Working Paper Series, N. 7, May 2007

Default priors from pseudo-likelihoods in the presence of nuisance parameter

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

Laura Ventura
Department of Statistical Sciences
University of Padua
Italy

Abstract: Consider a sampling parametric model with parameter $\theta = (\psi, \lambda)$, where ψ is the parameter of interest. Also in the Bayesian framework, an approach to the elimination of nuisance parameters is to use an appropriate pseudo-likelihood function $L_{PS}(\psi)$ for the parameter of interest only, in alternative to an integrated likelihood function.
The goal of this paper is to select a class of default priors $\pi_{PS}(\psi)$ for a parameter of interest using pseudo-likelihood functions. Developing Stein's (1985) and Tibshirani's (1989) results, our approach is to require that the resulting pseudo-posterior intervals, based on the pseudo-posterior distribution $\pi_{PS}(\psi|y) \propto \pi_{PS}(\psi)L_{PS}(\psi)$, have accurate frequentist coverage. Several illustrative examples are given and comparisons of $\pi_{PS}(\psi|y)$ are made to the posterior distributions based on the reference or Jeffreys priors. Some interesting conclusions emerge.

Keywords: Frequentist coverage probability, Jeffreys prior, Modified profile likelihood, Profile likelihood, Reference prior.

Contents

1 Introduction **1**

2 Background theory **3**

3 Default prior for ψ **3**

4 Modifications of the profile likelihood **4**

5 Simulation studies **6**

5.1 Gamma distribution 6

5.2 Inverse Gaussian distribution 7

5.3 Common ratio of Poisson means 9

5.4 Matched exponential samples 9

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: +39 049 827 4177
ventura@stat.unipd.it
http://www.stat.unipd.it/~ventura

Default priors from pseudo-likelihoods in the presence of nuisance parameter

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

Laura Ventura
Department of Statistical Sciences
University of Padua
Italy

Abstract: Consider a sampling parametric model with parameter $\theta = (\psi, \lambda)$, where ψ is the parameter of interest. Also in the Bayesian framework, an approach to the elimination of nuisance parameters is to use an appropriate pseudo-likelihood function $L_{PS}(\psi)$ for the parameter of interest only, in alternative to an integrated likelihood function.

The goal of this paper is to select a class of default priors $\pi_{PS}(\psi)$ for a parameter of interest using pseudo-likelihood functions. Developing Stein's (1985) and Tibshirani's (1989) results, our approach is to require that the resulting pseudo-posterior intervals, based on the pseudo-posterior distribution $\pi_{PS}(\psi|y) \propto \pi_{PS}(\psi)L_{PS}(\psi)$, have accurate frequentist coverage. Several illustrative examples are given and comparisons of $\pi_{PS}(\psi|y)$ are made to the posterior distributions based on the reference or Jeffreys priors. Some interesting conclusions emerge.

Keywords: Frequentist coverage probability, Jeffreys prior, Modified profile likelihood, Profile likelihood, Reference prior.

1 Introduction

Assume that $y = (y_1, \ldots, y_n)$ is a sample of n independent and identically distributed random variables from a model $p(y|\theta)$, with $\theta \in \Theta \subseteq \mathbb{R}^p$, partitioned as $\theta = (\psi, \lambda)$, where ψ is a scalar parameter of interest and λ is a $(p-1)$-dimensional nuisance parameter. Bayesian techniques for eliminating λ require a prior $\pi(\psi, \lambda) = \pi(\psi)\pi(\lambda|\psi)$ over Θ and are based on the marginal posterior for ψ, given by

$$\pi(\psi|y) = \int \pi(\psi, \lambda|y)\, d\lambda \propto \int L(\psi, \lambda)\pi(\psi, \lambda)\, d\lambda \propto L_I(\psi)\pi(\psi) \ , \tag{1}$$

where $L(\psi, \lambda)$ is the entire likelihood function and

$$L_I(\psi) = \int L(\psi, \lambda)\pi(\lambda|\psi)\, d\lambda \tag{2}$$

denotes the integrated likelihood for the parameter of interest ψ with respect to $\pi(\lambda|\psi)$. See, for example, Liseo (1993), Berger *et al.* (1999) and Severini (2000, Chapter 6) for further discussions on integrated likelihoods.

To obtain the integrated likelihood, the elicitation of $\pi(\lambda|\psi)$ may be difficult both in the subjective and objective Bayesian contexts, and computation of the integral in (2) can be heavy when the dimension of λ is high. Moreover, inferential procedures based on (1) are not robust with respect to model misspecifications, and for complex models it may be difficult to write the complete likelihood $L(\psi, \lambda)$. An alternative approach to the elimination of nuisance parameters is to resort to different pseudo-likelihood functions. In general, a pseudo-likelihood $L_{PS}(\psi)$ is a function of the parameter of interest only, and of the data, with properties similar to those of a genuine likelihood function. Some examples of pseudo-likelihoods for a parameter of interest are the marginal and the conditional, which require specific model structures, the profile and modified versions, and the quasi- likelihoods; see, e.g., Pace and Salvan (1997, Chapter 4) and Severini (2000, Chapters 8 and 9). Although the use of a pseudo-likelihood function in the Bayesian inference cannot be considered as orthodox in the Bayesian perspective, the use of alternative likelihoods is actually widely shared. Unlike integrated likelihoods, pseudo-likelihoods are usually based on maximizations rather than averaging. See, e.g., Efron (1993), Bertolino and Racugno (1992, 1994), Raftery *et al.* (1996), Cabras *et al.* (2006) and Greco *et al.* (2007). Papers which are more specifically related to the validation of a pseudo-posterior distribution based on an alternative likelihood are Monahan and Boos (1992), Severini (1999), Lazar (2003), Racugno *et al.* (2005), Pace *et al.* (2006) and Schennach (2006).

In this paper we consider the problem of selecting a default prior $\pi_{PS}(\psi)$ for a parameter of interest using pseudo-likelihood functions. Specifically, we seek a prior $\pi_{PS}(\psi)$ so that the resulting posterior interval for ψ, based on the pseudo-posterior distribution

$$\pi_{PS}(\psi|y) \propto L_{PS}(\psi)\pi_{PS}(\psi), \tag{3}$$

has a coverage error of only $O(n^{-1})$ in the frequentist sense. Extending Stein's (1985) and Tibshirani's (1989) results in the pseudo-likelihood framework, we show that the class of priors satisfying Stein's condition is proportional to the square root of an expected pseudo-information, where the latter is defined as the inverse of the asymptotic variance of the pseudo-maximum likelihood estimator. Through several illustrative examples, Bayesian techniques based on (1) with the reference or the Jeffreys priors are compared to the methods based on (3), and frequentist coverage of the implied confidence procedures are considered. Some interesting conclusions emerge.

Section 2 briefly introduces the concept of pseudo-likelihood in the presence of nuisance parameters and reviews its first order asymptotic properties. In Section 3 the general procedure to select a default prior $\pi_{PS}(\psi)$ using pseudo-likelihood functions is discussed. Section 4 states the relation between $\pi_{PS}(\psi)$ and Tibshirani's (1990) prior for orthogonal parameters, and gives the expression of $\pi_{PS}(\psi)$ for a modified profile likelihood. Finally, Section 5 is devoted to the discussion of several examples, that raise interesting questions.

2 Background theory

In order to draw inferences regarding the parameter of interest ψ in a given model, in alternative to $L_I(\psi)$, several useful pseudo-likelihoods for a parameter of interest can be considered. A pseudo-likelihood function for a parameter of interest ψ can be used as a genuine likelihood function, since in general it shares the same asymptotic properties. In particular, the pseudo-maximum likelihood estimator (MLE) $\hat{\psi}_{PS}$ is consistent and asymptotically normal, i.e.

$$\hat{\psi}_{PS} \dot{\sim} N(\psi, i_{PS}(\psi)^{-1}) \ ,$$

where $i_{PS}(\psi)$ can be interpreted as a pseudo-information, and the pseudo-likelihood ratio statistic $W_{PS}(\psi)$ has a null asymptotic chi-square distribution. Some well-known examples of pseudo-likelihoods are the conditional, the marginal, the approximate conditional, the profile, the modified profile, the generalized profile, the partial and the integrated likelihood functions. See also the Severini's integrated (Severini, 2007) and the quasi-profile likelihood functions (Adimari and Ventura, 2002).

In general, unlike as happens with a genuine likelihood, a pseudo-likelihood function does not satisfy the second Bartlett identity. In view of this, the pseudo-observed information $j_{PS}(\psi) = -\partial^2 \ell_{PS}(\psi)/\partial\psi\partial\psi^{\mathsf{T}}$, with $\ell_{PS}(\psi) = \log L_{PS}(\psi)$, is not connected in the usual way to the asymptotic ariance of the pseudo-MLE. When the information identity does not hold, the asymptotic variance of $\hat{\psi}_{PS}$ can be written as the Godambe information, i.e.

$$i_{PS}(\psi)^{-1} = E_\theta \left(-\frac{\partial^2 \ell_{PS}(\psi)}{\partial \psi^2} \right)^{-2} E_\theta \left(\left(\frac{\partial \ell_{PS}(\psi)}{\partial \psi} \right)^2 \right) \ , \tag{4}$$

where expectations are computed under $(\psi, \hat{\lambda}_\psi)$, with $\hat{\lambda}_\psi$ suitable consistent estimate of the nuisance parameter for fixed ψ.

The main asymptotic results of pseudo-likelihood based inference allow to state the asymptotic normality of a pseudo-posterior distribution $\pi_{PS}(\psi|y)$, extending results in, e.g., Bernardo and Smith (2000, Chapter 5) or Davison (2003, Chapter 11). Indeed, for large n, using the expansion $(\hat{\psi}_{PS} - \psi) \dot{\sim} i_{PS}(\psi)^{-1}(\partial \ell_{PS}(\psi)/\partial\psi) + o_p(n^{1/2})$, we can write

$$2(\ell_{PS}(\hat{\psi}_{PS}) - \ell_{PS}(\psi)) = i_{PS}(\psi)(\hat{\psi}_{PS} - \psi)^2 + o_p(1) \ ,$$

and an analogous expansion can be obtained for the log-prior $\log \pi_{PS}(\psi)$ around the prior mode. However, in large samples the loglikelihood contribution is typically much greater than that from the prior in (3) and thus for $\pi_{PS}(\psi|y)$ we have

$$\pi_{PS}(\psi|y) \dot{\sim} N(\hat{\psi}_{PS}, i_{PS}(\psi)) \ . \tag{5}$$

3 Default prior for ψ

In this section, we look for a prior $\pi_{PS}(\psi)$ for the parameter of interest only for which posterior probability regions, based on $\pi_{PS}(\psi)$ and a suitable pseudo-likelihood, have accurate frequentist coverage.

Following Stein (1985) and using (5), we seek a prior $\pi_{PS}(\psi)$ for the parameter of interest, so that

$$Pr_{\psi|y}\{\psi \in S_\alpha\} = \alpha \ , \tag{6}$$

where

$$S_\alpha = \{\psi : i_{PS}(\psi)^{1/2}(\psi - \hat{\psi}_{PS}) < \Phi^{-1}(\alpha)\} \ , \tag{7}$$

is a normal posterior region for ψ. Moreover, $Pr_{\psi|y}\{\cdot\}$ denotes posterior probability, and $\Phi^{-1}(\alpha)$ is the αth percentile of the standard normal distribution. We argue that

$$Pr_\theta\{\psi \in S_\alpha\} = \alpha + O(n^{-1}) \ , \tag{8}$$

for all ψ, where $Pr_\theta\{\cdot\}$ indicates probabilty under $p(y|\theta)$. Thus, a sufficient condition for $\pi_{PS}(\psi)$ to give (6) is

$$\pi_{PS}(\psi)\frac{\partial i_{PS}(\psi)^{-1/2}}{\partial \psi} + i_{PS}(\psi)^{-1/2}\frac{\partial \pi_{PS}(\psi)}{\partial \psi} = 0 \ . \tag{9}$$

It is then straightforward to show that the solutions to condition (9) take the form

$$\pi_{PS}(\psi) \propto i_{PS}(\psi)^{1/2} \ . \tag{10}$$

Expression (10) says that for a parameter of interest only, the default prior $\pi_{PS}(\psi)$ is proportional to the square root of the pseudo-information. Note that this result can be interpreted as a generalization to the situation of a pseudo-likelihood function of the procedure of the Jeffreys prior.

Using (10) and the corresponding pseudo-likelihood function for ψ, the pseudo-posterior distribution is

$$\pi_{PS}(\psi|y) \propto i_{PS}(\psi)^{1/2} \, L_{PS}(\psi) \ . \tag{11}$$

In view of (6), posterior intevals for ψ based on (11) should have accurate frequentist coverage.

4 Modifications of the profile likelihood

In this Section we show that (10) generalizes Tibshirani's result, which is related to the assumption of parameter orthogonality. Moreover, we give the expression of (10) for a modified profile likelihood.

In an orthogonal parameterization with respect to expected Fisher information, a suitable modification of the profile likelihood for inference about ψ gives the adjusted profile likelihood function (Cox and Reid, 1987)

$$L_{AC}(\psi) = L_p(\psi) \, |j_{\lambda\lambda}(\psi, \hat{\lambda}_\psi)|^{-1/2} \ , \tag{12}$$

where $L_p(\psi) = L(\psi, \hat{\lambda}_\psi)$ is the profile likelihood for ψ, $\hat{\lambda}_\psi$ is the restricted MLE of λ solution of $\partial \log L(\psi, \lambda)/\partial\lambda = 0$ and $j_{\lambda\lambda}(\psi, \lambda) = -\partial^2 \log L(\psi, \lambda)/\partial\lambda\partial\lambda^\top$ is the observed Fisher information for λ evaluated at $(\psi, \hat{\lambda}_\psi)$. The main consequences of parameter orthogonality are that $\hat{\psi}$ and $\hat{\lambda}$ are asymptotically independent and, in addition, that $\hat{\lambda}_\psi - \hat{\lambda} = O_p(n^{-1})$.

It is well-known that standard first-order methods for inference about ψ based on $L_p(\psi)$ can be seriously inaccurate in particular when the dimension of λ is substantial relative to n. Various modifications of the profile likelihood of the form

$$L_{mp}(\psi) = M(\psi)L_p(\psi) \ , \tag{13}$$

where $M(\psi)$ is a suitable smooth correction factor, have been proposed; see Barndorff-Nielsen and Cox (1994, Chapter 8) and Severini (2000, Chapter 9) for accounts. All the available adjustments to $L_p(\psi)$, including (12), are equivalent to second order and share the common feature of reducing the score bias to $O(n^{-1})$ (DiCiccio *et al.*, 1996).

The adjusted profile likelihood plays a central role also in the Bayesian setting. The easiest way to see this is to consider an expansion of $\log L(\psi, \lambda)$ as a function of λ about $\hat{\lambda}_\psi$. We get (see, e.g., Reid, 1995)

$$\begin{aligned} \pi(\psi|y) &\propto L(\psi, \hat{\lambda}_\psi)|j_{\lambda\lambda}(\psi, \hat{\lambda}_\psi)|^{-1/2}\pi(\psi)\pi(\hat{\lambda}_\psi|\psi) && (14) \\ &\propto L_{AC}(\psi)\pi(\psi)\pi(\hat{\lambda}_\psi|\psi) \ . && (15) \end{aligned}$$

This approximation to the marginal posterior was derived by Tierney and Kadane (1986), which point out that the relative error is $O(n^{-3/2})$. In (14) we have that the marginal posterior distribution is proportional to the profile likelihood $L_p(\psi)$ plus a correction term, where the correction takes into account the information about the nuisance parameter as ψ varies. The adjusted profile likelihood $L_{AC}(\psi)$ has been widely discussed in the Bayesian framework; see, e.g., Liseo (1993), Reid (1995) ans Sweeting (1995).

In an orthogonal parameterization, Tibshirani (1989) provided a general form for the class of priors satisfying Stein's condition. In particular, Tibshirani's prior can be assumed of the form

$$\pi_T(\psi, \lambda) \propto g(\lambda) i_{\psi\psi}(\psi, \lambda)^{1/2} \ , \tag{16}$$

where $g(\lambda) > 0$ is arbitrary and $i_{\psi\psi}(\psi, \lambda)$ is the (ψ, ψ) component of the Fisher information matrix $i(\psi, \lambda)$ based on $L(\psi, \lambda)$. Expression (16) is obtained so that the procedures implied by the objective prior $\pi_T(\psi, \lambda)$ have good frequentist properties. This result was derived specializing the results of Peers (1965) and Stein (1985); a rigorous derivation is given in Nicolaou (1993). Berger and Bernardo (1992) proved that the Tibshirani prior is equal to the reference prior for a particular ordering of the parameters.

Using (16) in (15), and using the orthogonality of ψ and λ so that we may replace $\hat{\lambda}_\psi$ by $\hat{\lambda}$ to the same order of approximation, we obtain the following pseudo-posterior distribution

$$\pi_{PS}(\psi|y) \propto i_{\psi\psi}(\psi, \hat{\lambda})^{1/2} \, L_{AC}(\psi) \ , \tag{17}$$

for which, in view of (16), the resulting posterior intervals for ψ have accurate frequentist coverage. It is straightforward to show that (17) is equivalent to (11), since in this case the pseudo-MLE $\hat{\psi}_{AC}$ is approximately normally distributed with mean ψ and asympotic variance $[i_{\psi\psi}(\psi,\lambda)]^{-1}$. In view of this, for orthogonal parameters, we can assume $i_{PS}(\psi) = i_{\psi\psi}(\psi,\hat{\lambda})$.

More generally, in the absence of an orthogonal parameterization, for a pseudo-likelihood function of the form (13), i.e. for a modified profile likelihood, we have that the pseudo-MLE $\hat{\psi}_{PS}$ is approximately normally distributed with mean ψ and asympotic variance $[i_{\psi\psi}(\psi,\lambda) - i_{\psi\lambda}(\psi,\lambda)(i_{\lambda\lambda}(\psi,\lambda))^{-1} i_{\lambda\psi}(\psi,\lambda)]^{-1}$. In this case,

$$i_{PS}(\psi) = i_{\psi\psi}(\psi,\hat{\lambda}_\psi) - i_{\psi\lambda}(\psi,\hat{\lambda}_\psi)(i_{\lambda\lambda}(\psi,\hat{\lambda}_\psi))^{-1} i_{\lambda\psi}(\psi,\hat{\lambda}_\psi) \ .$$

This expression holds for any $L_{PS}(\psi)$ of the form (13) or such that $\hat{\psi}_{PS} = \hat{\psi} + O_p(n^{-1})$, where $\hat{\psi}$ is the MLE of ψ.

5 Simulation studies

In this section we discuss four different examples. In particular, the first two examples focus on orthogonal parameters and on the pseudo-posterior (17), while the latter two concern the pseudo-posterior (11). In all examples coverage probabilities are all accurate and surprisingly most of the proposed pseudo-posteriors coincide with other well known posterior distributions.

5.1 Gamma distribution

Let $(y_1,\ldots,y_n)$ be n independent observations from a gamma distribution with density

$$f(y;\psi,\lambda) = \left(\frac{\lambda}{\psi}\right)^{-\psi} y^{(\psi-1)} \exp\left(-\frac{\psi}{\lambda}y\right)\Gamma(\psi)^{-1} \ .$$

This parameterization makes ψ and λ orthogonal. Assume the shape parameter ψ of interest. The likelihood function is

$$L(\psi,\lambda) = \left(\frac{\lambda}{\psi}\right)^{-n\psi} q^{\psi-1} \exp\left(-\frac{\psi}{\lambda}t\right)\Gamma(\psi)^{-n} \ ,$$

with $t = \sum_{i=1}^n y_i = n\bar{y}$ and $q = \prod_{i=1}^n y_i$. The restricted MLE for λ given ψ is $\hat{\lambda}_\psi = \hat{\lambda} = \bar{y}$, which does not depend on ψ, and

$$j_{\lambda\lambda}(\psi,\hat{\lambda}) = \frac{n\psi}{\bar{y}^2} \ ,$$

so that

$$L_{AC}(\psi) \ \propto \ \left(\frac{\bar{y}}{\psi}\right)^{-n\psi} q^{\psi-1} \exp\left(-\frac{\psi}{\bar{y}}\right)\Gamma(\psi)^{-n}\psi^{-1/2}.$$

The Jeffreys and the reference priors are (Liseo, 1993), respectively,

$$\begin{aligned} \pi_J(\psi,\lambda) \ &\propto \ \lambda^{-1}\sqrt{\psi\xi(\psi) - 1} \\ \pi_R(\psi,\lambda) \ &\propto \ \lambda^{-1}\sqrt{\xi(\psi) - \psi^{-1}} \ , \end{aligned}$$

	n=3	
(ψ,λ)	$\pi_R(\psi\|y)$ $\pi_T(\psi\|y)$	$\pi_J(\psi\|y)$
$(\frac{1}{2},1)$	0.949	0.888
(1,2)	0.946	0.888
(2,1)	0.952	0.888
(5,3)	0.950	0.885

	n=10	
(ψ,λ)	$\pi_R(\psi\|y)$ $\pi_T(\psi\|y)$	$\pi_J(\psi\|y)$
(1,2)	0.955	0.938
(2,1)	0.957	0.940

Table 1: Frequentist coverages of 95% posterior credible intervals in the Gamma example. The monte carlo error is $\sqrt{p(1-p)/5000}$, where p is the table entry.

where $\xi(\psi)$ is the Trigamma function, so that

$$\begin{aligned}\pi_J(\psi,\hat{\lambda}) &\propto \sqrt{\psi\xi(\psi)-1},\\ \pi_R(\psi,\hat{\lambda}) &= \sqrt{n}\sqrt{\xi(\psi)-\psi^{-1}} = i_{\psi\psi}(\psi,\hat{\lambda})^{1/2}\ .\end{aligned}$$

Note that the reference prior in this case equals the Tibshirani prior in (17), with $g(\lambda)=\lambda^{-1}$. Then the kernels of the posteriors are

$$\begin{aligned}\pi_J(\psi \mid y) &\propto \psi^{n\psi}q^{\psi}\exp\left(-\psi/\bar{y}\right)\Gamma(\psi)^{-n}\psi^{-1/2}\sqrt{\psi\xi(\psi)-1}\ ,\\ \pi_R(\psi \mid y) &= \pi_T(\psi \mid y)\\ &\propto \psi^{n\psi}q^{\psi}\exp\left(-\psi/\bar{y}\right)\Gamma(\psi)^{-n}\psi^{-1/2}\sqrt{\xi(\psi)-\psi^{-1}}\ .\end{aligned}$$

The frequentist coverages for 95% posterior credible intervals based on $\pi_J(\psi \mid y)$ and $\pi_R(\psi \mid y)=\pi_T(\psi \mid y)$ are given in Table 1 and are approximated with 5000 thousand samples. Simulations suggest that credible intervals of $\pi_T(\psi \mid y)$ have frequentist coverages.

5.2 Inverse Gaussian distribution

Let $(y_1,\ldots,y_n)$ be n independent observations from the Inverse Gaussian model, with density

$$f(y|\psi,\lambda)=\sqrt{\frac{\psi}{2\pi y^3}}\exp\left[\psi\left(\frac{1}{\lambda}-\frac{y}{2\lambda^2}-\frac{1}{2y}\right)\right]\ ,$$

where ψ and λ are orthogonal. The scale parameter ψ is of interest. The likelihood function is

$$L(\psi,\lambda)=\psi^{n/2}\exp\left[\psi\left(\frac{n}{\lambda}-\frac{t}{2\lambda^2}-\frac{a}{2}\right)\right]\ ,$$

with $t = n\bar{y}$ and $a = \sum_{i=1}^{n} \frac{1}{y_i}$. In this case $\hat{\lambda}_\psi = \hat{\lambda} = \bar{y}$,

$$j_{\lambda\lambda}(\psi, \hat{\lambda}) = \frac{n\psi}{\bar{y}^3}$$

and

$$L_{AC}(\psi) = \psi^{\frac{n-1}{2}} \exp(-\psi s),$$

where $s = \frac{n}{2}\left(\frac{a}{n} - \frac{1}{\bar{y}}\right)$. The Tibshirani's prior $\pi_T(\psi, \lambda)$ is

$$\pi_T(\psi, \lambda) \propto \psi^{-1} \ ,$$

and thus the pseudo-posterior distribution is

$$\pi_T(\psi \mid y) \sim \text{Gamma}\left(\frac{n-1}{2}, s\right) \ ,$$

with mean $\frac{n-1}{2s}$ and variance $\frac{n-1}{2s^2}$. Note that $\pi_T(\psi \mid y)$ is proper with at least $n = 2$, while the posterior credible intervals in Liseo (1993) need $n = 3$. The frequentist coverages for 95% posterior credible intervals based on $\pi_T(\psi \mid y)$ and the reference posteriors $\pi_R(\psi \mid y)$ and $\pi_{R1}(\psi \mid y)$ discussed in Liseo (1993) are given in Table 2 and are approximated with 5000 thousand samples. It can be noted that also for $n = 10$, comparisons are in favour of $\pi_T(\psi \mid y)$.

$n = 2$

(ψ, λ)	$\pi_T(\psi \mid y)$	$\pi_R(\psi \mid y)$	$\pi_{R1}(\psi \mid y)$
$(\frac{1}{2},1)$	0.948	–	–
(1,1)	0.948	–	–
(2,2)	0.951	–	–

$n = 10$

(ψ, λ)	$\pi_T(\psi \mid y)$	$\pi_R(\psi \mid y)$	$\pi_{R1}(\psi \mid y)$
$(1,\frac{1}{2})$	0.946	0.972	0.962
(1,1)	0.945	0.963	0.963
(2,2)	0.951	0.964	0.961
(3,3)	0.949	0.966	0.946

Table 2: Frequentist coverages of 95% posterior credible intervals for the Inverse Gaussian example. The monte carlo error is $\sqrt{p(1-p)/5000}$, where p is the table entry.

5.3 Common ratio of Poisson means

Let $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$ denote independent poisson random variables with mean $\psi\lambda$ and λ, respectively, with $\psi > 0$ and $\lambda > 0$. Let $x = \sum_{i=1}^n x_i$ and $y = \sum_{i=1}^n y_i$. The likelihood function for the model is given by

$$L(\psi, \lambda) = \exp\left(x \log \psi + (x+y) \log \lambda - n\lambda(\psi+1)\right) .$$

Severini'7 (2007) integrated likelihood is in this case given by

$$L_S(\psi) = \frac{\psi^x}{(\psi+1)^{x+y}} ,$$

which can combined with the default prior

$$\pi_{PS}(\psi) \propto \sqrt{\frac{2x}{\psi(\psi+1)^2}} \propto (\psi+1)^{-1}\psi^{-1/2} .$$

The pseudo-posterior (17) is thus

$$\pi_{PS}(\psi \mid x, y) \propto \frac{\psi^{x-1/2}}{(\psi+1)^{x+y+1}} ,$$

and the corresponding density is

$$\pi_{PS}(\psi \mid x, y) = \frac{\Gamma(x+y+1)}{\Gamma(x+1/2)\Gamma(y+1/2)} \frac{\psi^{x-1/2}}{(\psi+1)^{x+y+1}} .$$

For this problem, the reference integrated likelihood is

$$L_R(\psi) = \frac{\psi^x}{(\psi+1)^{x+y+1/2}}$$

and the reference prior for ψ slightly differs from $\pi_{PS}(\psi)$ as $\pi_R(\psi) \propto (\psi+1)^{-1/2}\psi^{-1/2}$. Nonetheless $\pi_{PS}(\psi \mid x, y)$ equals the reference posterior $\pi_R(\psi \mid y) \propto \pi_R(\psi) L_R(\psi)$, for which it is known that it satisfies equation (8). Table 3 gives the frequentist coverages for 95% posterior credible intervals based on $\pi_{PS}(\psi \mid y) = \pi_R(\psi \mid y)$, whcih are approximated with 5000 thousand samples.

5.4 Matched exponential samples

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be independent pairs of independent Gamma random variables, with fixed index m and unknown rate parameters λ_{Xj}, λ_{Yj}, respectively, $j = 1, \ldots, n$. Let $\psi = \lambda_{Xj}/\lambda_{Yj}$ denote the parameter of interest, which is assumed to be the same for all j, and take $\lambda_j = \lambda_{Yj}$ as the nuisance parameter.

Let $T_j = X_j/Y_j$. The marginal distribution of $T = (T_1, \ldots, T_n)$ depends on ψ only and hence it may be used to form a marginal likelihood function. In particular, the marginal likelihood is

$$L_m(\psi) = \exp\left\{ mn \log \psi - 2m \sum_{j=1}^n \log(\psi x_j + y_j) \right\} .$$

	$\psi = 4, \lambda = 4$
n	$\pi_{PS}(\psi \mid y) = \pi_R(\psi \mid y)$
1	0.948
2	0.948
4	0.951
6	0.953
10	0.950

Table 3: Frequentist coverages of 95% posterior credible intervals. The monte carlo error is $\sqrt{p(1-p)/5000}$, where p is the table entry.

The marginal likelihood $L_m(\psi)$ can be combined with the default prior

$$\pi_{PS}(\psi) \propto \sqrt{\frac{2m^2 n}{2(2m+1)\psi^2}} \propto \psi^{-1} \ .$$

The kernel of the pseudo-posterior (17) is

$$\pi_{PS}(\psi \mid x, y) \propto \exp\left\{ (mn-1)\log\psi - 2m\sum_{j=1}^{n} \log(\psi x_j + y_j) \right\} \ .$$

For comparison, consider the full likelihood for this experiment

$$L(\psi, \lambda_1, \lambda_2, \ldots, \lambda_n) = \psi^{mn} \prod_{j=1}^{n} \lambda_j^{2m} \exp\left\{ -\sum_{j=1}^{n} \lambda_j(\psi x_j + y_j) \right\} \ ,$$

and the joint Jeffreys's prior

$$\pi_J(\psi, \lambda_1, \lambda_2, \ldots, \lambda_n) \propto \psi^{-1} \prod_{j=1}^{n} \lambda_j^{-1} \ .$$

We have $\pi_J(\psi \mid x, y) = \pi_{PS}(\psi \mid x, y)$. The proof is straightforward by noting that the kernel of $\pi_J(\psi, \lambda_j \mid x, y)$, as a function of λ_j, is a $Gamma(2m, \psi x_j + y_j)$.

Table 4 gives the frequentist coverages for 95% posterior credible intervals based on $\pi_J(\psi \mid x, y) = \pi_{PS}(\psi \mid x, y)$, whcih are approximated with 5000 thousand samples.

References

Adimari, G., Ventura, L. (2002). Quasi-profile loglikelihoods for unbiased estimating functions. *Ann. Inst. Statist. Math.*, **54**, 235–244.

Berger, J.O. (1985). *Statistical Decision Theory and Bayesian Analysis*, 2nd edition. Springer, Berlin.

$\psi = 1, \lambda = 1$	
n	$\pi_{PS}(\psi \mid x, y)$
5	0.955
10	0.955
20	0.948
50	0.949
100	0.948

Table 4: Frequentist coverages of 95% posterior credible intervals. The monte carlo error is $\sqrt{p(1-p)/5000}$, where p is the table entry.

Berger, J.O., Bernardo, J.M. (1992). On the development of the reference prior method (with discussion). In *Bayesian Statistics* 4, Ed J.M. Bernardo, J.O. Berger, A.P. David and A.F.M. Smith, 33–60. Oxford University Press.

Berger, J.O., Liseo, B., Wolpert, R. (1999). Integrated likelihood methods for eliminating nuisance parameters. *Statist. Sc.*, **14**, 1–28.

Bernardo, J.M., Smith, A.F.M. (2000). *Bayesian Theory.* Wiley, New York.

Bertolino, F., Racugno, W. (1992). Analysis of the linear correlation coefficient using pseudo-likelihoods. *J. It. Statist. Soc.*, **1**, 33–50.

Bertolino, F., Racugno, W. (1994). Robust Bayesian analysis of variance and the χ^2-test by using marginal likelihoods. *The Statistician*, **43**, 191–201.

Cabras, S., Racugno, W., Ventura, L. (2006). Bayesian inference on the scalar skew-normal distribution. *Proceedings of COMPSTAT 2006, 17th Symposium of IASC-ERS*, 1373–1380.

Cox, D.R., Reid, N. (1987). Orthogonal parameters and approximate conditional inference (with discussion). *J. R. Statist. Soc.* B, **49**, 1–39.

Efron, B. (1993). Bayes and likelihood calculations from confidence intervals. *Biometrika*, **80**, 3–26.

Greco, L., Racugno, W., Ventura L. (2007). Robust likelihood functions in Bayesian inference. *J. Statist. Plan. Inf.*, to appear.

Lazar, N.A. (2003). Bayesian empirical likelihood. *Biometrika*, **90**, 319–326.

Liseo, B. (1993). Elimination of nuisance parameters with reference priors. *Biometrika*, **80**, 295–304.

Monahan, J.F., Boos, D.D. (1992). Proper likelihoods for Bayesian analysis. *Biometrika*, **79**, 271–278.

Nicolau, A. (1983). Bayesian intervals with good frequentist behaviour in the presence of nuisance parameters. *J. R. Statist. Soc.* B, **55**, 377–390.

Pace, L., Salvan, A. (1997). *Principles of Statistical Inference.* World Scientific, Singapore.

Pace, L., Salvan, A., Ventura, L. (2006). Likelihood based discrimination between separate scale and regression models. *J. Statist. Plan. Inf.*, **136**, 3539–3553.

Peers, H.W. (1965). On confidence points and Bayesian probability points in the case of several parameters. *J. R. Statist. Soc.* B, **275**, 9–16.

Racugno, W., Salvan, A., Ventura, L. (2005). On the use of pseudo-likelihoods in Bayesian variable selection. *Working Papers*, n. 2005.12, University of Padova, Italy.

Raftery, A.E., Madigan, D., Volinsky, C.T. (1996). Accounting for model uncertainty in survival analysis improves predictive performance. *Bayesian Statistics*, 323–349.

Reid, N. (1995). Likelihood and Bayesian approximation methods. *Bayesian Statistics*, **5**, 351–368.

Severini, T.A. (1999). On the relationship between Bayesian and non-Bayesian elimination of nuisance parameters. *Statistica Sinica*, **9**, 713–724.

Severini, T.A. (2000). *Likelihood Methods in Statistics.* Oxford University Press

Severini, T.A. (2007). Integrated likelihood functions for non-Bayesian inference. To appear.

Schennach, S.M. (2006). Bayesian exponentially tilted empirical likelihood. *Biometrika*, to appear.

Stein, C. (1985). On the coverage probability of confidence sets based on a prior distribution. In *Sequential Methods in Statistics*, Banach center publications, **16**. Warsaw: PWN-Polish scientific publishers.

Sweeting, T.J. (1987). Discussion of the paper D.R. Cox and N. Reid. *J. R. Statist. Soc.* B, **49**, 20–21.

Tibshirani, R. (1989). Noninformative priors for one parameter of many. *Biometrika*, **76**, 604–608.

Tierney, L.J., Kadane, J.B. (1986). Accurate approximations for posterior moments and marginal densities. *J. Amer. Statist. Assoc.*, **81**, 82–86.

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

Department of Statistical Sciences
University of Padua
Italy