerical procedure continues calculating points of the test statistic, $\{\mathbf{y}|\hat{\boldsymbol{\beta}}_\psi(\mathbf{y};\psi_0) = \hat{\boldsymbol{\beta}}_\psi^0\}$, and the associated density values that approximate $c\frac{L_p(\psi;\mathbf{s}|\mathbf{a})}{L_p(\hat{\psi};\mathbf{s}|\mathbf{a})}$, where c is a normalizing constant, in a similar manner. The table below provides a partial list of the results of this numerical procedure. The density values have been normalized by dividing each by the sum of all un–normalized values calculated during the numerical procedure and so, for example, un–normalized $f_1^+ = 6.7 \times 10^{-4}$ becomes normalized $f_1^+ \approx 0.006$.

position	-3	-2	-1	0	1	2	3
y_1	1.898	1.599	1.300	1.000	0.700	0.400	0.100
y_2	-0.797	-0.199	0.400	1.000	1.600	2.200	2.799
density, $f = c\frac{L_p(\psi;\mathbf{s}\|\mathbf{a})}{L_p(\hat{\psi};\mathbf{s}\|\mathbf{a})}$	0.000	0.000	0.000	0.001	0.006	0.035	0.125

position	4	5	6	7	8	9	10
y_1	-0.199	-0.499	-0.799	-1.099	-1.428	-1.747	-2.057
y_2	3.398	3.998	4.598	5.198	5.857	6.493	7.113
density, $f = c\frac{L_p(\psi;\mathbf{s}\|\mathbf{a})}{L_p(\hat{\psi};\mathbf{s}\|\mathbf{a})}$	0.286	0.416	0.384	0.159	0.054	0.012	0.002

Ignoring the numerical error, the points of the test statistic, $\{\mathbf{y}|\hat{\boldsymbol{\beta}}_\psi(\mathbf{y};\psi_0) = \hat{\boldsymbol{\beta}}^0_\psi\}$, given in the table above could be plotted as a one–dimensional line, $y_1+\frac{1}{2}y_2+\frac{3}{2} = 0$, in the two–dimensional variable space. The associated density values that approximate $c\frac{L_p(\psi;\mathbf{s}|\mathbf{a})}{L_p(\hat{\psi};\mathbf{s}|\mathbf{a})}$ identify a bell–shaped curve on this test statistic. The p–value, calculated as the smallest of the two sums of approximate density values to either side of the observed value at position zero (0), is found to be 0.0003. In other words, there is evidence against the hypothesized $\psi = 3$ value at the observed value $\mathbf{y}_0 = (1, 1)$ for this generalized linear problem.

5.2 Other Examples

Normal With Three Variables. In this version of the normal example, assume there are three variables, $n = 3$ (instead of 2, as in the first version) and two parameters, $r = 2$. The underlying distribution, F_i, $i = 1, 2, 3$, is chosen to be the non-standard normal with interest distribution mean θ_i and nuisance distribution standard deviation, $\sigma_i = 1$, $i = 1, 2, 3$. The one nuisance test parameter is given by $\beta_1^\psi = \lambda$. The systematic component is given by,

$$\mathbf{X}_{3\times 2} = \begin{pmatrix} 1 & -3 \\ -2 & 1 \\ 4 & 2 \end{pmatrix}.$$

Let the inverse link function be the identity, where $\theta_i = \eta_i$, $i = 1, 2, 3$. The p–value is calculated to test the interest parameter, $\psi = 3$ at the observed point $\mathbf{y}_0 = (y_1^0, y_2^0) = (1, 1)$.

Ignoring the numerical error, the points of the test statistic, $\{\mathbf{y}|\hat{\boldsymbol{\beta}}_\psi(\mathbf{y};\psi_0) = \hat{\boldsymbol{\beta}}^0_\psi\}$, could be plotted as a one–dimensional line,

$$\left(\frac{43}{25}b - \frac{44}{25}\right)\mathbf{e}_1 + \left(-\frac{16}{25}b - \frac{28}{25}\right)\mathbf{e}_1 + b\mathbf{e}_3$$

where b is a parameterizing constant and $\mathbf{e}_i$ are basis vectors, in the three–dimensional variable space. Also ignoring numerical error, the associated density values that approximate $c\frac{L_p(\psi;\mathbf{s}|\mathbf{a})}{L_p(\hat{\psi};\mathbf{s}|\mathbf{a})}$ identify a bell–shaped curve on this test statistic. The p–value is 0.0000; thus, there is no evidence against the hypothesized $\psi = 2$ at the observed variable value $\mathbf{y} = (1, 1, 1)$.

Exponential With Exponential Link. Let $n = 2$ and $r = 2$. The underlying distribution is

$$F_i(y_i; \theta_i) = 1 - \exp\{-y_i\theta_i\}, \quad y_i > 0, \quad i = 1, 2.$$

The initial point on the test statistic curve is the observed variable value, assumed to be $\mathbf{y}_0 = (y_1^0, y_2^0) = (0.25, 0.25)$. The interest test parameter is assumed to be $\psi = 0.7$ and the one nuisance test parameter is given by $\beta_1^\psi = \lambda$. Also,

$$\mathbf{X}_{2\times 2} = \begin{pmatrix} -1 & -2 \\ -1 & 2 \end{pmatrix}.$$

The inverse link function is the exponential, where $\theta_i = e^{\eta_i}$, $i = 1, 2$.

The test statistic is a one–dimensional line and the associated density is near bell–shaped on this test statistic. The p–value is 0.003; thus, there is no evidence against the hypothesized $\psi = 0.7$ at the observed variable value $\mathbf{y} = (0.25, 0.25)$. In contrast to the two examples above, this example could not easily be solved in an analytical way.

Exponential With Three Variables and Exponential Link. Let $n = 2$ and $r = 2$. The underlying distribution is

$$F_i(y_i; \theta_i) = 1 - \exp\{-y_i\theta_i\}, \quad y_i > 0, \quad i = 1, 2.$$

The initial point on the test statistic curve is the observed variable value, assumed to be $\mathbf{y}_0 = (y_1^0, y_2^0, y_3^0) = (0.4, 0.5, 0.55)$. The interest test parameter is assumed to be $\psi = 0.1$ and the one nuisance test parameter is given by λ. Also,

$$\mathbf{X}_{3\times 2} = \begin{pmatrix} -0.1 & -0.2 \\ -0.3 & -0.5 \\ -0.2 & -0.2 \end{pmatrix}.$$

The inverse link function is the exponential, where $\theta_i = e^{\eta_i}$, $i = 1, 2, 3$.

The test statistic is a one–dimensional *non–linear* curve and the associated density is near bell–shaped on this test statistic. The p–value is 0.388; thus, there is no evidence against the hypothesized $\psi = 0.1$ at the observed variable value $\mathbf{y}_0 = (0.4, 0.5, 0.55)$.

Other Examples. Other examples have also been investigated, including for the Erlang and half normal distributions. These examples and the computer program used to calculate them can be found on the authors' web site.

6 Discussion

The procedure proposed in this paper is more or less strictly numerical in nature and cannot be used to solve generalized linear model problems, except those of a simple nature, in an analytical way (although it might be possible that clever use of analytic symbolic language software packages, such as Maple, might produce results in more advanced problems). Even using the numerical procedure given in this paper, only problems with a small number of variables and parameters can be dealt with successfully before succumbing to numerical issues of, for example, matrix singularity. The success of this numerical procedure necessarily relies on highly sophisticated advanced numerical routines; in particular, the numerical procedure used in this paper could be improved by using more advanced procedures, such as the EM method (Lindsey 1996), say, to determine maximum likelihood estimates.

Rather than calculate the approximate density values and then sum these values to obtain the p–value, as is done in this paper, it seems more sensible to obtain the p–value directly from an accurate approximation to the distribution. This way, there is less chance that the numerical errors will overwhelm the analytical accuracy of the method; some work on this for the generalized linear model has been done in Fraser and Reid 2001.

Both the test statistic, $\{\mathbf{y}|\hat{\boldsymbol{\beta}}_\psi(\mathbf{y};\psi_0) = \hat{\boldsymbol{\beta}}_\psi^0\}$ and the associated p^* approximation of the likelihood ratio of the profile likelihood for ψ,

$$\frac{L_p(\psi; \mathbf{s}|\mathbf{a})}{L_p(\hat{\psi}; \mathbf{s}|\mathbf{a})},$$

are difficult to calculate in both an analytical and numerical sense. The complexity of calculation might be reduced by creating a test statistic, such as a r^*–like statistic (Barndorff–Nielsen 1991) which approximates the well–known normal distribution, say, rather than some function of the complicated p^* formula, as is the case in this paper.

Fraser and Reid (1995, 2001) have studied general cases of the ancillary statistic used in this paper, defined by

$$\mathbf{Ady} = \mathbf{0}.$$

Most importantly, and as stated previously, they appear to have shown this statistic is ancillary in a local sense. Not only this, it also appears that it is possible to condition the variable associated with the observed constrained maximum likelihood estimate of the interest parameter in the presence of the nuisance parameters, $\{\mathbf{y}|\hat{\boldsymbol{\beta}}_\psi(\mathbf{y};\psi_0) = \hat{\boldsymbol{\beta}}_\psi^0\}$, on the local tangent ancillary statistic $\mathbf{a}(\mathbf{y}) = \mathbf{y}$ under very general conditions because the latter is defined to be tangent to the former and so the two will intersect with respect to one another. Nonetheless, is this ancillary statistic necessarily the right one? Are there are other ancillary statistics that can be integrated into the procedure given in this paper and how would they compare to the one chosen here?

The main strength of the numerical method given in this paper is that it provides a possible technique to calculating the p–value in a broad class of problems. Also, possibly more importantly, the differential–geometric nature of the analysis may provide a fruitful direction to pursue. In particular, it is of interest that the test statistic is, in general, a one–dimensional *curve* (rather than, necessarily, a line) in variable space and that this curvature may have some bearing on some properties of the problem, such as with respect to ancillarity, sufficiency or information, say. The differential–geometric analysis used in this analysis is possibly more sensitive or adaptable to local influences of the nuisance parameter than other global analyses in some sense.

It may not necessarily be the case that the local analysis given in this paper will produce results that necessarily match up with previous global results and so it somewhat surprising when they do. For example, the canonical exponential family distribution, where $(y_1, \ldots, y_{r-1}) = \mathbf{y}_{\boldsymbol{\lambda}}$, $y_r = y_\psi$ and $n = r$, given by,

$$f(\mathbf{y}; \boldsymbol{\beta}) = \exp\{\, \mathbf{y}_{\boldsymbol{\lambda}} \boldsymbol{\lambda} + y_\psi \psi - \kappa(\boldsymbol{\beta}) \,\}\, h(\mathbf{y}),$$

can be factored in the following way,

$$f(\mathbf{y}; \boldsymbol{\beta}) = g_1(\mathbf{y}; \boldsymbol{\beta}) \cdot g_2(y_\psi \,|\, \mathbf{y}_{\boldsymbol{\lambda}}; \psi).$$

This factorization suggests (Fraser and Reid 1988) if ψ were an interest parameter in the presence of nuisance parameters, $\boldsymbol{\lambda}$, then it would make sense to use conditional density $g_2(y_\psi \,|\, \mathbf{y}_{\boldsymbol{\lambda}}; \psi)$ instead of the original density $f(\mathbf{y}; \boldsymbol{\beta})$ for inference purposes on ψ since g_2 is free of the nuisance parameters whereas f is not. In particular, notice, the test statistic used in density g_2 is the line, $y_r = y_\psi$. An analysis using the local tangent ancillary statistic $\mathbf{Ady} = \mathbf{0}$ described in this paper gives the same result. Further study is required to more carefully compare this method to other more traditional global methods.

7 Acknowledgements

The author thanks D.A.S. Fraser for his help.

8 References

R. Abraham, J.E. Marsden, T. and Ratiu (1983). *Manifolds, Tensor Analysis, and Applications*, Springer–Verlag, Berlin.

S. Amari (1985). *Differential–Geometrical Methods in Statistics*, Springer–Verlag, Berlin.

O. Barndorff–Nielsen (1978). *Information and Exponential Families in Statistical Theory*, John Wiley, New York.

O. Barndorff–Nielsen (1983). On a formula for the distribution of the maximum likelihood estimator, *Biometrika.* 70(2), pp 343–365.

O. Barndorff–Nielsen (1991). Modified signed log likelihood ratio, *Biometrika.* 78(3), pp 557–563.

O. Barndorff–Nielsen (1999). L-nonformation, L-ancillarity and L-sufficiency. *Theory of Probability and Its Applications.* 44, pp 225-229.

O. Barndorff–Nielsen and D.R. Cox (1994). *Inference and Asymptotics.* Chapman and Hall, London.

D.R. Cox and D.V. Hinkley (1974). *Theoretical Statistics.* Chapman and Hall, London.

D.R. Cox and E. Snell (1981). *Applied Statistics, Principles and Examples.* Chapman and Hall, London.

D.A.S. Fraser (1964). Local conditional sufficiency. *The Journal of the Royal Statistical Society Series B*, 26(1), pp 52–62.

D.A.S. Fraser, P. McDunnough, and N. Reid (1987). Some aspects of conditioning. *Proceedings of the Symposium in Honor of M.M. Ali*, University of Western Ontario.

D.A.S. Fraser and N. Reid (1988). Fibre analysis and conditional inference. in Matusita, M., editor, *Statistical Theory and Data Analysis II*, Teikyo University, North–Holland.

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

D.A.S. Fraser and N. Reid (2001). Ancillary information for statistical inference. *Proceedings of a Symposium on Empirical Bayes and Likelihood Inference.* Springer–Verlag, New York, too appear.

J.R.D. Kuhn (1994). *Parameter Forcing.* Ph.D. Thesis, University of Toronto.

J.K. Lindsey (1996). *Parametric Statistical Inference.* Clarendon Press, Oxford.

P. McCullagh and J. Nelder (1989). *Generalized Linear Models.* Chapman and Hall, London.

N. Reid (2000). Likelihood. *Journal of the American Statistical Association*, 95, pp 1335–1340.

Y. Zhu and N. Reid (1994). Information, ancillarity ad sufficiency in the presence of nuisance parameters. *Canadian Journal of Statistics*, 22, pp 111–123.