Working Paper Series, N. 11, September 2011

Department of Statistical Sciences
University of Padua
Italy

Objective Bayesian higher-order asymptotics in models with nuisance parameters

Laura Ventura
Department of Statistical Sciences
University of Padua
Italy

Nicola Sartori
Department of Statistical Sciences
University of Padua
Italy

Walter Racugno
Department of Mathematics
University of Cagliari
Italy

Abstract: We discuss higher-order approximations to the marginal posterior distribution for a scalar parameter of interest in the presence of nuisance parameters. These higher-order approximations are obtained using a suitable matching prior. The proposed procedure has several advantages since it does not require the elicitation on the nuisance parameter, neither numerical integration or MCMC simulation, and it enables us to perform accurate Bayesian inference even for very small sample sizes. Numerical illustrations are given for models of practical interest, such as linear non-normal models and logistic regression. We also illustrate how the proposed accurate approximation can routinely be applied in practice using results from likelihood asymptotics and the `R` package bundle `hoa`.

Keywords: Asymptotic expansion, Directed and modified directed likelihood, Matching prior, Modified profile likelihood, Tail area probability.

UNIVERSITÀ DEGLI STUDI DI PADOVA
DIPARTIMENTO DI SCIENZE STATISTICHE

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 4168
ventura@stat.unipd.it
http://www.stat.unipd.it/~name

Objective Bayesian higher-order asymptotics in models with nuisance parameters

Laura Ventura
Department of Statistical Sciences
University of Padua
Italy

Nicola Sartori
Department of Statistical Sciences
University of Padua
Italy

Walter Racugno
Department of Mathematics
University of Cagliari
Italy

Abstract: We discuss higher-order approximations to the marginal posterior distribution for a scalar parameter of interest in the presence of nuisance parameters. These higher-order approximations are obtained using a suitable matching prior. The proposed procedure has several advantages since it does not require the elicitation on the nuisance parameter, neither numerical integration or MCMC simulation, and it enables us to perform accurate Bayesian inference even for very small sample sizes. Numerical illustrations are given for models of practical interest, such as linear non-normal models and logistic regression. We also illustrate how the proposed accurate approximation can routinely be applied in practice using results from likelihood asymptotics and the `R` package bundle `hoa`.

Keywords: Asymptotic expansion, Directed and modified directed likelihood, Matching prior, Modified profile likelihood, Tail area probability.

1 Introduction

Let us consider a model with a scalar parameter of interest ψ, a d-dimensional nuisance parameter λ and likelihood function $L(\psi, \lambda) = L(\psi, \lambda; y)$, where $y = (y_1, \ldots, y_n)$ is a random sample of size n. Given a prior $\pi(\psi, \lambda)$ over the entire parameter, under general regularity conditions Bayesian inference about ψ is based on the marginal posterior distribution

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

For objective Bayesian inference, when agreement between Bayesian and non-Bayesian inference is of interest, the class of matching priors can be considered (see, for instance, Datta and Mukerjee, 2004, for a comphrensive review). In particular, in the presence of nuisance parameters, a suitable solution is discussed by Tibshirani (1989), which suggests a prior such that the resulting marginal posterior intervals have accurate frequentist coverage. However, this approach works only with an orthogonal parameterization and requires the computation of a multidimensional integral in (1), which can be heavy when the nuisance parameter is high-dimensional. To avoid these limitations, using results in Ventura *et al.* (2009), it is possible to write the marginal posterior distribution (1), based on the Tibshirani's prior, in the original non-orthogonal parameterization as

$$\pi_m(\psi|y) \propto \pi_{mp}(\psi)\, L_{mp}(\psi) \ , \tag{2}$$

where $L_{mp}(\psi)$ denotes a modified profile likelihood (see e.g. Severini, 2000, Chap. 9) and $\pi_{mp}(\psi)$ is the corresponding matching prior on the parameter of interest only (see also Ventura and Racugno, 2011).

In this paper we discuss higher-order asymptotic inference based on $\pi_m(\psi|y)$, starting from (2). In particular, we derive an explicit higher-order approximation for $\pi_m(\psi|y)$, in terms of a higher-order pivotal quantity. The higher-order approximations enables one to perform accurate Bayesian inference on ψ, even for small sample sizes. The proposed procedure has several advantages since it does not require the elicitation on the nuisance parameter, neither numerical integration or MCMC simulation. A remarkable further advantage in the use of such approximation is that its expression automatically includes the matching prior, without requiring its explicit computation.

Examples in the context of models of practical interest, such as linear non-normal models and logistic regression, are discussed. Moreover, we also show how the proposed accurate approximation can routinely be applied in practice using results from likelihood asymptotics and the `R` package bundle `hoa` (Brazzale *et al.*, 2007).

The outline of the paper is as follows. Background theory is briefly reviewed in Section 2. In Section 3 we discuss higher-order asymptotics for $\pi_m(\psi|y)$. Examples are illustrated in Section 4. Some final remarks conclude the paper.

2 Background theory

Assume that the likelihood is given in an orthogonal parameterization, denoted by (ψ, ϕ). In this parameterization, Tibshirani (1989) and Nicolau (1993) show that a class of matching priors for (ψ, ϕ), i.e. priors that ensure approximate frequentist validity of posterior credible sets, is given by

$$\pi_T(\psi, \phi) \propto i_{\psi\psi}(\psi, \phi)^{1/2}\ g(\phi) \ , \tag{3}$$

where $g(\phi)$ is an arbitrary function and $i_{\psi\psi}(\psi, \phi)$ is the (ψ, ψ) component of the Fisher information matrix $i(\psi, \phi)$ based on $L(\psi, \phi)$. The computation of the marginal posterior distribution (1) with this prior requires to fix the arbitrary function $g(\phi)$

in (3) and possible cumbersome numerical integration. These drawbacks can be avoided using results in Ventura *et al.* (2009, Appendix B). In particular, using the original non-orthogonal parameterization (ψ, λ), the marginal posterior distribution (1) based on the prior $\pi_T(\psi, \phi)$ can be written as in (2), with

$$\pi_{mp}(\psi) \propto i_{\psi\psi.\lambda}(\psi, \hat{\lambda}_\psi)^{1/2} \tag{4}$$

matching prior for ψ only (see Ventura *et al.*, 2009, Ventura and Racugno, 2011), and

$$L_{mp}(\psi) = L_p(\psi)\, M(\psi) \tag{5}$$

modified profile likelihood for ψ (see, e.g., Barndorff-Nielsen, 1983, Barndorff-Nielsen and Cox, 1994, Chap. 8, Severini, 2000, Chap. 9). In (4), $i_{\psi\psi.\lambda}(\psi, \lambda) = i_{\psi\psi}(\psi, \lambda) - i_{\psi\lambda}(\psi, \lambda) i_{\lambda\lambda}(\psi, \lambda)^{-1} i_{\lambda\psi}(\psi, \lambda)$ is the partial information, with $i_{\psi\psi}(\psi, \lambda)$, $i_{\psi\lambda}(\psi, \lambda)$, $i_{\lambda\lambda}(\psi, \lambda)$, and $i_{\lambda\psi}(\psi, \lambda)$ blocks of the expected Fisher information $i(\psi, \lambda)$ from $L(\psi, \lambda)$, and $\hat{\lambda}_\psi$ is the Maximum Likelihood Estimator (MLE) of λ for fixed ψ. In (5), $L_p(\psi) = L(\psi, \hat{\lambda}_\psi)$ is the profile likelihood and $M(\psi)$ is a suitably defined correction term. For instance, the modified profile likelihood of Barndorff-Nielsen (1983) uses

$$M(\psi) = \frac{|j_{\lambda\lambda}(\psi, \hat{\lambda}_\psi)|^{1/2} |j_{\lambda\lambda}(\hat{\psi}, \hat{\lambda})|^{1/2}}{|\ell_{\lambda;\hat{\lambda}}(\psi, \hat{\lambda}_\psi)|}\ , \tag{6}$$

where $j_{\lambda\lambda}(\psi, \lambda)$ is the (λ, λ)-block of the observed Fisher information $j(\psi, \lambda)$ from $\ell(\psi, \lambda) = \log L(\psi, \lambda)$, $\ell_{\lambda;\hat{\lambda}}(\psi, \lambda) = \partial \ell(\psi, \lambda)/(\partial\lambda\partial\hat{\lambda}^\top)$ is a sample space derivative, and $(\hat{\psi}, \hat{\lambda})$ is the MLE of (ψ, λ). Other expressions for $M(\psi)$ are discussed in Severini (2000, Chap. 9); see also Pace and Salvan (2006).

A quantity closely related to the modified profile likelihood is the modified directed likelihood

$$r_p^*(\psi) = r_p(\psi) + \frac{1}{r_p(\psi)} \log \frac{q(\psi)}{r_p(\psi)}\ , \tag{7}$$

where $r_p(\psi) = \mathrm{sign}(\hat{\psi} - \psi)\sqrt{2(\ell_p(\hat{\psi}) - \ell_p(\psi))}$ is the directed profile likelihood, with $\ell_p(\psi) = \log L_p(\psi)$, and $q(\psi)$ is a suitable quantity; see, e.g., Severini (2000, Chap. 7) for a review. For instance, Barndorff-Nielsen and Chamberlin (1994) use

$$q(\psi) = \frac{\ell_p'(\psi)}{j_p(\hat{\psi})^{1/2}} \frac{i_{\psi\psi.\lambda}(\hat{\psi}, \hat{\lambda})^{1/2}}{i_{\psi\psi.\lambda}(\psi, \hat{\lambda}_\psi)^{1/2}} \frac{|\ell_{\lambda;\hat{\lambda}}(\psi, \hat{\lambda}_\psi)|}{|j_{\lambda\lambda}(\hat{\psi}, \hat{\lambda})|^{1/2} |j_{\lambda\lambda}(\psi, \hat{\lambda}_\psi)|^{1/2}}\ , \tag{8}$$

with $\ell_p'(\psi) = \partial \ell_p(\psi)/\partial\psi$ and $j_p(\psi) = -\partial \ell_p'(\psi)/\partial\psi$ denoting the profile score function and the profile observed information, respectively. The modified directed likelihood (7) is well-known in the non-Bayesian framework as a higher-order pivotal quantity, with standard normal null distribution with third-order accuracy. Therefore, a confidence interval for ψ with approximate level $(1-\alpha)$ is

$$\{\psi : |r_p^*(\psi)| \leq z_{1-\alpha/2}\}\ , \tag{9}$$

where $z_{1-\alpha/2}$ is the $(1-\alpha/2)$-quantile of the standard normal distribution (see, e.g., Barndorff-Nielsen and Cox, 1994, and Severini, 2000, Chap. 7). In the next section, we show how $r_p^*(\psi)$ plays also a central role in the higher-order approximation for $\pi_m(\psi|y)$.

3 Higher-order asymptotics for $\pi_m(\psi|y)$

Let us focus on the posterior distribution (2), which can be written as

$$\pi_m(\psi|y) \propto \exp\left\{-\frac{1}{2}r_{mp}(\psi)^2 + \log \pi_{mp}(\psi)\right\} \ , \tag{10}$$

where

$$r_{mp}(\psi) = \mathrm{sgn}(\hat{\psi}_{mp} - \psi)[2(\ell_{mp}(\hat{\psi}_{mp}) - \ell_{mp}(\psi))]^{1/2} \ ,$$

$\hat{\psi}_{mp}$ is the maximizer of $\ell_{mp}(\psi) = \log L_{mp}(\psi)$ and $j_{mp}(\psi) = -\partial^2 \ell_{mp}(\psi)/\partial\psi^2$. Starting from (10), in Section 3.1 we prove the following approximation

$$\pi_m(\psi|y) \doteq \exp\left(-\frac{1}{2}\, r_p^*(\psi)^2\right)\left|\frac{s_p(\psi)}{r_p(\psi)}\right| \ ,$$

where $s_p(\psi) = \ell_p'(\psi)/j_p(\hat{\psi})^{1/2}$ is the profile score statistic, and the symbol "$\doteq$" means second-order asymptotic equivalence. Then, in Section 3.2 we discuss the related tail area approximation

$$\int_{-\infty}^{\psi_0} \pi_m(\psi|y)\, d\psi \ \equiv \ \Phi\left(r_p^*(\psi_0)\right) \ ,$$

where the symbol "$\equiv$" means third-order asymptotic equivalence, and $\Phi(\cdot)$ is the standard normal distribution function. Note that the prior $\pi_{mp}(\psi)$ is also a strong matching prior (Fraser and Reid, 2002, Ventura and Racugno, 2011), in the sense that a frequentist p-value coincides with a Bayesian posterior survivor probability to a high degree of approximation. In general, a strong matching prior for ψ only, which guarantees an equivalence between a frequentist p-value and a Bayesian tail area probability, can be expressed as $\pi^*(\psi) \propto r_p(\psi)/(M(\psi)q(\psi))$, where $M(\psi)$ is the correction term of the profile likelihood and $q(\psi)$ is the quantity involved in expression (7).

3.1 Approximation for $\pi_m(\psi|y)$

Consider the strong matching prior $\pi^*(\psi) \propto r_p(\psi)/(M(\psi)q(\psi))$. Then, using results in Sartori *et al.* (1999), we have

$$\begin{aligned}
\log(L_{mp}(\psi)\,\pi^*(\psi)) &= c - \frac{1}{2}(r_{mp}(\psi))^2 + \log \pi^*(\psi) \\
&= c - \frac{1}{2}(r_p(\psi))^2 - r_p(\psi)\,\mathrm{NP} + \log \pi^*(\psi) \\
&= c - \frac{1}{2}(r_p(\psi))^2 - r_p(\psi)\left[\mathrm{NP} + \frac{1}{r_p(\psi)}\log\frac{1}{\pi^*(\psi)}\right] \\
&= c - \frac{1}{2}(r_p(\psi))^2 - r_p(\psi)\,[\mathrm{NP} + \mathrm{INF}] \ ,
\end{aligned}$$

where NP$= -r_p(\psi)\log M(\psi)$ and INF$= r_p(\psi)^{-1}\log(q(\psi)M(\psi)/r_p(\psi))$ are known as the nuisance parameters and the information adjustments, respectively, in the modified directed likelihood decomposition (see, e.g., Barndorff-Nielsen and Cox, 1994, Sect. 6.6.4). Then,

$$\begin{aligned}\log(L_{mp}(\psi)\,\pi^*(\psi)) &= -\frac{1}{2}(r_p(\psi)+\mathrm{NP}+\mathrm{INF})^2 + O_p(n^{-1})\\ &= -\frac{1}{2}r_p^*(\psi)^2 + O_p(n^{-1})\ . \end{aligned} \tag{11}$$

Using (6) and (8) we have $\pi^*(\psi) \propto \pi_{mp}(\psi)\, r_p(\psi)/s_p(\psi)$, and therefore we obtain

$$\log \pi_m(\psi|y) = -\frac{1}{2}r_p^*(\psi)^2 + \log\frac{r_p(\psi)}{s_p(\psi)} + O_p(n^{-1})\ , \tag{12}$$

and thus

$$\pi_m(\psi|y) \,\dot{\propto}\, \exp\left(-\frac{1}{2}r_p^*(\psi)^2\right)\left|\frac{r_p(\psi)}{s_p(\psi)}\right|\ . \tag{13}$$

A remarkable advantage of this approximation is that its expression automatically includes the matching prior, without requiring its explicit computation.

Note that in (13) the modified directed likelihood (7) may be replaced by the modified directed likelihood of Barndorff-Nielsen (1991) or by the adjusted directed likelihoods discussed in Barndorff-Nielsen and Chamberlin (1994). Indeed, all these versions of the directed profile likelihood statistic are closely related to each other, in the sense that they are equivalent to order $O_p(n^{-1})$ (see Barndorff-Nielsen and Chamberlin, 1994, Sect. 5).

3.2 Tail area approximation

Accurate tail probabilities are easily computable by direct integration of (10). In particular, using results in Ventura and Racugno (2011), it can be shown that

$$\int_{-\infty}^{\psi_0} \pi_m(\psi|y)\, d\psi \;\equiv\; \Phi\left(r_p^*(\psi_0)\right) \tag{14}$$

where $r_p^*(\psi)$ is the modified directed likelihood given in (7), with formula (8) for $q(\psi)$.

In practice, using (14), an asymptotic Highest Posterior Density credible set (HPD) for ψ can be computed as $\{\psi : |r_p^*(\psi)| \leq z_{1-\alpha/2}\}$, i.e. as in (9). Therefore, this HPD credible set is also an accurate likelihood-based confidence interval for ψ with approximate level $(1-\alpha)$. The strong matching prior (4) is thus also an HPD matching prior for ψ (see Ventura and Racugno, 2011). Note also that from (14) the Maximum A Posteriori estimator (MAP) of (2), i.e. the value that maximizes the posterior density, can be computed as the solution $\hat{\psi}^*$ in ψ of the estimating equation $r_p^*(\psi) = 0$, and thus it coincides with the frequentist estimator defined as the zero-level confidence interval based on r_p^* (Skovgaard, 1989). In particular, the solution of $r_p^*(\psi) = 0$ is a refinement of the maximum likelihood estimator $\hat{\psi}$ (see Pace and Salvan, 1999, Giummolé and Ventura, 2002).

As a final remark, note that HPD credible sets, as well as MAPs, are often criticized in the literature because of paradoxical behaviour due to a lack of invariance under reparametrization, and because of this undesirable feature, many authors do not recommend their use. However, since the modified directed profile likelihood (7) is parameterisation invariant, the HDP $\{\psi : |r_p^*(\psi)| \leq z_{1-\alpha/2}\}$ and the MAP $\hat{\psi}^*$ are invariant, and they can be interpreted as the generalization of the proposals by Druilhet and Marin (2007), when nuisance parameters are present in the model.

4 Examples and numerical illustrations

The aim of this Section is to provide an illustration of the accuracy of the higher-order approximation (13) and of the related tail area approximation. In particular, in applications of practical interest, we study the accuracy of (13), in comparison to the first-order approximation

$$\pi_m^a(\psi|y) \sim N(\hat{\psi}, j_p(\hat{\psi})^{-1}) \ , \tag{15}$$

when using likelihood asymptotics tools and the `R` package bundle `hoa` (Brazzale *et al.*, 2007).

Example 1: Gamma distribution. Let $(y_1, \ldots, y_n)$ be a random sample from the gamma density $p(y; \psi, \lambda) = \lambda^\psi y^{\psi-1} \exp(-\lambda y)\Gamma(\psi)^{-1}$, $y > 0$, $\psi, \lambda > 0$. We assume ψ the parameter of interest, with the scale parameter λ as nuisance. The profile loglikelihood is $\ell_p(\psi) = \psi(t - n) - n \log \Gamma(\psi) + n\psi \log(\psi/\bar{y})$, with $t = \sum \log y_i$ and $\bar{y}$ sample mean, and the modified profile loglikelihood is $\ell_{mp}(\psi) = \ell_p(\psi) - 0.5 \log \psi$, since $M(\psi) = 1/\sqrt{\psi}$. Moreover, simple calculations give $i_{\psi\psi.\lambda}(\psi, \lambda) = (n/\psi)(\psi\rho(\psi) - 1) = j_p(\psi)$, with $\rho(\psi) = (\partial^2/\partial\psi^2) \log \Gamma(\psi)$. Note that the matching prior $\pi_{mp}(\psi)$ does not depend on the nuisance parameter λ. This is a general result when ψ is the index parameter of a group model and λ is the group element, and (2) corresponds to the Laplace approximation of the marginal posterior distribution (1) based on the Chang-Eaves reference prior discussed in Datta and Ghosh (1995).

The behaviour of (13) under the gamma model is illustrated through simulation studies based on 10000 Monte Carlo trials. Table 1 gives the empirical frequentist coverages for 95% posterior HPD credible sets and for the lower and upper 0.025 quantiles from $\pi_m^a(\psi|y)$ and from $\pi_m(\psi|y)$. From Table 1 we observe that, for every n, (13) clearly improve on $\pi_m^a(\psi|y)$. Larger sample sizes would show, as one would expect, rather little differences between the results of all the procedures.

We also evaluated the finite-sample properties of the MAP of (13). The MAPs of (13) and of the first-order approximation are compared in terms of the usual centering and dispersion measures, i.e., bias and standard deviation. From Table 2 it is clear that the MAP of (13) exhibits a smaller bias than the maximum profile estimator.

Example 2: Survival times. Let us consider the survival times t_i in weeks of $n = 17$ patients with leukaemia along with their white blood cell counts x_i at the time of diagnosis, $i = 1, \ldots, 17$ (see Cox and Snell, 1981, Example U). We use these data

	$n=5$	$n=10$	$n=15$	$n=20$
$\pi_m^a(\psi\|y)$	0.900	0.928	0.937	0.942
	(0.009,0.089)	(0.010,0.061)	(0.014,0.048)	(0.014,0.044)
$\pi_m(\psi\|y)$	0.948	0.951	0.950	0.950
	(0.024,0.027)	(0.022,0.027)	(0.026,0.024)	(0.024,0.026)

Table 1: Frequentist coverage probabilities of approximate 0.95% HPD and of the lower and upper 0.025 quantiles (in brackets), under the gamma model.

	$n=5$	$n=10$	$n=15$	$n=20$
$\pi_m^a(\psi\|y)$	1.21 (4.09)	0.34 (0.72)	0.20 (0.49)	0.14 (0.37)
$\pi_m(\psi\|y)$	0.03 (1.64)	0.01 (0.51)	0.00 (0.40)	0.00 (0.31)

Table 2: Bias (and standard deviations) of the MAPs of $\pi_m^a(\psi|y)$ and of $\pi_m(\psi|y)$, under the gamma model.

to illustrate the higher-order approximation (13) for a Weibull model with shape parameter κ and scale parameter $\lambda_i = \beta_1 \exp(\beta_2(x_i - \bar{x}))$, $i = 1, \ldots, n$. To this end, we use the fact that $y_i = \log t_i$ follows a non-normal regression and scale model of the form $y_i = x_i^\intercal \beta + \sigma\varepsilon_i$, with here $x_i^\intercal \beta = \log \beta_1 + \beta_2(x_i - \bar{x})$, $\sigma = 1/\kappa$ and $\varepsilon_i \sim f(\varepsilon_i) = \exp(\varepsilon_i - e^{\varepsilon_i})$ log-Weibull random variabile, called also extreme-value or Gumbel random variable.

The proposed Bayesian procedure for inference in non-normal regression models can be easily fitted by means of the `rsm` fitting routine, provided by the `marg` section of the library `HOA`, and (13) can be obtained from the corresponding `cond` method. Figure 1 gives the posterior distribution (13) together with the posterior (15) for $\psi = \beta_2$: the corresponding asymptotic HPD credible sets for ψ with approximate level $(1-\alpha) = 0.95$ are $(-3.06, -1.43)$ and $(-2.96, -1.63)$, respectively, and the two MAPs are -2.30 and -2.28, respectively.

As in Example 1, Tables 3 and 4 give the results of a simulation study based on 10000 Monte Carlo trials with $n = 10, 17, 34$. From Table 3 we observe that, even for small n, $\pi_m(\psi|y)$ for the regression coefficient $\psi = \beta_2$ clearly improves on $\pi_m^a(\psi|y)$. From Table 4 it can be noted that the MAP of (2) is more accurate than the maximum profile estimator.

	$n=10$	$n=17$	$n=34$
$\pi_m^a(\psi\|y)$	0.892	0.915	0.932
	(0.043,0.065)	(0.041,0.044)	(0.036,0.032)
$\pi_m(\psi\|y)$	0.948	0.951	0.951
	(0.029,0.023)	(0.021,0.023)	(0.025,0.024)

Table 3: Frequentist coverage probabilities of approximate 0.95% HPD and of the lower and upper 0.025 quantiles (in brackets), under the extreme-value model.

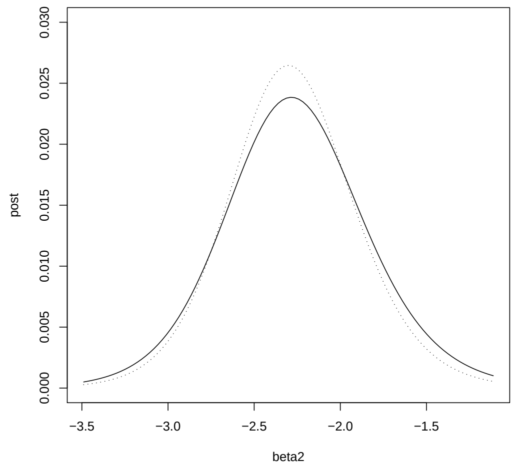

Figure 1: Posteriors $\pi_{mp}(\psi|y)$ (solid line) and first-order approximation $\pi_m^a(\psi|y)$ (dashed line) for $\psi = \beta_2$ in the Leukaemia data.

	$n = 10$	$n = 17$	$n = 34$
$\pi_m^a(\psi\|y)$	0.066 (1.53)	0.039 (0.80)	0.013 (0.64)
$\pi_m(\psi\|y)$	0.025 (1.46)	0.029 (0.77)	0.011 (0.62)

Table 4: Bias (and standard deviations) of the MAPs of $\pi_m^a(\psi|y)$ and of $\pi_m(\psi|y)$, under the extreme-value model.

Example 3: Binary data. Let us consider the `bank` dataset, which is made of four measurements on 100 genuine Swiss banknotes and 100 counterfeit ones (see Marin and Robert, 2007). The response variable y is the status of the banknote. The explanatory variables are the length of the bill x_1, the width of the left edge x_2, the width of the right edge x_3, and the bottom margin width x_4, all expressed in millimeters. A logit model is used to predict the type of banknote (i.e., to detect counterfeit banknotes) based on the four regressors x_1, x_2, x_3 and x_4.

Here, we focus our attention on inference on the scalar parameter $\psi = \beta_4$, i.e. the coefficient of x_4. The proposed Bayesian procedure for inference on ψ can be easily fitted by means of the `cond` method for `glm` objects (Brazzale *et al.*, 2007). The posterior distribution $\pi_m(\psi|y)$ is compared with the MCMC approximation (10^5 simulations) for the posteriors obtained with a flat prior and with a Zellner's uninformative G-prior (see Figure 2), discussed in Marin and Robert (2007). Moreover, Table 5 gives the empirical frequentist coverages for 95% posterior HPD credible sets and for the lower and upper 0.025 quantiles. We simulated 10^4 samples from the fitted model (MLE). For the MCMC approximation for $\pi_m^{\text{flat}}(\psi|y)$ and $\pi_m^{\text{G-prior}}(\psi|y)$ we used 10^4 replications.

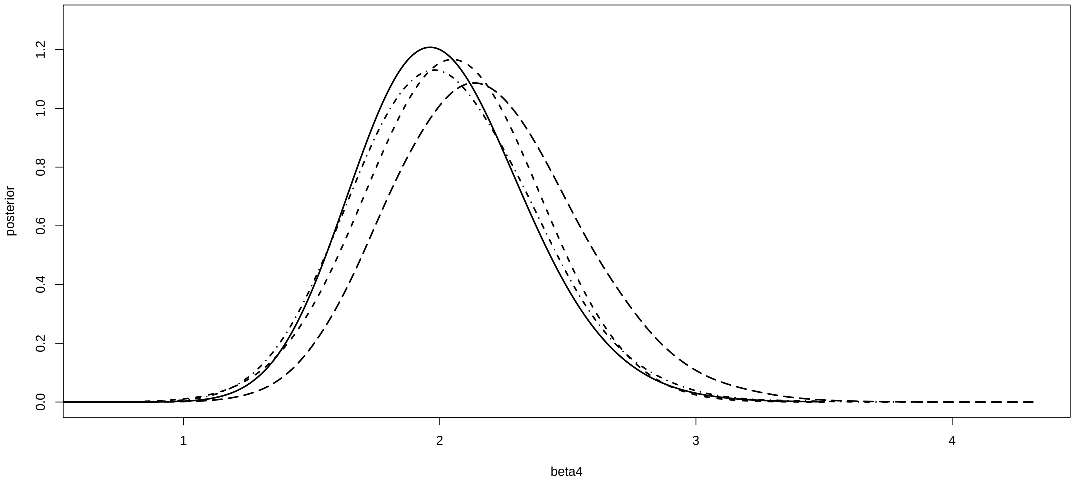

Figure 2: Higher-order (solid line) and first-order posteriors (dashed), MCMC posteriors with flat prior (long dash) and with uninformative G-prior for $\psi = \beta_4$ (dot-dashed). The corresponding 0.95% asymptotic HPD credible sets are $(1.38, 2.68)$, $(1.38, 2.72)$, $(1.48, 2.89)$ and $(1.37, 2.69)$.

$\pi_m^a(\psi\|y)$	$\pi_m(\psi\|y)$	$\pi_m^{\text{flat}}(\psi\|y)$	$\pi_m^{\text{G-prior}}(\psi\|y)$
0.956	0.951	0.906	0.943
(0.023, 0.021)	(0.025, 0.024)	(0.084, 0.010)	(0.032, 0.025)

Table 5: Frequentist coverage probabilities of approximate 0.95% HPD and of the lower and upper 0.025 quantiles (in brackets), under the logistic model.

5 Final remarks

For the purpose of making objective Bayesian inferences for a one-dimensional interest parameter, higher-order asymptotic theory is discussed. Advantages of the discussed approximations are that no elicitation on the nuisance parameter, neither no multidimensional integration or MCMC simulation are necessary in order to obtain $\pi_m(\psi|y)$, and no orthogonal parameterization is required in order to specify the matching prior. A further advantage of this approximation is that its expression automatically includes the matching prior, without requiring its explicit computation. From a practical point of view, the proposed approximation enables one to compute accurate invariant HPDs for ψ as $\{\psi : |r_p^*(\psi)| \leq z_{1-\alpha/2}\}$, i.e. as accurate likelihood-based confidence interval for ψ with approximate level $(1-\alpha)$.

Other possibile applications of the proposed methods are, among others, the binormal and bi-exponential stress-strength models $P(X < Y)$ (with software available at `homes.stat.unipd.it/ventura/` at `page=Software`), nonlinear regression models (with the `profile` method available for objects of class `nlreg`), linear mixed effects models, and generalized mixed effects models.

As a final remark, note that, while the statistic $r_p^*(\psi)$ requires the MLE $\hat{\psi}$ as an ingredient, the posterior (2) does not. In view of this, expression (2) can be useful in these situations where $\hat{\psi}$ can be infinite, using the suitably modified profile likelihood (see Severini, 2000, Chap. 9).

References

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

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

[3] Barndorff-Nielsen, O.E, Chamberlin, S.R. (1994). Stable and invariant adjusted directed likelihoods. *Biometrika*, **81**, 485–499.

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

[5] Brazzale, A.R., Davison, A.C., Reid, N. (2007). *Applied Asymptotics*. Cambridge University Press, Cambridge.

[6] Cox, D.R., Snell, E.J. (1981). *Applied statistics: Principles and examples*. Chapman and Hall, London.

[7] Datta, G.S., Ghosh, J.K. (1995). Noninformative priors for maximal invariant parameter in group models. *Test*, **4**, 95-114.

[8] Datta, G.S., Mukerjee, R. (2004). *Probability Matching Priors: Higher-order Asymptotics*, Lecture Notes in Statistics, Springer.

[9] Druilhet, P., Marin, J.M. (2007). Invariant HPD credible sets and MAP estimators. *Bayesian Analysis*, **2**, 681–692.

[10] Fraser, D.A.S., Reid, N. (2002). Strong matching of frequentist and Bayesian parametric inference. *J. Statist. Plan. Inf.*, **103**, 263–285.

[11] Giummolé, F., Ventura, L. (2002). Practical point estimation from higher-order pivots. *J. Statist. Comput. Simul.*, **72**, 419-430.

[12] Marin, J.M., Robert, C. (2007). *Bayesian Core: A Practical Approach to Computational Bayesian Statistics*. Springer.

[13] Nicolaou, A. (1993). Bayesian intervals with good frequentist behaviour in the presence of nuisance parameters, *J. Roy. Statist. Soc.* B, **55**, 377-390.

[14] Pace, L., Salvan, A. (1999). Point estimation based on confidence intervals: Exponential families. *J. Statist. Comput. Simul.*, **64**, 1-21.

[15] Pace, L., Salvan, A. (2006). Adjustments of the profile likelihood from a new perspective. *J. Statist. Plann. Inf.*, **136**, 3554–3564.

[16] Sartori, N., Bellio, R., Salvan, A., Pace, L. (1999). The directed modified profile likelihood in models with many nuisance parameters. *Biometrika*, **86**, 735-742.

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

[18] Skovgaard, I.M. (1989). A review of higher order likelihood inference. *Bull. Int. Statist. Inst.*, **53**, 331–351.

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

[20] Ventura, L., Cabras, S., Racugno, W. (2009). Prior distributions from pseudo-likelihoods in the presence of nuisance parameters. *J. Amer. Stat. Assoc.*, **104**, 768–774.

[21] Ventura, L., Racugno, W. (2011). Recent advances on Bayesian inference for $P(X < Y)$. *Bayesian Analysis*, **6**, 411–428.

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