Default prior distributions from quasi- and quasi-profile likelihoods

Laura Ventura  
Department of Statistical Sciences  
University of Padua  
Italy

Stefano Cabras, Walter Racugno  
Department of Mathematics and Informatics  
University of Cagliari  
Italy

Abstract: In some problems of practical interest, a standard Bayesian analysis can be difficult to perform. This is true, for example, when the class of sampling parametric models is unknown or if robustness with respect to data or to model misspecifications is required. These situations can be usefully handled by using a posterior distribution for the parameter of interest which is based on a pseudo-likelihood function derived from estimating equations, i.e. on a quasi-likelihood, and on a suitable prior distribution.

The aim of this paper is to propose and discuss the construction of a default prior distribution for a scalar parameter of interest to be used together with a quasi-likelihood function. We show that the proposed default prior can be interpreted as a Jeffreys-type prior, since it is proportional to the square-root of the expected information derived from the quasi-likelihood. The frequentist coverage of the credible regions, based on the proposed procedure, is studied through Monte Carlo simulations in the context of robustness theory and of generalized linear models with overdispersion.

Keywords: Estimating equation, Jeffreys’s prior, Matching prior, Nuisance parameter, Pseudo-likelihood, Robustness
Default prior distributions from quasi- and quasi-profile likelihoods

Contents

1 Introduction 1

2 Estimating equations and quasi-likelihoods 3

3 A default prior from quasi-likelihoods 4

4 Examples and simulation results 6
   4.1 Robust inference for scale-location model ...................... 6
   4.2 Overdispersion in count data ................................... 7

5 Final remarks 8

Department of Statistical Sciences
Via Cesare Battisti, 241
35121 Padova
Italy

tel: +39 049 8274168
fax: +39 049 8274170
http://www.stat.unipd.it

Corresponding author:
Laura Ventura
tel: +39 049 8274177
ventura@stat.unipd.it
http://homes.stat.unipd.it/
ventura/
Default prior distributions from quasi- and quasi-profile likelihoods

Laura Ventura
Department of Statistical Sciences
University of Padua
Italy

Stefano Cabras, Walter Racugno
Department of Mathematics and Informatics
University of Cagliari
Italy

Abstract: In some problems of practical interest, a standard Bayesian analysis can be difficult to perform. This is true, for example, when the class of sampling parametric models is unknown or if robustness with respect to data or to model misspecifications is required. These situations can be usefully handled by using a posterior distribution for the parameter of interest which is based on a pseudo-likelihood function derived from estimating equations, i.e. on a quasi-likelihood, and on a suitable prior distribution.

The aim of this paper is to propose and discuss the construction of a default prior distribution for a scalar parameter of interest to be used together with a quasi-likelihood function. We show that the proposed default prior can be interpreted as a Jeffreys-type prior, since it is proportional to the square-root of the expected information derived from the quasi-likelihood. The frequentist coverage of the credible regions, based on the proposed procedure, is studied through Monte Carlo simulations in the context of robustness theory and of generalized linear models with overdispersion.

Keywords: Estimating equation, Jeffrey’s prior, Matching prior, Nuisance parameter, Pseudo-likelihood, Robustness

1 Introduction

The theory and the use of estimating equations and of the related quasi- and quasi-profile likelihood functions has received much attention in recent years; see, among others, Liang and Zeger (1995), Barndorff-Nielsen (1995), Desmond (1997), Heyde (1997), Adimari and Ventura (2002), Severini (2002), Wang and Hanfelt (2003), Jorgensen and Knudsen (2004), and Bellio et al. (2008). In addition, Lin (2006) and Greco et al. (2008) discussed the use of quasi-likelihood functions in the Bayesian setting. In this paper we derive a default prior distribution for a parameter of interest to be used together with a quasi- or a quasi-profile likelihood function derived from estimating functions. The corresponding posterior distribution allows to deal with
those problems in which standard Bayesian analyses are difficult to perform or not available.

Let \( y = (y_1, \ldots, y_n) \) be a random sample of size \( n \) drawn from \( F_\theta = F(y; \theta) \), \( \theta \in \Theta \subseteq \mathbb{R}^p, p \geq 1 \). Bayesian inference on \( \theta \), or a scalar component of \( \theta \), is usually based on the likelihood function \( L(\theta) = L(\theta; y) \) and on a suitable prior \( \pi(\theta) \) on \( \theta \), which is often derived through default procedures, in order to avoid the difficulties related to the elicitation. However, complementary to likelihood-based procedures, in many situations of practical interest it is preferable to base inference on estimating equations and on the related quasi- and quasi-profile likelihood functions. This is true, for example, in the context of robustness theory when stability with respect to small deviations from the assumed model is required (see Hampel et al., 1986), or in the context of generalized linear models with overdispersion or random effects (see McCullagh and Nelder, 1989). For inference about the parameter of interest, a quasi-likelihood function \( L_Q(\theta) \) can be defined, with the standard limiting behaviour (McCullagh, 1991, Adimari and Ventura, 2002). In a Bayesian approach, if one treats \( L_Q(\theta) \) as a true likelihood then a posterior distribution for \( \theta \) can be considered, given by

\[
\pi_Q(\theta|y) \propto \pi(\theta)L_Q(\theta).
\] (1)

Some properties of \( \pi_Q(\theta|y) \) follow from the used \( L_Q(\theta) \). For instance, if \( L_Q(\theta) \) is derived from a robust estimating equation, then the posterior distribution (1) provides inferential procedures which are reliable when the underlying distribution lies in a neighborhood of \( F_\theta \) (Greco et al., 2008).

Although the use of a pseudo-likelihood function, in the Bayesian inference cannot be considered as orthodox, it is actually widely shared. Papers which discuss the use of pseudo-likelihoods in the Bayesian perspective are Efron (1993), Bertolino and Racugno (1992, 1994), Raftery et al. (1996), Fraser et al. (2003), Cabras et al. (2006), Lin (2006), Chang and Mukerjee (2006), Greco et al. (2008) and Pauli et al. (2008). Papers which are more specifically related to the validation of a posterior distribution based on an alternative likelihood are Monahan and Boos (1992), Severini (1999), Lazar (2003), Schennach (2005), Racugno et al. (2008) and Ventura et al. (2008).

In this paper, we show that the proposed default priors to be used in (1) assume an expression analogue to the Jeffreys priors, since \( \pi(\theta) \) turns to be proportional to the square root of the expected quasi-information. This information is related to the asymptotic variance of the quasi-maximum likelihood estimator. Such result agrees with the Welch and Peers (1963) solution in a model with a scalar parameter and with the Ventura et al. (2008) solution for a scalar parameter of interest in the presence of nuisance parameters. The proposed procedure is illustrated in the context of robust scale-location models and of log-linear models for count data with overdispersion. Monte Carlo studies are also provided in order to investigate the frequentist coverage of the credible regions based on \( \pi_Q(\theta|y) \).

The remainder of the paper is organized as follows. Section 2 briefly reviews the background theory on estimating equations and on the related quasi- and quasi-profile likelihood functions. In Section 3 the proposed default priors are discussed.
Section 2  Estimating equations and quasi-likelihoods

2 Estimating equations and quasi-likelihoods

Let $\Psi_\theta = \Psi(y; \theta) = \sum_{i=1}^n \psi(y_i; \theta)$ be an unbiased estimating function for $\theta$ based on $y$, i.e. such that $E_\theta(\Psi_\theta) = 0$, where $\psi(\cdot)$ is given function and $E_\theta(\cdot)$ denotes expectation under $F_\theta$. A general $M$-estimator of $\theta$ is defined as the solution $\hat{\theta}$ of the estimating equation $\Psi_\theta = 0$ (see, e.g., Hampel et al., 1986). The class of $M$-estimators includes well-known estimators: for example, if $\Psi$ is the likelihood score function, then $\hat{\theta}$ is the maximum likelihood estimator. Under regularity conditions assumed throughout this paper (see e.g. Huber, 1981), $M$-estimators are consistent and asymptotically normally distributed with mean $\theta$ and covariance matrix

$$V(\theta) = M^{-1} \Omega(M^{-1})^T,$$

(2)

where $M = M(\theta) = -E_\theta(\partial \Psi_\theta / \partial \theta^T)$ and $\Omega = \Omega(\theta) = E_\theta(\Psi_\theta \Psi_\theta^T)$. Moreover, if $\psi(y; \theta)$ is bounded, the $M$-estimator $\hat{\theta}$ is $B$-robust (see, e.g., Hampel et al., 1986), i.e. it has bounded influence function.

A quasi-likelihood for $\theta$ based on $\Psi_\theta$ (see e.g. McCullagh, 1991) is given by

$$L_Q(\theta) = \exp \left\{ \sum_{i=1}^n \int_{c_0}^{\theta} A(t) \psi(y_i; t) \, dt \right\},$$

(3)

where $A(\theta)^T = \Omega^{-1} M$ and $c_0$ is an arbitrary constant. When $p = 1$, a quasi-likelihood for $\theta$ is usually easy to derive. However, when $p > 1$, $L_Q(\theta)$ exists if and only if the matrix $M$ is symmetric. Matrix $A(\theta)$ in (3) allows to obtain a quasi-observed information with the usual relation with $V(\theta)$ and a quasi-likelihood ratio statistic $W_Q(\theta) = 2\{\ell_Q(\hat{\theta}) - \ell_Q(\theta)\}$ with a standard $\chi^2_p$ distribution, where $\ell_Q(\theta) = \log L_Q(\theta)$. Moreover, $A(\theta)$ does not modify the robustness properties of $\hat{\theta}$ because it does not change its influence function. According to the results obtained in Heritier and Ronchetti (1994), such robustness properties will carry over to quasi-likelihood based inferential procedures.

Assume now $\theta$ partitioned as $\theta = (\tau, \lambda)$, where $\tau$ is a scalar parameter of interest and $\lambda$ a $(p-1)$-dimensional nuisance parameter. Similarly, $\Psi_\theta = (\Psi_\tau, \Psi_\lambda)$, where $\Psi_\tau = \sum_{i=1}^n \psi_\tau(y_i; \tau, \lambda)$ and $\Psi_\lambda = \sum_{i=1}^n \psi_\lambda(y_i; \tau, \lambda)$ are the estimating functions corresponding to $\tau$ and $\lambda$, respectively. Let $\lambda_\tau$ be the $M$-estimate for $\lambda$ derived from $\Psi_\lambda$ for fixed $\tau$, i.e. from $\Psi_\lambda = 0$. A quasi-profile likelihood for $\tau$ (Adimari and Ventura, 2002) is given by

$$L_{Q\Phi}(\tau) = \exp \left\{ \sum_{i=1}^n \int_{c_0}^{\tau} w(t, \lambda_\tau) \psi_\tau(y_i; t, \lambda_\tau) \, dt \right\},$$

(4)

where

$$w(\tau, \lambda) = \frac{-\nu_{\tau\tau} + \nu_{\tau a} \nu_{\tau \lambda} \nu_{\lambda \tau}}{E_\theta(\Psi_\tau^2) + 2\nu_{\tau a} \nu_{\tau b} E_\theta(\Psi_\tau \Psi_b) + \nu_{\tau a} \nu_{\tau b} \nu_{\tau \lambda} \nu_{\lambda \tau} \nu_{\lambda \tau} \nu_{\lambda \lambda} E_\theta(\Psi_\tau \Psi_\lambda \Psi_b \Psi_d)}.$$

(5)

Section 4 illustrates two examples and provides Monte Carlo studies. Section 5 is devoted to some final remarks.
In (5) index notation has been used: components of $\lambda$ and $\Psi_\lambda$ are denoted by $\lambda^a$ and $\Psi_a$, respectively, and derivatives of $\Psi_\tau$ and $\Psi_a$ with respect to the components of $\lambda$ are $\Psi_{\tau a} = (\partial/\partial \lambda^a)\Psi_\tau$ and $\Psi_{ab} = (\partial/\partial \lambda^b)\Psi_a$, where $a, b = 1, \ldots, p - 1$. Moreover, $\nu_{\tau \tau} = E_\theta(\partial^2/\partial \tau^2)\Psi_\tau$, $\nu_{\tau a} = E_\theta(\Psi_{\tau a})$, $\nu_{ab} = E_\theta(\Psi_{ab})$, and $\kappa^{ab}$ is the inverse matrix of $-\nu_{ab}$. The quasi-profile likelihood (4) has all the desired standard first-order properties. In particular, the quasi-maximum likelihood estimator $\tilde{\tau}$ is consistent and asymptotically normal, with mean $\tau$ and variance

$$V_{\tilde{\tau}}(\tau) = (M^{\tau \tau})^2 \Omega_{\tau \tau} + 2M^{\tau \lambda} \Omega_{\lambda \tau} (M^{\tau \tau})^T + M^{\tau \lambda} \Omega_{\lambda \lambda} (M^{\tau \lambda})^T \bigg|_{\lambda = \tilde{\lambda}},$$

(6)

where $M^{\tau \tau}$ and $M^{\tau \lambda}$ denote the $(\tau, \tau)$-element and the $(\tau, \lambda)$-block of the inverse of $M(\theta)$, respectively, and $\Omega_{\tau \tau}, \Omega_{\lambda \tau}$ and $\Omega_{\lambda \lambda}$ are the blocks of $\Omega$. When the dimension of $\lambda$ is large relative to $n$ (see e.g. Di Ciccio et al., 1996, Liang and Zeger, 1995, Severini, 2002), it may be preferable to consider a modification of $L_{QP}(\tau)$ (see Bellio et al., 2008), which is given by

$$L_{QP M}(\tau) = L_{QP}(\tau) \exp \left\{-\int_{c_0}^{\tau} w(t, \tilde{\lambda}_t) m(t, \tilde{\lambda}_t) dt \right\},$$

(7)

where $m(\tau, \lambda)$ is of order $O(1)$ and is given by

$$m(\tau, \lambda) = \kappa^{ba} E_\theta(\Psi_b \Psi_{\tau a}) + \nu_{\tau a} \kappa^{ca} \nu_{\tau b} E_\theta(\Psi_d \Psi_{cb}) + \frac{1}{2} \nu_{\tau a} \nu_{\tau b} \kappa^{ca} \kappa^{db} E_\theta(\Psi_e \Psi_f) + \frac{1}{2} \nu_{\tau a} \nu_{\tau b} \kappa^{ca} \kappa^{db} E_\theta(\Psi_d \Psi_e).$$

(8)

Note that the first-order bias correction (8) involves only the first two derivatives with respect to $\lambda$ of $(\Psi_\tau, \Psi_\lambda)$.

## 3 A default prior from quasi-likelihoods

Let us assume that the parameter of interest $\tau$ is a scalar function of $\theta$, $\tau = \tau(\theta)$ say. If $\theta$ is scalar, we have $\tau = \theta$; otherwise, $\theta = (\tau, \lambda)$. Let us denote with $\ell^*(\tau)$ a pseudo-loglikelihood function for $\tau$, i.e. a function of the parameter of interest and of the data with properties similar to a genuine likelihood. In particular, in the context of this paper, $\ell^*(\tau)$ may be given by (3) when $\theta = \tau$, or by (4) or (7), when $\theta = (\tau, \lambda)$.

Let $\ell^*(\tau)$ be differentiable with derivatives $\ell^{*1}(\tau), \ell^{*2}(\tau), \ell^{*3}(\tau)$. For the expected values of these derivatives, we use the notation $\nu_2 = E_\theta(\ell^{*2}(\tau))$ and $\nu_{1,1} = E_\theta(\ell^{*1}(\tau)^2)$, which we assume of order $O(n)$. These assumptions are typically satisfied in practice, when the pseudo-score function $\ell^*(\tau)$ behaves asymptotically like the sum of $n$ independent random variables. Under these regularity conditions, $\ell^*(\tau)$ can be expanded in a Taylor series about its maxima, i.e. $\tilde{\tau}$, giving

$$\ell^*(\tau) = \ell^*(\tilde{\tau}) - \frac{1}{2}(\tau - \tilde{\tau})^2 j^*(\tilde{\tau}) + R_n,$$

where $j^*(\tau) = -\ell^{*2}(\tau)$ is the pseudo-observed information and $R_n$ is a remainder term of order $O_p(n^{-1/2})$. 
In general, unlike as it happens with a genuine likelihood, the identity $\nu_2 + \nu_1 = 0$, analogous to the information identity, does not hold. However, for the pseudo-likelihood functions considered in this paper, i.e. (3), (4) and (7), we have $\nu_2 + \nu_1 = o(n)$ and

$$j^*(\tau) = i^*(\tau) + o_p(n),$$

where $i^*(\tau) = \nu_2^2/\nu_1$ is the expected pseudo-information. The asymptotic variance of $\tilde{\tau}$ can be expressed as

$$i^*(\tau)^{-1} = \frac{\nu_1}{\nu_2^2}.$$  

(10)

In the context of the quasi-likelihood and the quasi-profile likelihood, (10) reduces to (2) and (6), respectively.

A default prior distribution for $\tau$ can be obtained defining an uniform distribution that takes into account the geometry of the parameter space (see e.g. Ghosh et al., 2007, Chapter 5). Let $\phi = \phi(\tau)$ be a smooth one-to-one transformation. Then

$$i^*_\phi(\phi_0) = i^*(\tau_0) \left( \frac{d\tau}{d\phi} \right)^2,$$

where $\tau_0$ is a fixed value, $\phi_0 = \phi(\tau_0)$ and $i^*_{\phi}(\cdot)$ denotes the expected pseudo-information in the new parameterization. If $\phi(\tau)$ is chosen so that $(d\tau/d\phi)^2 \propto (i^*(\tau_0))^{-1/2}$, then $i^*_\phi(\phi_0)$ will be a constant. Thus, the metric for which a locally uniform prior is approximately noninformative can be obtained from the relationship $d\phi/d\tau \propto i^*(\tau)^{1/2}$. This, in turn, implies that the corresponding default prior invariant under one-to-one transformation of $\tau$ is

$$\pi(\tau) \propto i^*(\tau)^{1/2}.$$  

(11)

This means that a parametrization invariant prior distribution for $\tau$, derived from a pseudo-likelihood function, is proportional to the square root of the pseudo-expected information. The expression of this prior is analogous to the Jeffreys prior.

In the context of quasi-likelihood functions, when $\theta = \tau$ the default prior (11) leads to the posterior distribution

$$\pi(\theta|y) \propto V(\theta)^{1/2} L_Q(\theta).$$

(12)

If $\theta = (\tau, \lambda)$, the posterior distribution based on a quasi-profile likelihood and (11) is given by

$$\pi(\tau|y) \propto V_{\tau\tau}(\tau)^{1/2} L_{Q\tau}(\tau),$$

or by

$$\pi(\tau|y) \propto V_{\tau\tau}(\tau)^{1/2} L_{QPM}(\tau).$$

(13)

(14)

These results agree with the Welch and Peers (1963) solution in a model with a scalar parameter and with the Ventura et al. (2008) solution for a scalar parameter of interest in the presence of nuisance parameters based on pseudo-likelihoods.
4 Examples and simulation results

In this section we discuss two examples to illustrate the use of (13) and (14) and to evaluate the frequentist coverage of the corresponding credible sets. Interest is focused on quasi-profile likelihood functions derived from estimating equations in the context of a robust scale–location model and of a log-linear model for count data with overdispersion. In the first example, a proper posterior distribution for the parameter of interest is available, but it is not robust with respect to model misspecifications (see Greco et al., 2008). In the second example a genuine likelihood function is not available and the comparison with a marginal posterior distribution is not possible. In both the simulation studies, the number of Monte Carlo replications is set to 5000.

4.1 Robust inference for scale-location model

Let $\theta = (\tau, \lambda)$, where $\tau \in \mathbb{R}$ is a location parameter and $\lambda > 0$ a scale parameter.

In this case we have $F(y; \tau, \lambda) = F_0((y - \tau)/\lambda)$, $\psi(y; \tau, \lambda) = \psi((y - \tau)/\lambda)$, $\Omega(\tau, \lambda) = n \int \psi(x)\psi(x)^T dF_0(x) = \Omega$, $M(\tau, \lambda) = (n/\lambda) \int \psi(x)\rho(x)^T dF_0(x) = M/\lambda$, where $\rho(x)$ is the column vector $(1, x)$, and $A(\tau, \lambda)^T = (1/\lambda)\Omega^{-1}M = A^T/\lambda$.

Consider robust inference about $\tau$ when $\lambda$ is the nuisance parameter. For a symmetric model, a well-known location and scale $M$-estimator is the Huber estimator (Hampel et al., 1986, Section 4.2). The Huber’s estimator for $(\tau, \lambda)$ is the solution of the unbiased estimating equations $\Psi_\tau = \sum_{i=1}^n \psi_k((y_i - \tau)/\lambda)$, where $\psi_k(x) = x \min\{1, k/|x|\}$ for some $k > 0$, and $\Psi_\lambda = \sum_{i=1}^n \psi_{k_1}((y_i - \tau)/\lambda)^2 - k_2$, where $k_1$ and $k_2$ are appropriate constants.

In general, when $\Psi_\tau$ and $\Psi_\lambda$ are odd and even functions, respectively, the quasi-profile likelihood (4) for $\tau$ is given by (Adimari and Ventura, 2001, Bellio et al., 2008)

$$L_{QP}(\tau) = \exp \left\{ A_{\tau \tau} \int_{c_{\tau}}^\tau \sum_{i=1}^n \frac{1}{\hat{\lambda}_t} \psi_{\tau} \left( \frac{y_i - t}{\hat{\lambda}_t} \right) dt \right\},$$

where $\hat{\lambda}_t$ is the estimate for $\lambda$ derived from $\Psi_\lambda$ when $\tau$ is considered as known. The factor $A_{\tau \tau} = M_{\tau \tau}/\Omega_{\tau \tau}$ is the diagonal element of the matrix $A$ corresponding to $\tau$. When the central model is the normal one and the Huber estimator is used, then the factor $A_{\tau \tau}$ has expression

$$A_{\tau \tau} = \frac{\Phi(k) - \Phi(-k)}{2(k^2\Phi(-k) - k\phi(k) + \Phi(k) - 1/2)},$$

where $\phi(x)$ and $\Phi(x)$ are the density and the cumulative density function of the standard normal model $N(0, 1)$, respectively.

It is straightforward to show that, in this case, the default prior (11) reduces to $\pi(\tau) \propto \hat{\lambda}_\tau^{-1}$, so that the posterior distribution (13) is given by

$$\pi(\tau|y) \propto L_{QP}(\tau) \hat{\lambda}_\tau^{-1}.$$  (16)

The normalizing constant of (16) is computationally demanding, because of the integral in (15), and it has been obtained using numerical integration after the
approximation of the posterior kernel $L_{QP}(\tau)\tilde{\lambda}_{\tau}^{-1}$ in a grid of 100 points with a fitted smooth spline on them.

Table 1 gives the frequentist coverages for 95% posterior credible intervals for the pseudo-posterior (16) compared with the classical confidence intervals for the mean of a normal distribution, which are also the posterior credible intervals with the usual Jeffreys’ prior $\pi_J(\tau, \lambda) \propto \lambda^{-2}$. The numerical study is carried out under two different distributions: the $N(0, 1)$, and the $N(0, 1)$ contaminated by a $N(0, 100)$. The contamination percentage is set at 10%. The Huber estimators are used with $k = k_1 = 1.5$ and $k_2 = 0$. From Table 1, it can be noted that inference based on $\pi(\tau|y)$ is quite satisfactory when the model is correctly specified and is preferable to $\pi_J(\tau|y)$ under departures from the standard normal model.

| $n$  | $\pi(\tau|y)$ | $\pi_J(\tau|y)$ | $\pi(\tau|y)$ | $\pi_J(\tau|y)$ |
|------|---------------|-----------------|---------------|-----------------|
| 5    | 0.993         | 0.966           | 0.957         | 0.975           |
| 15   | 0.984         | 0.951           | 0.954         | 0.976           |

Table 1: Frequentist coverages of 95% posterior credible intervals for the location-scale model.

### 4.2 Overdispersion in count data

Let us consider a log-linear regression with overdispersion for count data (see McCullagh and Nelder, 1989). The responses $y_i$ are realizations of independent random variables with mean $\mu_i = \exp(x_i^T\beta)$, $\beta \in \mathbb{R}^p$, $p \geq 1$, $i = 1, \ldots, n$. We focus on the situation where the variance is assumed to be a quadratic function of the mean, i.e. $V_i = \mu_i(1 + \alpha)$, with $\alpha > 0$, $i = 1, \ldots, n$.

An unbiased estimating function for $\beta$ is the score function from the Poisson likelihood, given by

$$ \Psi_{\beta} = \sum_{i=1}^{n} (y_i - \mu_i) x_i^T. $$

An estimating function for $\alpha$ can be obtained from the method of moments (Lawless, 1987), that is

$$ \Psi_{\alpha} = \sum_{i=1}^{n} \frac{(y_i - \mu_i)^2}{V_i} - (n - p). $$

Assume that the parameter of interest $\tau$ is a scalar regression coefficient, i.e. $\tau = \beta_j$ ($1 \leq j \leq p$). In the presence of overdispersion, a genuine likelihood for $\tau$ is not available, but it is possible to resort to a quasi-profile likelihood (see Bellio et al., 2008). Let $\Psi_{\tau}$ and $\Psi_{\lambda}$ be unbiased estimating functions for $\tau$ and $\lambda$, respectively, with $\tau = \beta_j$ and $\lambda = (\beta_1, \ldots, \beta_{j-1}, \beta_{j+1}, \ldots, \beta_d, \alpha)$. In this setting, for inference about $\tau$ we use (7), which accounts for the consequences of the overdispersion for inference on $\tau$. 
Since the assumptions on the variance are typical of a negative binomial model for the response, this distribution can be used for computing the adjustments in (7) and the prior (11).

We apply this procedure to the Ames Salmonella data (see Lawless, 1987, and Bellio et al., 2008) and we use this design matrix to assess the coverage of posterior credible intervals in a Monte Carlo study. In this example $y_i$, $i = 1, \ldots, n$, represents the number of revertant colonies on a plate, and covariates are based on the dose level of quinoline on the plate. We assume

$$\log(\mu_i) = \beta_0 + \beta_1 x_i + \tau \log(x_i + 10),$$

for $i = 1, \ldots, 18$.

The prior distribution (11) does not have in this case an analytical form and it can be approximated by averaging the observed pseudo-information for $\tau$ on 100 parametric bootstrap samples from the negative binomial model, with $\tau$ chosen in a grid of points and $\lambda$ fixed. This average is supposed to approximate $\hat{\tau}_{PS}(\tau)$ and so $\pi_{PS}(\tau)$ at point $\tau$.

In order to evaluate the frequentist coverages of 95% posterior credible intervals based on (14), we ran a simulation study generating data under a negative binomial model. In particular, the data were simulated according to the same design and parameter values similar to those obtained from the Ames Salmonella data (see Bellio et al., 2008). The frequentist coverage of credible intervals of $\pi(\tau|y)$ is 0.968, quite near to the nominal 95%. Other scenarios are showed in second and third rows of Table 2. We note that in all cases considered, despite the small dimension of the sample size, credible intervals of $\pi(\tau|y)$ have accurate frequentist coverages.

| $\tau$  | $\lambda = (\beta_0, \beta_1, \alpha)$ | $\pi(\tau|y)$ |
|---------|--------------------------------------|--------------|
| 0.3     | $(1,0,2)$                            | 0.968        |
| 0       | $(1,0,0.01)$                          | 0.956        |
| 0       | $(1,0,0.05)$                          | 0.945        |

**Table 2:** Frequentist coverages of 95% posterior credible intervals in the log-linear model for count data.

### 5 Final remarks

In this paper a Jeffreys-type prior for a parameter of interest is derived and discussed for Bayesian inference based quasi- and quasi-profile likelihood functions. Other examples of pseudo-likelihoods for which (9) is satisfied, and for which the default prior can be derived, are the marginal, the conditional, the profile and its modified versions, the approximate conditional, the empirical likelihood functions; see, e.g., Barndorff-Nielsen (1995), DiCiccio et al. (1996), Severini (1998), and Adimari and Ventura (2002). For the marginal, the conditional and the modified profile likelihoods in Ventura et al. (2008) it is shown that these Jeffreys-type priors have also matching properties.
Acknowledgments

This work was partially supported by grants from Ministero dell’Università e della Ricerca Scientifica e Tecnologica, Italy.

References


Working Paper Series
Department of Statistical Sciences, University of Padua

You may order paper copies of the working papers by emailing wp@stat.unipd.it

Most of the working papers can also be found at the following url: http://wp.stat.unipd.it