Working Paper Series, N. 12, July 2005

On the use of pseudo-likelihoods in Bayesian variable selection

Walter Racugno
Department of Matematхics
University of Cagliari
Italy

Alessandra Salvan and Laura Ventura
Department of Statistical Sciences
University of Padua
Italy

Abstract: In the presence of nuisance parameters, we discuss a one-parameter Bayesian analysis based on a pseudo-likelihood assuming a default prior distribution for the parameter of interest only. Although this way to proceed cannot always be considered as orthodox in the Bayesian perspective, it is of interest to evaluate whether the use of suitable pseudo-likelihoods may be proposed for Bayesian inference. Attention is focused in the context of regression models, in particular on inference about a scalar regression coefficient in various multiple regression models, i.e. scale and regression models with non-normal errors, non-linear normal heteroscedastic regression models, and log-linear models for count data with overdispersion. Some interesting conclusions emerge.

Keywords: Asymptotics; Bayesian inference; likelihood; nuisance parameter; regression models; semiparametric models.

Contents

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

On the use of pseudo-likelihoods in Bayesian variable selection

Walter Racugno
Department of Matemathics
University of Cagliari
Italy

Alessandra Salvan and Laura Ventura
Department of Statistical Sciences
University of Padua
Italy

Abstract: In the presence of nuisance parameters, we discuss a one-parameter Bayesian analysis based on a pseudo-likelihood assuming a default prior distribution for the parameter of interest only. Although this way to proceed cannot always be considered as orthodox in the Bayesian perspective, it is of interest to evaluate whether the use of suitable pseudo-likelihoods may be proposed for Bayesian inference. Attention is focused in the context of regression models, in particular on inference about a scalar regression coefficient in various multiple regression models, i.e. scale and regression models with non-normal errors, non-linear normal heteroscedastic regression models, and log-linear models for count data with overdispersion. Some interesting conclusions emerge.

Keywords: Asymptotics; Bayesian inference; likelihood; nuisance parameter; regression models; semiparametric models.

1 Introduction

Consider a statistical model indexed by a parameter θ which may be written as $\theta = (\psi, \lambda)$, where ψ is the parameter of interest and λ is a nuisance parameter. In general, both ψ and λ may be vectors, although often ψ is a scalar. In the absence of nuisance parameters, inference about ψ is easy to perform both in the likelihood and in the Bayesian approaches. However, the presence of λ often makes inference about ψ difficult and elimination of nuisance parameters is of central interest. In this case, Bayesian inference requires to assume a prior distribution $\pi(\lambda|\psi)$ for λ given ψ, and to determine the integrated likelihood (see e.g. Berger *et al.*, 1999)

$$L_I(\psi) = \int_\Lambda L(\psi, \lambda)\pi(\lambda|\psi)d\lambda \ , \qquad (1)$$

where $L(\psi, \lambda)$ denotes the complete likelihood function based on data y. Function $L_I(\psi)$ is then combined with the marginal prior for ψ to obtain its posterior distribution. This procedure for eliminating λ has some well-known optimality properties. However, the elicitation of $\pi(\lambda|\psi)$ may be difficult to state precisely both in the subjective and objective Bayesian contexts and computation of the integral in (1) can be heavy when the parameter spaces are high-dimensional. Examples, where both drawbacks can arise, are regression problems when variable selection is of interest. In this case, the parameter of interest is a regression coefficient and the nuisance parameter is given by the remaining regression coefficients and by variance parameters.

In the frequentist likelihood approach, elimination of nuisance parameters may be carried out using appropriate pseudo-likelihoods. 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. Examples of pseudo-likelihoods for a parameter of interest are the marginal, the conditional, the profile, the quasi-likelihoods, and modifications thereof. See Severini (2000, Chapters 8 and 9) for a review.

The aim of this contribution is to discuss, from a pratical point of view, in the context of regression models, a one-parameter Bayesian analysis based on a pseudo-likelihood $L_{ps}(\psi)$ assuming a default prior distribution for ψ only. This approach avoids the need for the detailed attention to the nuisance parameter as would be required in a global Bayesian approach based on the integrated likelihood (1). Although this way to proceed cannot always be considered as orthodox in the Bayesian perspective, it is of interest to evaluate whether the use of suitable pseudo-likelihoods may be proposed for Bayesian inference. However, it is natural to ask when a likelihood function other than the density which the data are assumed to be generated can be used as the likelihood portion in a Bayesian analysis. It may be noted that since the pseudo-likelihood functions considered here have first-order properties similar to a proper likelihood function, the resulting pseudo-posterior distribution is approximately normal with mean given by the maximum of $L_{ps}(\cdot)$ and variance related to minus the second derivative of the pseudo loglikelihood. Moreover, Monahan and Boss (1992) and Severini (1999) provide two criteria for evaluating whether or not an alternative likelihood is proper for Bayesian inference. Related work in this area is in Efron (1993), Bertolino and Racugno (1994), Fraser and Reid (1996), Fraser *et al.* (2003), Lazar (2003), Pace *et al.* (2005). We focus attention on inference about a scalar regression coefficient β^*, in various multiple regression models. In particular, we consider three classes of models: (i) scale and regression models with non-normal errors; (ii) non-linear normal heteroscedastic regression models; (iii) log-linear models for count data with overdispersion. While the first two situations focus on parametric statistical models, where the full likelihood function is available, the last one concerns a semi-parameteric model.

2 Pseudo-likelihoods for a parameter of interest

In frequentist likelihood inference, a pseudo-likelihood $L_{ps}(\psi)$ is a function of the data and ψ only, which shares some of the properties of a genuine likelihood function.

A notable instance is the profile likelihood

$$L_p(\psi) = L(\psi, \hat{\lambda}_\psi) \ , \qquad (2)$$

where $\hat{\lambda}_\psi$ denotes the maximum likelihood estimate of λ for fixed ψ. However, it is well-known that this solution may be unsatisfactory, particularly when the dimension of the nuisance parameter is large relative to the sample size n.

In the following, we briefly present three pseudo-likelihoods that are preferable to $L_p(\psi)$ in models (i), (ii) and (iii), respectively. All these pseudo-likelihood functions have the standard limiting behaviour and the corresponding pseudo-likelihood-type ratio statistics have the classical asymptotic distribution.

Even if the computation of these pseudo-likelihoods may seem rather complicated, it can be implemented in modern statistical environments, such as `R`. Moreover, a recently developed library for higher-order asymptotics (HOA, 2000) implements higher-order solutions, including some pseudo-likelihoods.

2.1 The marginal likelihood

An approach to constructing a pseudo-likelihood function for ψ is to base a likelihood function on the distribution of a statistic a with marginal distribution depending on ψ only. Clearly, the resulting marginal likelihood function is a genuine likelihood function for ψ. More precisely, suppose there exists a statistic a such that the density of the data y may be written as

$$p(y; \psi, \lambda) = p(a; \psi) p(y|a; \psi, \lambda) \ .$$

Inference about ψ may be based on the marginal likelihood function given by

$$L_m(\psi) = L_m(\psi; a) = p(a; \psi) \ ,$$

provided that $p(y|a; \psi, \lambda)$ does not contain useful information about ψ.

A marginal likelihood always exists when the parameter of interest is the index parameter of a composite transformation model, with λ having the role of the group parameter. In this case the maximal invariant statistic plays the role of a. Then, the marginal likelihood is (see e.g. Barndorff-Nielsen and Cox, 1994, Section 2.8)

$$L_m(\psi) = L_m(\psi; a) = \int_\Lambda L(\psi, \lambda) d\mu(\lambda) \, , \qquad (3)$$

where $\mu(\lambda)$ is the right-invariant Haar measure on the group of transformations, whose action on the parameter space leaves ψ unchanged. Note that $L_m(\psi)$ may be interpreted as the integrated likelihood with respect to the prior $\mu(\lambda)$.

Composite transformation models include scale and regression models (i) as a special case. The use of a marginal likelihood of form (3) for model selection between separate scale and regression models has been discussed in Pace *et al.* (2005).

2.2 The modified profile likelihood

Function $L_m(\psi)$ can only be applied in special circumstances. For instance, it cannot be computed for non-linear heteroscedastic regression models (*ii*). There exist other pseudo-likelihood functions that are generally applicable and which lead to accurate inferences on ψ when the dimension of the nuisance parameter is large with respect to the sample size. Given the profile likelihood function (2), a modified profile likelihood $L_{mp}(\psi)$ for ψ has the general form

$$L_{mp}(\psi) = L_p(\psi)M(\psi) \ , \tag{4}$$

where $M(\psi)$ is a suitable correction factor; see, for instance, Barndorff-Nielsen and Cox (1994, Chapter 8), Severini (2000, Chapter 9) and Fraser (2003). All the proposed adjustments are equivalent to second order and share the common feature of reducing the score bias to $O(n^{-1})$. It can be shown that the modified profile likelihood (4) is a higher-order approximation to a conditional likelihood or marginal likelihood, when either exists. For more general justifications, see Severini (1998) and Pace and Salvan (2005).

Severini (1999) discusses the interpretation of a version of $L_{mp}(\psi)$ as an integrated likelihood. Moreover, if $\pi(\theta)$ is a prior density for θ, the use of analytical approximations for integrals gives the following expansion for the posterior distribution

$$\pi(\psi; y) \propto L_{mp}(\psi)\pi(\psi, \hat{\lambda}_\psi)\{1 + O(n^{-1})\} \ , \tag{5}$$

see for instance Reid (1995, 2003).

2.3 The quasi-profile likelihood

In many situations inference based on estimating equations is preferable to a fully parametric specification. This is the case for example in the context of generalized linear models with overdispersion or with random effects. For inference about a scalar parameter of interest, extending (4) in the estimating functions setting, a modified quasi-profile likelihood can be defined, with the standard limit behaviour.

Let $G_\psi = G_\psi(y;\theta)$ and $G_\lambda = G_\lambda(y;\theta)$ be unbiased estimating functions corresponding to ψ and λ, respectively, and let $\tilde{\lambda}_\psi$ be the estimator derived from G_λ when ψ is considered as known, i.e. the solution of $G_\lambda(y;\psi,\tilde{\lambda}_\psi) = 0$. In this setting, for inference about ψ a modified quasi-profile loglikelihood can be considered (see Bellio *et al.*, 2005),

$$\ell_{mqp}(\psi) = \int_{\psi_0}^{\psi} w(t, \tilde{\lambda}_t) \left(G_\psi(y; t, \tilde{\lambda}_t) - m(t, \tilde{\lambda}_t) \right) \, dt \ , \tag{6}$$

where ψ_0 is an arbitrary constant. The scale adjustement is necessary to obtain modified quasi-profile likelihood-type tests based on (6) with the classical asymptotic distribution, while the additive adjustment is a first-order bias correction of the profile estimating equation for ψ. Indeed, $G_\psi(y;\psi,\tilde{\lambda}_\psi)$ is not unbiased and such a bias can lead to poor inference on the parameter of interest.

3 Examples

With an objective prior $\pi(\beta^*)$ for the interest regression coefficient β^* only, a one-parameter Bayesian inference based on a pseudo-likelihood $L_{ps}(\cdot)$ may be carried out with

$$\pi_{ps}(\beta^*; y) \propto \pi(\beta^*) L_{ps}(\beta^*) \ , \tag{7}$$

without any integration of form (1).

There are three main interesting cases in which (7) may be useful: (a) we are able to assign a prior distribution for the parameter of interest, but not for some or all the nuisance parameters; (b) it is difficult to explicit the full likelihood; (c) the parameter space of the nuisance parameter is high dimensional. In this section we present three examples indicanting how the pseudo-posterior distribution (7) proceeds for variable selection.

Example1: Non-normal linear regression models
Let us consider a scale and regression model

$$y = X\beta + \sigma\varepsilon \ , \tag{8}$$

where X is a $n \times p$ matrix of rank p, $\beta \in \mathbb{R}^p$ is a vector of regression coefficients, $\sigma > 0$ is a scale parameter, and ε is a vector of errors, with components $(\varepsilon_1, \ldots, \varepsilon_n)$ independent and identically distributed according to a density $p_\kappa(\cdot)$ not necessarily normal. Popular assumptions for $p_\kappa(\cdot)$ include the normal, Student's t, extreme value, logistic, skew-normal and Cauchy distributions.

Let β^* be a scalar component of β. If $\psi = \beta^*/\sigma$ is considered as the parameter of interest, the model is a composite transformation family with index parameter ψ. The maximal invariant statistic a can be written as $a = (a_1, \ldots, a_n)$, where $a_i = (y_i - x_i^{\top(-*)} \hat{\beta}^{(-*)})/\hat{\sigma}$, $\beta^{(-*)}$ is the vector β without its β^* component, $(\hat{\beta}^{(-*)}, \hat{\sigma})$ is the maximum likelihood estimate of $(\beta^{(-*)}, \sigma)$ and $x_i^{\top(-*)}$ is the ith row of X without the component associated to β^*, $i = 1, \ldots, n$. Then, the marginal likelihood for ψ based on a (see Fraser, 1979, Section 6.2) is

$$L_m(\psi) = \int_{\mathbb{R}^+} \int_{\mathbb{R}^{p-1}} \sigma^{-1} L(\beta, \sigma) \, d\beta^{(-*)} \, d\sigma \ , \tag{9}$$

where β^* is expressed as $\psi\sigma$. The marginal likelihood (9) has a Bayesian interpretation: it is an integrated likelihood for ψ using the Jeffreys prior density is $\pi(\beta, \sigma) \propto 1/\sigma$. Moreover, some calculations show that the posterior distribution derived from (9), i.e. $\pi_m(\psi; a) \propto \pi(\psi) L_m(\psi)$, corresponds exactly to the posterior distribution $\pi(\psi; y)$ derived from the integrated likelihood (1) based on the complete distribution of the data y and on the conditional prior $1/\sigma$. Indeed, with $\lambda = (\beta^{(-*)}, \sigma)$ and data y represented by $(\hat{\beta}^{(-*)}, \hat{\sigma}, a)$, we have

$$\int p(y|a; \psi, \lambda) \pi(\lambda|\psi) d\lambda = \hat{\sigma}^{-1} \ ,$$

so that

$$
\begin{aligned}
\pi(\psi;y) &\propto \pi(\psi)\int p(y;\psi,\lambda)\pi(\lambda|\psi)d\lambda \\
&= \pi(\psi)p(a;\psi)\int p(y|a;\psi,\lambda)\pi(\lambda|\psi)d\lambda \\
&\propto \pi(\psi)p(a;\psi) \propto \pi_m(\psi;a)\ .
\end{aligned}
$$

In view of this, for non-normal regression models there is no loss of information about ψ in using a posterior distribution $\pi_m(\psi;a)$ derived from the marginal likelihood instead of the proper posterior distribution $\pi(\psi;y)$. This seems to be a general property in composite transformation models (see also Reid, 1995, Section 5).

Both the marginal and the integrated likelihoods do not have an explicit analytic solution because the integrals involved must be solved numerically. Note that the marginal likelihood (9) can be computed in a closed form only for normal linear models (Fraser, 1979). However, higher-order likelihood theory applies rather naturally to this context and allows us to by-pass higher-order numerical integration. In particular, an approximation to the marginal likelihood function (and similarly to an integrated likelihood) is of the form

$$
L_m(\psi) = L_{mp}(\psi)(1+O(n^{-1}))\ , \tag{10}
$$

where the correction term $M(\psi)$ of the modified profile likelihood $L_{mp}(\psi)$ is very simple to compute (see, e.g., Pace *et al.*, 2005). This function is available through the `marg` library section of the `S-Plus` and `R` library `HOA` (HOA, 2000).

Example 2: Nonlinear heteroscedastic regression models.
We consider the general nonlinear heteroscedastic model (see e.g. Seber and Wild, 1989)

$$
y_{ij} = \mu(x_i;\beta) + V(x_i;\beta,\rho)\,\varepsilon_{ij}\ , \quad i=1,\ldots,q\ , \quad j=1,\ldots,m_i\ , \tag{11}
$$

where q is the number of design points, m_i is the number of replicates at design point x_i, the variable y_{ij} represents the response of the jth experimental unit in the ith group, and the errors ε_{ij} are independent $N(0,1)$ random variables. The mean response is given by the nonlinear function $\mu(x_i;\beta)$, called the *mean function*, that depends on an unknown p-dimensional regression coefficient β. The definition of the model is completed by specifying the *variance function* $V(x_i;\beta,\rho)$, where ρ is a vector of variance parameters.

Inference focuses on a scalar component of the unknown regression parameter. As outlined in Bellio and Brazzale (1999), first-order likelihood methods can be highly inaccurate, especially when the variance function $V(\cdot)$ depends on β. On the contrary, in such a situation, the modified profile likelihood of form (4) can give accurate inferences. For nonlinear heteroscedastic models this function is implemented in the `nlreg` library section of the `S-Plus` and `R` library `HOA` (HOA, 2000). The expression for the correction factor $M(\psi)$ of the modified profile likelihood in (4) is in this case simple to compute.

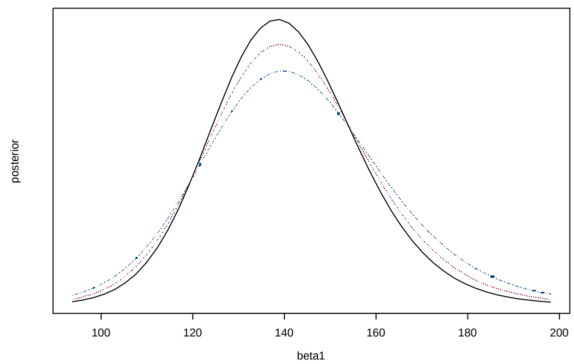


Figure 1: Posterior distributions for the Weed data: $\pi(\beta_1; y)$ (dashed), $\pi_{mp}(\beta_1; y)$ (solid), $\pi_p(\beta_1; y)$ (dotted).

To show a first application of (7) based on (4) we consider a real data set. In bioassays, a typical experiment is devoted to establish a dose-response relation by measuring the growth of a plant corresponding to different levels of doses of herbicide. The Weed Data (Seiden *et al.*, 1998) concern the callus area of a tissue culture of *Brassica napus* corresponding to different doses of a sulfonylurea herbicide. The experiment consists of $q = 8$ doses of and 5 replications at each level. The mean function is the four parameter log-logistic function $\mu(x; \beta) = \log\left(\beta_1 + (\beta_2 - \beta_1)/(1 + (x/\beta_4)^{\beta_3})\right)$, and the variance function is $V(x; \gamma, \sigma) = \sigma(1 + x^{\gamma})^2$. Figure 1 shows the posterior distributions for β_1 combining independent non-informative priors with $L_I(\beta_1)$ (Laplace approximation), $L_p(\beta_1)$ and $L_{mp}(\beta_1)$, as discussed in Bellio and Brazzale (1999). All the posteriors give approximately the same posterior mode for β_1: 137.83 from $\pi_{mp}(\beta_1; y)$, 138.92 from $\pi_p(\beta_1; y)$ and 136.83 from $\pi(\beta_1; y)$.

Let us consider a second example based on a real data set given in Belanger et al. (1996). The data concern a run of a radioimmunoassay to estimate the concentrations of a drug in samples of procine serum. The experiment consists of 16 observations made at 8 different drug levels with 2 replications at each level. The mean function is the four parameter logistic function $\mu(x; \beta) = \beta_1 + (\beta_2 - \beta_1)/(1+(x/\beta_4)^{\beta_3})$, and the variance function is $V(x; \beta, \gamma, \sigma) = \exp(\sigma)\mu(x; \beta)^{\gamma}$, where $\rho = (\sigma, \gamma)$, with $\sigma, \gamma \in \mathbb{R}$. Figure 2 shows the posterior distributions for β_4 combining independent non-informative priors with $L_I(\beta_4)$ (Laplace approximation), $L_p(\beta_4)$ and $L_{mp}(\beta_4)$. All the posteriors give approximately the same posterior mode for β_4: 334.95 from $\pi_{mp}(\beta_4; y)$, 335.46 from $\pi_p(\beta_4; y)$ and 3332.93 from $\pi(\beta_4; y)$.

Finally, we consider a simpler non-linear model with mean function $\mu(x; \beta) = \beta_0(1 - \exp(-\beta_1 x))$ and variance function is $V(x; \sigma) = \sigma$. Data have been generated with $n = 10, 20, 50$ and with $\beta_0 = 4.3$, $\beta_1 = 0.2$ and $\sigma = 0.27$. Figure 3 shows the posterior distributions for β_0 combining independent non-informative priors with $L_I(\beta_0)$ (Laplace approximation), $L_p(\beta_0)$ and $L_{mp}(\beta_0)$. Note that all the posterior distributions give approximately the same posterior mode for β_0, for all the values of n considered.

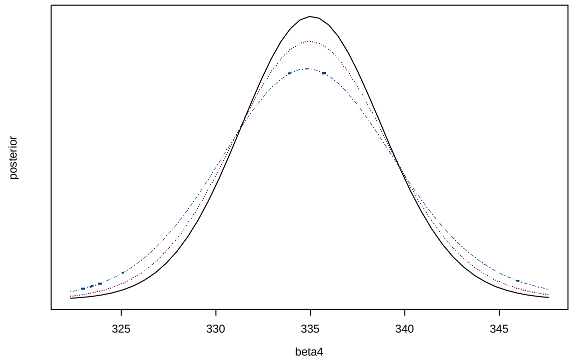
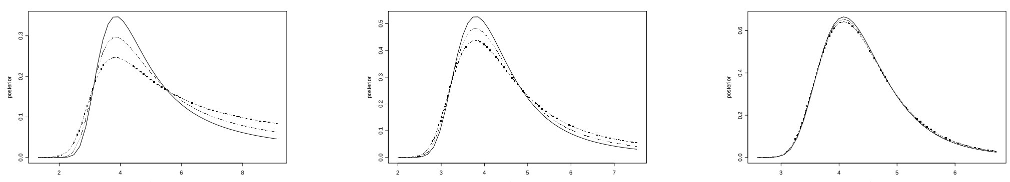


Figure 2: Posterior distributions for the Ria data: $\pi(\beta_4; y)$ (dashed), $\pi_{mp}(\beta_4; y)$ (solid), $\pi_p(\beta_4; y)$ (dotted).



(a) (b) (c)

Figure 3: Posterior distributions for simulated data with $n = 10$ (a), $n = 20$ (b), $n = 50$ (c): $\pi(\beta_0; y)$ (dashed), $\pi_{mp}(\beta_0; y)$ (solid) and $\pi_p(\beta_0; y)$ (dotted).

Example 3: Models for count data with overdispersion.
Let us consider a log-linear model for count data. The responses y_i are realizations of independent random variables with mean $\mu_i = \exp(x_i^\mathsf{T} \beta)$, $\beta \in \mathbb{R}^p, p \geq 1$, $i = 1, \ldots, n$. In many applications with discrete data, overdispersion can be encountered. Let us focus on the situation where the variance is assumed to be a quadratic function of μ_i of the form $V(x_i; \beta, \alpha) = \mu_i(1 + \alpha\mu_i)$, $\alpha \geq 0$. In this case, the estimating function for β is the score function from the Poisson likelihood $G_\beta(y; \beta) = \sum_{i=1}^n (y_i - \mu_i)\, x_i^\mathsf{T}$, which still provides an unbiased estimating equation. An estimating function for α is $G_\alpha(y; \beta, \alpha) = \sum_{i=1}^n \frac{(y_i - \mu_i)^2}{\mu_i(1+\alpha\mu_i)} - (n - p)$, as shown in Lawless (1987).

The use of $L_{mqp}(\psi)$ allows us to quantify the consequences of overdispersion for inference on a regression coefficient (see Bellio *et al.*, 2005). We apply this procedure to the Ames Salmonella data, already analysed by Lawless (1987). The response is the number of revertant colonies observed on a plate, and covariates are based on the dose level of quinoline on the plate (x). We assume the following model for the response $\log(\mu_i) = \beta_0 + \beta_1 x_i + \beta_2 \log(x_i + 10)$, $i = 1, \ldots, 18$, where the interest lies on $\psi = \beta_2$. Figure 4 gives the posterior distribution for β_2 obtained, with independent non-informative prior distributions, using the modified quasi-profile likelihood. In such a situation, a comparison with an integrated likelihood of form (1) is not easy

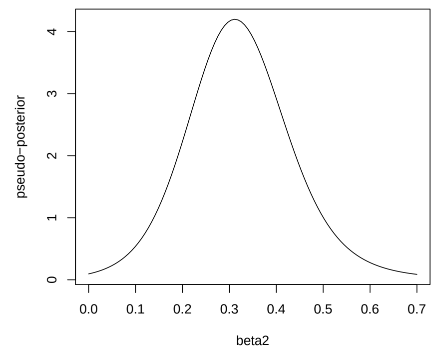


Figure 4: Posterior distribution for the Ames Salmonella Data using the modified quasi-profile likelihood.

to perform, since the complete likelihood for (β, α) is not available.

4 Discussion

From the examples discussed in the previous section, some interesting features emerge.

(1) For non-normal regression models, there is no loss of information about ψ in using a posterior distribution derived from the marginal posterior $\pi_m(\psi; a)$ instead of the proper posterior distribution $\pi(\psi; y)$ based on a noninformative prior.

(2) In the context of nonlinear heteroscedastic models, it emerges that all the posterior distributions give approximately the same posterior mode. On the other hand, the posterior based on $L_{mp}(\psi)$ appears to have lighter tails. This point needs further investigation. For example, more accurate computation of the integrated likelihood should be considered.

(3) Suitable pseudo-likelihoods may be very useful to deal with semi-parametric models, i.e. when the full likelihood is not available..

From the Bayesian point of view, the examples presented in this contribution indicate that non-Bayesian methods for eliminating nuisance parameters can be usefully incorporated into a one-parameter Bayesian analysis.

References

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

Belanger, B.A., Davidian, M., Giltinian, D.M. (1996), The effect of variance function estimation on nonlinear calibration inference in immunoassay data, *Biometrics*, **52**, 158–175.

Bellio, R., Brazzale, A.R. (1999), On the implementation of approximate conditional inference, *Statistica Applicata*, **11**, 251–271.

Bellio, R., Greco, L., Ventura, L. (2005), Adjusted quasi-profile likelihoods from estimating equations, *Working Paper*, n. 2005.10, Department od Statistics, University of Padova, Italy.

Berger, J.O., Liseo, B., Wolpert, R. (1999), Integrated likelihood methods for eliminating nuisance parameters, *Statistical Science*, **14**, 1-28.

Berger, J., Pericchi, L., Varshavsky, J. (1998), Bayes factors and marginal distributions in invariant situations, *Sankhya*, **60**, 307-321.

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

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

Fraser, D.A.S. (1979), *Inference and Linear Models*, McGraw-Hill.

Fraser, D.A.S. (2003), Likelihood for component parameters, *Biometrika*, **90**, 327–339.

Fraser, D.A.S., Reid, N. (1996), Bayes posteiors for scalar interest parameters, *Bayesian Statistics*, **5**, 581–585.

Fraser, D.A.S., Reid, N., and Yi, G.Y. (2003), Direct Bayes for interest parameters, *Bayesian Statistics*, **7**, to appear.

HOA (2000), Higher-Order Asymptotics. A library of S-Plus functions for higher-order asymptotic inference, http://www.stat.unipd.it/LIKASY.

Lawless, J.F. (1987), Negative Binomial and mixed Poisson regression, *Can. J. Statist.*, **15**, 9-225.

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

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

Pace, L., Salvan, A. (2005), A new motivation for adjustments of the profile likelihood, *J. Statist. Plan. Inf.*, to appear.

Pace, L., Salvan, A., Ventura, L. (2005), Likelihood based discrimination between separate scale and regression models, *J. Statist. Plan. Inference*, to appear.

Reid, N. (1995), Likelihood and Bayesian approximation methods, In *Bayesian Statistics 5*, Oxford University Press, 1–18.

Reid, N. (2003), Asymptotics and the theory of inference, *Ann. Statist.*, to appear.

Seber, G.A.F., Wild, C.J. (1989), *Nonlinear Regression*, New York: Wiley.

Seiden, P., Kappel, D., Streibig, J.C. (1998), Response of Brassica napus L. tissue culture to metsulfuron methyl and chlorsulfuron, *Weed Research*, **38**, 221–228.

Severini, T.A. (1998), Likelihood functions for inference in the presence of a nuisance parameter, *Biometrika*, **85**, 507-522.

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*, University Press, Oxford.