Bayesian Analysis (2014) **9**, Number 1, pp. 129–146

Marginal Posterior Simulation via Higher-order Tail Area Approximations

Erlis Ruli * , Nicola Sartori * and Laura Ventura *

Abstract. A new method for posterior simulation is proposed, based on the combination of higher-order asymptotic results with the inverse transform sampler. This method can be used to approximate marginal posterior distributions, and related quantities, for a scalar parameter of interest, even in the presence of nuisance parameters. Compared to standard Markov chain Monte Carlo methods, its main advantages are that it gives independent samples at a negligible computational cost, and it allows prior sensitivity analyses under the same Monte Carlo variation. The method is illustrated by a genetic linkage model, a normal regression with censored data and a logistic regression model.

Keywords: Asymptotic expansion, Bayesian computation, Inverse transform sampling, Marginal posterior distribution, MCMC, Modified likelihood root, Nuisance parameter, Sensitivity analysis.

1 Introduction

Consider a parametric statistical model with density $f(y;\boldsymbol{\theta})$, with $\boldsymbol{\theta}=(\psi,\boldsymbol{\lambda})$, $\boldsymbol{\theta}\in\boldsymbol{\Theta}\subseteq \mathbb{R}^d$ $(d>1)$, where ψ is a scalar parameter of interest and $\boldsymbol{\lambda}$ is a $(d-1)$-dimensional nuisance parameter. Let $\ell(\boldsymbol{\theta})=\ell(\psi,\boldsymbol{\lambda})=\ell(\psi,\boldsymbol{\lambda};\mathbf{y})$ denote the log-likelihood function based on data $\mathbf{y}=(y_1,\ldots,y_n)$, $\pi(\boldsymbol{\theta})=\pi(\psi,\boldsymbol{\lambda})$ be a prior distribution of $(\psi,\boldsymbol{\lambda})$ and let $\pi(\boldsymbol{\theta}|\mathbf{y})=\pi(\psi,\boldsymbol{\lambda}|\mathbf{y})\propto\pi(\psi,\boldsymbol{\lambda})\exp\{\ell(\psi,\boldsymbol{\lambda})\}$ be the posterior distribution of $(\psi,\boldsymbol{\lambda})$. Bayesian inference on ψ, in the presence of the nuisance parameter $\boldsymbol{\lambda}$, is based on the marginal posterior distribution

$$\pi(\psi|\mathbf{y}) = \int \pi(\psi,\boldsymbol{\lambda}|\mathbf{y})\,d\boldsymbol{\lambda}, \qquad (1)$$

which is typically approximated numerically, by means of Monte Carlo integration methods. In order to approximate (1), a variety of Markov chain Monte Carlo (MCMC) schemes have been proposed in the literature (see, e.g., Robert and Casella 2004). However, MCMC methods in practice may need to be specifically tailored to the particular model (e.g. choice of proposal, convergence checks, etc.) and they may have poor tail behavior, especially when d is large.

Parallel with these simulation-based procedures has been the development of analytical higher-order approximations for parametric inference in small samples (see, e.g., Brazzale and Davison 2008, and references therein). Using higher-order asymptotics it is possible to avoid the difficulties related to MCMC methods and obtain accurate

*Department of Statistics, University of Padova, Padova, Italy ruli@stat.unipd.it

© 2014 International Society for Bayesian Analysis DOI:10.1214/13-BA851

approximations to (1), and related tail area probabilities (see, e.g., Reid 1996, 2003; Sweeting 1996, and Brazzale et al. 2007). These methods are highly accurate in many situations, but are nevertheless underused compared to simulation-based procedures (Brazzale and Davison 2008).

The aim of this paper is to discuss a new posterior sampling scheme, which is obtained by combining higher-order tail area approximations with the inverse transform sampler (see, e.g., Robert and Casella 2004, Chap. 2). The proposed method, called HOTA (Higher-Order Tail Area), gives accurate approximations of marginal posterior distributions, and related quantities, also in the presence of nuisance parameters.

The HOTA sampling scheme is straightforward to implement, since it is available at little additional computational cost over simple first-order approximations. It is based on an asymptotic expansion of the log-posterior distribution around the posterior mode. In principle, the whole procedure requires as an input only the unnormalized log-posterior distribution. The method can be applied to a wide variety of regular statistical models, with the essential requirement of the posterior mode being unique (see Kass et al. 1990, for precise regularity conditions). When the posterior mode is close to the maximum likelihood estimate (MLE), then an asymptotic expansion around the MLE can be used. The latter approximation allows the use of standard maximum likelihood routines for Bayesian analysis.

The proposed simulation scheme gives independent samples at a negligible computational cost from a third-order approximation to the marginal posterior distribution. These are distinct advantages with respect to MCMC methods, which in general are time consuming and provide dependent samples. Nevertheless, MCMC techniques give samples from the full posterior distribution subject only to Monte Carlo error, provided convergence has been reached. On the other hand, HOTA has an easily bounded Monte Carlo error, while it has an asymptotic error for the approximation to the true marginal posterior distribution, which depends on the sample size. This approximation is typically highly accurate even for small n.

One possible use of the HOTA sampling scheme is for quick prior sensitivity analyses (Kass et al. 1989; Reid and Sun 2010). 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.

The use of higher-order approximations for posterior simulation is a novel approach in the Bayesian literature. An exception is given by Kharroubi and Sweeting (2010), where a multivariate signed root log-likelihood ratio is used to obtain a suitable importance function. The proposal by Kharroubi and Sweeting (2010) deals with issues related to the importance function and importance weights, which are avoided by the HOTA sampling scheme. Moreover, the performance of HOTA is independent of the ordering of the parameters, required by the multivariate signed root log-likelihood ratio. On the other hand, Kharroubi and Sweeting's method has the advantage that it generates samples from the full posterior distribution, which can be used to obtain an integrated likelihood for doing model selection or for computing predictive distributions.

The paper is organized as follows. Section 2 briefly reviews higher-order approximations for the marginal posterior distribution (1), and for the corresponding tail area. Section 3 describes the proposed HOTA sampling scheme and its implementation. Numerical examples and sensitivity analyses are discussed in Section 4. Finally, some concluding remarks are given in Section 5.

2 Background on higher-order asymptotics

Let $\widetilde{\ell}(\boldsymbol{\theta}) = \ell(\boldsymbol{\theta}) + \log \pi(\boldsymbol{\theta})$ be the unnormalized log-posterior, $\widetilde{\boldsymbol{\theta}} = (\widetilde{\psi}, \widetilde{\boldsymbol{\lambda}})$ the posterior mode, and $\widetilde{\boldsymbol{\lambda}}_\psi$ the posterior mode of $\boldsymbol{\lambda}$ for fixed ψ. The basic requirements for the approximations given in this section are that there exists an unique posterior mode and that the Hessian of $\widetilde{\ell}(\boldsymbol{\theta})$ evaluated at the the full mode or at the constrained posterior mode is negative definite (see for instance Kass et al. 1990). These assumptions are typically satisfied in many commonly used parametric models.

The marginal posterior distribution (1) can be approximated with the Laplace formula (Tierney and Kadane 1986). It gives

$$\pi(\psi|y) \doteq \frac{1}{\sqrt{2\pi}} \exp\{\widetilde{\ell}(\psi, \widetilde{\boldsymbol{\lambda}}_\psi) - \widetilde{\ell}(\widetilde{\psi}, \widetilde{\boldsymbol{\lambda}})\} \, \frac{|\widetilde{\jmath}(\widetilde{\psi}, \widetilde{\boldsymbol{\lambda}})|^{1/2}}{|\widetilde{\jmath}_{\boldsymbol{\lambda}\boldsymbol{\lambda}}(\psi, \widetilde{\boldsymbol{\lambda}}_\psi)|^{1/2}}, \tag{2}$$

where $\tilde{\jmath}(\psi, \boldsymbol{\lambda}) = -\partial^2 \widetilde{\ell}(\boldsymbol{\theta})/(\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T)$ is the $(d \times d)$ negative Hessian matrix of the log-posterior, $\widetilde{\jmath}_{\boldsymbol{\lambda}\boldsymbol{\lambda}}(\psi, \boldsymbol{\lambda})$ is the $(\boldsymbol{\lambda}, \boldsymbol{\lambda})$-block of $\widetilde{\jmath}(\psi, \boldsymbol{\lambda})$, and the symbol "$\doteq$" indicates accuracy with relative error of order $O(n^{-3/2})$. Moreover, the accuracy of (2) is uniform on compact sets of ψ.

From (2), a third-order approximation to the marginal posterior tail area (see, e.g., Davison 2003, Sec. 11.3.1) can be obtained as follows. Starting from

$$\int_{\psi_0}^{+\infty} \pi(\psi|y)\, d\psi \doteq \frac{1}{\sqrt{2\pi}} \int_{\psi_0}^{+\infty} \exp\{\widetilde{\ell}(\psi, \widetilde{\boldsymbol{\lambda}}_\psi) - \widetilde{\ell}(\widetilde{\psi}, \widetilde{\boldsymbol{\lambda}})\} \, \frac{|\widetilde{\jmath}(\widetilde{\psi}, \widetilde{\boldsymbol{\lambda}})|^{1/2}}{|\widetilde{\jmath}_{\boldsymbol{\lambda}\boldsymbol{\lambda}}(\psi, \widetilde{\boldsymbol{\lambda}}_\psi)|^{1/2}}\, d\psi \ , \tag{3}$$

change the variable of integration from ψ to $\widetilde{r}(\psi)$, where $\widetilde{r}(\psi) = \mathrm{sign}(\widetilde{\psi} - \psi)[2(\widetilde{\ell}(\widetilde{\psi}, \widetilde{\boldsymbol{\lambda}}) - \widetilde{\ell}(\psi, \widetilde{\boldsymbol{\lambda}}_\psi))]^{1/2}$. The Jacobian is $-\widetilde{\ell}_\psi(\psi)/\widetilde{r}(\psi)$, where $\widetilde{\ell}_\psi(\psi) = \partial\widetilde{\ell}(\psi, \widetilde{\boldsymbol{\lambda}}_\psi)/\partial\psi$. Hence, we obtain

$$\int_{\psi_0}^{+\infty} \pi(\psi|y)\, d\psi \doteq \frac{1}{\sqrt{2\pi}} \int_{\widetilde{r}(\psi_0)}^{+\infty} \exp\left\{-\frac{1}{2}\widetilde{r}^2\right\} \frac{\widetilde{r}}{\widetilde{\ell}_\psi(\psi)} \frac{|\widetilde{\jmath}(\widetilde{\psi}, \widetilde{\boldsymbol{\lambda}})|^{1/2}}{|\widetilde{\jmath}_{\boldsymbol{\lambda}\boldsymbol{\lambda}}(\psi, \widetilde{\boldsymbol{\lambda}}_\psi)|^{1/2}}\, d\,\widetilde{r}. \tag{4}$$

The final step is an additional change of variable from $\tilde{r}(\psi)$ to

$$\widetilde{r}^\star(\psi) \,=\, \widetilde{r}(\psi) + \frac{1}{\widetilde{r}(\psi)} \log \frac{\widetilde{q}_B(\psi)}{\widetilde{r}(\psi)}, \tag{5}$$

with

$$\widetilde{q}_B(\psi) \,=\, \widetilde{\ell}_\psi(\psi)\, \frac{|\widetilde{\jmath}_{\boldsymbol{\lambda}\boldsymbol{\lambda}}(\psi, \widetilde{\boldsymbol{\lambda}}_\psi)|^{1/2}}{|\widetilde{\jmath}(\widetilde{\psi}, \widetilde{\boldsymbol{\lambda}})|^{1/2}}. \tag{6}$$

Since the Jacobian of this transformation contributes only to the error of (4), it can be shown that the approximate posterior tail area is given by

$$\int_{\psi_0}^{+\infty} \pi(\psi|y)\, d\psi \doteq \frac{1}{\sqrt{2\pi}} \int_{\widetilde{r}^\star(\psi_0)}^{+\infty} \exp\left\{-\frac{1}{2}t^2\right\} dt \;=\; \Phi\{\widetilde{r}^\star(\psi_0)\}, \tag{7}$$

where $\Phi(\cdot)$ is the standard normal distribution function. Expression (7) relies entirely on simple posterior quantities. This is a remarkable computational advantage since these quantities can be computed with any software that performs optimizations and numerical derivatives.

When the posterior mode and the MLE are close, the tail area approximation (7) can be obtained also by considering expansions around the MLE with the same order of approximation error. In particular, let $\hat{\boldsymbol{\theta}} = (\hat{\psi}, \hat{\boldsymbol{\lambda}})$ be the MLE and let $\hat{\boldsymbol{\lambda}}_\psi$ be the constrained MLE of $\boldsymbol{\lambda}$ for fixed ψ. Moreover, let $\ell_p(\psi) = \ell(\psi, \hat{\boldsymbol{\lambda}}_\psi)$ be the profile log-likelihood, $j(\boldsymbol{\theta}) = -\partial^2 \ell(\boldsymbol{\theta})/(\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T)$ the observed Fisher information matrix and $j_p(\psi) = -\partial^2 \ell_p(\psi)/\partial\psi$ the profile information.

The marginal posterior distribution (1) approximated with the Laplace formula becomes (see, e.g., Reid 1996)

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

Then following the same steps as before, the higher-order tail area approximation for (1) can be written as

$$\int_{\psi_0}^{+\infty} \pi(\psi|\mathbf{y}) d\psi \quad \doteq \quad \Phi\{r_p^\star(\psi_0)\}, \tag{9}$$

where $r_p^\star(\psi) = r_p(\psi) + r_p(\psi)^{-1} \log\{q_B(\psi)/r_p(\psi)\}$ is the modified likelihood root, $r_p(\psi) = \text{sign}(\hat{\psi} - \psi)[2(\ell_p(\hat{\psi}) - \ell_p(\psi))]^{1/2}$ is the likelihood root, and

$$q_B(\psi) \;=\; \ell_p'(\psi)\, j_p(\psi)^{-1/2}\, \frac{|j_{\boldsymbol{\lambda}\boldsymbol{\lambda}}(\psi, \hat{\boldsymbol{\lambda}}_\psi)|^{1/2}}{|j_{\boldsymbol{\lambda}\boldsymbol{\lambda}}(\hat{\psi}, \hat{\boldsymbol{\lambda}})|^{1/2}}\, \frac{\pi(\hat{\psi}, \hat{\boldsymbol{\lambda}})}{\pi(\psi, \hat{\boldsymbol{\lambda}}_\psi)}\ , \tag{10}$$

with $\ell_p'(\psi) = \partial \ell_p(\psi)/\partial\psi$ the profile score. Obviously, when $\pi(\boldsymbol{\theta}) \propto 1$, expressions (7) and (9) coincide.

When the class of matching priors (see Tibshirani 1989) is considered in (1), the marginal posterior distribution for ψ can be expressed as (Ventura et al. 2009, 2013)

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

where $L_{mp}(\psi) = L_p(\psi) M(\psi)$ is the modified profile likelihood for a suitably defined correction term $M(\psi)$ (see, among others, Severini 2000, Chap. 9 and Pace and Salvan 2006), and the corresponding matching prior is

$$\pi_{mp}(\psi) \quad \propto \quad i_{\psi\psi.\boldsymbol{\lambda}}(\psi, \hat{\boldsymbol{\lambda}}_\psi)^{1/2}, \tag{12}$$

with $i_{\psi\psi.\boldsymbol{\lambda}}(\psi,\boldsymbol{\lambda}) = i_{\psi\psi}(\psi,\boldsymbol{\lambda}) - i_{\psi\boldsymbol{\lambda}}(\psi,\boldsymbol{\lambda}) i_{\boldsymbol{\lambda}\boldsymbol{\lambda}}(\psi,\boldsymbol{\lambda})^{-1} i_{\psi\boldsymbol{\lambda}}(\psi,\boldsymbol{\lambda})^T$ partial information, and $i_{\psi\psi}(\psi,\boldsymbol{\lambda})$, $i_{\psi\boldsymbol{\lambda}}(\psi,\boldsymbol{\lambda})$, and $i_{\boldsymbol{\lambda}\boldsymbol{\lambda}}(\psi,\boldsymbol{\lambda})$ blocks of the expected Fisher information $i(\psi,\boldsymbol{\lambda})$ from $\ell(\psi,\boldsymbol{\lambda})$. In Ventura and Racugno (2011) it is shown that (9) holds with $r_p^\star(\psi)$ given by the modified profile likelihood root of Barndorff-Nielsen and Chamberlin (1994); see also Barndorff-Nielsen and Cox (1994) and Severini (2000, Chap. 7). In particular, the quantity (10) has the form

$$q_B(\psi) \;=\; \frac{\ell_p'(\psi)}{j_p(\hat\psi)^{1/2}} \frac{i_{\psi\psi.\boldsymbol{\lambda}}(\hat\psi,\hat{\boldsymbol{\lambda}})^{1/2}}{i_{\psi\psi.\boldsymbol{\lambda}}(\psi,\hat{\boldsymbol{\lambda}}_\psi)^{1/2}} \frac{1}{M(\psi)} \;.$$

3 Posterior simulation via tail area approximations

Expressions (7) and (9) give accurate approximation of quantiles of the marginal posterior distribution, but it is not possible to use them to obtain posterior summaries, such as posterior moments or highest posterior density (HPD) regions. One possibility to obtain posterior summaries could be to integrate numerically (2) or (8). However, even though ψ is scalar, numerical integration may become time consuming since a large number of function evaluations is needed to obtain accurate estimates, especially when d is large. In fact, a first numerical integration is needed to compute the normalizing constant and then several numerical integrations are needed for each required posterior summary.

In the following we introduce the HOTA simulation scheme, which is based on the combination of (7) or (9) with the inverse transform sampling. Its main advantage is that it gives independent samples with negligible computational time. Indeed, its implementation only requires a few function evaluations (e.g., 50), independently of the number of simulations. As happens in every simulation method, the HOTA simulation scheme is subject to Monte Carlo error of order $O_p(T^{-1/2})$, where T is the number of Monte Carlo trials. On the other hand, since the samples are drawn independently, it is easy to control such Monte Carlo error by taking T large enough. Finally, it is important to note that HOTA samples from a third-order approximation of the marginal posterior distribution, whose accuracy depends on the sample size. However, the approximation is typically highly accurate even for small sample sizes.

The HOTA simulation scheme is summarized in Algorithm 1, and it can be implemented in two versions: HOTA$_\pi$, based on $\widetilde{r}^\star(\psi)$ and inversion of (7), and HOTA$_\ell$, based on $r_p^\star(\psi)$ and inversion of (9).

Algorithm 1 HOTA for marginal posterior simulation.

1: **for** $t = 1 \to T$ **do**
2: draw $z_t \sim \mathsf{N}(0,1)$
3: solve $\widetilde{r}^\star(\psi_t) = z_t$ (or $r_p^\star(\psi_t) = z_t$) in ψ_t
4: **end for**
5: store $\boldsymbol{\psi}$ as an approximate sample from $\pi(\psi|y)$.

To implement Algorithm 1, we need to invert $\widetilde{r}^\star(\psi)$ or $r_p^\star(\psi)$. This may be done as follows. Generate $\mathbf{z} = (z_1, \ldots, z_T)$ from the standard normal distribution. Fix a grid of values ψ_j, $j = 1, \ldots, N$, for a moderate value of N (e.g. 50-100), equally spaced in the interval $[\psi_{(1)}, \psi_{(T)}]$. The extremes of the grid can be found by solving numerically $z_{(T)} = \widetilde{r}^\star(\psi_{(1)})$ and $z_{(1)} = \widetilde{r}^\star(\psi_{(T)})$ (or the corresponding for $r_p^\star(\psi)$), where $z_{(1)}$ and $z_{(T)}$ are the minimum and maximum of $\mathbf{z}$, respectively. Then, evaluate $\widetilde{r}^\star(\psi)$ (or $r_p^\star(\psi)$) over the grid of ψ values. Finally, apply a spline interpolator to $(\widetilde{r}^\star(\psi_j), \psi_j)$, $j = 1, \ldots, N$, and then obtain the predicted value of ψ_t, corresponding to the values z_t, for $t = 1, \ldots, T$.

Typically, both $\widetilde{r}^\star(\psi)$ and $r_p^\star(\psi)$ are monotonically decreasing functions in ψ and have a numerical discontinuity at $\widetilde{\psi}$ and $\hat{\psi}$, respectively. This is not a concern for practical purposes as it can be avoided by the numerical interpolation described above (see Brazzale et al. 2007, Sec. 9.3). To this purpose, it may be necessary to exclude values of ψ in a δ-neighborhood of $\widetilde{\psi}$ or $\hat{\psi}$ (for instance $\hat{\psi} \pm \delta\, j_p(\hat{\psi})^{-1/2}$), for some small δ. Other techniques (secant method, Brent's method, etc.) can be used to invert $\widetilde{r}^\star(\psi)$ or $r_p^\star(\psi)$. Nevertheless, this would be computationally more demanding than the proposed spline interpolation, without solving the numerical instability around the posterior mode, or MLE.

Constrained maximization and computation of the required Hessians are generally straightforward to obtain numerically, whenever code for the likelihood or unnormalized posterior is available. For many statistical models with diffuse priors, built-in `R` functions (see R Core Team 2012) can sometimes be used to obtain full and constrained likelihood maximization as well as the related profile quantities required for HOTA$_\ell$. For instance, the `glm` function in `R` can handle many generalized linear models, and it offers the `offset` option for constrained estimation. Therefore, if the model in question belongs to the `glm` class, then all the quantities required in HOTA$_\ell$ can be extracted from it.

When the posterior mode and the MLE are substantially different, HOTA$_\pi$ is generally recommended. In this case, maximum likelihood routines can be used to find appropriate starting values for posterior optimization. For instance, if the model is in the `glm` class, starting values for the constrained posterior optimization can be obtained from the `glm` command along with the `offset` used to fix the parameter of interest. More generally, starting values for constrained optimization can be obtained by a linear expansion around the maximum (Cox and Wermuth 1990)

$$\hat{\boldsymbol{\lambda}}_\psi^{start} = \hat{\boldsymbol{\lambda}} + j_{\boldsymbol{\lambda}\boldsymbol{\lambda}}(\hat{\psi}, \hat{\boldsymbol{\lambda}})^{-1} j_{\boldsymbol{\lambda}\psi}(\hat{\psi}, \hat{\boldsymbol{\lambda}})(\hat{\psi} - \psi). \tag{13}$$

These are the strategies used in the examples of Section 4.

Algorithm 1 approximates (1) by simulating independently from the higher-order tail area approximations (7) or (9). In this respect, it has an obvious advantage over MCMC methods, which usually are more time consuming. Moreover, MCMC methods typically require more attention from the practitioner (e.g. choice of proposal, convergence checks, etc.). A pitfall of HOTA is that its theoretical approximation error (i.e. $O(n^{-3/2})$) is bounded by the sample size n. Nonetheless, as it will be shown in Section 4, HOTA gives typically very accurate approximations, even in small samples.

4 Examples

The aim of this section is to illustrate the performance of the HOTA method by three examples. In all but the first example, HOTA is compared with a trustworthy posterior approximation technique, namely the random walk Metropolis, which is one of the MCMC methods most widely used in practice. Prior sensitivity analysis is also considered with HOTA and compared also with MCMC. Prior sensitivity with HOTA is based on the same set of independent random variates, thus giving a comparison of different priors, given the same Monte Carlo error.

In general MCMC methods give autocorrelated samples and it is important to check that the chain has converged to its ergodic distribution (see, e.g., Gelman et al. 2003). In the examples, a multivariate normal proposal is used, suitably scaled in order to have an acceptance rate of 30-40%. Chains of simulations are run for a very large number of iterations, are thinned and the initial observations are discarded. In addition, the convergence is checked by the routines of the `coda` package of `R`. In each example, 10^5 final MCMC samples are considered, all with moderate autocorrelation. These MCMC samples will be considered as the gold standard, even though they are only an approximation of the exact posterior distribution.

The functions $\widetilde{r}^\star(\psi)$ and $r_p^\star(\psi)$ are inverted by spline interpolation applied to a grid of 50 values evenly spaced with $\delta = 0.3$. A sample of 10^5 is taken from all the approximate marginal posteriors. Required derivatives are computed numerically. This may be an other source of approximation error, difficult to quantify in practice. Nonetheless, we stress that this is an issue for many statistical applications since numerical derivatives are ubiquitous in statistics. Fortunately, there are many routines which performs accurate numerical derivatives; for instance, the `numDeriv` `R` package (see Gilbert and Varadhan 2012). The `R` code used in the examples is available at `http://homes.stat.unipd.it/ventura/?page=Software&lang=IT`.

4.1 Example 1: Genetic linkage model

The following scalar parameter problem has been studied also in Kharroubi and Sweeting (2010), among others. It concerns a genetic linkage model in which n individuals are distributed multinomially into four categories with cell probabilities $(\frac{1}{2}+\frac{\theta}{4}, \frac{1}{4}(1-\theta), \frac{1}{4}(1-\theta), \frac{\theta}{4})$, with $\theta \in (0,1)$. There are $n = 20$ animals with cell counts $y = (14, 0, 1, 5)$. Under a uniform prior, the posterior of θ is proportional to the likelihood and is given by

$$\pi(\theta|\mathbf{y}) \propto (2+\theta)^{14}(1-\theta)\theta^5, \qquad \theta \in (0,1).$$

There are no nuisance parameters and tail area approximations (7) and (9) coincide and simplify to

$$\int_{\theta_0}^{+\infty} \pi(\theta|y)d\theta \quad \doteq \quad \Phi\{\widetilde{r}^{\star}(\theta_0)\},$$

where $\widetilde{r}^{\star}(\theta) = r(\theta) + r(\theta)^{-1}\log\{q_B(\theta)/r(\theta)\}$, $q_B(\theta) = \ell'(\theta)j(\hat{\theta})^{-1/2}$, $\ell'(\theta) = d\ell(\theta)/d\theta$, and $r(\theta) = \text{sign}(\hat{\theta} - \theta)[2(\ell(\hat{\theta}) - \ell(\theta))]^{1/2}$. In view of this, HOTA$_\ell$ and HOTA$_\pi$ coincide.

Figure R.1 shows the posterior distribution computed with the HOTA algorithm and the exact posterior distribution $\pi(\theta|\mathbf{y})$.

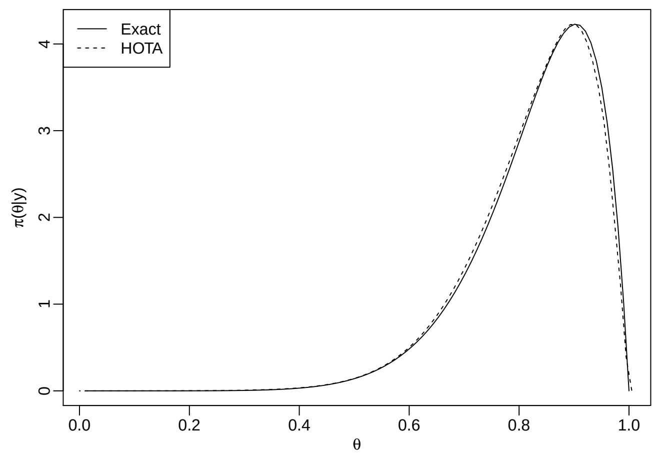

Figure R.1: Genetic linkage model: exact and HOTA simulated posterior distributions.

The exact posterior distribution appears to be extremely skewed to the right, with a long left tail, and in this case one might expect HOTA algorithm to fail. On the contrary, it gets very close to the exact posterior, even thought the sample size is only $n = 20$. In order to further explore the accuracy of the approximation, the two posteriors are compared also in terms of some summary statistics (mean, standard deviation, 2.5 percentile, median, 97.5 percentile and 0.95 HPD credible set) in Table 1. The HOTA

Posterior	Mean	St. Dev.	$Q_{0.025}$	Median	$Q_{0.975}$	0.95 HPD
Exact	0.831	0.108	0.570	0.852	0.978	(0.620, 0.994)
HOTA	0.827	0.108	0.566	0.848	0.976	(0.617, 0.994)

Table R.1: Genetic linkage model: numerical summaries of the exact and HOTA posterior distributions.

results are very close to those based on the exact posterior.

4.2 Example 2: Censored regression

The data consist of temperature accelerated life tests on electrical insulation in $n = 40$ motorettes (Davison 2003, Table 11.10). Ten motorettes were tested at each of four temperatures in degrees Centigrade (150°, 17°, 190° and 220°), the test termination (censoring) time being different at each temperature. This data were analysed from a Bayesian perspective in Kharroubi and Sweeting (2010), among others.

The following linear model is considered

$$y_i = \beta_0 + \beta_1 x_i + \sigma\,\varepsilon_i,$$

where ε_i are independent standard normal random variables, $i = 1, \dots, n$. The response is the $\log_{10}$(failuretime), with time in hours, and $x = 1000/(\text{temperature}+273.2)$. Given a type I censoring mechanism with censoring indicator δ_i, $i = 1, \dots, n$, the joint density of (Y_i, δ_i) given the parameter $\boldsymbol{\theta} = (\beta_0, \beta_1, \sigma)$ is (see, e.g., Pace and Salvan 1997, pp. 21–22)

$$p(y_i, \delta_i; \boldsymbol{\theta}) \;=\; \phi\left(\frac{y_i - \beta_0 - \beta_1 x_i}{\sigma}\right)^{1-\delta_i} \left(1 - \Phi\left\{\frac{y_i - \beta_0 - \beta_1 x_i}{\sigma}\right\}\right)^{\delta_i},$$

where $\phi(\cdot)$ is the standard normal density. 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 loglikelihood is

$$\ell(\boldsymbol{\theta}) = -m \log \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^{m} (y_i - \beta_0 - \beta_1 x_i)^2 + \sum_{i=m+1}^{n} \log\left\{1 - \Phi\left(\frac{u_i - \beta_0 - \beta_1 x_i}{\sigma}\right)\right\}. \tag{14}$$

For illustrative purposes several prior specifications are considered. The first is given by the flat prior $\pi_F(\boldsymbol{\theta})$. The second prior is a Normal-Half Cauchy distribution $\pi_{NHC}(\boldsymbol{\theta})$, given by independent components, which are respectively $\mathrm{N}(0, k)$ for the components of (β_0, β_1) and Half Cauchy with scale s for σ, with $(k, s) = (5, 0.1)$. The third prior is the Zellner's G-prior $\pi_G(\boldsymbol{\theta})$ (see, e.g., Marin and Robert 2007), which is the product of σ^{-1} and a bivariate normal density with mean $\boldsymbol{a}$ and covariance matrix $c\sigma^2(X^T X)^{-1}$, where X is the design matrix with the first column being a vector of ones. For simplicity we assume $\boldsymbol{a} = \mathbf{0}$ and $c = 100$. Several proposals exist for fixing c, but we choose 100 since this result can be interpreted as giving to the prior a weight of 1% of the data (see Marin and Robert 2007).

The posterior distributions obtained with these priors do not have a closed form solution, and numerical integration is needed in order to compute $\pi(\psi|\mathbf{y})$, and related quantities, with ψ being a separate component of $\boldsymbol{\theta}$.

Figure R.2 shows a sensitivity study on the effect of the three different priors on the posterior distributions based on HOTA_π. Note that the same set of random variates has been used in all cases, therefore what is shown are the differences between posteriors, given the same Monte Carlo error. See also Tables 2 and 3 for some numerical summaries for β_1 and σ, respectively.

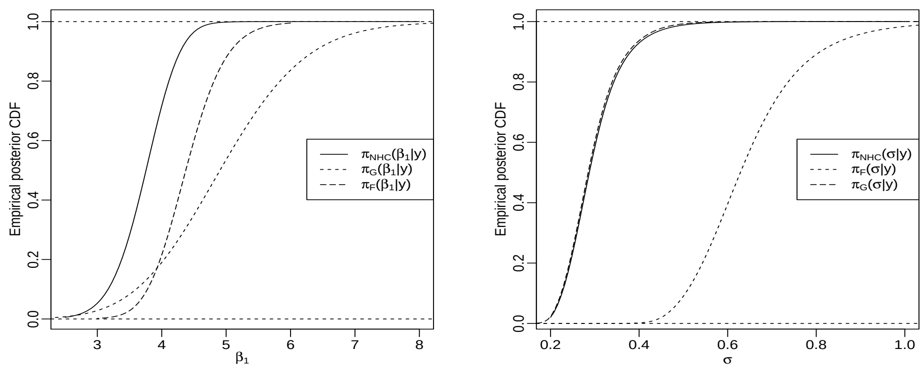

Figure R.2: Censored regression model: empirical marginal posterior CDFs for β_1 (left) and σ (right), computed with HOTA$_\pi$. The hyperparameters for $\pi_{NHC}(\boldsymbol{\theta})$ are $k = 5$ and $s = 0.1$.

Figure R.3 presents a graphical comparison between MCMC, HOTA$_\pi$ (based on $\widetilde{r}^\star(\psi)$) and HOTA$_\ell$ (based on $r_p^\star(\psi)$) in terms of the approximate posterior cumulative distribution functions (CDF) for β_1 (left column) and σ (right column). Results with HOTA$_\pi$ are always in close agreement with those of MCMC. On the contrary, the accuracy of HOTA$_\ell$ may not be satisfactory with non-flat priors, as also confirmed by the summary statistics in Tables 2 and 3.

Posterior	Method	Mean	St Dev.	$Q_{0.025}$	Median	$Q_{0.975}$	0.95 HPD
$\pi_F(\beta_1\|\mathbf{y})$	MCMC	4.409	0.518	3.461	4.382	5.512	(3.425, 5.470)
	HOTA$_\ell$	4.401	0.521	3.459	4.370	5.521	(3.398, 5.443)
	HOTA$_\pi$	4.401	0.521	3.459	4.370	5.521	(3.398, 5.443)
$\pi_{NHC}(\beta_1\|\mathbf{y})$	MCMC	3.731	0.447	2.802	3.746	4.571	(2.827, 4.594)
	HOTA$_\ell$	3.739	0.437	2.823	3.755	4.549	(2.889, 4.611)
$k = 5$, $s = 0.1$	HOTA$_\pi$	3.739	0.443	2.818	3.754	4.569	(2.840, 4.589)
$\pi_G(\beta_1\|\mathbf{y})$	MCMC	4.955	1.114	2.907	4.908	7.304	(2.906, 7.304)
	HOTA$_\ell$	5.885	3.078	1.182	5.388	13.173	(0.781, 12.389)
	HOTA$_\pi$	4.955	1.099	2.939	4.897	7.285	(2.838, 7.119)

Table R.2: Censored regression model: numerical summaries of the marginal posteriors of β_1 with $\pi_F(\boldsymbol{\theta})$, $\pi_{NHC}(\boldsymbol{\theta})$ and $\pi_G(\boldsymbol{\theta})$, computed with MCMC, HOTA$_\ell$ and HOTA$_\pi$.

Posterior	Method	Mean	St Dev.	$Q_{0.025}$	Median	$Q_{0.975}$	0.95 HPD
$\pi_F(\sigma\|\mathbf{y})$	MCMC	-1.240	0.201	-1.60	-1.253	-0.811	(-1.616, -0.832)
	HOTA$_\ell$	-1.240	0.202	-1.601	-1.251	-0.808	(-1.624, -0.837)
	HOTA$_\pi$	-1.240	0.202	-1.601	-1.251	-0.808	(-1.624, -0.837)
$\pi_{NHC}(\sigma\|\mathbf{y})$	MCMC	0.299	0.064	0.201	0.288	0.452	(0.193, 0.431)
	HOTA$_\ell$	0.277	0.052	0.196	0.270	0.398	(0.189, 0.384)
$k=5$, $s=0.1$	HOTA$_\pi$	0.298	0.064	0.203	0.287	0.452	(0.190, 0.426)
$\pi_G(\sigma\|\mathbf{y})$	MCMC	0.649	0.127	0.454	0.630	0.941	(0.434, 0.899)
	HOTA$_\ell$	1.327	0.306	0.875	1.278	2.058	(0.815, 1.936)
	HOTA$_\pi$	0.647	0.125	0.456	0.628	0.941	(0.430, 0.894)

Table R.3: Censored regression model: numerical summaries of the marginal posteriors of σ, with $\pi_F(\boldsymbol{\theta})$, $\pi_{NHC}(\boldsymbol{\theta})$ and $\pi_G(\boldsymbol{\theta})$, computed with MCMC, HOTA$_\ell$ and HOTA$_\pi$.

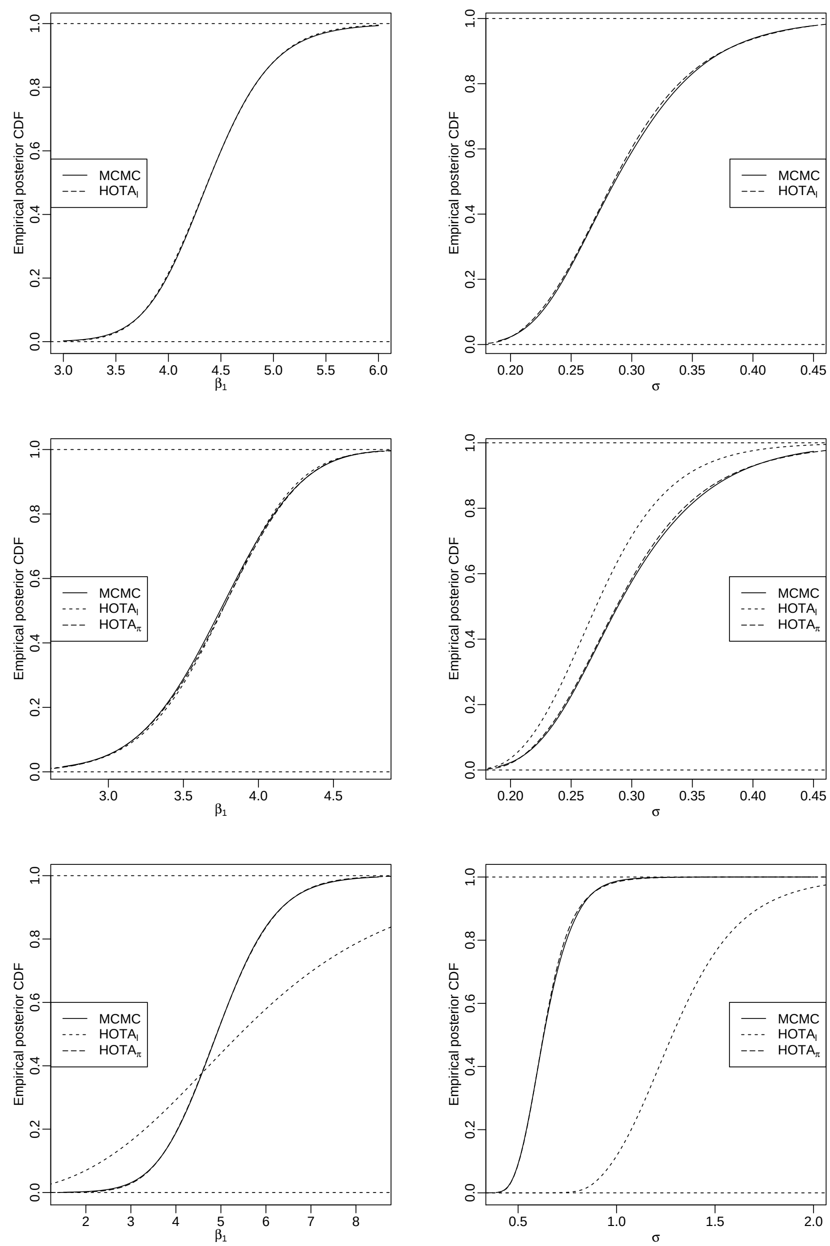

Figure R.3: Censored regression model: empirical marginal posterior CDFs for β_1 (left column) and σ (right column). The three rows correspond to priors $\pi_F(\boldsymbol{\theta})$, $\pi_{NHC}(\boldsymbol{\theta})$ ($k=5$, $s=0.1$) and $\pi_G(\boldsymbol{\theta})$, respectively. In the first line, HOTA$_\pi$ coincides with

4.3 Example 3: Logistic regression

In this example we consider a logistic regression model applied to the `urine` dataset analysed in Brazzale et al. (2007, Chap. 4), among others. This dataset concerns calcium oxalate crystals in samples of urine. The response is an indicator of the presence of such crystals, and the explanatory variables are: specific gravity (`gravity`) (i.e. the density of urine relative to water), pH (`ph`), osmolarity (`osmo`, mOsm), conductivity (`conduct`, mMho), urea concentration (`urea`, millimoles per litre), and calcium concentration (`calc`, millimoles per litre). After dropping two incomplete cases, the dataset consists of 77 observations. Let X denote the $(n \times 7)$ design matrix composed by a vector of ones and the six covariates, and let $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_6)$ be regression parameters, where β_0 is the intercept.

Different prior specifications are considered: a flat prior $\pi_F(\boldsymbol{\beta}) \propto 1$, a multivariate normal prior $\pi_N(\boldsymbol{\beta})$ with independent components $N(a, k)$, with $a = 0$ and $k = 5$, as well as the Zellner's G-prior (see Marin and Robert 2007, Chap. 4), given by

$$\pi_G(\boldsymbol{\beta}) \propto |X^T X|^{1/2} \Gamma(13/4) (\boldsymbol{\beta}^T (X^T X) \boldsymbol{\beta})^{-13/4} \pi^{-7/2}.$$

The choice of these priors has only the aim of illustrating our method and not to suggest their use for Bayesian data analysis.

Figure R.4 shows a sensitivity study on the effect of different priors on the posterior distributions based on HOTA$_\pi$. Here, we also consider the matching prior (12), given by

$$\pi_{mp}(\beta_r) \propto j_p(\beta_r)^{1/2}, \quad \text{for } r = 0, \ldots, 6.$$

With this prior the posterior distribution is approximated by HOTA$_\ell$. See also Tables 4 and 5 for some numerical summaries for β_4 and β_6, respectively.

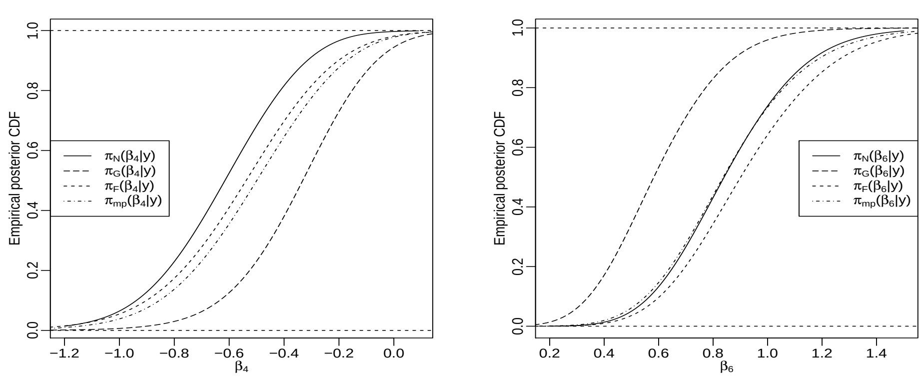

Figure R.4: Logistic regression model: empirical marginal posterior CDFs for β_4 (left) and β_6 (right), computed with HOTA$_\pi$.

Figure R.5 presents a graphical comparison between MCMC, HOTA$_\pi$ and HOTA$_\ell$ in terms of the approximate posterior cumulative distribution functions (CDF) for β_4 (left column) and β_6 (right column). The same comments about Figure R.3 apply here, with the difference that the accuracy of HOTA$_\ell$ is better than the one for the previous example when non-flat priors are used. See also Tables 4 and 5.

Posterior	Method	Mean	St Dev.	$Q_{0.025}$	Median	$Q_{0.975}$	0.95 HPD
$\pi_{mp}(\beta_4\|\mathbf{y})$	HOTA$_\ell$	-0.508	0.270	-1.063	-0.497	-0.007	(-1.010, 0.033)
$\pi_F(\beta_4\|\mathbf{y})$	MCMC	-0.591	0.256	-1.116	-0.585	-0.114	(-1.089, -0.095)
	HOTA$_\ell$	-0.547	0.278	-1.117	-0.537	-0.032	(-1.063, -0.009)
	HOTA$_\pi$	-0.547	0.278	-1.117	-0.537	-0.032	(-1.063, -0.009)
$\pi_N(\beta_4\|\mathbf{y})$	MCMC	-0.619	0.248	-1.132	-0.607	-0.163	(-1.117, -0.155)
	HOTA$_\ell$	-0.645	0.214	-1.073	-0.641	-0.239	(-1.035, -0.206)
$k = 5$	HOTA$_\pi$	-0.623	0.246	-1.127	-0.613	-0.169	(-1.079, -0.133)
$\pi_G(\beta_4\|\mathbf{y})$	MCMC	-0.335	0.227	-0.816	-0.323	0.068	(-0.793, 0.081)
	HOTA$_\ell$	-0.348	0.236	-0.837	-0.336	0.081	(-0.773, 0.114)
	HOTA$_\pi$	-0.343	0.228	-0.819	-0.330	0.070	(-0.790, 0.102)

Table R.4: Logistic regression model: numerical summaries of the marginal posterior of β_4, with $\pi_{mp}(\beta_4)$, $\pi_F(\boldsymbol{\beta})$, $\pi_N(\boldsymbol{\beta})$, and $\pi_G(\boldsymbol{\beta})$ approximated by MCMC, HOTA$_\ell$ and HOTA$_\pi$.

Posterior	Method	Mean	St Dev.	$Q_{0.025}$	Median	$Q_{0.975}$	0.95 HPD
$\pi_{mp}(\beta_6\|\mathbf{y})$	HOTA$_\ell$	0.859	0.255	0.424	0.839	1.417	(0.414, 1.399)
$\pi_F(\beta_6\|\mathbf{y})$	MCMC	0.883	0.250	0.454	0.863	1.425	(0.435, 1.391)
	HOTA$_\ell$	0.924	0.264	0.472	0.903	1.500	(0.461, 1.482)
	HOTA$_\pi$	0.924	0.264	0.472	0.903	1.500	(0.461, 1.482)
$\pi_N(\beta_6\|\mathbf{y})$	MCMC	0.863	0.241	0.447	0.845	1.386	(0.419, 1.347)
	HOTA$_\ell$	0.829	0.217	0.445	0.817	1.289	(0.436, 1.277)
$k = 5$	HOTA$_\pi$	0.859	0.239	0.445	0.842	1.373	(0.435, 1.357)
$\pi_G(\beta_6\|\mathbf{y})$	MCMC	0.604	0.204	0.259	0.586	1.054	(0.241, 1.024)
	HOTA$_\ell$	0.591	0.197	0.237	0.573	1.030	(0.229, 0.995)
	HOTA$_\pi$	0.600	0.212	0.264	0.584	1.060	(0.235, 1.045)

Table R.5: Logistic regression model: numerical summaries of the marginal posterior of β_6, with $\pi_{mp}(\beta_6)$, $\pi_F(\boldsymbol{\beta})$, $\pi_N(\boldsymbol{\beta})$, and $\pi_G(\boldsymbol{\beta})$ approximated by MCMC, HOTA$_\ell$ and HOTA$_\pi$.

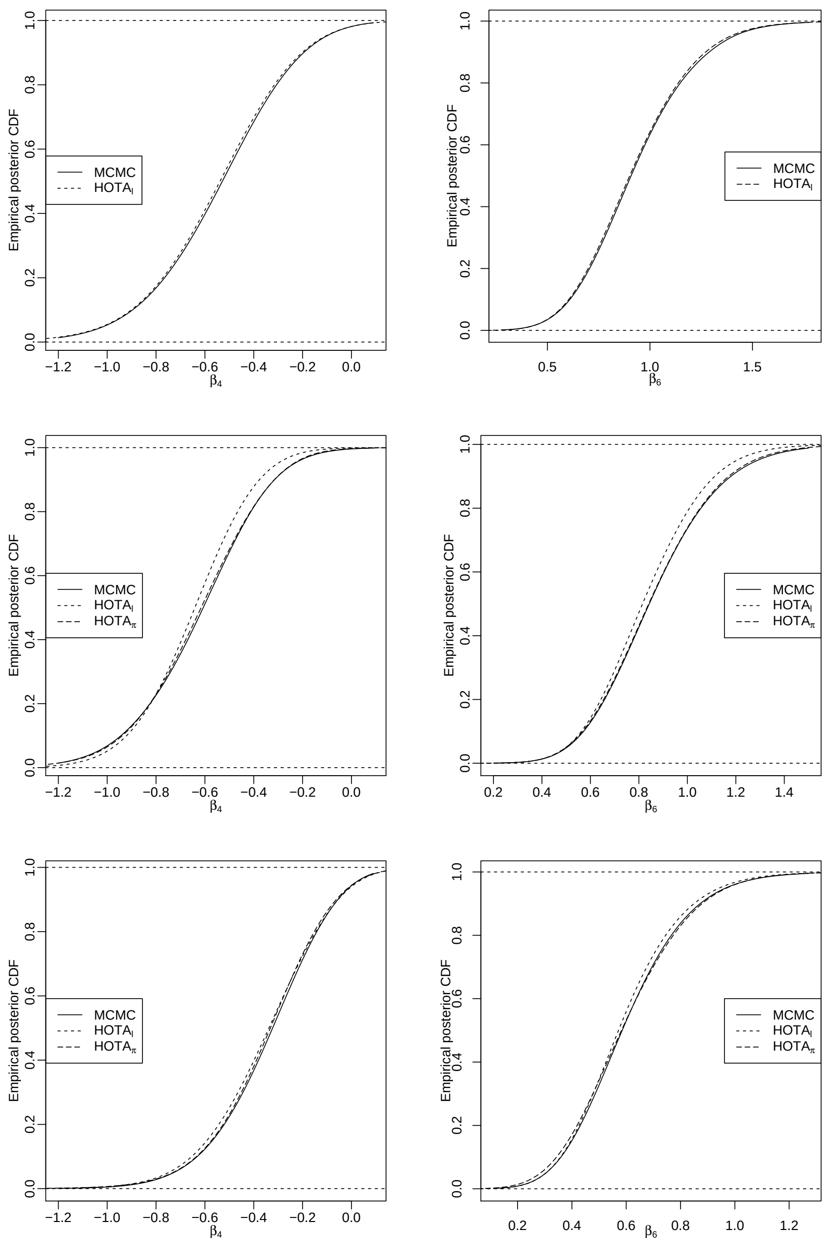

Figure R.5: Logistic regression model: empirical marginal posterior CDFs for β_4 (left column) and β_6 (right column). The three rows correspond to priors $\pi_F(\boldsymbol{\beta})$, $\pi_N(\boldsymbol{\beta})$, with $k = 5$, and $\pi_G(\boldsymbol{\beta})$ respectively.

5 Concluding remarks

The HOTA simulation method for Bayesian approximation combines higher-order tail area approximations with the inverse transform sampler. This sampling method gives accurate approximations of marginal posterior distributions for a scalar parameter of interest.

The accuracy of the two versions of the HOTA algorithm may be different and, in particular, may depend on the chosen prior. In this respect, the version based on the expansion around the posterior mode is a safer choice, since the approximation makes explicit use of the prior information. On the contrary, the accuracy of the version based on the expansion around the MLE, although easier to compute, could be affected by the difference between the likelihood and the posterior, which is indeed the effect of the prior. Therefore, in general we would recommend the use of HOTA_π, since the effect of the prior on the posterior depends on many aspects, such as the nature and range of the parameter, and it is not straightforward to assess such effect in advance. On the other hand, both approximations rely on small-sample results, in the sense that as the sample size increases the effect of the prior vanishes, implying that the two approximations will tend to coincide.

Bayesian robustness with respect to the prior can be easily handled with the HOTA sampling scheme. Indeed, higher-order approximations make it straightforward to assess the influence of the prior, and the effect of changing priors on the posterior quantities (see also Reid and Sun 2010). Moreover, with HOTA the effect of the prior on the posterior distribution can be appreciated under the same Monte Carlo variation. Finally, default priors, such as the matching prior used in Example 3, could be easily handled by the method and could be used as a benchmark for Bayesian robustness.

The proposed use of higher-order asymptotics for Bayesian simulation opens to other interesting applications. For instance, the HOTA procedure could be used in conjunction with MCMC methods, e.g. to simulate from conditional posteriors within Gibbs sampling. Moreover, HOTA could be applied also to other Bayesian procedures, such as model selection or prediction. In the latter cases, simulations from the full model are required. Although in the present article we are not concerned with these topics, the extension of our method in these directions is currently under investigation.

The `R` code used in the examples is available at `homes.stat.unipd.it/ventura`. A more general `R` package is under preparation, which will implement the HOTA method for regular parametric models, requiring only the unnormalized log-posterior function as input.

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. We thank the referees for their useful comments that substantially improved the paper.

References

Barndorff-Nielsen, O. E. and Chamberlin, S. R. (1994). "Stable and invariant adjusted directed likelihoods." *Biometrika*, 81(3): 485–499. 133

Barndorff-Nielsen, O. E. and Cox, D. R. (1994). *Inference and Asymptotics*. Monographs on Statistics and Applied Probability. Taylor & Francis. 133

Brazzale, A. R. and Davison, A. C. (2008). "Accurate parametric inference for small samples." *Statistical Science*, 23(4): 465–484. 129, 130

Brazzale, A. R., Davison, A. C., and Reid, N. (2007). *Applied Asymptotics*. Cambridge: Cambridge University Press. 130, 134, 140

Cox, D. R. and Wermuth, N. (1990). "An approximation to maximum likelihood estimates in reduced models." *Biometrika*, 77(4): 747–761. 134

Davison, A. (2003). *Statistical Models*. Cambridge: Cambridge University Press. 131, 137

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2003). *Bayesian Data Analysis*. Chapman & Hall/CRC. 135

Gilbert, P. and Varadhan, R. (2012). *numDeriv: Accurate Numerical Derivatives*. R package version 2012.9-1.
URL `http://CRAN.R-project.org/package=numDeriv` 135

Kass, R. E., Tierney, L., and Kadane, J. B. (1989). "Approximate methods for assessing influence and sensitivity in Bayesian analysis." *Biometrika*, 76(4): 663–674. 130

— (1990). "The validity of posterior expansions based on Laplace's method." In Geisser, S., Hodges, J., Press, S., and Zellner, A. (eds.), *Bayesian and Likelihood Methods in Statistics and Econometrics: Essays in Honor of George A. Barnard*, 473–487. North Holland. 130, 131

Kharroubi, S. A. and Sweeting, T. J. (2010). "Posterior simulation via the signed root log-likelihood ratio." *Bayesian Analysis*, 5(4): 787–815. 130, 135, 137

Marin, J. and Robert, C. (2007). *Bayesian Core: A Practical Approach to Computational Bayesian Statistics*. New York: Springer. 137, 140

Pace, L. and Salvan, A. (1997). *Principles of statistical inference: from a Neo-Fisherian perspective*. Singapore: World Scientific. 137

— (2006). "Adjustments of the profile likelihood from a new perspective." *Journal of Statistical Planning and Inference*, 136: 3554–3564. 132

R Core Team (2012). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
URL `http://www.R-project.org/` 134

Reid, N. (1996). "Likelihood and Bayesian approximation methods." In *Bayesian Statistics*, volume 5, 351–368. Oxford University Press. 130, 132

— (2003). "Asymptotics and the theory of inference." *The Annals of Statistics*, 31(6): 1695–1731. 130

Reid, N. and Sun, Y. (2010). "Assessing Sensitivity to Priors Using Higher Order Approximations." *Communications in Statistics–Theory and Methods*, 39(8-9): 1373–1386. 130, 143

Robert, C. and Casella, G. (2004). *Monte Carlo Statistical Methods*. New York: Springer, 2nd edition. 129, 130

Severini, T. A. (2000). *Likelihood Methods in Statistics*. New York: Oxford University Press. 132, 133

Sweeting, T. J. (1996). "Approximate Bayesian computation based on signed roots of log-density ratios (with discussion)." In *Bayesian Statistics*, volume 5, 427–444. Oxford University Press. 130

Tibshirani, R. (1989). "Noninformative priors for one parameter of many." *Biometrika*, 76(3): 604–608. 132

Tierney, L. and Kadane, J. B. (1986). "Accurate approximations for posterior moments and marginal densities." *Journal of the American Statistical Association*, 81(393): 82–86. 131

Ventura, L., Cabras, S., and Racugno, W. (2009). "Prior distributions from pseudo-likelihoods in the presence of nuisance parameters." *Journal of the American Statistical Association*, 104(486): 768–774. 132

Ventura, L. and Racugno, W. (2011). "Recent advances on Bayesian inference for $P(X < Y)$." *Bayesian Analysis*, 6: 411–428. 133

Ventura, L., Sartori, N., and Racugno, W. (2013). "Objective Bayesian higher-order asymptotics in models with nuisance parameters." *Computational Statistics & Data Analysis*, 60: 90–96. 132