Approximate Bayesian computation with modified log-likelihood ratios

Laura Ventura & Nancy Reid

METRON

ISSN 0026-1424

METRON
DOI 10.1007/s40300-014-0041-4

Springer

METRON
DOI 10.1007/s40300-014-0041-4

Approximate Bayesian computation with modified log-likelihood ratios

Laura Ventura · Nancy Reid

Received: 30 December 2013 / Accepted: 11 April 2014

Abstract The aim of this contribution is to discuss approximate Bayesian computation based on the asymptotic theory of modified likelihood roots and log-likelihood ratios. Results on third-order approximations for univariate posterior distributions, also in the presence of nuisance parameters, are reviewed and the computation of asymptotic credible sets for a vector parameter of interest is illustrated. All these approximations are available at little additional computational cost over simple first-order approximations. Some illustrative examples are discussed, with particular attention to the use of matching priors.

Keywords Bayesian simulation · Credible set · Higher-order asymptotics · 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

Asymptotic arguments are widely used in Bayesian inference, and in recent years there have been considerable developments of so-called higher-order asymptotics. The aim of this contribution is to discuss recent advances in approximate Bayesian computation based on the theory of higher-order asymptotics. The theory provides very accurate approximations to posterior distributions, and to various summary quantities of interest, including tail areas and credible regions. The approximations are based on modifications of the usual log-likelihood ratio statistic. It is argued that analytic approximations still have an important role to play in Bayesian statistics.

L. Ventura (✉)
Department of Statistical Sciences, University of Padova, Padua, Italy
e-mail: ventura@stat.unipd.it

N. Reid
Department of Statistics, University of Toronto, Toronto, ON, Canada
e-mail: reid@utstat.utoronto.ca

Published online: 01 May 2014

Higher-order approximations for posterior distributions have been widely discussed in Bayesian inference; see, among others, [4,7,8,18,19,24,26,27,32], 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. Moreover, they are particularly useful for sensitivity analyses [11,20,21].

In this paper, recent results on asymptotic approximations based on modifications of the log-likelihood ratio are reviewed and their use in Bayesian computation is illustrated. A first result shows that third-order approximations give rise to a simple simulation scheme [21] for Bayesian computation of marginal posterior distributions for a scalar parameter of interest. Its main advantage over Markov chain Monte Carlo methods is that samples are drawn independently, so that much less computational time is needed. We then show how third-order tail area approximations can be used for testing precise a null hypothesis [5,16]. Finally, we also review how approximate Bayesian computations based on modified log-likelihood ratios can be generalized for a vector parameter of interest [31]. As is the case with the approximations for univariate posterior distributions, the results are based on the asymptotic theory of modified log-likelihood ratios and require only routine maximization output for implementation. The theory is illustrated by some examples, with particular attention to the use of matching priors.

The paper is organized as follows. Section 2 illustrates higher-order Bayesian approximations for inference about a scalar parameter of interest; tail areas, simulation methods, and testing are described. Section 3 indicates how these ideas generalize to the multi-parameter case. Some concluding remarks are given in Sect. 4.

2 Scalar parameter of interest

2.1 Modified likelihood roots

Consider a sampling model $f(y;\theta)$ with scalar parameter $\theta \in \Theta \subseteq \mathbb{R}$, and let $L(\theta) = L(\theta; y) = \exp\{\ell(\theta)\}$ be the likelihood function based on data $y = (y_1, \ldots, y_n)$. 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 to an integral of the form

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

is required. The derivation of a tail area approximation is simple in the scalar case. See, among others, [18,19,24,26,27], [6, Chap. 11], [4, Chap. 8], and references therein. The first step is to consider in (1) the Laplace approximation to the normalizing constant, which gives an approximation to $\pi(\theta|y)$:

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

where $\hat\theta$ is the maximum likelihood estimator (MLE) of θ and $j(\theta) = -\ell''(\theta)$ is the observed information function. The notation $\doteq$ indicates that the approximation has relative error $O(n^{-1})$ as $n \to \infty$, in so-called moderate deviation regions, where $|\theta - \hat\theta| < \delta/\sqrt{n}$.

Expressions for the $O(n^{-1})$ and $O(n^{-2})$ terms are given explicitly in [29]. The posterior survival function for θ is then approximated to the same order by

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

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

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

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

where $r_0 = r(\theta_0)$ and the positive quantity $b(r) = |j(\hat\theta)|^{1/2}\{\pi(\theta)/\pi(\hat\theta)\}\{r(\theta)/\ell'(\theta)\}$ is a function of r. This change of variable expresses the posterior density of r as

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

The final step is a further 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

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

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

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

The notation $\doteqdot$ indicates that the approximation is accurate to order $O(n^{-3/2})$ in moderate deviation regions; see, e.g., [8] or [22, Chap. 2]. The improvement in the accuracy from the density, $O(n^{-1})$, to the distribution function, $O(n^{-3/2})$, follows from detailed examination of the Taylor series expansions relating r, q, and hence r^*.

If $\theta = (\psi, \lambda)$, where ψ is a scalar parameter of interest and λ is a $(d-1)$-dimensional nuisance parameter, then a similar argument can be applied to the marginal posterior density for ψ,

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

The Laplace approximation to (7) is (see, e.g., [18,19,29])

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

where $\ell_p(\psi) = \log L(\psi, \hat{\lambda}_\psi)$ is the profile log-likelihood for ψ, with $\hat{\lambda}_\psi$ the constrained MLE of λ given ψ, $j_p(\psi) = -\partial^2 \ell_p(\psi)/\partial\psi^2$ is the observed information function corresponding to the profile log-likelihood, and $j_{\lambda\lambda}(\psi,\lambda)$ is the (λ,λ)-block of the observed information function from the full log-likelihood $\ell(\psi,\lambda)$. Expression (8) has the same structure as (2), and can be integrated using similar arguments:

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

where

$$\begin{aligned}
r_p &= r_p(\psi) = \text{sign}(\hat{\psi} - \psi)[2(\ell_p(\hat{\psi}) - \ell_p(\psi))]^{1/2}, \\
b(r_p) &= r_p \frac{|j_p(\hat{\psi})|^{1/2}}{\ell'_p(\hat{\psi})} \frac{|j_{\lambda\lambda}(\hat{\psi},\hat{\lambda})|^{1/2}}{|j_{\lambda\lambda}(\psi,\hat{\lambda}_\psi)|^{1/2}} \frac{\pi(\psi,\hat{\lambda}_\psi)}{\pi(\hat{\psi},\hat{\lambda})}, \\
r_p^* &= r_p^*(\psi) = r_p(\psi) + \frac{1}{r_p(\psi)} \log \frac{q_B(\psi)}{r_p(\psi)}, \\
q_B(\psi) &= \ell'_p(\psi)\ |j_p(\hat{\psi})|^{-1/2} \frac{|j_{\lambda\lambda}(\psi,\hat{\lambda}_\psi)|^{1/2}}{|j_{\lambda\lambda}(\hat{\psi},\hat{\lambda})|^{1/2}} \frac{\pi(\hat{\psi},\hat{\lambda})}{\pi(\psi,\hat{\lambda}_\psi)}\,.
\end{aligned}$$

The improvement in the order of the remainder term from $O(n^{-1})$ in (2) to $O(n^{-3/2})$ in (8) is due to the similarity of the integrals in the numerator and denominator [29]. For details of the derivation of (9) see, e.g., [8], although they derive a different version of the approximation, $\Phi(r_p) + \phi(r_p)(1/r_p - 1/q_B)$, accurate to the same order. Formula (9) gives an explicit expression for the posterior quantiles. The derivation implicitly assumes enough regularity in the model to ensure that r_p, and hence r_p^* is a monotone function of ψ. Thus the normal approximation to the distribution of $r_p^*(\psi)$ gives an equi-tailed credible interval for ψ computed as

$$CI = \left\{\psi : |r_p^*(\psi)| \le z_{1-\alpha/2}\right\}, \quad (10)$$

where $z_{1-\alpha/2}$ is the $(1-\alpha/2)$-quantile of the standard normal distribution. From (9) the posterior median of (7) can be computed as the solution $\hat{\psi}^*$ in ψ of the estimating equation $r_p^*(\psi) = 0$.

A version of the tail area approximation (9) can also be developed by expanding the log-posterior density around the posterior mode, with the same order of approximation error, but possibly better finite sample performance (see, e.g., [21]).

2.2 Approximations with matching priors

The order of the approximations of the previous section refers to the posterior distribution function, and may depend more or less strongly on the choice of prior. A so-called matching prior ensures that the credible intervals based on the posterior marginal distribution also have frequentist coverage to $O(n^{-1})$ (see, e.g., [8,25,28]). Using a matching prior, [30,32] showed that approximation (7) to the marginal posterior density for ψ can be expressed to $O(n^{-1})$ as

$$\pi_m(\psi|y) \propto L_{mp}(\psi)\, \pi_{mp}(\psi), \tag{11}$$

where $L_{mp}(\psi) = L_p(\psi)M(\psi)$ is the modified profile likelihood for a suitably defined correction term $M(\psi)$ (discussed, for example, in [22, Chap. 9] and in [15, Chap. 11]) and

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

where $i_{\psi\psi.\lambda}(\psi,\lambda) = i_{\psi\psi}(\psi,\lambda) - i_{\psi\lambda}(\psi,\lambda)i_{\lambda\lambda}(\psi,\lambda)^{-1}i_{\lambda\psi}(\psi,\lambda)$. The use of the matching prior $\pi_{mp}(\psi)$ has the advantage that it does not require elicitation of a prior distribution, and the approximation in (11) can be used to avoid numerical integration over λ, or MCMC simulation, in order to obtain an $O(n^{-1})$ approximation to the marginal posterior density.

Following [32], the marginal posterior density (11) can also be written, to second-order, as

$$\pi_m(\psi|y) \propto \exp\left(-\frac{1}{2}r_p^*(\psi)^2\right)\left|\frac{s_p(\psi)}{r_p(\psi)}\right|, \tag{12}$$

where $s_p(\psi) = \ell'_p(\psi)/j_p(\hat{\psi})^{1/2}$ is the profile score statistic, and $r_p^*(\psi)$ has the form (10), with $q_B(\psi) = q_p(\psi)$, where

$$q_p(\psi) = \frac{\ell'_p(\psi)}{j_p(\hat{\psi})^{1/2}}\frac{i_{\psi\psi.\lambda}(\hat{\psi},\hat{\lambda})^{1/2}}{i_{\psi\psi.\lambda}(\psi,\hat{\lambda}_\psi)^{1/2}}\frac{1}{M(\psi)}. \tag{13}$$

This version of $r_p^*(\psi)$ was derived in [2] as an $O(n^{-1})$ approximation to the non-Bayesian version of $r_p^*(\psi)$ of [1], thus verifying that $\pi_{mp}(\psi)$ is also a strong matching prior, in the sense of [9].

A remarkable advantage of Eq. (12) is that its expression automatically includes the matching prior, without requiring its explicit computation. Moreover, in (12) the modified directed likelihood $r_p^*(\psi)$ may be replaced by the modified directed likelihood of [1] or by the adjusted directed likelihoods discussed in [2]; see also [3, Chap. 6], and [22, Chap. 7]. Indeed, all these versions are closely related to each other in the sense that they are equivalent to second order (see [2, Sect. 5]).

Finally, from (12) accurate tail area probabilities are computable as in (9) with $q_B(\psi) = q_p(\psi)$. In this situation, the credible interval (10) for ψ coincides with an accurate higher-order likelihood-based confidence interval for ψ with approximate level $(1-\alpha)$. Moreover,

the posterior median coincides with the frequentist estimator defined as the zero-level confidence interval based on $r_p^*(\psi)$ [23]. Such an estimator has been shown to be a refinement of the MLE $\hat{\psi}$ [10].

Example 1: Nonlinear regression. In a nonlinear regression model the responses $y_1, \ldots, y_n$ are related to explanatory variables x_i as

$$y_i = \mu(x_i; \beta) + \sigma\, \varepsilon_i, \quad i = 1, \ldots, n, \tag{14}$$

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 $f(\cdot)$. The standard normal density is widely used, especially for dose-response curves in bioassays. A more general form is

$$y_{ij} = \mu(x_i; \beta) + \sigma_i\, \varepsilon_{ij}, \quad i = 1, \ldots, m, \quad j = 1, \ldots, n_i, \tag{15}$$

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. The variance can be modelled as $\sigma_i^2 = \sigma^2 V(x_i; \beta, g)$, where σ^2 and the $q \times 1$ vector g are variance parameters and $V(\cdot)$ is a given function.

A study on a radioimmunoassay (RIA) taken to estimate the concentrations of a drug in samples of porcine serum is discussed in [4, Sect. 5.4]. The experiment consists of 16 observations made at 8 different drug levels with 2 replications at each level. The data are available in the data frame ria of the nlreg 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 the marginal posterior (11) based on the matching prior $\pi_{mp}(\psi)$ can be performed using the profile method available for objects of class nlreg, of the library HOA [4]. Figure 1 (left) gives the plot of the posterior distribution (12) 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 Fig. 1 (right).

2.3 Tail area approximations for Bayesian simulation

Starting from the higher-order tail area approximation (HOTA) (9), it is possible to develop a sampling scheme that gives rise to an accurate computation of marginal posterior densities, and related quantities, such as posterior summaries [21].

The implementation of the HOTA sampling scheme is available at little additional computational cost over simple first-order approximations, and it has the advantage over MCMC methods that samples are drawn independently, resulting in much lower computation time. Starting from (7), the simulation algorithm can be summarized as follows. For $t = 1, \ldots, T$:

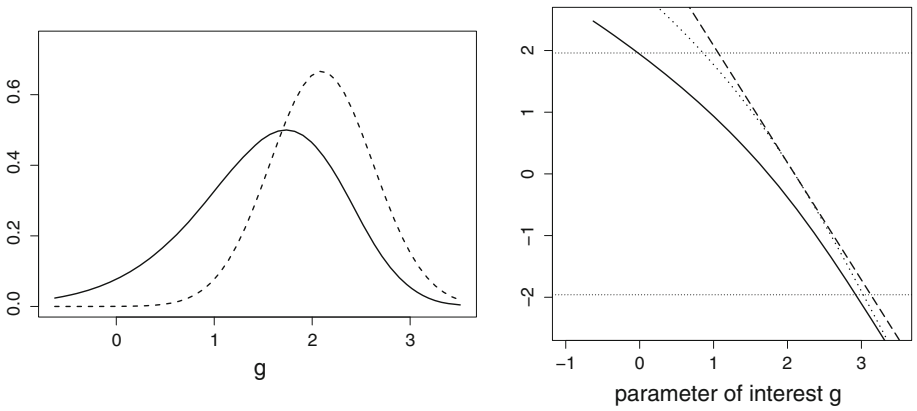

Fig. 1 RIA data. *Left* approximations to the posterior marginal density $\pi_m(\psi|y)$, for $\psi = g$, using the matching prior and (12) (*solid line*) and using the normal approximation (*dashed line*). *Right* the pivotal quantities $r_p^*(\psi)$ (*solid line*), $r_p(\psi)$ (*dotted line*) and Wald (*dashed line*); the *horizontal lines* are the quantiles $\pm z_{0.975}$

1. Draw $z_t \sim N(0, 1)$;
2. Find ψ_t as the solution of $r_p^*(\psi_t) = z_t$.

Then $(\psi_1, \ldots, \psi_T)$ is an approximate sample 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 [21] for details).

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 (9). In this respect, it has an obvious advantage over MCMC algorithms, which usually require more tuning by the practitioner. Another possible use of the HOTA sampling scheme is to provide quick prior sensitivity analyses [11,20]. Indeed, it is possible to easily assess the effect of different priors on marginal posterior distributions, given the same Monte Carlo error. This is not generally true for MCMC or importance sampling methods, which in general have to be tuned for the specific model and prior.

Example 1: Nonlinear regression (cont). The computation, with the HOTA algorithm, of the approximate posterior densities, using (11) and using the normal approximation $\pi_m^I(\psi|y)$ is illustrated in Fig. 2 for the parameters β_1 and g. The total computation time was 4 s. Moreover, we can use $q_B(\psi)$ in (9) to find approximate posterior marginal densities (and related quantities) for any fixed prior, and this gives a quick way to judge sensitivity to the prior (see [20,21]). In Fig. 2 we also show the approximation marginal posterior densities $\pi_f(\psi|y)$ based on a flat prior.

Based on the HOTA simulations, the 0.95 highest posterior density (HPD) interval for β_1 is (1.05, 2.46) using the matching prior; is (1.02, 2.37) using the flat prior; and is (1.42, 2.18) based on the normal approximation to the posterior. The HOTA posterior for g gives an HPD interval of (0.14, 3.03) using the matching prior and (0.13, 3.09) using the flat prior. While the resulting intervals are not very sensitive to the choice of prior, it is clear that the shape of the variance function is not very well determined in this example. The lower bounds for g based on the HPD interval and the equi-tailed interval (see Fig. 1) are also fairly different, reflecting the skewness in the left tail of the posterior marginal.

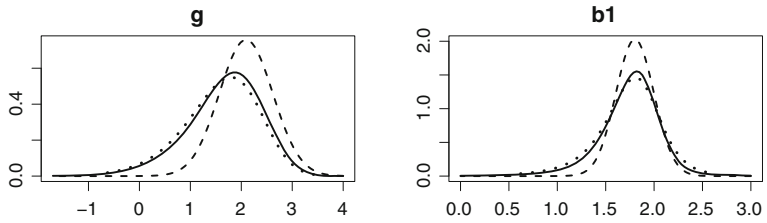

Fig. 2 RIA data and HOTA algorithm: posteriors $\pi_m(\psi|y)$ based on (11) (*solid line*), first-order approximation $\pi_m^I(\psi|y)$ (*dashed line*) and $\pi_f(\psi|y)$ based on a flat prior (*bold dotted line*) for $\psi = g$ (*left*) and for $\psi = \beta_1$ (*right*)

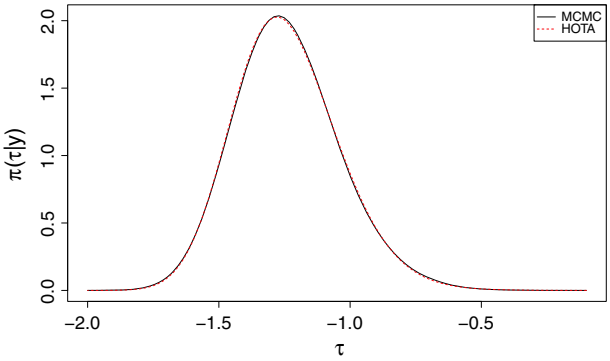

Fig. 3 Censored regression: the HOTA and MCMC simulation methods give nearly identical marginal posterior distributions for the parameter τ

Example 2: Censored regression. This example is discussed in [21]; 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

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

where y_i is the $\log_{10}$(failure time) with time in hours, $x_i = 1{,}000/(\text{temperature} + 273.2)$ 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 $n - m$ are censored at times u_i, the log-likelihood 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 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

Table 1 Censored regression: comparison of marginal posterior densities obtained using Markov chain Monte Carlo (MCMC) and HOTA, for the three parameters in the model

Method	Posterior	Mean	SD	$Q_{0.025}$	Median	$Q_{0.975}$	0.95 HPD
MCMC	$\pi_m(\tau\|y)$	−1.240	0.201	−1.600	−1.253	−0.811	(−1.616, −0.832)
HOTA	$\pi_m(\tau\|y)$	−1.240	0.202	−1.601	−1.251	−0.808	(−1.624,−0.837)
MCMC	$\pi_m(\beta_0\|y)$	−6.204	1.117	−8.570	−6.139	−4.149	(−8.413, −4.010)
HOTA	$\pi_m(\beta_0\|y)$	−6.191	1.128	−8.596	−6.134	−4.130	(−8.475, −4.038)
MCMC	$\pi_m(\beta_1\|y)$	4.409	0.518	3.461	4.382	5.512	(3.425, 5.470)
HOTA	$\pi_m(\beta_1\|y)$	4.401	0.521	3.459	4.370	5.521	(3.398, 5.443)

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 on the two methods are in good agreement.

For the computation with HOTA, grids of 50 points were chosen for each of the parameters and the total number of simulations was $T = 10^5$. The overall computation time on a laptop with 4 GB RAM was 1.8 s. For MCMC with 10^6 simulations which were thinned by taking every 10th observation to reduce the autocorrelation, the computation time was 95 s.

2.4 Tail area approximations for measuring 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 of [16] can be considered; see also [14] and [17], and references therein.

Following [5], consider the set

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

Then the Pereira and Stern posterior evidence EV in favor of H_0 is

$$EV = 1 - \Pr(\psi \in T(y)|y), \tag{16}$$

and large values of EV indicated consistency with the null hypothesis H_0. Using (9), a simple and accurate higher-order approximation of (16) is

$$EV \doteqdot 1 - \Phi(r_p^*(\psi_0)) + \Phi(r_p^*(\psi_0^*)), \tag{17}$$

where ψ_0^* is such that $\pi_m(\psi_0^*|y) = \pi_m(\psi_0|y)$. With respect to the original definition of EV [16], (17) is simpler to compute, in particular when the dimension of the nuisance parameter is large. When matching priors are used, as above, the approximation (17) has the further advantage that it does not require the elicitation of the prior for the nuisance parameter.

Note that

$$\Phi(r_p^*(\psi_0)) - \Phi(r_p^*(\psi_0^*)) \doteqdot \int_{\psi_0}^{\psi_0^*} \pi_m(\psi|y)\, d\psi = Pr(\psi \in T(y)|y) = 1 - EV, \tag{18}$$

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

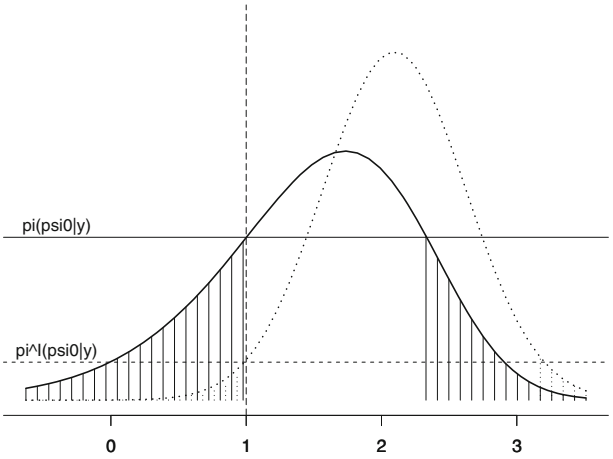

Fig. 4 RIA data: Posteriors $\pi_m(\psi|y)$ (*solid line*) and $\pi_m^I(\psi|y)$ (*dashed line*) for $\psi = g$. The EV measures of evidence for the precise hypothesis $H_0 : \psi = 1$ are the *dashed areas*

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 Fig. 4 for the marginal posterior distribution (11) and for the first-order approximation $\pi_m^I(\psi|y)$.

The computation of (16) for the HOTA posterior distributions gives $EV = 0.26$ using (17), whereas $EV = 0.04$ if the normal approximation to the posterior is used, the latter suggesting evidence against the value $g = 1$, which as we have seen does not seem supported by the data. This illustrates an important advantage of third-order asymptotics with respect to first-order results.

3 Approximations for multidimensional parameters

3.1 No nuisance parameters

Suppose that $\theta \in \Theta \subseteq \mathbb{R}^d$, with $d > 1$. In this section, we illustrate the approximations based on modifications of the log-likelihood ratios. As in the scalar parameter case, the derivation of these approximations can be based on three steps (see [24,31]).

Paralleling the derivation of the tail area approximations discussed in Sect. 2, in the first step the Laplace approximation of $\pi(\theta|y)$ is considered, i.e.,

$$\pi(\theta|y) \doteq (2\pi)^{-d/2}|j(\hat{\theta})|^{1/2}\frac{\pi(\theta)}{\pi(\hat{\theta})}\exp\left\{-\frac{1}{2}W(\theta)\right\}. \tag{19}$$

The second step is a change of variable from θ to a statistic $r_m = r_m(\theta)$, which is asymptotically multivariate standard normal to $O(n^{-1/2})$ and satisfies $W(\theta) = 2\left(\ell(\hat{\theta}) - \ell(\theta)\right) = r_m^{\mathsf{T}} r_m$. To this end, the signed root log-likelihood ratio transformation $r_m(\theta)$ defined in [26,27] can be considered; see also [13]. 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 $\hat{\theta}_{\theta^i}^{(i+1)}$ be the partial MLE of $\theta^{(i+1)}$ given θ^i, and let $\hat{\theta}_{j,\theta^i}$ be the jth component of $(\theta^i, \hat{\theta}_{\theta^i}^{(i+1)})$,

for $j > i$. The signed root log-likelihood ratio transformation is thus given by

$$r_m(\theta) = (r_{m1}, \ldots, r_{md}), \tag{20}$$

with

$$r_{mi} = \text{sign}(\theta_i - \hat{\theta}_{i,\theta^{i-1}}) \left\{ 2 \left[\ell \left(\theta^{i-1}, \hat{\theta}^{(i)}_{\theta^{i-1}} \right) - \ell \left(\theta^i, \hat{\theta}^{(i+1)}_{\theta^i} \right) \right] \right\}^{1/2}, \tag{21}$$

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 $i = 1, \ldots, d$. Moreover, the Jacobian matrix $(dr_m/d\theta)$ is lower triangular, and in particular

$$\left| \frac{dr_m}{d\theta} \right| = \prod_{i=1}^{d} \left| \frac{\ell_i \left(\theta^i, \hat{\theta}^{(i+1)}_{\theta^i} \right)}{r_{mi}} \right|, \tag{22}$$

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 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 $r_m^{\mathsf{T}} \delta(r_m) = \log g(r_m)$, so that

$$-(r_m - \delta(r_m))^{\mathsf{T}} (r_m - \delta(r_m)) = -r_m^{\mathsf{T}} r_m + 2 \log g(r_m) + O(n^{-2}). \tag{23}$$

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

$$w^*_m = w^*_m(\theta) = r_m^{\mathsf{T}} r_m - 2 \log g\{r_m(\theta)\}, \tag{24}$$

with

$$g\{r_m(\theta)\} = |j(\hat{\theta})|^{1/2} \frac{\pi(\theta)}{\pi(\hat{\theta})} \left[\prod_{i=1}^{d} \left| \frac{\ell_i \left(\theta^i, \hat{\theta}^{(i+1)}_{\theta^i} \right)}{r_{mi}} \right| \right]^{-1}. \tag{25}$$

The asymptotic distribution of w^*_m is χ^2_d with relative error $O(n^{-1})$ in a large deviation region (see [24]).

To obtain a statistic which generalizes the scalar version (6), [24] suggests the asymptotically equivalent approximation

$$w^{**}_m = w^{**}_m(\theta) = r_m^{\mathsf{T}} r_m \left(1 - \frac{\log g(r_m)}{r_m^{\mathsf{T}} r_m} \right)^2. \tag{26}$$

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

Paralleling the scalar parameter case, from (24) or (26) a credible region for θ with approximately $100(1-\alpha)\%$ coverage in repeated sampling, can be computed as

$$CR = \left\{ \theta : w^{**} \leq \chi^2_{d;1-\alpha} \right\}, \tag{27}$$

based on w^{**}, or equivalently on w^*. This region can be interpreted as the extension to the multidimensional case of the set (10). See [31] for some simulation studies on CR.

Example 3: Nonlinear regression. Let us consider a nonlinear regression model of the form (15), with $\mu(x_i; \beta) = \beta_1(1 - \exp(-\beta_2 x_i))$ and with $\sigma_i^2 = 0.29$ [6, Sect. 10.1] discusses this model for the calcium data, for which the response is the calcium uptake of cells in hot calcium suspension, and the covariate is time in minutes; $n_i = 3$, for $i = 1, \ldots, 9$.

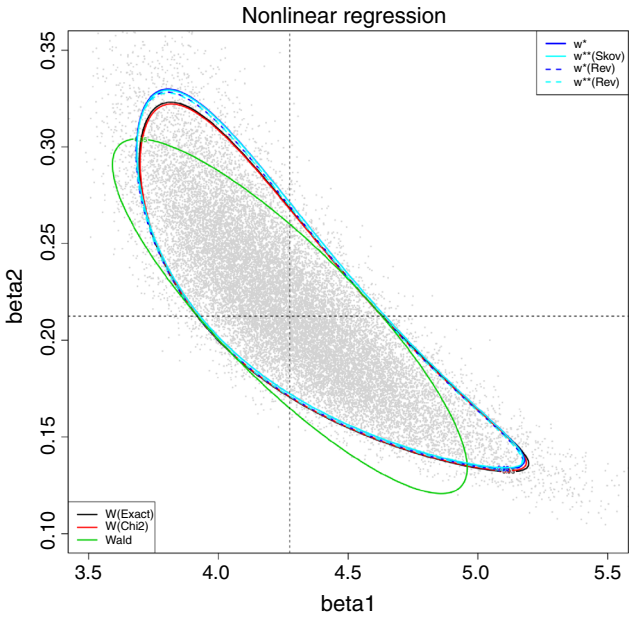

Fig. 5 Calcium data: credible regions for (β_1, β_2)

Figure 5 gives the contours of several credible regions for (β_1, β_2), i.e., the first-order credible regions (Wald)

$$CR_N = \left\{\theta : (\theta - \tilde{\theta})^{\mathsf{T}} j(\tilde{\theta})(\theta - \tilde{\theta}) \leq \chi^2_{d;1-\alpha}\right\}, \tag{28}$$

where $\tilde{\theta}$ is the posterior mode and $j(\tilde{\theta}) = -\partial \log \pi(\theta|y)/(\partial\theta\partial\theta^{\mathsf{T}})$, the likelihood-type credible regions (W(Chi2))

$$CR_L = \left\{\theta : -2\log\frac{\pi(\theta|y)}{\pi(\tilde{\theta}|y)} \leq \chi^2_{d;1-\alpha}\right\}, \tag{29}$$

the 95 % HPD credible region (W(Exact)) computed with MCMC simulation, and two versions derived from w_m^* and w_m^{**}. The latter are labelled W*, W** for one ordering of the parameters in the transformation (20), and W*(Rev), W**(Rev) for an inversion of the parameter order. The posterior probability of CR_N is 0.9129, of CR_L is 0.9478, and of (27) is 0.95133. These results indicate that the accuracy of the credible region CR is very high.

3.2 With nuisance parameters

Suppose now that $\theta = (\psi, \lambda)$, with ψ the parameter of interest of dimension k and λ the nuisance parameter of dimension $d - k$, and consider the approximations based on modifications of the log-likelihood ratios.

In the first step we use the Laplace approximation (8) of the marginal posterior $\pi_m(\psi|y)$, given by

$$\pi_m(\psi|y) \doteq (2\pi)^{-k/2}|j_p(\hat{\psi})|^{1/2}\frac{\pi(\psi,\hat{\lambda}_\psi)}{\pi(\hat{\psi},\hat{\lambda})}\exp\left\{-\frac{1}{2}W_p(\psi)\right\}\frac{|j_{\lambda\lambda}(\hat{\psi},\hat{\lambda})|^{1/2}}{|j_{\lambda\lambda}(\psi,\hat{\lambda}_\psi)|^{1/2}}, \tag{30}$$

where $W_p(\psi) = 2(\ell_p(\hat{\psi}) - \ell_p(\psi))$ is the profile log likelihood ratio.

The second step is a change of variable from ψ to a statistic $r_{mp} = r_{mp}(\psi)$, such that it is asymptotically multivariate standard normal to $O(n^{-1/2})$ and such that for the profile log-likelihood ratio we have $W_p(\psi) = r_{mp}^{\mathsf{T}} r_{mp}$. To this end, the profile version of the signed root log-likelihood ratio transformation $r_m(\theta)$ can be considered.

Let $\theta = (\psi_1, \ldots, \psi_k, \lambda) = (\psi^i, \psi^{(i+1)}, \lambda)$, where $\psi^i = (\psi_1, \ldots, \psi_i)$, $i \leq k$, is the vector of the first i components of ψ and $\psi^{(i+1)} = (\psi_{i+1}, \ldots, \psi_k)$. Let $\hat{\psi}^{(i+1)}_{\psi^i}$ and $\hat{\lambda}_{\psi^i}$ be the partial MLEs of $\psi^{(i+1)}$ and λ, respectively, given ψ^i, and let $\hat{\psi}_{j,\psi^i}$ be the jth component of $(\psi^i, \hat{\psi}^{(i+1)}_{\psi^i}, \hat{\lambda}_{\psi^i})$, for $i < j$. The profile signed root log-likelihood ratio transformation can be written as

$$r_{mp}(\psi) = (r_{mp1}, \ldots, r_{mpk}), \tag{31}$$

with

$$r_{mpi} = \text{sign}(\psi_i - \hat{\psi}_{i,\psi^{i-1}}) \left\{ 2 \left[\ell \left(\psi^{i-1}, \hat{\psi}^{(i)}_{\psi^{i-1}}, \hat{\lambda}_{\psi^{i-1}} \right) - \ell \left(\psi^i, \hat{\psi}^{(i+1)}_{\psi^i}, \hat{\lambda}_{\psi^i} \right) \right] \right\}^{1/2}, \tag{32}$$

for $i = 1, \ldots, k$. The determinant of the Jacobian matrix $(dr_{mp}/d\psi)$ is

$$\left| \frac{dr_{mp}}{d\psi} \right| = \prod_{i=1}^{k} \left| \frac{\ell_i \left(\psi^i, \hat{\psi}^{(i+1)}_{\psi^i}, \hat{\lambda}_{\psi^i} \right)}{r_{mpi}} \right|. \tag{33}$$

The last step is again a change of variable from r_{mp} to a more accurate version of the form $r^*_{mp} = r^*_{mp}(\psi) = r_{mp} - \delta(r_{mp})$, with $\delta = \delta(r_{mp})$ chosen to satisfy $r_{mp}^{\mathsf{T}} \delta(r_{mp}) = \log g(r_{mp})$, so that

$$-(r_{mp} - \delta(r_{mp}))^{\mathsf{T}} (r_{mp} - \delta(r_{mp})) = -r_{mp}^{\mathsf{T}} r_{mp} + 2 \log g(r_{mp}) + O(n^{-2}). \tag{34}$$

Actually, we only need the existence of $\delta(r_{mp})$ to calculate

$$w^*_{mp} = w^*_{mp}(\psi) = W_p(\psi) - 2 \log g(r_{mp}(\psi)), \tag{35}$$

with

$$g(r_{mp}(\psi)) = |j_p(\hat{\psi})|^{1/2} \frac{\pi(\psi, \hat{\lambda}_\psi)}{\pi(\hat{\psi}, \hat{\lambda})} \frac{|j_{\lambda\lambda}(\hat{\psi}, \hat{\lambda})|^{1/2}}{|j_{\lambda\lambda}(\psi, \hat{\lambda}_\psi)|^{1/2}} \left[\prod_{i=1}^{k} \left| \frac{\ell_i \left(\psi^i, \hat{\psi}^{(i+1)}_{\psi^i}, \hat{\lambda}_{\psi^i} \right)}{r_{mpi}} \right| \right]^{-1}. \tag{36}$$

The asymptotic distribution of w^*_{mp} is χ^2_k with error $O(n^{-1})$. Alternatively, the asymptotically equivalent approximation

$$w^{**}_{mp} = w^{**}_{mp}(\psi) = r_{mp}^{\mathsf{T}} r_{mp} \left(1 - \frac{\log g(r_{mp})}{r_{mp}^{\mathsf{T}} r_{mp}} \right)^2 \tag{37}$$

can be used. A credible region for ψ can be computed as $CR = \{\psi : w^{**}_{mp}(\psi) \leq \chi^2_{k;(1-\alpha)}\}$, as in the previous section. Simulation studies of $w^*_{mp}(\psi)$ and $w^{**}_{mp}(\psi)$ are under study.

4 Remarks

This paper aims to outline how approximate computational tools have a role to play in the modern era of Bayesian statistics, where high computational 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 log-likelihood 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., [20,21]).

A key feature of the approximations discussed and developed in this paper is that they do not require the calculation of log-likelihood 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 [12], for example, for precise regularity conditions; the most important is that the posterior density have a unique mode, and is smoothly differentiable in a moderate deviation region of the mode.

Acknowledgments 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.

References

1. Barndorff-Nielsen, O.E.: Modified signed log likelihood ratio. Biometrika **78**, 557–563 (1991)
2. Barndorff-Nielsen, O.E., Chamberlin, S.R.: Stable and invariant adjusted directed likelihoods. Biometrika **81**, 485–499 (1994)
3. Barndorff-Nielsen, O.E., Cox, D.R.: Inference and asymptotics. Chapman and Hall, London (1994)
4. Brazzale, A.R., Davison, A.C., Reid, N.: Applied asymptotics. Cambridge University Press, Cambridge (2007)
5. Cabras, S., Ventura, L., Racugno, W.: Higher-order asymptotics for Bayesian significance tests for precise null hypotheses in the presence of nuisance parameters. Manuscript under revision (2013)
6. Davison, A.C.: Statistical models. Cambridge University Press, Cambridge (2003)
7. DiCiccio, T.J., Field, C.A., Fraser, D.A.S.: Approximations of marginal tail probabilities and inference for scalar parameters. Biometrika **77**, 77–95 (1990)
8. DiCiccio, T.J., Martin, A.M.: Simple modifications for signed roots of likelihood ratio statistics. J. R. Stat. Soc. B **55**, 305–316 (1993)
9. Fraser, D.A.S., Reid, N.: Strong matching of frequentist and bayesian parametric inference. J. Stat. Plan. Inference **103**, 263–285 (2002)
10. Giummolé, F., Ventura, L.: Practical point estimation from higher-order pivots. J. Stat. Comput. Simul. **72**, 419–430 (2002)
11. Kass, R.E., Tierney, L., Kadane, J.: Approximate methods for assessing influence and sensitivity in bayesian analysis. Biometrika **76**, 663–674 (1989)
12. Kass, R.E., Tierney, L., Kadane, J.: The validity of posterior expansions based on Laplace's method. In: Bayesian and likelihood methods in statistics and econometrics, pp. 473–488. Elsevier Science Publishers, North-Holland (1990)
13. Kharroubi, S.A., Sweeting, T.J.: Posterior simulation via the signed root log-likelihood ratio. Bayesian Anal. **5**, 787–816 (2010)
14. Madruga, M., Pereira, C., Stern, J.: Bayesian evidence test for precise hypotheses. J. Stat. Plan. Inference **117**, 185–198 (2003)
15. Pace, L., Salvan, A.: Adjustments of the profile likelihood from a new perspective. J. Stat. Plan. Inference **136**, 3554–3564 (2006)
16. Pereira, C.A., Stern, J.M.: Evidence and credibility: full bayesian significance test for precise hypotheses. Entropy **1**, 99–110 (1999)

17. Pereira, C.A., Stern, J.M., Wechslerz, S.: Can a significance test be genuinely bayesian? Bayesian Anal. **3**, 79–100 (2008)
18. Reid, N.: Likelihood and bayesian approximation methods. Bayesian Stat. **5**, 351–368 (1995)
19. Reid, N.: The 2000 wald memorial lectures: asymptotics and the theory of inference. Ann. Statist. **31**, 1695–1731 (2003)
20. Reid, N., Sun, Y.: Assessing sensitivity to priors using higher order approximations. Commun. Stat. Theory Methods **39**, 1373–1386 (2010)
21. Ruli, E., Sartori, N., Ventura, L.: Marginal posterior simulation via higher-order tail area approximations. Bayesian Anal. **9**, 129–146 (2014)
22. Severini, T.A.: Likelihood methods in statistics. Oxford University Press, Oxford (2000)
23. Skovgaard, I.M.: A review of higher order likelihood inference. Bull. Int. Stat. Inst. **53**, 331–351 (1989)
24. Skovgaard, I.M.: Scand. J. Stat. **28**, 3–32 (2001)
25. Staicu, A.M., Reid, N.: On probability matching priors. Can. J. Stat. **36**, 613–622 (2008)
26. Sweeting, T.J.: A framework for bayesian and likelihood approximations in statistics. Biometrika **82**, 1–23 (1995)
27. Sweeting, T.J.: Approximate bayesian computation based on signed roots of log-density ratios (with discussion). Bayesian Stat. **5**, 427–444 (1996)
28. Tibshirani, R.: Noninformative priors for one parameter of many. Biometrika **76**, 604–608 (1989)
29. Tierney, L.J., Kadane, J.B.: Accurate approximations for posterior moments and marginal densities. J. Am. Stat. Assoc. **81**, 82–86 (1986)
30. Ventura, L., Cabras, S., Racugno, W.: Prior distributions from pseudo-likelihoods in the presence of nuisance parameters. J. Am. Stat. Assoc. **104**, 768–774 (2009)
31. Ventura, L., Ruli, E., Racugno, W.: A note on approximate bayesian credible sets based on modified log-likelihood ratios. Stat. Prob. Lett. **83**, 2467–2472 (2013)
32. Ventura, L., Sartori, N., Racugno, W.: Objective bayesian higher-order asymptotics in models with nuisance parameters. Comput. Stat. Data Anal. **60**, 90–96 (2013)