Working Paper Series, N. 2, January 2013

Advances in approximate Bayesian computation based on modified loglikelihood ratios

Laura Ventura
Department of Statistical Sciences
University of Padua
Italy

Erlis Ruli
Department of Statistical Sciences
University of Padua
Italy

Walter Racugno
Department of Mathematics
University of Cagliari
Italy

Abstract: Asymptotic arguments are widely used in Bayesian inference, and in recent years there has been considerable developments of the so-called higher-order asymptotics. This theory provides very accurate approximations to posterior distributions, and to related quantities, in a variety of parametric statistical problems, even for small sample sizes.
The aim of this contribution is to discuss recent advances in approximate Bayesian computations based on the asymptotic theory of modified loglikelihood ratios, both from theoretical and practical point of views. Results on third-order approximations for univariate posterior distributions, also in the presence of nuisance parameters, are reviewed and a new formula for a vector parameter of interest is presented.
All these approximations may routinely be applied in practice for Bayesian inference, since they require little more than standard likelihood quantities for their implementation, and hence they may be available at little additional computational cost over simple first-order approximations. Moreover, these approximations give rise to a simple simulation scheme, alternative to MCMC, for Bayesian computation of marginal posterior distributions for a scalar parameter of interest. In addition, they can be used for testing precise null hypothesis and to define accurate Bayesian credible sets. Some illustrative examples are discussed, with particular attention to the use of matching priors.

Keywords: Asymptotic expansion, Bayesian simulation, Credible set, Laplace approximation, Marginal posterior distribution, Matching priors, Modified likelihood root, Nuisance parameter, Pereira-Stern measure of evidence, Precise null hypothesis, Tail area probability.

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://homes.stat.unipd.it/ventura

Advances in approximate Bayesian computation based on modified loglikelihood ratios

Laura Ventura
Department of Statistical Sciences
University of Padua
Italy

Erlis Ruli
Department of Statistical Sciences
University of Padua
Italy

Walter Racugno
Department of Mathematics
University of Cagliari
Italy

Abstract: Asymptotic arguments are widely used in Bayesian inference, and in recent years there has been considerable developments of the so-called higher-order asymptotics. This theory provides very accurate approximations to posterior distributions, and to related quantities, in a variety of parametric statistical problems, even for small sample sizes.

The aim of this contribution is to discuss recent advances in approximate Bayesian computations based on the asymptotic theory of modified loglikelihood ratios, both from theoretical and practical point of views. Results on third-order approximations for univariate posterior distributions, also in the presence of nuisance parameters, are reviewed and a new formula for a vector parameter of interest is presented.

All these approximations may routinely be applied in practice for Bayesian inference, since they require little more than standard likelihood quantities for their implementation, and hence they may be available at little additional computational cost over simple first-order approximations. Moreover, these approximations give rise to a simple simulation scheme, alternative to MCMC, for Bayesian computation of marginal posterior distributions for a scalar parameter of interest. In addition, they can be used for testing precise null hypothesis and to define accurate Bayesian credible sets. Some illustrative examples are discussed, with particular attention to the use of matching priors.

Keywords: Asymptotic expansion, Bayesian simulation, Credible set, Laplace approximation, Marginal posterior distribution, Matching priors, Modified likelihood root, Nuisance parameter, Pereira-Stern measure of evidence, Precise null hypothesis, Tail area probability.

1 Introduction

The aim of this contribution is to discuss recent advances in approximate Bayesian computation based on the asymptotic theory of modified loglikelihood ratios and of modified likelihood roots. This theory provides asymptotic formulae for various posterior quantities of interest, including tail areas, quantiles and credible regions.

Higher-order approximations for posterior distributions based on modifications of likelihood roots have been widely discussed in the Bayesian literature; see, among others, DiCiccio *et al.* (1990), DiCiccio and Martin (1991), Reid (1995, 2003),

Sweeting (1995, 1996), Skovgaard (2001), Brazzale *et al.* (2007), Ventura *et al.* (2013), and references therein. One appealing feature of these approximations is that they may routinely be applied in practical Bayesian inference, since they require little more than standard likelihood quantities for their implementation, and hence they may be available at little additional computational cost over simple first-order approximations.

In this paper, two recent results on third-order approximations for univariate posterior distributions based on modifications of the likelihood root are reviewed and their use in Bayesian computation is illustrated. The first result shows that third-order approximations give rise to a simple simulation scheme (Ruli *et al.*, 2012), alternative to MCMC, for Bayesian computation of marginal posterior distributions for a scalar parameter of interest. Its main advantage, over MCMC methods, is that samples are drawn independently and much lower computational time is needed. The second result shows that third-order tail area approximations can be used for testing precise null hypothesis (see Pereira and Stern, 1999, Cabras *et al.*, 2013) and to define accurate Bayesian credible sets. Particular attention is devoted to illustrations with matching priors; see Ventura *et al.* (2013), and references therein. Indeed, matching priors have several advantages, since they do not require the elicitation on the nuisance parameters, neither numerical integration or MCMC simulation.

Moreover, in this paper we indicate how approximate Bayesian computations based on modified loglikelihood ratios can be generalized for a vector parameter of interest. As is the case with the approximations for univariate posterior distributions, the proposed results are based on the asymptotic theory of modified loglikelihood ratios and they require only routine maximization output for its implementation.

The paper is organized as follows. Section 2 reviews results and applications of higher-order Bayesian approximations for a scalar parameter of interest, even in the presence of nuisance parameters. Section 3 indicates how these ideas generalize to the multiparameter case and illustrates some numerical examples. Finally, some concluding remarks are given in Section 4.

2 Approximations for a scalar parameter of interest

Consider a sampling model $f(y;\theta)$ with scalar parameter $\theta \in \Theta \subseteq \mathbb{R}$, and let $L(\theta) = L(\theta; y) = \exp\{\ell(\theta)\}$ denote the likelihood function based on data y. Given a prior density $\pi(\theta)$ for θ, Bayesian inference is based on the posterior density $\pi(\theta|y) \propto \pi(\theta)\, L(\theta)$. In several applications, an approximation is often required to an integral of the form

$$\int_{-\infty}^{\theta_0} \pi(\theta|y)\, d\theta = Pr(\theta \leq \theta_0|y)\ , \tag{1}$$

i.e. to a tail area. The derivation of a tail area approximation is simple in the scalar parameter setting (see, among others, Reid, 1995, 2003, Sweeting, 1995, 1996, Skovgaard, 2001, Davison, 2003, Chap. 11, Brazzale *et al.*, 2007, Chap. 8, and references therein). The first step to derive a higher-order approximation for posterior probabilities is to consider in (1) the Laplace expansion of $\pi(\theta|y)$, given by

$$\pi(\theta|y) \ \cong \ \frac{1}{\sqrt{2\pi}}\, |j(\hat\theta)|^{1/2}\, \frac{\pi(\theta)}{\pi(\hat\theta)} \exp\left\{\ell(\theta) - \ell(\hat\theta)\right\}\ , \tag{2}$$

where $\hat\theta$ is the maximum likelihood estimator (MLE) of θ, $j(\theta)$ is the observed information, and the approximation is accurate to order $O(n^{-1})$ (see, e.g., Tierney

and Kadane, 1986, Sweeting, 1995, 1996). We obtain

$$\int_{-\infty}^{\theta_0} \pi(\theta|y)\, d\theta \;\cong\; \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\theta_0} |j(\hat\theta)|^{1/2} \frac{\pi(\theta)}{\pi(\hat\theta)} \exp\left\{-\frac{1}{2} r(\theta)^2\right\} d\theta \;, \tag{3}$$

where $r(\theta) = \text{sign}(\hat\theta - \theta) W(\theta)^{1/2}$ is the likelihood root, with $W(\theta) = 2(\ell(\hat\theta) - \ell(\theta))$ loglikelihood ratio statistic.

The second step is to change the variable of integration from θ to $r = r(\theta)$. A motivation for considering such a transformation is that, in terms of r^2, the quantity $\exp(-r^2/2)$ is the kernel of the standard normal density $N(0,1)$. Note that the Jacobian is $dr(\theta)/d\theta = -\ell'(\theta)/r(\theta)$, with $\ell'(\theta)$ score function. We obtain

$$\pi(r|y) \;\cong\; \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2} r^2 + \log b(r)\right\} \;,$$

where the positive quantity

$$b(r) = |j(\hat\theta)|^{1/2} \frac{\pi(\theta)}{\pi(\hat\theta)} \frac{r(\theta)}{\ell'(\theta)}$$

is regarded as a function of r, and thus we have

$$\int_{-\infty}^{\theta_0} \pi(\theta|y)\, d\theta \;\cong\; \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{r_0} \exp\left\{-\frac{1}{2} r^2 + \log b(r)\right\} dr \;, \tag{4}$$

where $r_0 = r(\theta_0)$.

The last step is again a change of variable, from r to $r^* = r^*(\theta) = r - r^{-1} \log b(r)$, so that

$$-(r^*)^2 = -r^2 + 2 \log b(r) - \left(r^{-1} \log b(r)\right)^2 \;.$$

The Jacobian of the transformation and the third term in $-(r^*)^2$ contribute only to the error of (4), and it can be shown that (see, e.g., DiCiccio and Martin, 1991, Sweeting, 1995, 1996, Severini, 2000, Chap. 2)

$$\int_{-\infty}^{\theta_0} \pi(\theta|y)\, d\theta \;\doteq\; \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{r_0^*} \exp\left\{-\frac{1}{2} (r^*)^2\right\} dr^* = \Phi(r_0^*) \;, \tag{5}$$

where $\Phi(\cdot)$ is the standard normal distribution function,

$$r_0^* = r^*(\theta_0) = r_0 + \frac{1}{r_0} \log \frac{q_0}{r_0} \;, \tag{6}$$

with $q_0 = q(\theta_0)$ and

$$q(\theta) = \frac{r}{b(r)} = \ell'(\theta) |j(\hat\theta)|^{-1/2} \frac{\pi(\hat\theta)}{\pi(\theta)} \;,$$

and the symbol "$\doteq$" indicates that the approximation is accurate to order $O(n^{-3/2})$, i.e. to third order.

The tail area approximation (5) holds only for a scalar parameter θ. Now let us assume that $\theta = (\psi, \lambda)$, where ψ is a scalar parameter of interest and λ is a $(d-1)$-dimensional nuisance parameter. Consider the marginal posterior distribution

$$\pi_m(\psi|y) \;=\; \int \pi(\psi, \lambda|y)\, d\lambda \;. \tag{7}$$

Applying the Laplace expansion to (7), we obtain (see, e.g., Tierney and Kadane, 1986, Reid, 1995, 2003)

$$\pi_m(\psi|y) \doteq \frac{1}{\sqrt{2\pi}} |j_p(\hat{\psi})|^{1/2} \exp\{\ell_p(\psi) - \ell_p(\hat{\psi})\} \frac{|j_{\lambda\lambda}(\hat{\psi},\hat{\lambda})|^{1/2}}{|j_{\lambda\lambda}(\psi,\hat{\lambda}_\psi)|^{1/2}} \frac{\pi(\psi,\hat{\lambda}_\psi)}{\pi(\hat{\psi},\hat{\lambda})} , \qquad (8)$$

where $\ell_p(\psi) = \log L(\psi,\hat{\lambda}_\psi)$ is the profile loglikelihood for ψ, with $\hat{\lambda}_\psi$ constrained MLE of λ given ψ, $j_p(\psi) = -\partial^2\ell_p(\psi)/\partial\psi^2$ is the observed information corresponding to the profile loglikelihood, and $j_{\lambda\lambda}(\psi,\lambda)$ is the (λ,λ)-block of the observed information $j(\psi,\lambda)$. Note that the approximation (8) depends on simple likelihood quantities evaluated at $(\hat{\psi},\hat{\lambda})$ and at $(\psi,\hat{\lambda}_\psi)$.

Expression (8) has the same structure as (2), and thus it is readily integrated to give approximate posterior probabilities. More precisely, paralleling the scalar parameter case, it can be integrated as follows (see DiCiccio *et al.*, 1990, DiCiccio and Martin, 1991, and Reid, 2003)

$$\begin{aligned}
\int_{-\infty}^{\psi_0} \pi_m(\psi|y)\, d\psi &\doteq \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\psi_0} |j_p(\hat{\psi})|^{1/2} \exp\{\ell_p(\psi) - \ell_p(\hat{\psi})\} \frac{|j_{\lambda\lambda}(\hat{\psi},\hat{\lambda})|^{1/2}}{|j_{\lambda\lambda}(\psi,\hat{\lambda}_\psi)|^{1/2}} \frac{\pi(\psi,\hat{\lambda}_\psi)}{\pi(\hat{\psi},\hat{\lambda})} d\psi \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{r_p(\psi_0)} \exp\left(-\frac{1}{2}r_p^2\right) r_p \frac{|j_p(\hat{\psi})|^{1/2}}{\ell_p'(\hat{\psi})} \frac{|j_{\lambda\lambda}(\hat{\psi},\hat{\lambda})|^{1/2}}{|j_{\lambda\lambda}(\psi,\hat{\lambda}_\psi)|^{1/2}} \frac{\pi(\psi,\hat{\lambda}_\psi)}{\pi(\hat{\psi},\hat{\lambda})} dr_p \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{r_p(\psi_0)} \exp\left(-\frac{1}{2}r_p^2 + \log b(r_p)\right) dr_p , \qquad (9)
\end{aligned}$$

where $r_p = r_p(\psi) = \text{sign}(\hat{\psi}-\psi)[2(\ell_p(\hat{\psi}) - \ell_p(\psi))]^{1/2}$ is profile likelihood root and

$$b(r_p) = r_p \frac{|j_p(\hat{\psi})|^{1/2}}{\ell_p'(\hat{\psi})} \frac{|j_{\lambda\lambda}(\hat{\psi},\hat{\lambda})|^{1/2}}{|j_{\lambda\lambda}(\psi,\hat{\lambda}_\psi)|^{1/2}} \frac{\pi(\psi,\hat{\lambda}_\psi)}{\pi(\hat{\psi},\hat{\lambda})} .$$

The next step is to consider a change of variable from r_p to $r_p^* = r_p^*(\psi) = r_p - r_p^{-1}\log b(r_p)$, so that $-(r_p^*)^2 = -r_p^2 + 2\log b(r) - (r_p^{-1}\log b(r_p))^2$. Since the Jacobian of the transformation and the third term in $-(r_p^*)^2$ contribute only to the error of (9), it can be shown that

$$\int_{-\infty}^{\psi_0} \pi_m(\psi|y)\, d\psi \doteq \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{r_p^*(\psi_0)} \exp\left(-\frac{1}{2}(r_p^*)^2\right) dr_p^* = \Phi\left(r_p^*(\psi_0)\right) , \qquad (10)$$

where

$$r_p^* = r_p^*(\psi) = r_p(\psi) + \frac{1}{r_p(\psi)} \log \frac{q_B(\psi)}{r_p(\psi)} , \qquad (11)$$

is a modification of the profile likelihood root, with

$$q_B(\psi) = \ell_p'(\psi)\ |j_p(\hat{\psi})|^{-1/2} \frac{|j_{\lambda\lambda}(\psi,\hat{\lambda}_\psi)|^{1/2}}{|j_{\lambda\lambda}(\hat{\psi},\hat{\lambda})|^{1/2}} \frac{\pi(\hat{\psi},\hat{\lambda})}{\pi(\psi,\hat{\lambda}_\psi)} .$$

Formula (10) gives an explicit expression for the posterior quantiles, and $1-\Phi(r_p^*(\psi_0))$ is the Bayesian survivor probability.

Moreover, from $r_p^*(\psi)$ or, equivalently, from $(r_p^*(\psi))^2$, it is possible to define a credible interval for ψ. In particular, an accurate asymptotic credible interval for ψ can be computed as

$$CI = \left\{\psi : |r_p^*(\psi)| \le z_{1-\alpha/2}\right\} , \qquad (12)$$

where $z_{1-\alpha/2}$ is the $(1-\alpha/2)$-quantile of the standard normal distribution. Note that (12) is an equi-tailed credible interval for ψ. Note also that from (10) the median posterior estimator (MPE) of (7) can be computed as the solution $\hat{\psi}^*$ in ψ of the estimating equation $r_p^*(\psi) = 0$.

2.1 The special case of approximations with matching priors

When the particular class of matching priors for posterior quantiles is considered in (7) (see Tibshirani, 1989), the marginal posterior distribution for ψ can be expressed as (see Ventura *et al.*, 2009, 2013)

$$\pi_m(\psi|y) \quad \propto \quad L_{mp}(\psi)\ \pi_{mp}(\psi)\ , \tag{13}$$

where $L_{mp}(\psi) = L_p(\psi)M(\psi)$ is the modified profile likelihood for a suitably defined correction term $M(\psi)$ (see, among others, Severini, 2000, Chap. 9, Pace and Salvan, 2006), and $\pi_{mp}(\psi) \propto i_{\psi\psi.\lambda}(\psi, \hat{\lambda}_\psi)^{1/2}$ is the corresponding matching prior, with $i_{\psi\psi.\lambda}(\psi,\lambda) = i_{\psi\psi}(\psi,\lambda) - i_{\psi\lambda}(\psi,\lambda)i_{\lambda\lambda}(\psi,\lambda)^{-1}i_{\lambda\psi}(\psi,\lambda)$ partial information, and $i_{\psi\psi}(\psi,\lambda)$, $i_{\psi\lambda}(\psi,\lambda)$, $i_{\lambda\lambda}(\psi,\lambda)$, and $i_{\lambda\psi}(\psi,\lambda)$ blocks of the expected Fisher information $i(\psi,\lambda)$ from $\ell(\psi,\lambda)$. The matching prior $\pi_{mp}(\psi)$ has the advantages that it does not require the elicitation on the nuisance parameters, neither numerical integration or MCMC simulation in order to compute the marginal posterior distribution (13) for ψ.

Accurate tail area probabilities are computable by direct integration of (13). In particular, in Ventura and Racugno (2011) it is shown that (10) holds with $r_p^*(\psi)$ given by the modified profile likelihood root of Barndorff-Nielsen and Chamberlin (1994); see also Barndorf-Nielsen and Cox (1994), and Severini (2000, Chap. 7). More precisely, the quantity $r_p^*(\psi)$ has the form (11), with $q_B(\psi) = q_p(\psi)$, where

$$q_p(\psi) = \frac{\ell_p'(\psi)}{j_p(\hat{\psi})^{1/2}} \frac{i_{\psi\psi.\lambda}(\hat{\psi},\hat{\lambda})^{1/2}}{i_{\psi\psi.\lambda}(\psi,\hat{\lambda}_\psi)^{1/2}} \frac{1}{M(\psi)}\ . \tag{14}$$

When using (14), the credible interval (12) for ψ coincides with an accurate higher-order likelihood-based confidence interval for ψ with approximate level $(1-\alpha)$. Following Sweeting (1999), this credible interval is an invariant third-order Bayes-confidence interval. Moreover, the MPE coincides with the frequentist estimator defined as the zero-level confidence interval based on $r_p^*(\psi)$ (Skovgaard, 1989). Such estimator has been shown to be a refinement of the maximum likelihood estimator $\hat{\psi}$ (see Pace and Salvan, 1999, Giummolé and Ventura, 2002). Finally, following Ventura *et al.* (2013), it can be shown that the marginal posterior distribution for ψ given by (13) can also be written, to second-order, in the following form

$$\pi_m(\psi|y) \ \tilde{\propto}\ \exp\left(-\frac{1}{2}r_p^*(\psi)^2\right)\left|\frac{s_p(\psi)}{r_p(\psi)}\right|\ , \tag{15}$$

where $s_p(\psi) = \ell_p'(\psi)/j_p(\hat{\psi})^{1/2}$ is the profile score statistic. A remarkable advantage of this approximation is that its expression automatically includes the matching prior, without requiring its explicit computation.

Example 1: Nonlinear regression. Models for data with continuous response values, i.e. linear and nonlinear regression with normal and non-normal errors, are nowadays widely used in many fields (see, e.g., Davison, 2003). In a nonlinear regression model the responses $y_1, \ldots, y_n$ are related to explanatory variables x_i as

$$y_i = \mu(x_i;\beta) + \sigma\,\varepsilon_i\ , \quad i = 1, \ldots, n\ , \tag{16}$$

where x_i is a known $p \times 1$ vector, the unknown parameters are the $p \times 1$ vector β and the scale parameter $\sigma > 0$, $\mu(x_i; \beta)$ is the mean function, and the ε_i are independent and generated from a known continuous density function $f(\cdot)$. If $\varepsilon_i \sim N(0, 1)$, this model is usually called the nonlinear regression model, and it is widely used, especially for dose-response curves in bioassays. A more general form is

$$y_{ij} = \mu(x_i; \beta) + \sigma_i \, \varepsilon_{ij} \ , \quad i = 1, \ldots, m \ , \quad j = 1, \ldots, n_i \ , \tag{17}$$

where m is the number of design points x_i, n_i is the number of replicates at design points, y_{ij} represents the response of the jth experimental unit at the ith design point, and the ε_{ij} are $N(0, 1)$ variates. Moreover, $\sigma_i^2 = \sigma^2 V(x_i; \beta, g)$, where σ^2 and the $q \times 1$ vector g are variance parameters and $V(\cdot)$ is a given function.

Brazzale *et al.* (2007, Sect. 5.4) discuss a study on a radioimmunoassay (RIA) taken to estimate the concentrations of a drug in samples of porcine serum. The experiment consists of 16 observations made at eight different drug levels with two replications at each level. The data are available in the data frame ria of the lnreg package: count (y) represents the observed percentage of radioactive gamma counts, and conc (x) the drug concentration (ng/ml). The concentration-response relationship is modeled by means of the four-parameter logistic function

$$\mu(x; \beta) = \beta_1 + \frac{\beta_2 - \beta_1}{1 + (x/\beta_4)^{2\beta_3}} \ , \quad x \geq 0 \ ,$$

and the variance of the associated error distribution may be captured by a power-of-the-mean variance function, i.e. $V(x_i; \beta, g) = \mu(x_i; \beta)^g$, where g is a scalar variance parameter.

The computation of $\pi_m(\psi|y)$ can be performed using the profile method available for objects of class nlreg, of the library HOA (Brazzale *et al.*, 2007). Figure 1 (left) gives the plot of the posterior distribution (15) and of the first order approximation

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

for the parameter of interest $\psi = g$. The corresponding third-order and first-order asymptotic 95% equi-tailed credible intervals are $(-0.02, 2.92)$ and $(1.06, 3.13)$, respectively. These credible intervals can be easily computed from the output of the profile method as shown in Figure 1 (right).

2.2 Tail area approximations for Bayesian simulation

Starting from higher-order tail area approximations from (7), it is possible to develop a sampling scheme that give rise to an accurate computation of marginal posterior densities, and related quantities, such as posterior summaries (Ruli *et al.*, 2012).

The implementation of the higher-order tail area approximation (HOTA) sampling scheme is available at little additional computation cost over simple first-order approximations, and it has the advantage over MCMC methods that samples are drawn independently in much lower computation time.

Starting from (7), the simulation algorithm can be summarized as follows. For $t = 1, \ldots, T$:

[1] draw $z_t \sim N(0, 1)$;

[2] find ψ_t as the solution of $r_p^*(\psi_t) = z_t$.

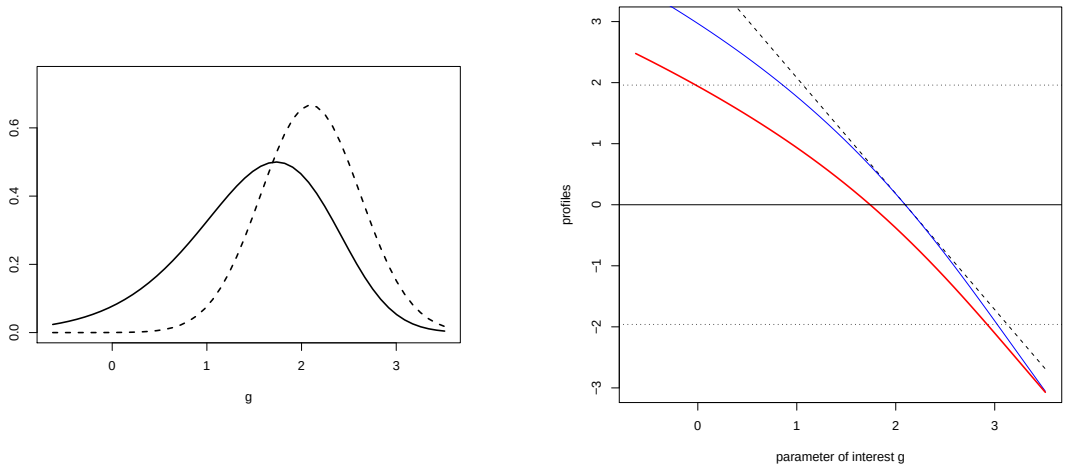

Figure 1: *RIA data. Left: Posteriors $\pi_m(\psi|y)$ (solid line) and first-order approximation $\pi_m^I(\psi|y)$ (dashed line) for $\psi = g$. Right: $r_p^*(\psi)$ (bold line), $r_p(\psi)$ (solid line) and Wald (dashed line) statitics; the horizontal lines are the quantiles $\pm z_{0.975}$.*

We obtain a sample $(\psi_1, \ldots, \psi_T)$ from the marginal density $\pi_m(\psi|y)$.

Note that the main computational effort involved in the HOTA scheme is the solution of the equation $r_p^*(\psi_t) = z_t$ for each sample value z_t of $r_p^*(\psi)$. A numerical procedure is usually required in order to solve this equation (see Ruli *et al.*, 2012).

The HOTA simulation procedure is essentially an inverse method of sampling and it gives independent samples from (7) by inverting the cumulative distribution function approximation (10). In this respect, it has an obvious advantage over MCMC methods which usually requires more attention from the practitioner. Moreover, HOTA is almost automatically obtained from likelihood quantities, which opens the possibility of doing Bayesian inference with maximum likelihood routines.

Example 1: Nonlinear regression (cont). The computation of the posteriors (13) and $\pi_m^I(\psi|y)$ with the HOTA algorithm is illustrated in Figure 2 for the parameters g and β_1. The overall computation time was 4 seconds.

Based on the HOTA posterior, the 0.95 HPD for g is (0.14,3.03). While, for the parameter β_1, the 0.95 HPD based on the HOTA posterior and $\pi_m^I(\psi|y)$ are (1.05,2.46) and (1.42,2.18), respectively.

Example 2: Censored regression. This example is discussed in Ruli *et al.* (2012); see also the references therein. The dataset consists on temperature accelerated life tests on electrical insulation in $n = 40$ motorettes. Ten motorettes were tested at each of four temperatures in degrees Centigrade (150°, 170°, 190° and 220°), the test termination (censoring) time being different at each temperature. The model is (Ruli *et al.*, 2012)

$$y_i = \beta_0 + \beta_1 x_i + \sigma\epsilon_i, \qquad i = 1, \ldots, n \ ,$$

where y_i is the $\log_{10}$(failuretime) with time in hours, $x_i = 1000/(\text{temperature}+273.2)$ and ϵ_i are independent standard normal errors. Reordering the data so that the first m observations are uncensored, with observed log-failure times y_i, and the remaining

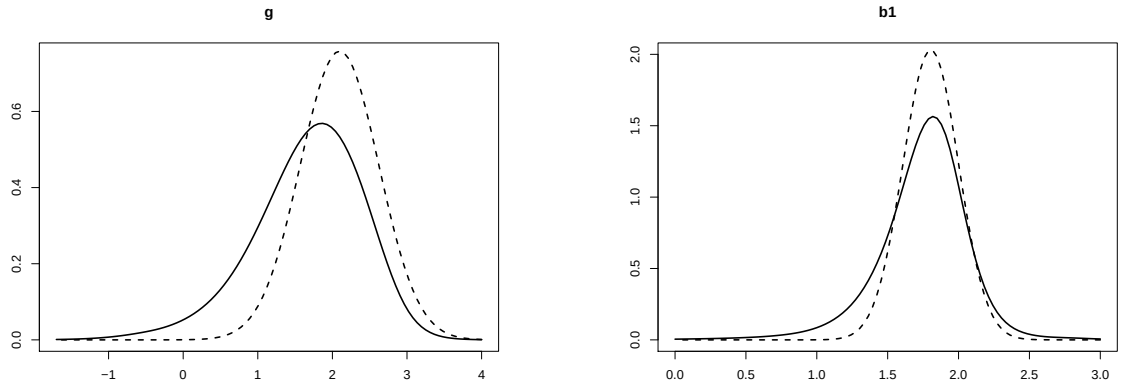

Figure 2: *RIA data and HOTA algorithm. Left: Posteriors $\pi_m(\psi|y)$ (solid line) and first-order approximation $\pi_m^I(\psi|y)$ (dashed line) for $\psi = g$. Right: Posteriors $\pi_m(\psi|y)$ (solid line) and first-order approximation $\pi_m^I(\psi|y)$ (dashed line) for $\psi = \beta_1$.*

$n - m$ are censored at times u_i, the loglikelihood function for $\theta = (\beta_0, \beta_1, \sigma)$ is

$$\ell(\theta) = -m \log \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^{m} (y_i - \beta_0 - \beta_1 x_i)^2 + \sum_{i=m+1}^{n} \log \left\{ 1 - \Phi \left(\frac{u_i - \beta_0 - \beta_1 x_i}{\sigma} \right) \right\},$$

For (β_0, β_1, τ), with $\tau = \log \sigma$, the non-informative prior $\pi(\beta_0, \beta_1, \tau) \propto 1$ is assumed. The posterior distribution $\pi(\beta_0, \beta_1, \tau|y)$ does not have a closed form expression and direct integration is not possible in order to compute $\pi_m(\psi|y)$ and related quantities, where ψ is one of the parameters of the model. Therefore numerical or analytical approximations are needed.

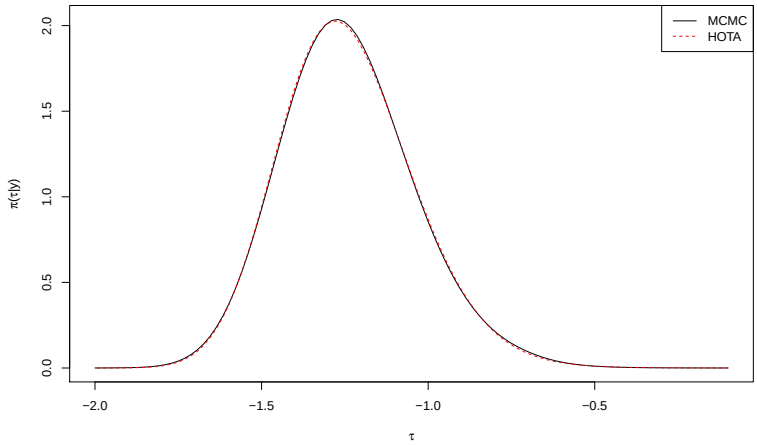

Figure 3: Censored regression: HOTA and MCMC marginal posterior distributions for the parameter τ.

Figure 3 illustrates the HOTA and MCMC marginal posterior distributions of τ. The HOTA sampling scheme and MCMC give similar results. The same pattern holds also for the other parameters (not shown here). Table 1 gives some summary statistics (mean, standard deviation, 2.5 percentile, median, 97.5 percentile and 0.95 HPD credible set) calculated over the three marginal posterior distributions. The results based of the two methods are in good agreement.

For the computation with HOTA, grids of 50 points were chosen for all the parameters and the total number of simulations was $T = 10^5$. The overall computation

Method	Posterior	Mean	St Dev.	$Q_{0.025}$	Median	$Q_{0.975}$	0.95 HPD
MCMC	$\pi_m(\tau\|y)$	-1.24	0.201	-1.60	-1.253	-0.811	(-1.616, -0.832)
HOTA	$\pi_m(\tau\|y)$	-1.24	0.202	-1.601	-1.251	-0.808	(-1.624, -0.837)
MCMC	$\pi_m(\beta_0\|y)$	-6.204	1.117	-8.57	-6.139	-4.149	(-8.413, -4.01)
HOTA	$\pi_m(\beta_0\|y)$	-6.191	1.128	-8.596	-6.134	-4.13	(-8.475, -4.038)
MCMC	$\pi_m(\beta_1\|y)$	4.409	0.518	3.461	4.382	5.512	(3.425, 5.47)
HOTA	$\pi_m(\beta_1\|y)$	4.401	0.521	3.459	4.37	5.521	(3.398, 5.443)

Table 1: Numerical summaries of the MCMC and HOTA marginal posterior distributions.

time on a laptop with 4 GB RAM was 1.8 seconds, while for MCMC with 10^6 simulations after thinning every 10, total 10^5 observations with low autocorrelation the computation time was 95 seconds.

2.3 Tail area approximations for a measure of evidence

Suppose we are interested in testing the precise (or sharp) null hypothesis $H_0 : \psi = \psi_0$ versus $H_1 : \psi \neq \psi_0$. In order to avoid the Jeffreys-Lindley paradox, the measure of evidence (EV) of the full Bayesian significance test (FBST) of Pereira and Stern (1999, 2001) can be considered; see also Madruga *et al.* (2001, 2003) and Pereira *et al.* (2008).

Following Cabras *et al.* (2013), consider the set

$$T(y) = \left\{\psi : \pi_m(\psi|y) \geq \sup_{\psi_0} \pi_m(\psi|y)\right\} \ .$$

Then, the Pereira and Stern posterior evidence EV in favor of H_0 can be computed as (see Figure 4)

$$EV = 1 - Pr(\psi \in T(y)|y) \ . \tag{18}$$

The null hypothesis H_0 is accepted whenever EV is large enough. Using (10), a simple and accurate higher-order approximation of (18) is

$$EV \doteq 1 + \Phi(r_p^*(\psi_0)) - \Phi(r_p^*(\psi_0^*)) \ , \tag{19}$$

where ψ_0^* is such that $\pi_m(\psi_0^*|y) = \pi_m(\psi_0|y)$. With respect to the original definition of EV (Pereira and Stern, 1999), (19) is simpler to compute, in particular when the dimension of the nuisance parameter is large. When in particular matching priors are assumed (see Subsection 2.1), the approximation (19) presents the further advantage that it does not require the elicitation of the prior on the nuisance parameter.

Note that

$$\begin{aligned} \Phi(r_p^*(\psi_0)) - \Phi(r_p^*(\psi_0^*)) &\doteq \int_{\psi_0}^{\psi_0^*} \pi_m(\psi|y)\, d\psi \\ &= Pr(\psi \in T(y)|y) = 1 - EV \ , \end{aligned} \tag{20}$$

gives the posterior probability of the HPD credible interval (ψ_0, ψ_0^*).

Example 1: Nonlinear regression (cont). Suppose that for the RIA data we are interested in testing $H_0 : \psi = 1$ versus $H_0 : \psi \neq 1$, for $\psi = g$. The Pereira and Stern posterior evidence in favor of H_0 is illustrated in Figure 5 for the marginal posterior distribution (15) and for the first-order approximation $\pi_m^I(\psi|y)$.

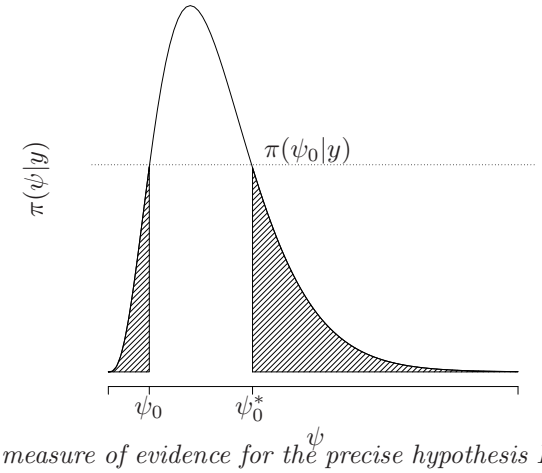

Figure 4: *The EV measure of evidence for the precise hypothesis $H_0 : \psi = \psi_0$ is the shaded area.*

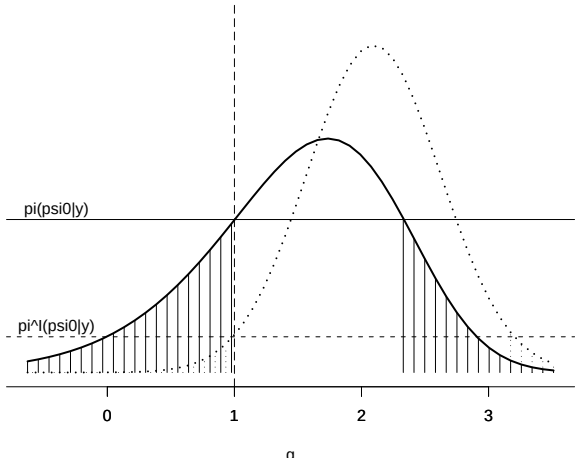

Figure 5: *RIA data. Posteriors $\pi_m(\psi|y)$ (solid line) and $\pi_m^I(\psi|y)$ (dashed line) for $\psi = g$. The EV measure of evidences for the precise hypothesis $H_0 : \psi = 1$ are the dashed areas.*

The computation of (18) for the HOTA posterior distributions gives $EV = 0.26$ to third-order, i.e. when using (19), and $EV = 0.04$ to first-order. The inferential conclusions about H_0 are quite different, with the first-order statement which is very unsatisfactory. This illustrates an important advantage of third-order asymptotics with respect to first-order results.

Example 2: Censored regression (cont). Suppose that for the data on temperature asselerated life tests we are interested in testing $H_0 : \psi = 0.2$ versus $H_0 : \psi \neq 0.2$, for $\psi = g$. The computation of (18) for the HOTA posterior distribution gives $EV = 0.035$ to third-order, i.e. when using (19), and $EV = 0.21$ to first-order. Even in this example, the first-order statement is unsatisfactory.

3 Approximate Bayesian computation for multidimensional parameters

Suppose that $\theta \in \Theta \subseteq \mathbb{R}^d$, with $d > 1$. In a similar way to Section 2, we study the approximations based on modifications of the loglikelihood ratios. As in the scalar parameter case, the derivation of these approximations can be based of the following three steps (see Skovgaard, 2001):

Step 1: computation of the Laplace approximation of $\pi(\theta|y)$;

Step 2: change of the variable of integration from θ to $r_m = r_m(\theta)$, such that for the loglikelihood ratio we have $W(\theta) = 2\left(\ell(\hat{\theta}) - \ell(\theta)\right) = r_m^{\mathsf{T}} r_m$;

Step 3: change of the variable of integration from r_m to a more accurate version of the form $r_m^* = r_m^*(\theta) = r_m - \delta(r_m)$, with $\delta = \delta(r_m)$ chosen to satisfy $r_m^{\mathsf{T}}\delta(r_m) = \log g(r_m)$ for a suitably defined term $g(r_m)$, so that

$$(r_m - \delta)^{\mathsf{T}}(r_m - \delta) = r_m^{\mathsf{T}} r_m - 2\log g(r_m) + O(n^{-2})$$

is asymptotically χ^2_d.

In order to compute Step 2, we need to reach a statistic $r_m = r_m(\theta)$ for which $r_m^{\mathsf{T}} r_m = W(\theta)$. Let us consider the signed root loglikelihood ratio transformation $r_m(\theta)$ defined in Sweeting (1995, 1996); see also Sweeting and Kharroubi (2003) and Kharroubi and Sweeting (2010).

Let $\theta = (\theta_1, \ldots, \theta_d) = (\theta^i, \theta^{(i+1)})$, where $\theta^i = (\theta_1, \ldots, \theta_i)$ is the vector of the first i components of θ and $\theta^{(i+1)} = (\theta_{i+1}, \ldots, \theta_d)$. Let $\hat{\theta}^{(i+1)}_{\theta^i}$ be the partial MLE of $\theta^{(i+1)}$ given θ^i, and let $\hat{\theta}_{j,\theta^i}$ be the jth component of $(\theta^i, \hat{\theta}^{(i+1)}_{\theta^i})$, for $j > i$. The signed root loglikelihood ratio transformation is thus given by

$$r_m(\theta) = (r_{m1}, \ldots, r_{md}) \ , \tag{21}$$

with

$$r_{mi} = \text{sign}(\theta_i - \hat{\theta}_{i,\theta^{i-1}}) \left\{ 2\left[\ell\left(\theta^{i-1}, \hat{\theta}^{(i)}_{\theta^{i-1}}\right) - \ell\left(\theta^i, \hat{\theta}^{(i+1)}_{\theta^i}\right)\right]\right\}^{1/2} \ , \tag{22}$$

for $i = 1, \ldots, d$. Notice that r_{mi} is a function of the first i components $\theta^i = (\theta_1, \ldots, \theta_i)$ of θ, for $i = 1, \ldots, d$. It follows that $r_m(\theta)$ is a one-to-one data-dependent trasformation of θ and, following Sweeting (1995, 1996), we have that (21) is such that $\exp\left\{-\frac{1}{2} r_m^{\mathsf{T}} r_m\right\} = L(\theta)/L(\hat{\theta})$. Moreover, $r_m(\theta)$ is asymptotically multivariate standard normal to $O(n^{-1/2})$ (Sweeting, 1995).

Paralleling the derivation of the tail area approximations discussed in Section 2, as a first step let us consider the Laplace approximation of $\pi(\theta|y)$, i.e.

$$\begin{aligned} \pi(\theta|y) &\cong (2\pi)^{-d/2} |j(\hat{\theta})|^{1/2} \frac{\pi(\theta)}{\pi(\hat{\theta})} \exp\left\{\ell(\theta) - \ell(\hat{\theta})\right\} \\ &= (2\pi)^{-d/2} |j(\hat{\theta})|^{1/2} \frac{\pi(\theta)}{\pi(\hat{\theta})} \exp\left\{-\frac{1}{2} W(\theta)\right\} \ . \end{aligned} \tag{23}$$

The second step in approximating the tail area probability is to change the variable of integration from θ to the statistic $r_m = r_m(\theta)$ given in (21). The Jacobian matrix

$dr_m/d\theta$ is lower triangular (Sweeting, 1995, 1996), and in particular it holds

$$\left|\frac{dr_m}{d\theta}\right| = \prod_{i=1}^{d}\left|\frac{\ell_i\left(\theta^i, \hat{\theta}^{(i+1)}_{\theta^i}\right)}{r_{mi}}\right| , \tag{24}$$

where $\ell_i(\theta)$ is the ith component of the score vector $\partial\ell(\theta)/\partial\theta$, $i = 1, \ldots, d$. The last step is again a change of variable. Following Skovgaard (2001), we perturb r_m to $r^*_m = r^*_m(\theta) = r_m - \delta(r_m)$, with $\delta(r_m)$ chosen to satisfy $r_m^{\top}\delta(r_m) = \log g(r_m)$, so that

$$-(r_m - \delta(r_m))^{\top}(r_m - \delta(r_m)) = -r_m^{\top}r_m + 2\log g(r_m) + O(n^{-2}) . \tag{25}$$

A differentiable transformation of this kind may be written explicitly by choosing

$$\delta(r_m) = \log g(r_m)\frac{d\log g(r_m)/dr_m}{(d\log g(r_m)/dr_m)^{\top}r_m} .$$

Actually, in order to compute (25), we only need the existence of $\delta(r_m)$ to calculate

$$w^*_m = w^*_m(\theta) = r_m^{\top}r_m - 2\log g(r_m(\theta)) , \tag{26}$$

with

$$g(r_m(\theta)) = |j(\hat{\theta})|^{1/2}\frac{\pi(\theta)}{\pi(\hat{\theta})}\left[\prod_{i=1}^{d}\left|\frac{\ell_i\left(\theta^i, \hat{\theta}^{(i+1)}_{\theta^i}\right)}{r_{mi}}\right|\right]^{-1} . \tag{27}$$

The asymptotic distribution of w^*_m is χ^2_d with relative error $O(n^{-1})$ in a large deviation region (see Skovgaard, 2001). To obtain a statistic which generalizes the scalar version (6), Skovgaard (2001) suggests to use the asymptotically equivalent approximation

$$w^{**}_m = w^{**}_m(\theta) = r_m^{\top}r_m\left(1 - \frac{\log g(r_m)}{r_m^{\top}r_m}\right)^2 . \tag{28}$$

Note that, for $d = 1$, the quantity $g(\theta)$ reduces to $g(\theta) = r(\theta)/q(\theta)$, and thus we have $w^{**}(\theta) = (r - (1/r)\log g(\theta))^2 = (r^*)^2$.

Example 3: Normal distribution. Consider a random sample $y = (y_1, \ldots, y_n)$ from a $N(\mu, \sigma^2)$ distribution, with $\theta = (\mu, \sigma^2)$ unknown. In the following we assume two different prior distributions of θ, that are the improper prior $\pi_1(\theta) \propto 1/\sigma^2$ and the normal-gamma prior $\pi_2(\theta)$. In this case, all the quantites involved in the computation of w^* and w^{**} are easy to compute.

Let $0 < \alpha < 1$. Paralleling the scalar parameter case, the goal is to study a credible region (CR) such that it has approximately $100(1-\alpha)\%$ coverage in repeated sampling, improving standard first-order approximations. In the following, we focus on the credible region

$$CR = \left\{\theta : w^{**} \leq \chi^2_{d;1-\alpha}\right\} , \tag{29}$$

based on w^{**}, on equivalently on w^*. This region can be interpreted as the extension to the multidimensional case of the set (12). To judge the coverage quality of CR, a simulation study based on 10000 Monte Carlo trials has been performed. Table 2

	π_1			π_2		
$1-\alpha$	0.90	0.95	0.99	0.90	0.95	0.99
	$n=10$			$n=10$		
CR_N	0.7280	0.7830	0.8685	0.5905	0.6470	0.7402
CR_L	0.8540	0.9130	0.9770	0.7871	0.8688	0.9578
CR	0.9075	0.9510	0.9925	0.9020	0.9517	0.9904
	$n=15$			$n=15$		
CR_N	0.7615	0.8280	0.900	0.6698	0.7302	0.8189
CR_L	0.8485	0.9225	0.984	0.8276	0.8992	0.9738
CR	0.8935	0.9500	0.990	0.9050	0.9544	0.9916
	$n=30$			$n=30$		
CR_N	0.8275	0.889	0.9495	0.7688	0.8250	0.9031
CR_L	0.8775	0.936	0.9840	0.8606	0.9242	0.9824
CR	0.8980	0.948	0.9875	0.9023	0.9533	0.9888
	$n=50$			$n=50$		
CR_N	0.8630	0.9240	0.9730	0.8160	0.8761	0.9436
CR_L	0.8965	0.9435	0.9890	0.8791	0.9346	0.9836
CR	0.9045	0.9525	0.9890	0.9011	0.9514	0.9897

Table 2: Normal distribution: Empirical coverage probabilities of credible regions, for several values of α and n.

gives the empirical frequentist coverages for $(1-\alpha)$ posterior credible regions (29) in comparison to the first-order credible regions

$$CR_N = \left\{ \theta : (\theta - \tilde{\theta})^{\mathsf{T}} j(\tilde{\theta})(\theta - \tilde{\theta}) \leq \chi^2_{d;1-\alpha} \right\} , \tag{30}$$

where $\tilde{\theta}$ is the posterior mode and $j(\tilde{\theta}) = -\partial \log \pi(\theta|y)/(\partial\theta\partial\theta^{\mathsf{T}})$, and the likelihood-type credible regions

$$CR_L = \left\{ \theta : -2 \log \frac{\pi(\theta|y)}{\pi(\tilde{\theta}|y)} \leq \chi^2_{d;1-\alpha} \right\} . \tag{31}$$

From Table 2 we note that, for every n, (29) clearly improves on (30) and (31). Larger sample sizes would show, as one would expect, rather little differences between the results of all the procedures. Finally, remark that the results of the simulation study do not change (results not reported here) when inverting the parameter order in the signed root loglikelihood ratio transformation (21).

Example 4: Gamma distribution. Consider a random sample $y = (y_1, \ldots, y_n)$ from a Gamma distribution, with both the shape and scale parameters unknown. Let $\theta = (\log\sigma, \log\kappa)$, with σ scale parameter and κ shape parameter. In the following, we assume two different prior distributions of θ, that are the flat prior $\pi_1(\theta) \propto 1$ and $\pi_2(\theta) \propto N(\mu, 10) \times N(\mu, 10)$, where μ is a given hyperparameter.

As in the previous example, to judge the coverage quality of CR, a simulation study based on 2000 Monte Carlo trials has been performed. Table 3 gives the empirical frequentist coverages for 0.95 posterior credible regions (29) in comparison to the first-order credible regions CR_N and CR_L. From Table 3 we note that, for every n, (29) clearly improves on (30) and (31). Moreover, note that also in this example the results of the simulation study do not change (results not reported here) when inverting the parameter order in the signed root loglikelihood ratio transformation (21). Finally, observe that there is some degradation in the coverage accuracy for parameter values in regions of low prior density.

	π_1			$\pi_2(\mu=0)$			$\pi_2(\mu=3)$		
$1-\alpha$	0.90	0.95	0.99	0.90	0.95	0.99	0.90	0.95	0.99
	$n=5$			$n=5$			$n=5$		
CR_N	0.8188	0.7991	0.8801	0.7642	0.8288	0.9040	0.6630	0.7324	0.8374
CR_L	0.8405	0.9084	0.9755	0.8624	0.9265	0.9837	0.7787	0.8659	0.9594
CR	0.9018	0.9500	0.9895	0.9166	0.9612	0.9933	0.8753	0.9338	0.9864
	$n=10$			$n=10$			$n=10$		
CR_N	0.8188	0.8779	0.9445	0.8281	0.8868	0.9495	0.7764	0.8381	0.9215
CR_L	0.8748	0.9336	0.9832	0.8826	0.9385	0.9854	0.8402	0.9115	0.9764
CR	0.9028	0.9519	0.9893	0.9084	0.9564	0.9908	0.8866	0.9424	0.9872

Table 3: Gamma distribution: Empirical coverage probabilities of credible regions, for several values of α and n.

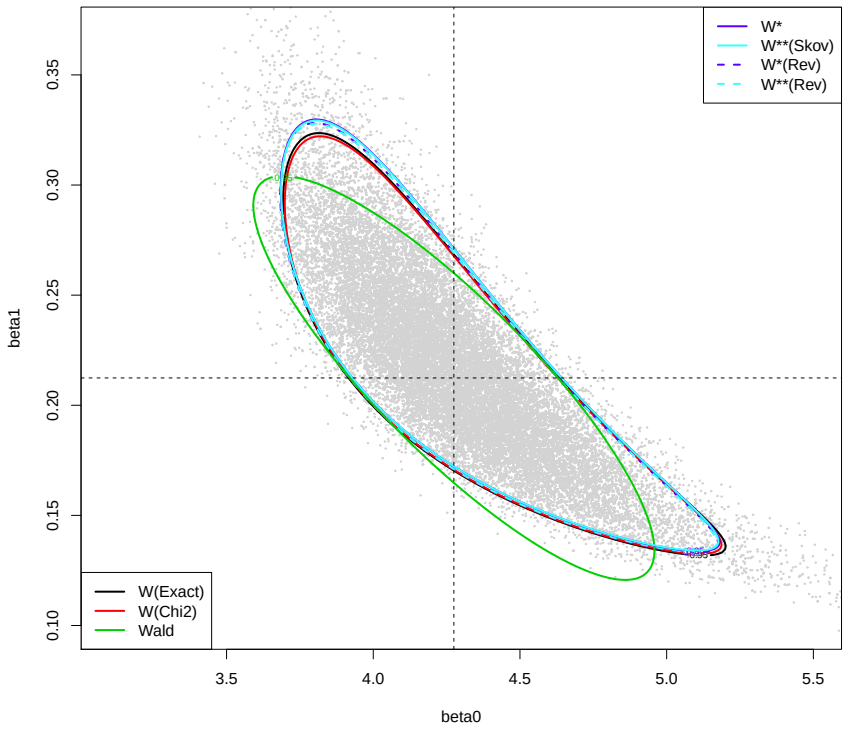

Figure 6: Calcium data: credible regions for (β_1, β_2).

Example 5: Nonlinear regression. Let us consider a nonlinear regression model of the form (17), with $\mu(x_i;\beta) = \beta_1(1-\exp(-\beta_2 x_i))$ and with $\sigma_i^2 = 0.29$. Davison (2003, Sect. 10.1) discusses this model for the calcium data, i.e. data on the calcium uptake on time by cells that has been in hot calcium suspension, with $n_i = 3$, for $i = 1, \ldots, 9$.

Figure 6 gives the contours of several credible regions for (β_1, β_2), i.e. CR_N (Wald), CR_L (W(Chi2)), the 95 HPD credible region (W(Exact)), CR based on w_m^* (W*) and on w_m^{**} (W**), and CR based on w_m^* (W*(Rev)) and on w_m^{**} (W**(Rev)) when inverting the parameter order in the signed root loglikelihood ratio transformation (21). The posterior probabilities of CR_N is 0.9129, of CR_L is 0.9478, and of CR is 0.95133. These results indicate that the accuracy of CR is very high.

4 Remarks

This paper aims to outline how approximate computational tools may have a role to play in the modern era of Bayesian statistics, where high computationl power allows the use of stochastic simulation techniques to obtain exact (i.e. simulation consistent) answers. In problems with a large number of nuisance parameters or to obtain credible regions for a vector parameter, approximate Bayesian computations based on loglikelihood ratios provide important quantities of the posterior distribution with very little computational effort, in a fraction of the time required for a full simulation approach. Moreover, sensitivity and influence analyses may also be carried out quickly within this framework (see, e.g., Ruli *et al.*, 2012, and Ventura *et al.*, 2013).

A key feature of the approximations discussed and developed in this paper is that they do not require the calculation of loglikelihood derivatives beyond the second order for their implementation. Although the approximations described in this paper are derived from asymptotic considerations, they perform extremely well in moderate or even small sample situations. On the other hand, the approximations are only available in regular models (see Kass *et al.*, 1990), for precise regularity conditions.

References

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

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

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

Cabras, S., Racugno, W., Ventura, L. (2013). Higher-order asymptotics for Bayesian significance tests for precise null hypotheses in the presence of nuisance parameters. Manuscript.

Davison, A.C. (2003). *Statistical Models*. Cambridge University Press, Cambridge.

DiCiccio, T.J., Field, C.A., Fraser, D.A.S. (1990). Approximations of marginal tail probabilities and inference for scalar parameters. *Biometrika*, **77**, 77–95.

DiCiccio, T.J., Martin, A.M. (1991). Approximations of marginal tail probabilities for a class of smooth functions with applications to Bayesian and conditional inference. *Biometrika*, **78**, 891–902.

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

Kass, R.E., Tierney, L., Kadane, J. (1990). The validity of posterior expansions based on Laplace's method. In: *Bayesian and Likelihood Methods in Statistics and Econometrics*. Elsevier Science Publishers, North-Holland, 473–488.

Kharroubi, S.A, Sweeting, T.J. (2010). Posterior simulation via the signed root log-likelihood ratio. *Bayesian Analysis*, **5**, 787–816.

Madruga, M., Esteves, L., Wechsler, S. (2001). On the Bayesianity of Pereira-Stern tests. *Test*, **10**, 291–299.

Madruga, M., Pereira, C., Stern, J. (2003). Bayesian Evidence Test for Precise Hypotheses. *Journal of Statistical Planning and Inference*, **117**, 185–198.

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

Pace, L., Salvan, A. (2006). Adjustments of the profile likelihood from a new perspective. *Journal of Statistical Planning and Inference*, **136**, 3554–3564.

Pereira, C.A., Stern, J.M. (1999). Evidence and Credibility: Full Bayesian Significance Test for Precise Hypotheses. *Entropy*, **1**, 99–110.

Pereira, C.A., Stern, J.M. (2001). Model Selection: Full Bayesian Approach. *Environmetrics*, **12**, 559–568.

Pereira, C.A., Stern, J.M., Wechslerz, S. (2008). Can a significance test be genuinely Bayesian? *Bayesian Analysis*, **3**, 79–100.

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

Reid, N. (2003). The 2000 Wald memorial lectures: Asymptotics and the theory of inference. *Annals of Statistics*, **31**, 1695–1731.

Ruli, E., Sartori, N., Ventura, L. (2012). A note on marginal posterior simulation via higher-order tail area approximations. *Working Papers*, 2012.8, Department of Statistical Sciences, University of Padova, Italy.

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

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

Skovgaard, I.M. (2001). Likelihood asymptotics. *Scan. J. Statist.*, **28**, 3–32.

Sweeting, T.J. (1995). A framework for Bayesian and likelihood approximations in statistics. *Biometrika*, **82**, 1–23.

Sweeting, T.J. (1996). Approximate Bayesian computation based on signed roots of log-density ratios (with Discussion). *Bayesian Statistics*, **5**, Ed. J.M. Bernardo, J.O. Berger, A.P. Dawid, and A.F.M. Smith, 427–44. Oxford University Press.

Sweeting, T.J. (1999). On the construction of Bayes-confidence regions. *J. R. Statist. Soc.* B, **61**, 849–861.

Sweeting, T.J., Kharroubi, S.A. (2003). Some new formulae for posterior expectations and Bartlett corrections. *Test*, **12**, 497–521.

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. *Journal of the American Statistical Association*, **81**, 82–86.

Ventura, L., Cabras, S., Racugno, W. (2009). Prior distributions from pseudo-likelihoods in the presence of nuisance parameters. *Journal of the American Statistical Association*, **104**, 768–774.

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

Ventura, L., Sartori, N., Racugno, W. (2013). Objective Bayesian higher-order asymptotics in models with nuisance parameters. *Computational Statistics and Data Analysis*, **60**, 90-96.

Acknowledgements

This work was supported by a grant from the University of Padua (Progetti di Ricerca di Ateneo 2011) and by the Cariparo Foundation Excellence-grant 2011/2012.

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