Numerical Approximation of the P–Value

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

December 30, 1999

Key Words: interest and nuisance parameters, ancillarity, sufficiency, generalized linear model, likelihood

Abstract: Parameter forcing is the central idea behind a method to approximate the p-value for an interest parameter in a Generalized Linear Model [12]. The method involves a numerical procedure which constructs both a one-dimensional test statistic curve and a distribution along this curve.

1 Introduction

The likelihood function plays a large part in statistical inference [4],[11]. Consider the case when 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

$$\mathrm{L}(\boldsymbol{\beta}; \mathbf{y}) = \mathrm{L}_1(\boldsymbol{\lambda}, \mathbf{t}_1)\mathrm{L}_2(\psi; \mathbf{t}_2)$$

where $\mathbf{t}_1$ and $\mathbf{t}_2$ are observed values of the statistics $\mathbf{T}_1$ and $\mathbf{T}_2$, it seems reasonable to use $\mathrm{L}_2(\psi; \mathbf{t}_2)$ for inference about the interest parameter, ψ.

An appropriate factorization of a likelihood can often be achieved using some combination of sufficient and ancillary statistics. For example, it would be ideal to find a sufficient statistic, $\mathbf{T}$, for the nuisance parameters, $\boldsymbol{\lambda}$, such that

$$\mathrm{L}(\boldsymbol{\beta}; \mathbf{y}) = \mathrm{L}_1(\boldsymbol{\lambda}; \mathbf{t})\mathrm{L}_2(\psi; \mathbf{y}|\mathbf{t})$$

or an ancillary statistic, $\mathbf{T}$, for the nuisance parameters, $\boldsymbol{\lambda}$, such that

$$\mathrm{L}(\boldsymbol{\beta}; \mathbf{y}) = \mathrm{L}_1(\psi; \mathbf{t})\mathrm{L}_2(\boldsymbol{\lambda}; \mathbf{y}|\mathbf{t}).$$

Then, the conditional likelihood, $\mathrm{L}_2(\psi; \mathbf{y}|\mathbf{t})$, in the first case, or the marginal likelihood, $\mathrm{L}_1(\psi; \mathbf{t})$, in the second case, could be used for inference purposes [1],[7],[10],[13].

A one–dimensional statistic, which displays ancillary–like properties for the nuisance parameters, along with a possible approximate distribution for this statistic, is described in this paper. The main idea behind this statistic and its distribution is called the parameter forcing method and, accordingly, the statistic is called the parameter forcing statistic and the distribution along this curve is called the parameter forcing distribution. Although applicable to more general problems, the parameter forcing method is utilized, in this paper, for a generalized version of the Generalized Linear Model.

A numerical procedure is employed to carry out the parameter forcing method because the system of differential equations used to describe the parameter forcing statistic and the formula used to approximate the parameter forcing distribution are, in general, too complicated to solve by analytical means.

The parameter forcing statistic is analogous to the r^* statistic, [3], and the parameter forcing distribution is analogous to the p^* formula, [2].

This paper begins with a discussion of the system of differential equations that describe the parameter forcing statistic in the context of a Generalized Linear Model. The next section describes the parameter forcing distribution. The last section describes results obtained from the parameter forcing numerical procedure.

2 Parameter Forcing Statistic For Generalized Linear Model

After describing the version of the Generalized Linear Model used for this analysis, the system of differential equations that define the parameter forcing statistic is identified and explained.

Generalized Linear Model and Inference For Interest Parameter Consider a sequence of independent random variables $(y_1, \ldots, y_n)$,

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

where each distribution function F_i is continuous and either stochastically increasing or decreasing in the parameter θ_i. The known F_i distribution functions are said to be *underlying* or *process* distribution functions. The parameters $\theta_i \in \Re, i = 1, \ldots, n$ are unknown *interest distribution function* parameters. The parameters $\xi_i = (\xi_1^i, \ldots, \xi_p^i) \in \Re^p$, $i = 1, \ldots, n$ are unknown *nuisance distribution function* parameters.

The version of the Generalized Linear Model discussed in this paper consists of three components. The *random component* consists of a sequence of independent random variables $(y_1, \ldots, y_n)$,

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

where each distribution F_i is continuous and either stochastically increasing or decreasing 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, $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$. Finally, 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.

Parameter Forcing Statistic There are two pieces to the system of differential equations that describe the parameter forcing statistic. One is called the nuisance–free system and the other is called the ancillary–direction system. Although both possess ancillary–like properties, the latter system also possesses a "parameter–forcing" property.

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 *one–forms* (or 1–forms) of the variables; $\mathrm{d}\theta_i$, $i = 1,\ldots,n$ are one–forms for the parameters. A (column) n–vector of variable *one-forms* is given by $\mathbf{dy} = (\mathrm{d}y_1,\ldots,\mathrm{d}y_n)^t$. A (column) n–vector of parameter one-forms is given by $\mathbf{d}\beta = (\mathrm{d}\beta_1,\ldots,\mathrm{d}\beta_n)^t$. The one–form of a variable, say, represents, roughly speaking, the incremental change in the variable.

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

Since the systematic component of the Generalized Linear Model is given by $\eta = \mathbf{X}\boldsymbol{\beta}$, with inverse link functions, $\theta_i = g(\eta_i)$, $\quad i = 1,\ldots,n$, the following notation, with regard to constrained 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,

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

Of course, 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, $\hat{\boldsymbol{\beta}}_\psi$.

Nuisance–Free System of Equations The nuisance–free system of equations, [8],[10], are given by,

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

where,

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

and,

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

Local Ancillary–Like Property of Nuisance–Free System. Details of the derivation of $\mathbf{Ady} = \mathbf{0}$ can be found in Kuhn, [10]. However, the essential idea behind this system is that the *score* of $\tilde{\ell} = \sum_{i=1}^{n} \ell_i$ 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 this score is *zero*, locally, at the constrained maximum likelihood estimator value, there is *no* change in the log–likelihood, $\ell(\hat{\theta}_\psi; \mathbf{y})$, for a change in the nuisance parameters, $\boldsymbol{\lambda}$ at this value. In other words, the log–likelihood is locally constant with respect to the $(r-1)$ nuisance parameters at the constrained maximum likelihood estimator value. It is for this reason, this system describes a statistic that has ancillary–like properties and, consequently, is called the local nuisance–free system of equations.

Example of Use of Nuisance–Free System of Equations: Canonical Exponential Distribution 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, [8], 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 statistic used in g_2 is the line, $y_r = y_\psi$.

This same result can be arrived at, using the local nuisance–free system of equations. Letting $\theta_i = \eta_i = \beta_i$, $i = 1, \ldots, r$, where, as above, $r = n$, and so,

$$\ell(\theta_i; y_i) = \mathbf{y}_{\boldsymbol{\lambda}}\boldsymbol{\lambda} + y_\psi \psi - \kappa(\beta) + \ln h(\mathbf{y}),$$

and

$$\ell_{\theta_i;}(\theta_i; y_i) = y_i - \frac{\partial \kappa(\beta)}{\partial \beta_i}, \quad \ell_{\theta_i; y_i}(\theta_i; y_i) = 1,$$

and $\frac{\partial g(\eta_i)}{\partial \eta_i} = 1$, for $i = 1, \ldots, r$. Thus,

$$\begin{aligned} \mathbf{A} &= \mathbf{X}^t_{r\times(r-1)}\mathbf{D}_\psi \mathbf{E}_\psi \\ &= \begin{pmatrix} 1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & \ddots & \vdots & 0 \\ \vdots & \ddots & \ddots & 0 & 0 \\ 0 & \ldots & 0 & 1 & 0 \end{pmatrix}, \end{aligned}$$

or, $\mathrm{d}y_i = 0$, which implies $y_i = c$, $i = 1, \ldots, r-1$, where c is a constant. In other words, the $\mathbf{A}\mathbf{d}\mathbf{y} = \mathbf{0}$ system of equations, here, suggests, as above, the line $y_r = y_\psi$ as the nuisance–free curve on which to define a distribution. Since the nuisance–free system produces the usual statistic in this case, this, thus, provides some justification for the use of the nuisance–free system of equations.

Since the number of variables and parameters which appear in the canonical exponential density are equal, $n = r$, the nuisance–free system, in this case, will give the *same* curve as the parameter forcing system of equations that include both the nuisance–free and the yet–to–be–explained ancillary directions system of equations. In general, geometrically, the nuisance–free system of equations reduces the original n–dimensional sample space to an $n - r + 1$ dimensional statistic.

Ancillary Directions System of Equations Under various technical conditions, described in Kuhn [10], and, in particular, the ancillary directions assumption, [6], given by,

$\mathrm{d}F_i = 0$, $i = 1, \ldots, n$, the ancillary directions system of equations can be written as,

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

where,

$$\mathbf{V} = -\mathbf{FEX}_{n\times r},$$

where,

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

and $\mathbf{d}\boldsymbol{\beta} = (\mathrm{d}\beta_1, \ldots, \mathrm{d}\beta_n)^t$ and $\mathbf{dy} = (\mathrm{d}y_1, \ldots, \mathrm{d}y_n)^t$.

Parameter Forcing Property of Ancillary Direction System Details of the derivation of $\mathbf{dy} = \mathbf{V}\mathbf{d}\boldsymbol{\beta}$ can be found in Kuhn, [10]. However, the essential idea behind this system lies with the ancillary directions assumption; that is,

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

This assumption restricts the "movement" of the parameters, θ_i, and variables, 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). Hence, the term *parameter forcing.* More than this, Fraser and Reid [9], suggest this system of equations is, at least, locally and approximately, ancillary.

Example of Use of Ancillary Directions System of Equations If

$$f(y_i; \theta_i) = \exp\{\, -y_i\theta_i + \ln(\theta_i)\,\}, \quad i = 1, \ldots, n,$$

for $y_i > 0$ and where, $\theta_i = g(\eta_i) = \exp(\eta_i)$, $i = 1, \ldots, n$, and where,

$$\begin{pmatrix} \eta_1 \\ \vdots \\ \eta_n \end{pmatrix} = -\begin{pmatrix} 1 & x_1 - \bar{x} \\ \vdots & \vdots \\ 1 & x_n - \bar{x} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix};$$

then

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

and

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

and so

$$\begin{aligned}\mathbf{V} &= -\mathbf{FEX}_{n\times r}\\ &= -\begin{pmatrix} y_1x_{11} & y_1x_{12}\\ \vdots & \vdots\\ y_nx_{n1} & y_nx_{n2}\end{pmatrix}\\ &= \begin{pmatrix} v_{11} & v_{12}\\ \vdots & \vdots\\ v_{n1} & v_{n2}\end{pmatrix} = \begin{pmatrix} \mathbf{v}_1 & \mathbf{v}_2 \end{pmatrix}.\end{aligned}$$

The tangent subspace for this example is a two–dimensional plane defined by the two n–dimensional column vectors, $\mathbf{v}_1$ and $\mathbf{v}_2$; the possible ancillary directions are defined in this plane.

In this exponential case, the explicit analytic calculation of the ancillary directions matrix $\mathbf{V}$ from $\mathbf{V}\mathrm{d}\boldsymbol{\beta} = \mathrm{d}\mathbf{y}$, is relatively easy to do. However, the determination of $F_{y_i;}$, $F_{;\theta_i}$ and the various maximum likelihood estimators in $\mathbf{V}$ are generally quite tricky to do. It is because of these tedious analytical calculations, that a computer program has been developed to handle them numerically.

The practical purpose of the ancillary directions system is to reduce the $(n-r+1)$–dimensional variable space, defined by the nuisance–free system of equations, to a 1–dimensional subspace of this space. This one–dimensional curve is called the parameter forcing statistic.

Parameter Forcing System of Equations The parameter forcing system of equations is the *intersection* of the ancillary directions system and the local nuisance–free system, given by,

$$\begin{aligned}&\begin{pmatrix} \mathbf{A}_{(r-1)\times n}(\hat{\theta}_\psi) & \mathbf{0}_{(r-1)\times r}\\ -\mathbf{I}_{n\times n} & \mathbf{V}_{n\times r}(\hat{\theta})\end{pmatrix}\begin{pmatrix} \mathrm{d}\mathbf{y}\\ \mathrm{d}\beta\end{pmatrix}\\ =\ & \mathbf{\Phi}_{(n+r-1)\times(n+r)}(\mathbf{u})\mathbf{du} = \begin{pmatrix} 0\\ \vdots\\ 0\end{pmatrix},\end{aligned}$$

where the $\mathbf{\Phi}$ matrix is of dimension $(n+r-1)\times(n+r)$, $\mathbf{A}_{(r-1)\times n}(\hat{\theta}_\psi) = \mathbf{A} = \mathbf{X}^t_{n\times(r-1)}\mathbf{D}_\psi\mathbf{E}_\psi$ is the local nuisance–free matrix and $\mathbf{V}_{n\times r}(\hat{\theta}) = \mathbf{V} = -\mathbf{FEX}_{n\times r}$, is the ancillary directions matrix.

Example of Use of Parameter Forcing System of Equations: Normal Consider the problem where,

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

where, $\sigma_1 = \sigma_2 = \sigma_3 = 1$ and $\theta_i = \eta_i$, $i = 1,3$, 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 point $\mathbf{y} = (1,1,1)$ and for hypothesized interest parameter value, $\psi = 2$, it can be shown the parameter forcing system of equations is given by,

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

$$= \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

After some algebraic manipulation, it can be shown the nuisance–free part of the parameter forcing system, $\mathbf{Ady} = \mathbf{0}$, can be described either as,

$$\begin{aligned} & y_1 - 2y_2 - 3y_3 + (-14\hat{\lambda}_{\psi_0} - \psi_0) \\ = \; & y_1 - 2y_2 - 3y_3 - 14(-0.429) - 2 \\ = \; & y_1 - 2y_2 - 3y_3 + 4 = 0, \end{aligned}$$

or, equivalently, for $\mathbf{e}_1 = (1,0,0)^t$, $\mathbf{e}_2 = (0,1,0)^t$, and $\mathbf{e}_3 = (0,0,1)^t$, as,

$$b_1\mathbf{e}_1 + b_2\mathbf{e}_2 + \frac{1}{3}(b_1 - 2b_2 + 4)\mathbf{e}_3.$$

The ancillary directions system, $\mathbf{Vdy} = \mathbf{d}\boldsymbol{\beta}$, can be immediately described either as,

$$(b_1 - 3b_2)\mathbf{e}_1 + (-2b_1 + b_2)\mathbf{e}_2 + (-3b_1 + 2b_2)\mathbf{e}_3,$$

or, equivalently,

$$-\frac{7}{5}y_1 - \frac{11}{5}y_2 + y_3 = 0,$$

where the linear functional part of this last equation annihilates the two basis vectors of the ancillary directions system.

The two ancillary direction basis vectors, $\mathbf{v}_1 = (1,-2,-3)^t$ and $\mathbf{v}_2 = (-3,1,2)^t$, span a two–dimensional subspace (a plane) of the three–dimensional variable space. The nuisance–free system also defines a plane in the variable space. The intersection of these two planes is the parameter forcing path, a line, which can be described as,

$$\left(\frac{43}{25}b - \frac{44}{25}\right)\mathbf{e}_1 + \left(-\frac{16}{25}b - \frac{28}{25}\right)\mathbf{e}_2 + b\mathbf{e}_3.$$

In this case, since the number of variables, n, is greater than the number of parameters, r, $n = 3 > r = 2$, the two–dimensional variable subspace defined by the ancillary directions system, $\mathbf{Vdy} = \mathbf{d}\boldsymbol{\beta}$, restricts the two–dimensional nuisance–free system of equations to the one–dimensional parameter forcing path. The ancillary directions system acts to reduce the size of the sample space to the number of parameters so that the local nuisance–free system can then be used to define the parameter forcing path.

The parameter forcing system defines a 1–dimensional tangent *line* (local point of view) at a point on the 1–dimensional parameter forcing system *curve or path* manifold (global point of view). Under mild (that is, easy–to–satisfy) technical conditions, the tangent line can be integrated up to define the parameter forcing system path.

The 1–dimensional parameter forcing system curve could be thought of as the result of an *intersection* the r–dimensional ancillary directions system and the $(r-1)$–dimensional local nuisance–free system. An equivalent global geometric interpretation is that the parameter path results from a two–step reduction in the dimension of the variable space. The variable space is first reduced to the r–dimensional ancillary directions space and then, *within* this space, a second reduction of $(r-1)$–dimensions, given by the nuisance–free system, leads to the 1–dimensional parameter forcing path. From this second point of view, the parameter forcing system of equations could be interpreted as a 1–dimensional path within or *conditional* with respect to the ancillary directions space.

3 Parameter Forcing Distribution

The parameter forcing distribution is given by,

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

where,

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

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

$$\begin{aligned}
\mathbf{C} &= \text{diag}_n\left\{\frac{\partial g(\eta_i)}{\partial\eta_i\partial\eta_i}|_{\hat{\eta}_i}\right\},\\
\mathbf{C}_\psi &= \text{diag}_n\left\{\frac{\partial g(\eta_i)}{\partial\eta_i\partial\eta_i}|_{\hat{\eta}_i^\psi}\right\},\\
\mathbf{D}_\psi &= \text{diag}_n\left\{\ell^i_{\theta_i;y_i}|_{\hat{\theta}^i_\psi}\right\},\\
\mathbf{L} &= \text{diag}_n\left\{\ell^i_{\theta_i;}(\hat{\theta};\mathbf{y})\right\},\\
\mathbf{L}_\psi &= \text{diag}_n\left\{\ell^i_{\theta_i;}(\hat{\theta}_\psi;\mathbf{y})\right\},\\
\mathbf{J} &= \text{diag}_n\left\{\ell^i_{\theta_i\theta_i;}(\hat{\theta};\mathbf{y})\right\},\\
\mathbf{J}_\psi &= \text{diag}_n\left\{\ell^i_{\theta_i\theta_i;}(\hat{\theta}_\psi;\mathbf{y})\right\}.
\end{aligned}$$

The derivation of this parameter forcing distribution uses quotient differentials; and further details are given in Kuhn [5].

Example of Calculation of Parameter Forcing Distribution: Exponential Revisited The parameter forcing conditional density approximation is calculated at the first two points, under the following conditions. The density is,

$$f_i=\theta_i e^{-y_i\theta_i},\quad y_i>0,\quad i=1,2.$$

The initial point on the parameter forcing path is the observed variable value, assumed to be $\mathbf{y}_0=(y^0_1,y^0_2)=(0.25,0.25)$. The interest test parameter is $\beta^\psi_2=\psi=0.7$ and the one nuisance test parameter is given by $\beta^\psi_1=\lambda_1$. Also,

$$\mathbf{X}_{n\times r}=\mathbf{X}_{2\times 2}=\mathbf{X}=\begin{pmatrix} x_{11} & x_{12}\\ x_{21} & x_{22}\end{pmatrix}=\begin{pmatrix} -1 & -2\\ -1 & 2\end{pmatrix},$$

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

At the initial point, $\mathbf{y}_0 = (0.25, 0.25)$,

$$\begin{aligned}
\tilde{\ell}(\hat{\theta};\mathbf{y}) &= 0.773,\\
\tilde{\ell}(\hat{\theta}_\psi;\mathbf{y}) &= -0.759,\\
|\tilde{\ell}_{\beta;\mathbf{V}}(\hat{\theta};\mathbf{y})| &= 16,\\
|\tilde{\ell}_{\boldsymbol{\lambda};\mathbf{V}}(\hat{\theta}_\psi;\mathbf{y})| &= 16.553,\\
|\hat{\jmath}|^{1/2} &= 4,\\
|\hat{\jmath}_{\boldsymbol{\lambda}}|^{1/2} &= 1.414,
\end{aligned}$$

and so, if c = 1, the parameter forcing distribution evaluated at this point is $f = 0.030$.

Using the parameter forcing system of equations to get to the second point, $(y_1, y_2) = (0.309, 0.246)$,

$$\begin{aligned}
\tilde{\ell}(\hat{\theta};\mathbf{y}) &= 0.576,\\
\tilde{\ell}(\hat{\theta}_\psi;\mathbf{y}) &= -0.756,\\
|\tilde{\ell}_{\beta;\mathbf{V}}(\hat{\theta};\mathbf{y})| &= 16,\\
|\tilde{\ell}_{\boldsymbol{\lambda};\mathbf{V}}(\hat{\theta}_\psi;\mathbf{y})| &= 15.810,\\
|\hat{\jmath}|^{1/2} &= 4,\\
|\hat{\jmath}_{\boldsymbol{\lambda}}|^{1/2} &= 1.715.
\end{aligned}$$

Consequently, if c = 1, $f = 0.046$.

This procedure is repeated until the parameter forcing path and corresponding parameter forcing conditional density approximations at each point along the path is complete. These un–normalized values are then normalized by dividing each by their sum. A p–value, for example, is determined by adding all the normalized values at or more extreme than some given observed point on the parameter forcing statistic.

4 Example of Numerical Calculation of Parameter Forcing Statistic and Distribution: Normal With Three Variables

In this version of the normal example, $n = 3$ and $r = 2$. The underlying density, 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 initial point on the parameter forcing path is the observed variable value, assumed to be $\mathbf{y}_0 = (y_1^0, y_2^0, y_3^0) = (1, 1, 1)$. The interest test parameter is assumed to be $\beta_2^\psi = \psi = 2$ and the one nuisance test parameter is given by $\beta_1^\psi = \lambda_1$. The systematic component is given by,

$$\mathbf{X}_{n\times r} = \mathbf{X}_{3\times 2} = \mathbf{X} = \begin{pmatrix} 1 & -3 \\ -2 & 1 \\ -3 & -2 \end{pmatrix},$$

and so, $\mathbf{X}_{n\times(r-1)} = (x_{11}, x_{21}, x_{31})^t = (1, -2, -3)^t$. Let the inverse link function be the identity, where $\theta_i = \eta_i$, $i = 1, 2, 3$.

The parameter forcing path is a one–dimensional line in a three dimensional variable space. This linear parameter forcing path can be analytically (exactly) determined as,

$$\left(\frac{43}{25}b - \frac{44}{25}\right)\mathbf{e}_1 + \left(-\frac{16}{25}b - \frac{28}{25}\right)\mathbf{e}_2 + b\mathbf{e}_3,$$

or,

$$(1.72b - 1.76)\mathbf{e}_1 + (-0.64b + 1.12)\mathbf{e}_2 + b\mathbf{e}_3,$$

where $\mathbf{e}_1 = (1,0,0)^t$, $\mathbf{e}_2 = (0,1,0)^t$ and $\mathbf{e}_2 = (0,0,1)^t$. By way of comparison, the numerical parameter forcing path finds, using the two calculated variable points, $\mathbf{y}_0 = (1,1,1)^t$ and $\mathbf{y}_3 = (-1.300, 1.859, -0.333)^t$,

$$(2.300b + 1)\mathbf{e}_1 + (-0.859b + 1)\mathbf{e}_2 + (1.333b + 1)\mathbf{e}_3,$$

which is equivalent to

$$(1.720b + 1)\mathbf{e}_1 + (-0.642b + 1)\mathbf{e}_2 + (0.997b + 1)\mathbf{e}_3.$$

In other words, the numerically determined parameter forcing statistic is close to the analytically determined parameter forcing statistic.

The p–value, determined by adding all the normalized parameter forcing distribution values at or more extreme than the observed variable value, $\mathbf{y}_0 = (1,1,1)$ on the parameter forcing statistic, is found to be 0.0000. Thus, there is evidence against the hypothesized $\beta_3 = 2$ at the observed variable value $\mathbf{y} = (1,1,1)$.

5 Summary, Conclusions and Future Work

The parameter forcing method is used to approximate the p–value for an interest parameter in the Generalized Linear Model. More specifically, this paper has described a one–dimensional statistic, called a parameter forcing statistic, which displays ancillary–like properties for the nuisance parameters, along with a possible approximate distribution for this statistic, called the parameter forcing distribution.

Not only do various properties (ancillarity, in particular) of this method need to be more carefully pursued but also the present numerical procedure needs to be improved to be able to deal with problems with larger dimensional parameter and sample spaces.

Nonetheless, the parameter forcing method appears to apply to a broad class of models, including the generalized Generalized Linear Models described above. More than this, the procedure is *numerical*, rather than *analytical*, and so could be used by those less familiar with statistical techniques.

6 Bibliography

[1] Barndorff–Nielsen, O. *Information and Exponential Families in Statistical Theory*, John Wiley, New York, 1978.

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

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

[4] Cox, D.R. and Hinkley, D.V. *Theoretical Statistics*, Chapman and Hall, London, 1974.

[5] Cox, D.R. and Snell, E. *Applied Statistics, Principles and Examples*, Chapman and Hall, London, 1981.

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

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

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

[9] Fraser, D.A.S. and Reid, N., Ancillaries and third order significance, *Technical Report 9303*, University of Toronto, Department of Statistics, April 1993.

[10] Kuhn, J.R.D., *Parameter Forcing*, Ph.D. Thesis, University of Toronto, 1994.

[11] Lindsey, J.K., *Parametric Statistical Inference*, Clarendon Press, Oxford, 1996.

[12] McCullagh, P. and Nelder, J. *Generalized Linear Models*, Chapman and Hall, London, 1989.

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