Comput Stat (2018) 33:213–234
https://doi.org/10.1007/s00180-017-0772-9

ORIGINAL PAPER

Posterior simulation via the exponentially tilted signed root log-likelihood ratio

Samer A. Kharroubi[1]

Received: 2 December 2016 / Accepted: 8 October 2017 / Published online: 11 October 2017

Abstract We explore the use of importance sampling based on exponentially tilted signed root log-likelihood ratios for Bayesian computation. Approximations based on exponentially tilted signed root log-likelihood ratios are used in two distinct ways; firstly, to define an importance function with antithetic variates and, secondly, to define suitable control variates for variance reduction. These considerations give rise to alternative simulation-consistent schemes to other importance sampling techniques (for example, conventional and/or adaptive importance sampling) for Bayesian computation in moderately parameterized regular problems. The schemes based on control variates can also be viewed as usefully supplementing computations based on asymptotic approximations by supplying external estimates of error. The methods are illustrated by a censored regression model and a more challenging 12-parameter nonlinear repeated measures model for bacterial clearance.

Keywords Bayesian computation · Control variates · Exponential tilting · Signed root log-likelihood ratio · Importance sampling · Variance reduction

1 Introduction

Accurate asymptotic formulae for posterior expectations and predictive distributions were obtained in Sweeting (1996) and Sweeting and Kharroubi (2003). These formulae are correct to the same asymptotic order as the corresponding Tierney and Kadane (1986) expressions but possess the advantage that the major computational effort

✉ Samer A. Kharroubi
sk157@aub.edu.lb

1 Department of Nutrition and Food Sciences, Faculty of Agricultural and Food Sciences, American University of Beirut, P.O. Box 11-0236, Riad El Solh, Beirut 1107-2020, Lebanon

only needs to be carried out once when many expectations are required, such as in the computation of a predictive density, or of posterior expected losses in a decision problem. In Kharroubi and Sweeting (2010) asymptotic formulae based on signed roots formed the basis for an importance sampling scheme for Bayesian computation.

Experience with applying these formulae has demonstrated that they can provide excellent approximations in practice. One difficulty that can arise, however, is the need for repeated computation of conditional maxima of the likelihood function in multiparameter cases. This can be particularly problematic when it is required to invert the signed root log-likelihood ratio, since a maximization procedure has to be included within a nonlinear inversion routine, as pointed out in Sweeting (1996) and the ensuing discussion. Although not a major problem in low dimensions, this can cause computational difficulties in higher dimensions, especially when many signed root inversions are required, such as in importance sampling scheme of Kharroubi and Sweeting (2010).

Recently, Kharroubi and Sweeting (2016) used exponential tilting to develop alternative asymptotic approximations for posterior expectations, predictive distributions and marginal posterior distributions that do not require any conditional maximization. The new tilted approximations are shown to be particularly relevant for the signed root based importance sampler and control variates of Kharroubi and Sweeting (2010). A number of researchers have obtained useful asymptotic approximations in statistics based on exponential tilting. Schennach (2005) derives a Bayesian exponentially tilted empirical likelihood and Schennach (2007) carries out further studies and comparisons involving the exponentially tilted estimator. Cerquetti (2007) develops Bayesian nonparametric inference in which the prior is obtained from an exponentially tilted Poisson-Kingman model, following Pitman (2003). Sweeting and Kharroubi (2005) used exponential tilting to enable the application of a predictive distribution formula in cases where the likelihood function does not possess a local maximum.

In the present paper we follow on from the work of Kharroubi and Sweeting (2016) to explore the use of exponentially tilted signed root based importance sampler in conjunction with antithetic variates for Bayesian computation in moderately parameterized models. It is further shown that by incorporating control variates we can achieve a substantial reduction in sampling variability as compared to straight importance sampling. The potential of the methodology is explored throughout the paper using a censored regression model and a more challenging 12-parameter nonlinear repeated measures model for bacterial clearance. Although the methodology is not universally applicable, in cases where it is applicable it has the obvious advantage over other importance sampling methods (for example, conventional/adaptive importance sampling or population Monte Carlo) that samples are drawn independently. Furthermore, relatively short runs of the schemes based on control variates could be used to provide external simulation-based validation of the accuracy of asymptotic approximation formulae.

The organization of the paper is as follows. In Sect. 2 we present the basics of importance sampling and introduce the notion of exponential tilting. In Sect. 3 we review briefly some of the exponentially tilted methods and results in Kharroubi and Sweeting (2016). This is necessary for the development in the subsequent sections and in order to facilitate comparisons with the modified results. We then develop a modified signed-root based importance sampling schemes based on exponential tilting and in conjunction with antithetic variates. We proceed in Sect. 4 to develop suitable control variates based on exponential tilting. The theory of Sects. 3 and 4 is illustrated by a censored regression model and a nonlinear repeated measures model for bacterial clearance. Some concluding remarks are given in Sect. 5.

2 Background (with examples)

In this section we present briefly an overview of importance sampling and exponential tilting, including some illustrative examples. This is necessary in to order to ease the development in the subsequent sections. We first go over the basics of importance sampling in Sect. 2.1 and then proceed with explaining exponential tilting in Sect. 2.2.

2.1 Basic importance sampling

Suppose that our problem is to find $\ell = E_f\{h(X)\} = \int_\Omega h(x)f(x)dx$ where X is an $\mathbb{R}$-valued random variable with probability density function f on Ω. If g is a positive probability density function on Ω, then

$$\ell = \int_\Omega h(x)f(x)dx = \int_\Omega \frac{h(x)f(x)}{g(x)}\, g(x)dx = E_g\left\{\frac{h(X)f(X)}{g(X)}\right\},$$

where $E_g(.)$ denotes expectation for $X \sim g$. The importance sampling of $\ell = E_f\{h(x)\}$ is

$$\hat{\ell}_{IS} = \frac{1}{n}\sum_{j=1}^{n} \frac{h(X_i)f(X_i)}{g(X_i)}, \quad X_i \sim g.$$

A variety of Monte Carlo importance sampling strategies have been proposed in the literature. See, for example, Van Dijk et al. (1986), Evans and Swartz (1995, 2000), Hesterberg (1995) and Owen and Zhou (2000). Recently, some important relevant advances in importance sampling includes Vehtari et al. (2016), Elvira et al. (2016) and Bugallo et al. (2017) .

2.2 Exponential tilting

Trying to find the best choice of g is essentially an infinite dimensional optimization problem which is generally very difficult. In practice, it is often easy to try to turn the problem into a finite dimensional one. This can be done by restricting the choice of g to a parametric family of densities $\{g(x;\theta)\}_{\theta\in\Theta}$.

The most useful way to create such a family of densities is by so-called exponential tilting (Efron 1981; Fuh et al. 2013). Recall that the moment generating function of a random variable $X \sim f$ is given by

$$M(\theta) = E_f \exp\{\theta X\}.$$

The cumulant generating function of a random variable is given by

$$\kappa(\theta) = \log E_f \exp\{\theta X\} = \log M(\theta)$$

Using this, one can define a family of densities by

$$f(x;\theta) = \exp\{\theta X - \kappa(\theta)\} f(x),$$

where $\theta \in \Theta = \{\theta : M(\theta) < \infty\}$.

Exponential tilting can be applied to every distribution for which a cumulant generating function is defined. Let us for example consider the normal distribution.

Example Tilting the normal density

The moment generating function of a normal distribution with mean μ and variance σ^2 is given by

$$M(\theta) = \exp\left\{\mu\theta X + \frac{\sigma^2\theta^2}{2}\right\}.$$

So,

$$\kappa(\theta) = \mu\theta X + \frac{\sigma^2\theta^2}{2}.$$

This means that the tilted normal distribution is of the form

$$f(x;\theta) \propto \exp\left\{-\frac{1}{2\sigma^2}\left(x - (\mu + \sigma^2\theta^2)\right)\right\},$$

which is the pdf of a normal distribution with mean $\mu + \sigma^2\theta$ and variance σ^2. Thus, exponentially tilting a normal distribution is equivalent to shifting its mean by $\sigma^2\theta$.

Using an exponential tilt, one obtains the estimator

$$\hat{\ell}_\theta = \frac{1}{n}\sum_{j=1}^{n} \frac{f(X_i)}{f(X_i;\theta)}\ h(X_i) = \frac{1}{n}\sum_{j=1}^{n} \exp\{\kappa(\theta) - \theta X_i\}\, h(X_i). \tag{1}$$

The obvious question is how to choose θ. One common way to choose θ is so that the mean of the tilted density is the point of interest (Fuh et al. 2013). For example, suppose we wish to estimate $\ell = P(X > \gamma)$ where $X \sim N(\mu, \sigma^2)$. We tilt the density so that $E_\theta X = \gamma$, this ensures we have samples in the region of interest. That is, we choose $\mu + \sigma^2\theta = \gamma$, so $\theta = (\gamma - \mu)/\sigma^2$. Using this, we can then estimate $P(X > \gamma)$ via the importance sampling estimator (1). For the purpose of illustration, Fig. 1 presents the original and exponentially tilted densities in the case where $X \sim N(0, 1)$ and $\gamma = 4$.

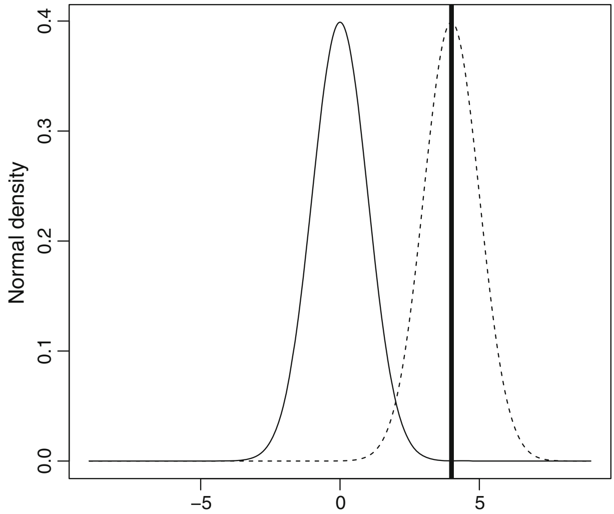

Fig. 1 Tilting the normal density: $N(0, 1)$ density (dashed line), $N(\gamma, 1)$ tilted density (dotted line)

3 Importance sampling with antithetic variates based on exponentially tilted signed roots

In this section we develop the exponentially tilted signed root based importance sampling scheme with antithetic variates. We begin by reviewing the construction of exponentially tilted signed root log-likelihood ratios in Sect. 3.1, which forms the basis of the remainder of this section. The construction of the exponentially tilted importance sampler is described in Sect. 3.2 and antithetic variate technique for variance reduction is discussed in Sect. 3.3. The computation of marginal posterior densities is given in Sect. 3.4.

3.1 Exponentially tilted signed root log-likelihood ratios

We begin by reviewing the construction of signed root log-likelihood ratios using exponential tilting. We suppose that the data x arise from a parametric model with unknown parameter $\theta = (\theta^1, \ldots, \theta^d) \in \Omega \subset \mathbb{R}^d$, where $d \geq 1$, and that the associated likelihood function $L(\theta)$ is available. We further suppose the availability of a prior density $\lambda(\theta)$ of θ, assumed to be continuous and positive throughout Ω. Then the posterior density of θ is

$$p(\theta|x) = c^{-1} L(\theta)\lambda(\theta), \tag{2}$$

where $c = \int L(\theta)\lambda(\theta)\, d\theta$. Under the assumption that the likelihood function possesses unique local conditional maxima, there will exist a one-to-one transformation of θ to a set of signed root log-likelihood ratios, which forms the basis for excellent asymptotic approximations; see, for example, Sweeting (1996). If, however, the requirement of unique local conditional maxima is not satisfied then it may be possible to achieve this requirement by applying a suitable initial exponential tilt to the likelihood function.

Let $l(\theta) = \log L(\theta)$ be the log-likelihood function. The score vector is $\dot{l}(\theta) = dl(\theta)/d\theta = (l_1(\theta), \ldots, l_d(\theta))^T$, where $l_j(\theta) \equiv \partial l(\theta)/\partial\theta^j$, $j = 1, \ldots, d$. Further define $\ddot{l}(\theta) = d^2 l(\theta)/d\theta^2$, $j(\theta) = -\ddot{l}(\theta)$, and $J = j(\hat{\theta})$, the Fisher information matrix. As in Sweeting and Kharroubi (2003), for $1 \leq i \leq d$ we let $\theta_i = (\theta^1, \ldots, \theta^i)$ be the vector of the first i components of θ and $\theta^{(i)} = (\theta^i, \ldots, \theta^d)$ the vector of the last $d - i + 1$ components.

We suppose that, for each $1 < i \leq d$ and fixed θ_{i-1}, there exists a local conditional maximizer $\hat{\theta}^{(i)}(\theta_{i-1}) = (\hat{\theta}^i(\theta_{i-1}) \ldots, \hat{\theta}^d(\theta_{i-1}))$ of $L(\theta)$. It follows from Kharroubi and Sweeting (2016) that the first-order term in the Taylor expansion of the jth component of the conditional maximizer of $l(\theta)$ given θ_i about $\theta_i = \hat{\theta}_i$ is given by

$$\bar{\theta}^j(\theta_i) = \hat{\theta}^j + \sum_{k=1}^{i} \hat{\theta}^j_{ik}(\theta^k - \hat{\theta}^k) \tag{3}$$

for $1 \leq i < j \leq d$, where $\hat{\theta}^j_{ik} = \partial\hat{\theta}^j(\theta_i)/\partial\theta^k$ evaluated at $\theta_i = \hat{\theta}_i$. Further write $\bar{\theta}^{(j)}(\theta_i) = (\bar{\theta}^j(\theta_i), \ldots, \bar{\theta}^d(\theta_i))$ for $1 \leq i < j \leq d$. The idea is to modify the likelihood function so that the approximate conditional maximizer $\bar{\theta}^{(i+1)}(\theta_i)$ of $L(\theta)$ given θ_i is the exact conditional maximizer of the modified likelihood. To gain a feeling for the construction of this modified likelihood, consider first the case $d = 2$ for which $\bar{\theta}^2(\theta^1) = \hat{\theta}^2 - (J_{21}/J^{22})(\theta^1 - \hat{\theta}^1)$. Define the exponentially tilted log-likelihood function by

$$\bar{l}(\theta^1, \theta^2) = l(\theta^1, \theta^2) - l_2\{\theta^1, \bar{\theta}^2(\theta^1)\}\{\theta^2 - \bar{\theta}^2(\theta^1)\}.$$

On partial differentiation of $\bar{l}(\theta)$ with respect to θ^2 we see that the maximum of $\bar{l}(\theta)$ for fixed θ^1 occurs at $\theta^2 = \bar{\theta}^2(\theta^1)$. Furthermore the maximum of $\bar{l}\{\theta^1, \bar{\theta}^2(\theta^1)\}$ occurs at $\hat{\theta}^1$ since $\bar{l}\{\theta^1, \bar{\theta}^2(\theta^1)\} = l\{\theta^1, \bar{\theta}^2(\theta^1)\}$ and $\bar{\theta}^2(\hat{\theta}^1) = \hat{\theta}^2$.

As in Kharroubi and Sweeting (2016), the following convention will prove useful: for any function $g(\theta)$ and $1 \leq i < d$, $g(\theta_i)$ will denote $g(\theta_i, \bar{\theta}^{(i+1)}(\theta_i))$, where $g(\theta_0)$ is understood to be $g(\hat{\theta})$. That is, we simply replace $\hat{\theta}^{(i+1)}(\theta_i)$ by $\bar{\theta}^{(i+1)}(\theta_i)$.

We now define the exponentially tilted log-likelihood function $\bar{l}(\theta)$ to be

$$\bar{l}(\theta) = l(\theta) - \sum_{j=1}^{d} h^j(\theta_{j-1})\{\theta^j - \bar{\theta}^j(\theta_{j-1})\}, \tag{4}$$

where

$$h^i(\theta) = \sum_{j=1}^{d} c_i^j l_j(\theta) = c_i^T \dot{l}(\theta), 1 \leq i \leq d,$$

$c_i = (c_i^1, \ldots, c_i^d)^T$ and $c_i^j = 0\,(j < i)$, $c_i^i = 1$, $c_i^j = \hat{\theta}_{ii}^j\,(j > i)$. It is shown in Kharroubi and Sweeting (2016) that, for $1 < i \leq d$, $\bar{\theta}^{(i)}(\theta_{i-1})$ is the unique local conditional maximizer of $\bar{l}(\theta)$ given θ_{i-1}.

Now define the functions $H^i(\theta_i) = \exp\{h^i(\theta_{i-1})\{\theta^i - \bar{\theta}^i(\theta_{i-1})\}$ for $1 \leq i \leq d$ and let $H(\theta) = \prod_{i=1}^d H^i(\theta_i)$. Then from (4) the exponentially tilted likelihood function is given by $\bar{L}(\theta) = L(\theta)/H(\theta)$. We therefore need to modify the prior by absorbing the factor $H(\theta)$ into the prior $\lambda(\theta)$, which produces the exponentially tilted prior density $\bar{\lambda}(\theta) = \lambda(\theta)H(\theta)$. The posterior density (2) can then be written as $p(\theta|x) = c^{-1}\bar{L}(\theta)\bar{\pi}(\theta)$. It is shown in Kharroubi and Sweeting (2016) that the factor $H(\theta)$ is $O(1)$ so that the asymptotic analysis of Sweeting (1996) and Sweeting and Kharroubi (2003) continues to apply. That is, we can apply all the formulae given there with L and λ replaced by $\bar{L}$ and $\bar{\lambda}$. Notice also that we do not actually require the existence of the conditional maxima $\hat{\theta}^{(i)}(\theta_{i-1})$, needed for the non-tilted formulae in Sweeting (1996) and Sweeting and Kharroubi (2003), in order to implement these formulae.

Specifically, for $1 \leq i \leq d$ define the tilted log-likelihood ratios

$$\bar{w}^i(\theta_i) = 2\{\bar{l}(\theta_{i-1}) - \bar{l}(\theta_i)\} = 2\{l(\theta_{i-1}) - l(\theta_i) + h^i(\theta_{i-1})(\theta^i - \bar{\theta}^i(\theta_{i-1}))\}$$

and the tilted signed root log-likelihood ratios

$$\bar{r}^i(\theta_i) = \operatorname{sign}\{\theta^i - \bar{\theta}^i(\theta_{i-1})\}\{\bar{w}^i(\theta_i)\}^{1/2}$$

and write $\bar{r}(\theta) = (\bar{r}^1(\theta_1), \ldots, \bar{r}^d(\theta_d))$.

3.2 Construction of exponentially tilted importance sampler

Consider the posterior expectation $\mu = E\{v(\theta)|x\}$ of the smooth function $v(\theta)$. Let $g(\theta)$ be any density function from which it is easy to sample and let E_g denote expectation under the density g. Then, since

$$\mu = c^{-1}E_g\left\{v(\theta)\frac{L(\theta)\lambda(\theta)}{g(\theta)}\right\},$$

it follows that the constant of proportionality c in Eq. (2) and the posterior expectation μ are consistently estimated by

$$\hat{c} = (km)^{-1}\sum_{j=1}^m u_j \tag{5}$$

and

$$\hat{\mu} = \sum_{j=1}^m v(\theta_{[j]})w_j \tag{6}$$

respectively, where $\theta_{[1]}, \ldots, \theta_{[m]}$ are m independent draws from $g(\theta)$, $u_j = L(\theta_{[j]})\lambda(\theta_{[j]})/h(\theta_{[j]})$ are the importance weights, $w_j = u_j/\sum_{i=1}^m u_i$ the normalized importance weights and $g(\theta) = kh(\theta)$.

The usual strategy is to choose a suitable density $g(\theta)$ that is close to, but more dispersed than, $p(\theta|x)$ in order that the importance sampling is stable. As discussed in Kharroubi and Sweeting (2010), in general importance functions based on local approximations to the posterior distribution can have poor tail behaviour, making them unsuitable for importance sampling. However, the signed root log likelihood ratio succeeds in capturing the entire shape of the likelihood function and would be expected to lead to a very reliable form of importance sampling. A theoretical result on the tail behaviour of $g(\theta)$ is given in Kharroubi and Sweeting (2010). All of this also applies to the tilted signed root log likelihood ratio. We can therefore construct an importance sampler exactly as in Kharroubi and Sweeting (2010) simply by replacing $r(\theta)$ there by $\bar{r}(\theta)$. Thus, for $1 \leq i \leq d$, $\theta^i \equiv \theta^i(\bar{R}_i)$ are defined by inversion of $\bar{r}^i(\theta_i) = \bar{R}^i$ for fixed θ_{i-1}, where the $\bar{R}^i$ are independently sampled from the standard normal distribution.

Since $\bar{r}^i(\theta_i)$ is a function of the first i components of θ, the Jacobian matrix $d\bar{r}/d\theta$ is lower triangular, so that

$$\left|\frac{d\bar{r}}{d\theta}\right| = \prod_{i=1}^{d} \frac{\partial \bar{r}^i(\theta)}{\partial \theta^i} = \prod_{i=1}^{d} \frac{-\bar{l}_i(\theta_i)}{\bar{r}^i(\theta_i)}\,.$$

It now follows from the usual multivariate transformation formula for densities that the density of θ is

$$g(\theta) = (2\pi)^{-d/2}\frac{\bar{L}(\theta)}{\bar{L}(\hat{\theta})}\prod_{i=1}^{d} \frac{-\bar{l}_i(\theta_i)}{\bar{r}^i(\theta_i)}\,. \tag{7}$$

We remark that no difficulty arises here when $\theta^i = \hat{\theta}^i(\theta_{i-1})$, since L'Hôpital's rule gives $\bar{r}^i(\theta_i)/\bar{l}_i(\theta_i) \to \{-\bar{k}^i(\theta_{i-1})\}^{-1/2}$ as $\theta^i \to \hat{\theta}^i(\theta_{i-1})$, where $\bar{k}^i(\theta) = -\partial^2\bar{l}(\theta)/(\partial\theta^i)^2$. The exponentially tilted signed root importance sampler algorithm is then as follows:

1. Generate $\bar{R}^i \sim N(0, 1)$, independently for $i = 1, \ldots, d$.
2. Obtain $\theta^i = \theta^i(\bar{R}_i)$ sequentially for $i = 1, \ldots, d$ as the solutions of the equations $\bar{r}^i(\theta_i) = \bar{R}^i$.
3. Repeat steps 1 and 2 m times to obtain $\bar{R}_{[1]}, \ldots, \bar{R}_{[m]}$ and the corresponding sample $\theta_{[1]}, \ldots, \theta_{[m]}$ from the importance density (7).

The simulation-consistent estimators of c and μ are then given by (5) and (6) respectively with $k = (2\pi)^{d/2}\bar{L}(\hat{\theta})$ and $u_j = u(\theta_{[j]})$, where

$$u(\theta) = \bar{\lambda}(\theta)\prod_{i=1}^{d}\left\{\frac{\bar{r}^i(\theta_i)}{-\bar{l}_i(\theta_i)}\right\}.$$

These formulae require no conditional maximization procedures for their implementation.

The precisions of (5) and (6) can be assessed via their Monte Carlo standard errors, which are

$$\text{s.e.}(\hat{c}) = \hat{c}\left\{\sum_{j=1}^{m}(w_j - m^{-1})^2\right\}^{1/2} \tag{8}$$

and

$$\text{s.e.}(\hat{\mu}) = \left\{\sum_{j=1}^{m} w_j^2(v(\theta_{[j]}) - \hat{\mu})^2\right\}^{1/2}. \tag{9}$$

Formula (9) follows from the delta method. Rather than pre-specifying the number of Monte Carlo samples m, (8) can be used to determine m automatically by instead pre-specifying a value for the coefficient of variation $\text{s.e.}(\hat{c})/\hat{c}$ of $\hat{c}$.

3.3 Antithetic variates

It is natural to seek to reduce the Monte Carlo variability in (5) and (6) so that accurate results can be obtained with a smaller computational effort. There are a number of variance reduction techniques that can be used in conjunction with importance sampling, including antithetic variates and control variates; see, for example, Hammersley and Handscomb (1964). We discuss the former here and the latter in Sect. 4. Discussion of a wider class of techniques can be found in Ripley (1987). Antithetic variates are a special case of the general technique of systematic sampling, as discussed in Evans and Swartz (2000). The basic idea is to induce a symmetry in the integrand that is possessed by the importance sampler. This technique is particularly useful when the basic importance sampler provides a relatively poor approximation to the integrand.

Here there is a natural antithetic variate, namely $\tilde{\bar{R}} = -\bar{R}$, which also has the multivariate standard normal distribution. Since $\text{corr}(\bar{R}^i, \tilde{\bar{R}}^i) = -1$ and $\theta^i(\bar{r})$ is a monotone function of $\bar{r}^i$ for fixed $\bar{r}_{i-1}$, we would expect $\theta(\bar{R}^i)$ and $\theta(\tilde{\bar{R}}^i)$ to be highly negatively correlated, as required for the method of antithetic variates to be effective. Suppose then that $\theta_{[j]} = \theta(\bar{R}_{[j]})$ and $\tilde{\theta}_{[j]} = \theta(\tilde{\bar{R}}_{[j]})$ for $j = 1, \ldots, m$. Then, from (5) and (6), the Monte Carlo estimators of the normalizing constant and posterior expectation under antithetic importance sampling are

$$\hat{c} = (km)^{-1}\sum_{j=1}^{m}\frac{1}{2}(u_j + \tilde{u}_j) \tag{10}$$

and

$$\hat{\mu} = \sum_{j=1}^{m}\left\{v(\theta_{[j]})w_j + v(\tilde{\theta}_{[j]})\tilde{w}_j\right\} \tag{11}$$

respectively, where $\tilde{u}_j = L(\tilde{\theta}_{[j]})\lambda(\tilde{\theta}_{[j]})/h(\tilde{\theta}_{[j]})$ are the importance weights, $w_j = u_j/\sum_{i=1}^{m}(u_i + \tilde{u}_i)$ and $\tilde{w}_j = \tilde{u}_j/\sum_{i=1}^{m}(u_i + \tilde{u}_i)$ are the normalized importance weights.

Standard manipulations give estimators of the standard errors of (10) and (11) under antithetic importance sampling as

$$\text{s.e.}(\hat{c}) = \hat{c}\{m\hat{\text{var}}(w_j + \tilde{w}_j)\}^{1/2}$$

and

$$s \equiv \text{s.e.}(\hat{\mu}) = \{m\hat{\text{var}}(v(\theta_{[j]})w_j + v(\tilde{\theta}_{[j]})\tilde{w}_j)\}^{1/2},$$

where $\hat{\text{var}}$ denotes estimated Monte Carlo variance.

3.4 Marginal densities

Here we further exploit the sample produced by the simulation algorithm to obtain numerical estimates of marginal densities of one or more parameters in multiparameter settings.

It follows as in Kharroubi and Sweeting (2010) that a simulation-consistent estimator of the marginal posterior density $p(\theta_i|x)$ of the first i components of θ is given by

$$\bar{p}(\theta_i|x) = \frac{(2\pi)^{(d-i)/2}}{\hat{c}m}\sum_{j=1}^{m}\frac{\bar{L}(\theta_{[j]i})\bar{L}(\theta_i, \tilde{\theta}_{[j]}^{(i+1)})\bar{\lambda}(\theta_i, \tilde{\theta}_{[j]}^{(i+1)})}{\bar{L}(\theta_{[j]})}\prod_{k=i+1}^{d}\frac{-\bar{R}_{[j]}^{k}}{\bar{l}_k(\theta_{[j]k})}, \quad (12)$$

where, for $j = 1, \ldots, m$,

$$\tilde{\theta}_{[j]}^{(i+1)} = \theta_{[j]}^{(i+1)} + \bar{\theta}^{(i+1)}(\theta_i) - \bar{\theta}^{(i+1)}(\theta_{[j]i}). \quad (13)$$

Formula (12) produces a smooth functional form for the marginal posterior density of θ_i and, again, requires no maximization procedure for its implementation. If a specific parameter is of interest then this may be included as the first component of θ so that its marginal posterior density can be computed from Eq. (12). Of course, a kernel density estimator of the posterior density of any parametric function can be readily computed from the importance sample.

Finally, it is clearly essential to employ an efficient and reliable method for inverting the signed root log-likelihood ratio. Fortunately the inversion consists of a sequence of one-dimensional inversions for $\theta^1, \theta^2, \ldots, \theta^d$. At each stage an initial value for a standard Newton procedure may be obtained by fitting a cubic approximation to $r(\theta_i)$, as a function of θ^i, that passes through the values zero and $\pm\sqrt{d}$, as described in Appendix B of Kharroubi and Sweeting (2010).

Example 1 Normal regression model with censored data

We revisit the censored regression model for the failure data given by Crawford (1970), as described in Schmee and Hahn (1979). This has been used as a running example in Sweeting and Kharroubi (2003, 2005) and Kharroubi and Sweeting (2010) to illustrate various posterior quantities of interest, including posterior distribution functions, posterior moments and marginal posterior densities, in addition to a data augmentation scheme and importance sampling based on signed-root log-likelihood ratios. The data arise from temperature accelerated life tests on electrical insulation in 40 motorettes. Ten motorettes were tested at each of four temperatures in degrees Centigrade (150°, 170°, 190° and 220°), resulting in a total of $l = 17$ failed units and $n - l = 23$ unfailed (i.e. censored) units.

The model for the data takes the form

$$X_i = \beta_0 + \beta_1 v_i + \sigma \epsilon_i, \quad i = 1, \ldots, 40,$$

where X_i is the $\log_{10}$(ith failure time), $v_i = 1000/(\text{temperature} + 273.2)$ and the errors ϵ_i are independent standard normal. The time to failure is in hours.

Reordering the data so that the first l observations are uncensored and the remaining $n - l$ are censored, the loglikelihood function has the form

$$-l \log\sigma - \frac{1}{2} \sum_{i=1}^{l} \left(\frac{X_i - \beta_0 - \beta_1 v_i}{\sigma} \right)^2 + \sum_{i=l+1}^{n} \log \left\{ 1 - \Phi \left(\frac{Z_i - \beta_0 - \beta_1 v_i}{\sigma} \right) \right\},$$

where X_i are the observed failure times, Z_i are the censored event times and Φ is the standard normal distribution function. For computational convenience we work with the parameterisation $\theta = (\beta_0, \beta_1, \phi)$, where $\phi = \log \sigma$. From Sweeting and Kharroubi (2005), the MLE is found to be $\hat{\theta} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\phi}) = (-6.0193, 4.3112, -1.3502)$ and

$$J = \begin{pmatrix} 427.66 & 931.31 & -65.39 \\ 931.31 & 2033.55 & -145.49 \\ -65.39 & -145.491 & 41.29 \end{pmatrix}.$$

For the purpose of illustration, we examine the accuracy and computational efficiency gain of formulae (6) and (11). The signed root based simulation algorithm using exponential tilting was run with $m = 1000$. In the case of the noninformative prior $\lambda(\theta) \propto 1$ the posterior expectation of $v(\theta) = \beta_0 + \beta_1 + \exp(\phi)$ obtained from (6) was found to be – 1.4980. The estimated standard error (9) is found to be 0.0175. The corresponding expectation and standard error obtained from the signed root based simulation algorithm in Kharroubi and Sweeting (2010) were – 1.4995 and 0.0263. The exact expectation and standard error obtained by running the importance sampler incorporating antithetic sampling and control variates scheme in Sect. 4 with $m = 10000$ were – 1.4986 and 0.0115 respectively. The exact expectation and standard error here are in line with those obtained via the adjusted importance sampling technique. In particular, we worked with the standard population Monte Carlo (PMC) where the total number of evaluations of the posterior density was fixed to 10000, as this is usually the most costly step in MC algorithms (Bugallo et al. 2017). The exact expectation and standard error were found to be – 1.4986 and 0.0116 respectively

We now apply the method of antithetic variates to this example. The signed root based simulation algorithm using exponential tilting was run with $m = 500$ and the

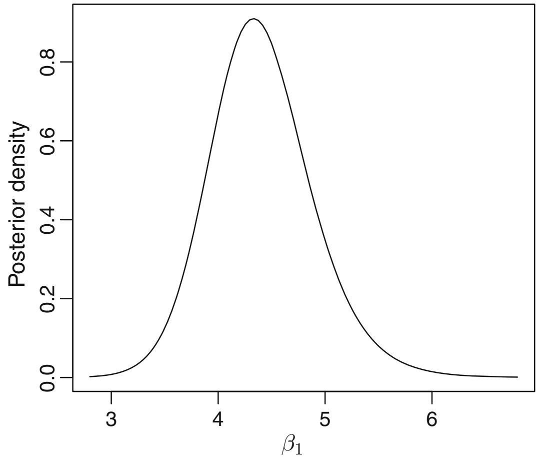

Fig. 2 Marginal posterior density of β_1 for the censored normal regression model: importance sampling (dotted line), exact (dashed line)

posterior expectation of the function $v(\theta) = \beta_0 + \beta_1 + \exp(\phi)$ obtained from (11) was found to be -1.4982. The estimated standard error of (11) was found to be 0.0166. We observe that the posterior expectation estimate shows a slight improvement over those obtained using formula (6) even with very few Monte Carlo draws. The method of antithetic variates here provides a small but worthwhile reduction in variability. We return to these results in Sect. 4 when we use a control variate.

Suppose next that we are primarily interested in the regression coefficient β_1. It is then preferable to work in a parameterization with β_1 the first component of θ, such as $\theta = (\beta_1, \beta_0, \phi)$. The marginal posterior density of β_1 may then be readily computed from formula (12). This density, represented by the dotted line in Fig. 2, is plotted along with the true marginal posterior distribution (solid line) calculated via the longer run described above. We see that the estimated distribution (12) is hardly distinguishable from the exact.

Example 2 Nonlinear repeated measures model for bacterial clearance

As a second example, we revisit the nonlinear repeated measures model for bacterial clearance data given by Smith et al. (2009). This has been used as a running example in Kharroubi and Sweeting (2016) to illustrate, respectively, new tilted Laplace approximations for posterior expectations, predictive distributions and marginal posterior distributions and importance sampling based on signed-root log-likelihood ratios. The data are in the form of radioactivity measurements on patients from three groups: healthy subjects (HC), subjects with Crohn's disease (CD), and subjects with ulcerative colitis (UC). The measurements were taken at time points 0, 4, 24, 48 and 72 h following inoculation with ^{32}P-labelled *E. coli*. Data from 14 healthy subjects, 14 patients with CD and three with UC were used. A nested replicated repeated measures design was adopted.

Let X_{gijk} denote the radioactivity measurement from the kth arm of the ith subject in group g at time t_j. Here $g = 1, 2, 3,\ i = 1, \ldots, n_g,\ j = 1, \ldots, 5,\ k = 1, \ldots, n_{gi}$, where n_g is the number of subjects in group g and n_{gi} is the number of readings at each time point of subject i in group g. The model may be written as

$$\log X_{gijk} = \alpha_{gi} - \beta_{gi} t_j + f(t_j) + \epsilon_{gijk},$$

where

$$\begin{pmatrix} \alpha_{gi} \\ \beta_{gi} \end{pmatrix} \sim N_2\left(\begin{pmatrix} \alpha_g \\ \beta_g \end{pmatrix}, \Sigma\right),$$
$$\epsilon_{gijk} \sim N(0, \sigma^2),$$

$(\alpha_{gi}, \beta_{gi})^T$, ϵ_{gijk} are all independent and

$$f(t) = \log(1 + e^{\zeta - \eta t}).$$

Here α_g is the additive group effect, β_g is the linear effect of time in group g, α_{gi}, β_{gi} are the corresponding quantities for subject i in group g and Σ is the common covariance matrix of the random effects α_{gi}, β_{gi}. The form of $f(t)$ produces a biexponential model for the clearance curves, which reflects the biphasic nonexponential kinetics. The likelihood function for this 12 parameter model is the product of the independent multivariate normal densities of the vectors of responses for each individual.

We use the parameterization $\theta = (\alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2, \beta_3, \zeta, \eta, \log \sigma_\alpha, \log \sigma_\beta, \lambda, \log \sigma)$, where σ_α, σ_β are the standard deviations of α and β and $\lambda = \frac{1}{2}\log\{(1+\rho)/(1-\rho)\}$ is the Fisher transformation of their correlation ρ. The maximum likelihood estimate is found to be $\hat{\theta} = (6.048, 6.752, 6.258, 0.01449, 0.00271, 0.02070, 0.6230, 0.5044, \text{-}0.615, \text{-}5.471, -0.112, -1.141)$. When there are many parameters in the model, as here, it is useful to carry out an initial tilt of the likelihood equal to $\exp\{-\dot{l}(\hat{\theta})^T \theta\}$ in order to produce a modified likelihood for which $\hat{\theta}$ is the exact maximizer. This avoids any potential problems in calculating signed roots close to zero. This initial tilt costs nothing and can be applied automatically. An initial tilt is also useful when the exact local maximum is hard to locate, or even when it does not exist; see Sweeting and Kharroubi (2005).

We now investigate exponentially tilted signed root based importance sampling and antithetic variates technique for this more challenging 12-parameter model. The simulation algorithm was run with $m = 500$. An initial run indicated the occasional presence of a secondary mode in the conditional tilted likelihood for the parameter $\log \sigma_\beta$. However, an initial tilt of $\exp(-2.3 \log \sigma_\beta)$, determined as outlined above, successfully removed this mode and allowed the generation of the importance sample. In the case of the noninformative prior $\lambda(\theta) \propto 1$, the posterior expectation (6) of the difference $\delta \equiv \beta_2 - \beta_1$ between the final bacterial clearance rates of CD and HC patients is found to be -0.0116. The estimated standard error (9) is found to be 0.00013. The corresponding expectation and standard error obtained using the method of antithetic variates were found to be -0.0115 and 0.0001 respectively. The exact value of -0.0114 was obtained by running the importance sampler incorporating antithetic sampling and control variates scheme in Sect. 4 until a coefficient of variation of 0.005 for $\hat{c}$ was achieved, resulting in a sample of size $m = 6328$. The standard PMC technique yields the posterior expectation and standard error of -0.0114 and 0.0001 respectively. Clearly, antithetic variate technique provides accurate approximation here.

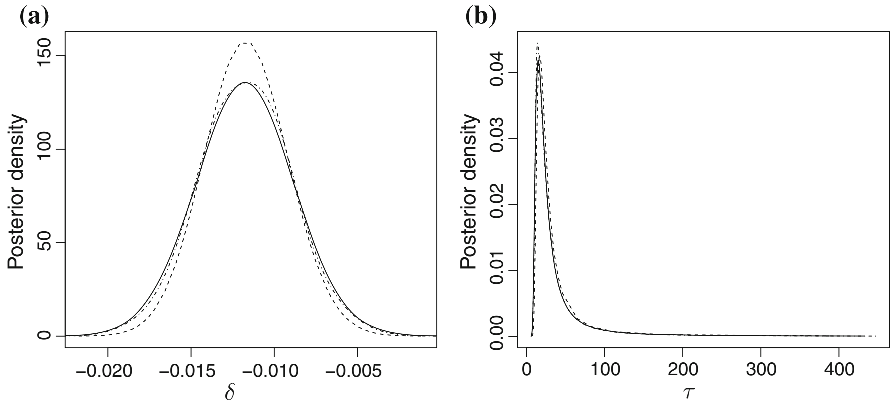

Fig. 3 Marginal posterior density of **a** δ and **b** τ for the bacterial clearance model: importance sampling (dotted line), exact (dashed line), control variates (dashed dot line)

The full marginal posterior distribution of δ can be obtained by making δ the first component of θ. Based on the sample $\theta_{[j]}$ and weights u_j, weighted kernel density estimate of the marginal posterior density of δ is readily computed. This density is shown as the dashed curve in Fig. 3a. The solid curve is the exact posterior density obtained using the longer run described above, from which we see that formula (12) slightly overstates the posterior precision of δ. It is perhaps worth mentioning that the posterior density using the standard PMC technique is hardly distinguishable from that obtained using the longer run described above.

Consider next the total clearance time τ to the point where 95% of inoculated material would be cleared for CD patients, which is approximately $\{\log 20 - \log(1 + e^{\zeta})\}/(24\beta_2)$. For the posterior distribution of τ it is necessary to include the (known) constraint $\beta_2 > 0$. This is most easily achieved by taking $1/\tau$ to be the first component of θ. The posterior density of $1/\tau$ can then be obtained from (12) along with the constraint and straightforwardly transformed to the density of τ, which is shown as the dashed curve in Fig. 3b. This is very close to the exact posterior density, which is shown as the solid curve. The posterior distribution of τ is highly skewed, producing very wide credible intervals. For example, the approximate equitailed 95% posterior credible interval for τ is (10.6, 224.5). The exact interval from the longer run described above is (10.9, 204.3). Not surprisingly the extreme tail behaviour here is hard to approximate accurately. As is the case with the results above, both the posterior distribution and equitailed 95% posterior credible interval for τ using the standard PMC technique are hardly distinguishable from those obtained using the longer run described above.

4 Control variates based on exponentially tilted signed roots

In this section we show how to construct control variates based on exponentially tilted signed roots in order to achieve further variance reduction in an exponentially tilted signed root based importance sampling scheme. We begin in Sect. 4.1 by reviewing some asymptotic approximations based on exponentially tilted signed roots.

These tilted approximations will form the basis of the control variates constructed in Sects. 4.2–4.4.

4.1 Approximations based on exponentially tilted signed roots

We begin by reviewing some results in Kharroubi and Sweeting (2016) associated with exponentially tilted signed root log-likelihood ratios. The reader is referred to this paper for detailed derivations.

Considering θ as a random variable, define the random variable $\bar{R} = \bar{r}(\theta)$. We consider the likelihood $\bar{L} = \bar{L}_n$ to be one in a sequence of likelihoods $\bar{L}_1, \bar{L}_2, \ldots$ arising from a sequence of data distributions indexed by n. Typically n will be sample size for independent data, but could refer to time in a time series, for example. In regular cases, as shown in Kharroubi and Sweeting (2016), the posterior density $f(\bar{r})$ of $\bar{R}$ satisfies

$$f(\bar{r}) = k^{-1}\phi(\bar{r})\prod_{i=1}^{d}\bar{q}^i(\bar{r}_i), \tag{14}$$

where $\phi(\cdot)$ is the $d-$dimensional standard normal density, k is a normalizing constant and

$$\bar{q}^i(\bar{r}_i) = \left\{\frac{-\bar{r}^i}{\bar{l}_i(\theta_i)}\right\}\left\{\frac{\bar{\nu}_i(\theta_i)}{\bar{\nu}_{i-1}(\theta_{i-1})}\right\}, \tag{15}$$

where $\bar{\nu}_i(\theta) = \bar{\lambda}(\theta)|\bar{j}^{(i+1)}(\theta)|^{-1/2}$ and $\bar{j}^{(i)}(\theta)$ is the submatrix of $\bar{j}(\theta)$ corresponding to $\theta^{(i)}$ (setting $|\bar{j}^{(d+1)}(\theta)| = 1$). It is also shown in Kharroubi and Sweeting (2016) that, under suitable assumptions on the likelihood function, (15) has the asymptotic form

$$\bar{q}^i(\bar{r}_i) = 1 + \bar{a}^i(\bar{r}_{i-1})r^i + \bar{b}^i(\bar{r}_{i-1})(\bar{r}^i)^2 + \bar{c}^i(\bar{r}_{i-1})(\bar{r}^i)^3 + O(n^{-2}), \tag{16}$$

where $\bar{a}^i(\bar{r}_{i-1}) = O(n^{-1/2})$, $\bar{b}^i(\bar{r}_{i-1}) = O(n^{-1})$ and $\bar{c}^i(\bar{r}_{i-1}) = O(n^{-3/2})$. Here $\bar{q}^i$ is considered as a function of $\bar{r}^i$ for fixed $\bar{r}_{i-1}$.

Now let $v(\theta)$ be a smooth function of θ and denote by k^* the new normalizing constant associated with $\left\{\prod_{i=1}^{d}\bar{q}^{*i}(\bar{r}_i)\right\}\phi(\bar{r})$, where $\bar{q}^{*i}(\bar{r}_i) = \{v(\theta_i)/v(\theta_{i-1})\}\bar{q}^i(\bar{r}_i)$. Finally, for $1 \le i < d$ define the constants

$$k_i = \int \prod_{j=i+1}^{d} \bar{q}^j(\bar{r}_j)\phi(\bar{r}^{(i+1)})d\bar{r}^{(i+1)}.$$

The following results give expressions for the constant of proportionality, posterior expectation of $v(\theta)$ and marginal posterior density of the first i components of θ. The proof is given in Appendix D of Kharroubi and Sweeting (2010).

$$c = (2\pi)^{d/2}|J|^{-1/2}L(\hat{\theta})\lambda(\hat{\theta})k \tag{17}$$

$$\mu = v(\hat{\theta})k^*/k \tag{18}$$

$$p(\theta_i|y) = c^{-1}(2\pi)^{(d-i)/2}L(\theta_i)\nu_i(\theta_i)k_i \tag{19}$$

Formulae (17), (18) and (19) are all exact. In order to obtain approximations amenable to the application of control variates in the next section we need to obtain asymptotic approximations to the constants k, k^* and k_i appearing in these formulae. Suitable approximations, $\bar{t}_d, \bar{t}_d^*$ and $\bar{t}_{d_i}$ respectively are presented in Kharroubi and Sweeting (2016). These approximations are asymptotically accurate to $O(n^{-2})$ and lead to the approximations

$$\bar{c}_{asy} = (2\pi)^{d/2}|\bar{J}|^{-1/2}L(\hat{\theta})\lambda(\hat{\theta})\bar{t}_d \tag{20}$$

$$\bar{\mu}_{asy} = v(\hat{\theta})\bar{t}_d^*/\bar{t}_d \tag{21}$$

$$\bar{p}_{asy}(\theta_i|y) = \bar{c}_{asy}^{-1}(2\pi)^{(d-i)/2}\bar{L}(\theta_i)\bar{v}_i(\theta_i)\bar{t}_{d_i} \tag{22}$$

4.2 Control variates

In this section we recast the control variates scheme described in Kharroubi and Sweeting (2010) in terms of the control variates based on the exponentially tilted likelihood function in order to achieve further variance reduction.

The use of control variates requires there to be a closely related integral whose value is known. In Evans and Swartz (1995, 2000) for example, a first-order normal approximation to the posterior density multiplied by a Laplace approximation is used as a control variate in an importance sampling scheme. Here we base our control variates on formulas (20), (21) and (22) for the constant of proportionality, posterior expectation of $v(\theta)$ and marginal posterior density of the first i components of θ respectively, with the definitions of $\bar{t}_d, \bar{t}_d^*$ and $\bar{t}_{d_i}$ in Kharroubi and Sweeting (2016).

We begin with the constant of proportionality, c, given by formula (17). We split the integral for the constant k in (14) into two terms as

$$k = \int u(\bar{r})\phi(\bar{r})\,d\bar{r} + \bar{C}, \tag{23}$$

where

$$\bar{C} = \int \left\{ \prod_{i=1}^{d} \bar{q}^i(\bar{r}_i) - u(\bar{r}) \right\} \phi(\bar{r})\,d\bar{r}\,.$$

We then integrate the first term analytically and estimate the second term by importance sampling. Since q has the asymptotic form (16), we choose

$$u(\bar{r}) = 1 + \sum_{i=1}^{d} \bar{a}^i \bar{r}^i + \sum_{i=1}^{d} \bar{b}^i (\bar{r}^i)^2 + \sum_{i=1}^{d-1} \sum_{k=i+1}^{d} \bar{a}^i \bar{a}^k \bar{r}^i \bar{r}^k \tag{24}$$

for a suitable choice of coefficients $\bar{a}^i$ and $\bar{b}^i$. A particular choice based on an asymptotic analysis is given in Appendix E of Kharroubi and Sweeting (2010). With this choice the first integral in (23) is simply $\bar{t}_d$, so that $k = \bar{t}_d + \bar{C}$. Therefore, from (17) we see that

$$c = \bar{c}_{asy}(1 + \bar{t}_d^{-1}\bar{C})\,. \tag{25}$$

Formula (25) exhibits the constant of proportionality, c, as approximation (20) multiplied by a correction term $1 + \bar{t}_d^{-1}C$.

We next proceed to consider the posterior expectation $\mu = E\{v(\theta)|y\}$. Similarly, we break the integral defining k^* into two terms to give

$$k^* = \int u^*(\bar{r})\phi(\bar{r})\,d\bar{r} + \bar{C}^*,$$

where now

$$\bar{C}^* = \int \left\{\prod_{i=1}^{d} q^{*i}(\bar{r}_i) - u^*(\bar{r})\right\}\phi(\bar{r})\,d\bar{r}$$

and

$$u^*(\bar{r}) = 1 + \sum_{i=1}^{d}\bar{a}^{*i}\bar{r}^i + \sum_{i=1}^{d}\bar{b}^{*i}(\bar{r}^i)^2 + \sum_{i=1}^{d-1}\sum_{k=i+1}^{d}\bar{a}^{*i}\bar{a}^{*k}\bar{r}^i\bar{r}^k\,. \tag{26}$$

Again, a particular choice of constants based on asymptotic considerations is given in Kharroubi and Sweeting (2010). With this choice the first integral is t^*, giving

$$\mu = \mu_{asy}\left\{\frac{1 + (\bar{t}_d^*)^{-1}\bar{C}^*}{1 + \bar{t}_d^{-1}\bar{C}}\right\}. \tag{27}$$

Thus $(1 + (\bar{t}_d^*)^{-1}\bar{C}^*)/(1 + \bar{t}_d^{-1}\bar{C})$ may be regarded as a multiplicative correction term to the asymptotic approximation (21).

Next, we use exponentially tilted importance sampling to estimate $\bar{C}$ and $\bar{C}^*$. Using the m values $\theta_{[1]}, \ldots, \theta_{[m]}$ from $g(\theta)$ produced by the simulation algorithm in Sect. 3.2, we find that $\bar{C}$ and $\bar{C}^*$ are consistently estimated by

$$\hat{\bar{C}} = \frac{1}{m}\sum_{j=1}^{m}\left\{\frac{-|\bar{J}|^{1/2}\bar{\lambda}(\theta_{[j]})}{\bar{\lambda}(\hat{\theta})}\prod_{i=1}^{d}\frac{\bar{R}^i_{[j]}}{\bar{l}_i(\theta_{[j]i})} - u(\bar{R}_{[j]})\right\} \tag{28}$$

and

$$\hat{\bar{C}}^* = \frac{1}{m}\sum_{j=1}^{m}\left\{\frac{-|\bar{J}|^{1/2}v(\theta_{[j]})\bar{\lambda}(\theta_{[j]})}{v(\hat{\theta})\bar{\lambda}(\hat{\theta})}\prod_{i=1}^{d}\frac{\bar{R}^i_{[j]}}{\bar{l}_i(\theta_{[j]i})} - u^*(\bar{R}_{[j]})\right\} \tag{29}$$

respectively. Substituting (28) and (29) into (25) and (27) we finally obtain

$$\hat{c} = \bar{c}_{asy}(1 + \bar{t}_d^{-1}\hat{\bar{C}}) \tag{30}$$

and

$$\hat{\mu} = \bar{\mu}_{asy} \left\{ \frac{1 + (\bar{t}_d^*)^{-1} \hat{\bar{C}}^*}{1 + \bar{t}_d^{-1} \hat{\bar{C}}} \right\} \tag{31}$$

as our tilted Monte Carlo estimators of the normalizing constant and posterior expectation respectively, neither of which require any conditional maximization. Straightforward manipulations give the approximation

$$s = \bar{\mu}_{asy} [\hat{\text{var}}\{(\bar{t}_d^*)^{-1} \hat{\bar{C}}^* - \bar{t}_d^{-1} \hat{\bar{C}}\}]^{1/2}, \tag{32}$$

to the standard error of (31).

4.3 Antithetic variates

As is the case in Sect. 3.3, we can apply the method of antithetic variates. Suppose that $\theta_{[1]}, \ldots, \theta_{[m]} \sim g(\theta)$ based on $\bar{R}$ and $\tilde{\theta}_{[1]}, \ldots, \tilde{\theta}_{[m]} \sim g(\tilde{\theta})$ based on $\tilde{\bar{R}} = -\bar{R}$. Then, from (30), (31) and (32), the estimators of the normalizing constant, posterior expectation and standard error using antithetic and control variates are

$$\hat{c} = \bar{c}_{asy} (1 + (2\bar{t}_d)^{-1} (\hat{\bar{C}} + \tilde{\bar{C}})\} \tag{33}$$

$$\hat{\mu} = \bar{\mu}_{asy} \left\{ \frac{1 + (2\bar{t}_d^*)^{-1} (\hat{\bar{C}}^* + \tilde{\bar{C}}^*)}{1 + (2\bar{t}_d)^{-1} (\hat{\bar{C}} + \tilde{\bar{C}})} \right\} \tag{34}$$

and

$$s = \bar{\mu}_{asy} [\hat{\text{var}}\{(2\bar{t}_d^*)^{-1} (\hat{\bar{C}}^* + \tilde{\bar{C}}^*) - (2\bar{t}_d)^{-1} (\hat{\bar{C}} + \tilde{\bar{C}})\}]^{1/2} \tag{35}$$

respectively, where $\tilde{\bar{C}}$ and $\tilde{\bar{C}}^*$ are (28) and (29) with θ and $\bar{R}$ replaced by $\tilde{\theta}$ and $\tilde{\bar{R}}$ respectively.

4.4 Marginal densities

Finally we derive an approximate expression for the marginal density $p(\theta_i)$ of the first i components of θ using control variates based on exponentially tilted formula (22). Again, we break the integral k_i in (19) into two terms,

$$k_i = \int u_i(\bar{r}) \phi(r^{(i+1)}) \, d\bar{r}^{(i+1)} + \bar{C}_i,$$

where

$$\bar{C}_i = \int \left\{ \prod_{k=i+1}^{d} \bar{q}^k(\bar{r}_k) - u_i(\bar{r}) \right\} \phi(\bar{r}^{(i+1)}) \, d\bar{r}^{(i+1)} \tag{36}$$

and

$$u_i(\bar{r}) = 1 + \sum_{k=i+1}^{d} \bar{a}^k \bar{r}^k + \sum_{k=i+1}^{d} \bar{b}^k (\bar{r}^k)^2 + \sum_{k=i+1}^{d-1} \sum_{s=k+1}^{d} \bar{a}^k \bar{a}^s \bar{r}^k \bar{r}^s$$

with $\bar{a}^i$ and $\bar{b}^i$ defined in Kharroubi and Sweeting (2010). Then the first integral is equal to $\bar{t_d}_i$ defined in Kharroubi and Sweeting (2016). This implies that

$$p(\theta_i|x) = c^{-1}(2\pi)^{(d-i)/2} L(\theta_i)\nu_i(\theta_i)(\bar{t}_{d_i} + \bar{C}_i)\,. \tag{37}$$

To estimate the first integral in (36), we use the device of Sect. 3.4, which gives

$$\hat{\bar{C}}_i = \frac{1}{\bar{L}(\theta_i)\bar{\nu}_i(\theta_i)} \frac{1}{m} \sum_{j=1}^{m} \frac{\bar{L}(\theta_{[j]i})\bar{L}(\theta_i, \tilde{\theta}_{[j]}^{(i+1)})\bar{\lambda}(\theta_i, \tilde{\theta}_{[j]}^{(i+1)})}{\bar{L}(\theta_{[j]})} \prod_{k=i+1}^{d} \frac{-\bar{R}_{[j]}^k}{\bar{l}_k(\theta_{[j]k})}$$
$$-\frac{1}{m} \sum_{j=1}^{m} u_i(R_{[j]})\,.$$

where $\tilde{\theta}_{[j]}^{(i+1)}$ is given by (13). From (37) and (30) we finally obtain

$$\hat{p}(\theta_i|x) = \bar{p}_{asy}(\theta_i|x)\{1 + (\bar{t_d}_i)^{-1}\hat{\bar{C}}_i\}\{1 + \bar{t}_d^{-1}\hat{\bar{C}}\}^{-1} \tag{38}$$

as an exponentially tilted simulation-consistent estimator of $p(\theta_i|y)$. Again, formula (38) exhibits the marginal posterior density of θ_i as the exponentially tilted asymptotic approximation (22) multiplied by a correction term. As is the case with the marginal density formula in Sect. 3.4, the implementation of formula (38) requires no conditional maximization.

Example 1 Normal regression model with censored data (continued)

We start by examining the accuracy and computational efficiency gain of formula (31). The signed root based simulation algorithm using exponential tilting obtained in Sect. 3 with $m = 1000$ yields the posterior expectation estimate of $v(\theta) = \beta_0 + \beta_1 + \exp(\phi)$ of -1.4984. The estimated standard error (32) is found to be 0.0145. The approximate posterior expectation of $v(\theta)$ obtained from (21) is -1.5085. These results, which should be compared to those in Sect. 3, show a marked reduction in sampling variability, which can also be seen from the standard error estimates. Antithetic variates may also be applied in an obvious way, which results in a slight improvement in this example.

We finally examine the β_1 marginal. Reparameterising with β_1 as the first component of θ the full marginal posterior distribution of β_1 may be computed from formula (38) with $i = 1$. This density is shown in Fig. 4. On comparing Fig. 4 with Fig. 2, this method is seen to behave well and gives an excellent marginal approximation with very few Monte Carlo draws. Antithetic variates may also be applied in an obvious way, which results in a slight improvement in this example.

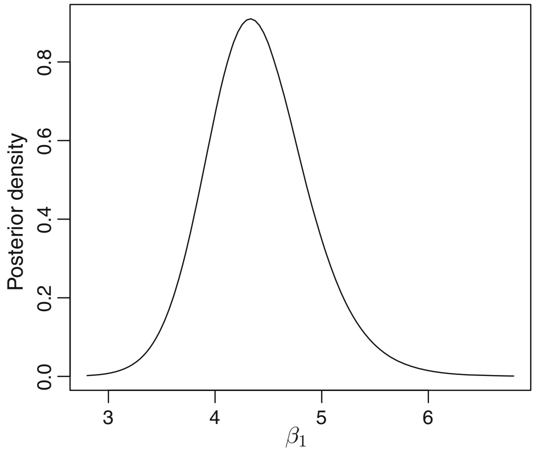

Fig. 4 Marginal posterior density of β_1 for the censored normal regression model via control variates

Example 2 Nonlinear repeated measures model for bacterial clearance (continued)

We investigate the method of control variates for this more challenging 12-parameter model. The signed root based simulation algorithm using exponential tilting obtained in Sect. 3 with $m = 500$ produces the posterior expectation (31) of $\delta \equiv \beta_2 - \beta_1$ of -0.0114, which is identical to the exact obtained previously in Sect. 3 and may be compared to the asymptotic value of -0.0118 obtained from (21). The estimated standard error (32) is found to be 0.0005. Clearly, the control variate technique provides an extremely accurate approximations here.

On reparameterization, as in Sect. 3, and based on the sample $\theta_{[j]}$ and weights u_j, weighted kernel density estimates of the marginal posterior densities of δ and $1/\tau$ are readily computed using formula (38) and the posterior density of τ obtained by transformation. These densities are shown as the dashed-dotted curves in Fig. 3. The exact posterior densities, shown as the solid curves in Fig. 3, were based on the longer run incorporating antithetic sampling. The equitailed 95% posterior credible interval for τ was (10.8, 209.8), which compares favourably with the exact interval (10.9, 204.3). An additional accuracy and small decrease in sampling variability can also be achieved by applying antithetic variates.

5 Discussion

In this paper we have explored the use of importance sampling based on exponentially tilted signed root log-likelihood ratios for Bayesian computation and have provided some numerical illustrations. The proposed exponentially tilted signed root based importance sampling algorithm in conjunction with antithetic variates has been shown to be useful for Bayesian computation in parameterized models of moderate dimension. The results of Sect. 4 can be regarded as providing external simulation-based error estimates for asymptotic approximations, which have not been available before. The results could therefore provide the basis for an extension to a computational Bayesian package based on asymptotic formulae, including simulation-based checks on the accuracy of asymptotic approximations and/or simulation-based tuning of these approximations.

The curse of dimensionality remains a stumbling block with the methods here. The theory and examples indicate that there is good reason to believe that they will often perform extremely well in moderately-dimensioned problems. More extensive investigation is required, however, to study the full range of application and to make proper comparisons between various competing methods.

The increase in precision reported for control variates is substantial in comparison with the straight importance sampling methods. Although additional computer code will be needed for the control variate methods, they will generally require less computing time than straight importance sampling since we anticipate smaller Monte Carlo sample sizes.

Exponential tilting has been shown to be particularly relevant for signed root based importance sampling and control variates and we have successfully implemented this for models containing many parameters. The procedure works well since the log-likelihood ratio decomposition means that we can construct a sequence of one-dimensional importance functions, each based on an approximation that captures the shape of the associated one-dimensional conditional posterior distribution. Many other importance samplers have been suggested for Bayesian inference in the literature. As mentioned earlier, in general importance functions based on local approximations to the posterior distribution can have poor tail behaviour, and the choice of parameterization will be crucial. Our procedure essentially only requires the availability of a sufficiently well-behaved likelihood function and has the significant advantage of not needing to be tailored to the specific model under study. For models containing latent variables for which the likelihood cannot be written down in closed form, the signed root based importance sampling scheme could be incorporated into a suitable form of poor man's data augmentation, as described in Sweeting and Kharroubi (2005). Alternatively, signed root based importance sampling could be carried out within a block Gibbs sampler that alternated between simulation from the conditional posterior distribution of θ given the latent variable z and of z given θ.

Our overall conclusion is that signed root based importance sampling approximations for posterior inference incorporating exponential tilting are computationally stable and both antithetic and control variates based on these approximations are efficient and are viable alternatives to other importance sampling methods (for example, conventional/adaptive importance sampling or population Monte Carlo) for sufficiently regular models. A set of [R] programs to implement the approach in this article is available on request. The main code is generic and the user only needs to supply the necessary code specific to his/her model.

Acknowledgements The author would particularly like to thank Professor Trevor J Sweeting for all his continual support, useful guidance and invaluable insights during his time working on this manuscript.

References

Bugallo MF, Elvira V, Martino L, Luengo D, Miguez J, Djuric PM (2017) Adaptive importance sampling: the past, the present, and the future. IEEE Signal Process Mag 34(4):60–79

Cerquetti A (2007) A note on Bayesian nonparametric priors derived from exponentially tilted Poisson–Kingman models. Stat Probab Lett 77:1705–1711
Crawford DE (1970) Analysis of incomplete life test data on motorettes. Insul Circuits 16:43–48
Efron B (1981) Nonparameteric standard errors and confidence intervals. Can J Stat 9:139–58
Elvira V, Martino L, Luengo D, Bugallo MF (2016) Heretical multiple importance sampling. IEEE Signal Process Lett 23(10):1474–1478
Evans M, Swartz T (1995) Methods for approximating integrals in statistics with special emphasis on Bayesian integration problems. Stat Sci 10:254–272
Evans M, Swartz T (2000) Approximating integrals via Monte Carlo and deterministic methods. Oxford University Press, Oxford
Fuh CD, Teng HW, Wang RH (2013) Efficient importance sampling for rare event simulation with applications. arXiv:1302.0583 [stat.ME]
Hammersley JM, Handscomb DC (1964) Monte Carlo methods. Methuen, London
Hesterberg T (1995) Weighted average importance sampling and defensive mixture distributions. Technometrics 37:185–194
Kharroubi SA, Sweeting TJ (2010) Posterior simulation via signed root log-likelihood ratios. Bayesian Anal 5(4):787–816
Kharroubi SA, Sweeting TJ (2016) Exponential tilting in Bayesian asymptotics. Biometrika 103(2):337–349
Owen A, Zhou Y (2000) Safe and effective importance sampling. J Am Stat Assoc 95(449):135–143
Pitman J (2003) Poisson–Kingman partitions. In: Goldstein DR (ed) A festschrift for terry speed, IMS lecture notes monograph series, vol 40. Institute of Mathematical Statistics, Hayward, pp 1–34
Ripley B (1987) Stochastic simulation. Wiley, New York
Schennach SM (2005) Bayesian exponentially tilted empirical likelihood. Biometrika 92:31–46
Schennach SM (2007) Point estimation with exponentially tilted empirical likelihood. Ann Stat 35:634–672
Schmee J, Hahn GJ (1979) A simple method for regression analysis with censored data. Technometrics 21:417–432
Smith AM, Rahman FZ, Hayee BH, Graham SJ, Marks DJB, Sewell GW, Palmer CD, Wilde J, Foxwell BMJ, Gloger IS, Sweeting T, Marsh M, Walker AP, Bloom SL, Segal AW (2009) Disordered macrophage cytokine secretion underlies impaired acute inflammation and bacterial clearance in Crohn's disease. J Exp Med 206:1883–1897
Sweeting TJ (1996) Approximate Bayesian computation based on signed roots of log-density ratios (with discussion). In: Bernardo JM, Berger JO, Dawid AP, Smith AFM (eds) Bayesian statistics, vol 5. Oxford University Press, Oxford, pp 427–444
Sweeting TJ, Kharroubi SA (2003) Some new formulae for posterior expectations and Bartlett corrections. Test 12(497–521):2003
Sweeting TJ, Kharroubi SA (2005) Application of a predictive distribution formula to Bayesian computation for incomplete data models. Stat Comput 15:167–178
Tierney L, Kadane J (1986) Accurate approximations for posterior moments and marginal densities. J Am Stat Assoc 81:82–86
Van Dijk HK, Kloek T, Louter AS (1986) An algorithm for the computation of posterior moments and densities using simple importance sampling. Statistician 35:83–90
Vehtari A, Gelman A, Gabry J (2016) Pareto smoothed importance sampling. arXiv:1507.02646