Bayesian Analysis (2010) **5**, Number 4, pp. 787–816

Posterior Simulation via the Signed Root Log-Likelihood Ratio

S. A. Kharroubi* and T. J. Sweeting†

Abstract. We explore the use of importance sampling based on signed root log-likelihood ratios for Bayesian computation. Approximations based on signed root log-likelihood ratios are used in two distinct ways; firstly, to define an importance function and, secondly, to define suitable control variates for variance reduction. These considerations give rise to alternative simulation-consistent schemes to MCMC 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 genetic linkage model and a censored regression model.

Keywords: Bayesian computation; Control variates; Importance sampling; Signed root log-likelihood ratio; Variance reduction

1 Introduction

Many authors have obtained useful asymptotic approximations in statistics based on signed root log-likelihood ratios. Much of this work has concerned approximations for sampling distributions of various quantities, but excellent approximations for Bayesian inference can also be obtained in this way. For example, marginal posterior approximations are obtained in DiCiccio *et al.* (1990), DiCiccio and Martin (1991), DiCiccio and Field (1991) and Sweeting (1992, 1995). One appealing feature of these approximations is that they generally require little more than standard likelihood or posterior maximization computer output for their implementation and hence may be available at little additional computational cost over simple first-order approximations.

Such approximations will eventually break down, however, when the dimension of the parameter space becomes large, or when the data are not sufficiently informative. In such cases the most widely used computational methods are based on Monte Carlo simulation. A number of researchers have investigated the use of hybrid methods that combine the best features of simulation, or numerical integration, and asymptotics; see, for example, Sweeting (1996) and the ensuing discussion. Sweeting and Kharroubi (2005) explored the application of a posterior predictive distribution formula based on signed root log-likelihood ratios as a stable importance function for use within poor man's data augmentation schemes and as a proposal distribution within a Metropolis-Hastings algorithm.

*Department of Mathematics, University of York, York, UK, mailto:sak503@york.ac.uk
†Department of Statistical Science, University College London, London, UK, mailto:trevor@stats.ucl.ac.uk

 DOI:10.1214/10-BA528

A variety of Monte Carlo importance sampling strategies have been proposed in the literature. These are sometimes based on the first-order normality of the posterior distribution; see, for example, Van Dijk et al (1986) and Evans and Swartz (1995a,b, 2000). However, such importance samplers may have poor tail behaviour. As noted in Sweeting (1996), distributional approximations based on signed root log-likelihood ratios often have very good tail properties, which suggests investigating importance sampling schemes based on these approximations.

In the present paper we develop signed root based importance sampling schemes for moderately parameterized models that give rise to exact (that is, simulation-consistent) computation of various posterior quantities of interest, including posterior distribution functions, posterior moments and marginal posterior densities. A key property is that the importance functions will often possess good tail behaviour. 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 two examples. Although the methodology is not universally applicable, in cases where it is applicable it has the obvious advantage over MCMC methods 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.

We begin in §2 by introducing importance sampling based on the signed root log-likelihood ratio. Following a brief review of some results in Sweeting (1995, 1996) and Sweeting and Kharroubi (2003), we proceed in §3 to develop suitable control variates based on asymptotic considerations. Some concluding remarks are given in §4 and various technical details are included in Appendices A - D.

2 Importance sampling based on signed roots

In this section we develop the signed root based importance sampling scheme. We begin in §2.1 by reviewing the construction of signed root log-likelihood ratios, which forms the basis of the remainder of this section. The construction of the importance sampler is described in §2.2 and the computation of marginal posterior densities discussed in §2.3.

2.1 Signed root log-likelihood ratios

We begin by reviewing the construction of signed root log-likelihood ratios. Let $L(\theta)$ be the likelihood function associated with data y from a parametric model with $\theta \in \Omega \subset \mathcal{R}^d$, where $d \geq 1$, and let $\lambda(\theta)$ be the prior density of θ, which is assumed to be continuous and positive throughout Ω. Then the posterior density of θ is

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

where $c = \int L(\theta)\lambda(\theta)\,d\theta$. The basic requirement for the methods in this paper is that there exists a unique local maximum of the likelihood function, as occurs in many

commonly used parametric models. This gives rise to a one-to-one transformation of θ to a signed root log-likelihood ratio, which forms the basis for excellent asymptotic approximations; see, for example, Sweeting (1996). In §4 we briefly discuss a possible modification when a unique local maximum does not exist.

We employ the notation in Sweeting (1995, 1996) and Sweeting and Kharroubi (2003) and use superscripts to donate the components of θ, so that $\theta = (\theta^1, \ldots, \theta^d)$. We further 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 assume that, for each $1 \leq i \leq d$ and fixed θ_{i-1}, there exists a unique local maximizer $\hat{\theta}^{(i)}(\theta_{i-1})$ of $L(\theta)$. Here $\hat{\theta}^{(1)}(\theta_0)$ is understood to be the overall maximum likelihood estimate $\hat{\theta} = (\hat{\theta}^1, \ldots, \hat{\theta}^d)$. For $j > i$, $\hat{\theta}^j(\theta_i)$ will denote the jth component of $(\theta_i, \hat{\theta}^{(i+1)}(\theta_i))$. The following convention will prove useful: for any function $g(\theta)$ and $1 \leq i < d$, $g(\theta_i)$ will denote $g(\theta_i, \hat{\theta}^{(i+1)}(\theta_i))$, where $g(\theta_0)$ is understood to be $g(\hat{\theta})$. With this convention our assumption is that, for $1 \leq i \leq d$, $\hat{\theta}^{(i)}(\theta_{i-1})$ is the unique solution of the conditional likelihood equation $l_i(\theta) = 0$, where $l(\theta) = \log L(\theta)$ is the log-likelihood function and $l_i(\theta) = \partial l(\theta)/\partial\theta^i$.

We define $l'(\theta) = \partial l(\theta)/\partial\theta = (l_1(\theta), \ldots, l_d(\theta))^T$, $j(\theta) = -d^2 l(\theta)/d\theta^2$ and $J = j(\hat{\theta})$, the observed information. Finally, we define the signed root log-likelihood ratio transformation $r(\theta) = (r^1(\theta_1), \ldots, r^d(\theta_d))$ by

$$r^i(\theta_i) = \text{sign}\{\theta^i - \hat{\theta}^i(\theta_{i-1})\}[2\{l(\theta_{i-1}) - l(\theta_i)\}]^{1/2} \tag{2}$$

for $i = 1, \ldots, d$. Notice that r^i is a function of the first i components $\theta_i = (\theta^1, \ldots, \theta^i)$ of θ. It then follows from the above assumption that $r(\theta)$ is a one-to-one data-dependent transformation of θ. This transformation is invariant under smooth one-to-one transformations of θ of the form $\phi^i = \phi^i(\theta_i)$, $i = 1, \ldots d$. In the scalar case, of course, $r(\theta)$ is invariant under all smooth one-to-one transformations.

2.2 Construction of the importance sampler

Consider the posterior expectation $\mu = E_p\{v(\theta)|y\}$ of the smooth function $v(\theta)$, where E_p denotes expectation under the posterior density $p(\theta|y)$. 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

$$\mu = E_g\left\{v(\theta)\frac{p(\theta|y)}{g(\theta)}\right\} \tag{3}$$

it follows that μ is consistently estimated via importance sampling by

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

where $\theta_{[1]}, \ldots, \theta_{[m]}$ are m independent draws from $g(\theta)$ and $w_j = p(\theta_{[j]}|y)/g(\theta_{[j]})$ are the importance weights. The usual strategy is to choose a suitable density $g(\theta)$ that is

close to, but more dispersed than, $p(\theta)$ in order that importance sampling is stable. See Hammersley and Handscomb (1964), for example, for a general discussion of importance sampling in statistics and van Dijk et al (1986), Glynn and Iglehart (1989), Hesterberg (1990) and Wolpert (1991) for applications in Bayesian statistics.

We see from the definition (2) of $r(\theta)$ that $\exp\{-\frac{1}{2}\|r(\theta)\|^2\} = L(\theta)/L(\hat{\theta})$, implying that, when expressed in terms of $r(\theta)$, the likelihood function is of exactly multivariate standard normal form. This suggests that the posterior distribution of $r(\theta)$ will be close to multivariate standard normal. In fact the posterior distribution of $r(\theta)$ for independent and identically distributed (i.i.d.) observations is asymptotically multivariate standard normal to $O(n^{-1/2})$, where n is the sample size (Sweeting, 1995). But asymptotics provide a stronger property than this first-order result. It turns out that the *shape* of the posterior distribution of $r(\theta)$ is actually normal to $O(n^{-3/2})$ (Sweeting, 1995), only requiring a location-scale transformation to be multivariate standard normal to that order. This provides strong asymptotic motivation for basing an importance sampling scheme on $r(\theta)$. Additional motivation is provided by the non-asymptotic result discussed later in this section.

We use $r(\theta)$ to construct our importance sampler as follows. Suppose that R is randomly generated from the d-dimensional standard normal distribution. Then, for $i = 1, \ldots, d$, $\theta^i = \theta^i(R_i)$ is defined by inversion of $r^i(\theta_i) = R^i$ for fixed θ_{i-1}. Since $r^i(\theta_i)$ is a function of the first i components of θ, the Jacobian matrix $dr/d\theta$ is lower triangular, so that

$$\left|\frac{dr}{d\theta}\right| = \prod_{i=1}^{d} \frac{\partial r^i(\theta)}{\partial \theta^i} = \prod_{i=1}^{d} \frac{-l_i(\theta_i)}{r^i(\theta_i)} .$$

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

$$g(\theta) = (2\pi)^{-d/2} \frac{L(\theta)}{L(\hat{\theta})} \prod_{i=1}^{d} \frac{-l_i(\theta_i)}{r^i(\theta_i)} . \tag{4}$$

We remark that no difficulty arises here when $\theta^i = \hat{\theta}^i(\theta_{i-1})$, since L'Hôpital's rule gives $r^i(\theta_i)/l_i(\theta_i) \to \{-k^i(\theta_{i-1})\}^{-1/2}$ as $\theta^i \to \hat{\theta}^i(\theta_{i-1})$, where $k^i(\theta) = -\partial^2 l(\theta)/(\partial \theta^i)^2$.

Substituting (1) and (4) into (3), we see that the posterior expectation of $v(\theta)$ is

$$\mu = E_p\{v(\theta)|y\} = (2\pi)^{d/2} c^{-1} L(\hat{\theta}) E_g\{v(\theta)h(\theta)\} ,$$

where $h(\theta) = \lambda(\theta) \prod_{i=1}^{d} \{-r^i(\theta_i)/l_i(\theta_i)\}$ is a weight function. In particular, taking $v(\theta) = 1$ we see that the constant of proportionality in (1) is

$$c = (2\pi)^{d/2} L(\hat{\theta}) E_g\{h(\theta)\} \tag{5}$$

from which it follows that

$$\mu = \frac{E_g\{v(\theta)h(\theta)\}}{E_g\{h(\theta)\}} . \tag{6}$$

Notice that, since the construction of $r(\theta)$ is invariant under reparameterization in the scalar parameter case, we obtain an invariant form of importance sampling in that case. This is an important propery as the performance of the scheme will not depend on the particular form of parameterization chosen. In the multiparameter case the partial invariance property described in §2.1 holds. An alternative strategy would be to absorb the prior density $\lambda(\theta)$ into the likelihood before constructing the signed root transformation (2). This construction might improve importance sampling performance, especially if the prior knowledge is substantial, but such a scheme will suffer from a lack of invariance.

It might seem at first sight that we need to use an importance function for $r(\theta)$ more dispersed than the asymptotic standard normal form. However, this is not necessarily the case, as indicated by Lemmas 1 and 2 in Appendix A, which give fairly weak conditions under which the importance estimator of μ will converge at rate $m^{-1/2}$. They also provide measures of precision for this estimator via the observed Monte Carlo standard error.

We now summarise the signed root based importance sampling algorithm for estimating the expectations in (5) and (6).

1. Generate $R^i \sim N(0,1)$, independently for $i = 1, \ldots, d$.
2. Obtain $\theta^i = \theta^i(R_i)$ sequentially for $i = 1, \ldots, d$ as the solutions of the equations $r^i(\theta_i) = R^i$.
3. Repeat steps 1 and 2 m times to obtain $R_{[1]}, \ldots, R_{[m]}$ and the corresponding sample $\theta_{[1]}, \ldots, \theta_{[m]}$ from the importance density (4).

Now approximate $E_g\{h(\theta)\}$ by $\bar{h} = m^{-1}\sum_{j=1}^{m} h(\theta_{[j]})$. Then from (5) and (6) we obtain the consistent estimators

$$\hat{c} = (2\pi)^{d/2} L(\hat{\theta})\bar{h} \tag{7}$$

of c and and

$$\hat{\mu} = \sum_{j=1}^{m} v(\theta_{[j]}) w_j \tag{8}$$

of μ, where $w_j = (m\bar{h})^{-1} h(\theta_{[j]})$, $j = 1, \ldots, m$ are the importance weights, and $\sum_j w_j = 1$.

In order to examine the precision of (7) and (8) we need approximations to their Monte Carlo standard errors. Straightforward manipulations give the approximations

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

and

$$s \equiv \text{s.e.}(\hat{\mu}) = \{m\hat{\text{var}}(v(\theta_{[j]})w_j)\}^{1/2}, \tag{9}$$

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

The main computational effort involved in the importance sampling scheme is the solution of the equation $r(\theta) = r$ for each sampled value r of R. Since $r_i(\theta)$ is a function

of the first i coordinates of θ, the equations $r^i(\theta_i) = r^i$ can be solved sequentially for $i = 1, \ldots, d$ using a univariate update formula at each stage. Provided that good initial values are used, this is usually not too hard in moderately parameterized problems using some version of Newton's method. One such scheme based on results in Sweeting (1996) is given in Appendix B. It is important to have an efficient method of solving $r(\theta) = r$ in the multiparameter case, particularly as conditional maximization is involved at each step of the iteration. The scheme in Appendix B has worked satisfactorily in the moderately parameterized examples that we have tried, but alternative schemes can be devised. Some further possibilities are discussed in Appendix B.

It is natural to seek to reduce the Monte Carlo variability in (8) 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 §3. 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 $\tilde{R} = -R$, which also has the multivariate standard normal distribution. Since $\mathrm{corr}(R^i, \tilde{R}^i) = -1$ and $\theta^i(r)$ is a monotone function of r^i for fixed r_{i-1}, we would expect $\theta(R^i)$ and $\theta(\tilde{R}^i)$ to be highly negatively correlated, as required for the method of antithetic variates to be effective. Suppose then that $\theta_{[j]} = \theta(R_{[j]})$ and $\tilde{\theta}_{[j]} = \theta(\tilde{R}_{[j]})$ for $j = 1, \ldots, m$. Then, from (7) and (8), the Monte Carlo estimators of the normalizing constant (5) and posterior expectation (6) under antithetic importance sampling are

$$\hat{c} = (2\pi)^{d/2} L(\hat{\theta}) \frac{1}{2}(\bar{h} + \tilde{h}) \tag{10}$$

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 $\tilde{h} = m^{-1} \sum_{j=1}^{m} h(\tilde{\theta}_{[j]})$, $w_j = \{m(\bar{h} + \tilde{h})\}^{-1} h(\theta_{[j]})$ and $\tilde{w}_j = \{m(\bar{h} + \tilde{h})\}^{-1} h(\tilde{\theta}_{[j]})$. 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

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

EXAMPLE 1. *Genetic linkage model*

We illustrate the above methods with the genetic linkage model given in Rao (1973), in which n animals are distributed multinomially into four categories with cell probabilities $\left(\frac{1}{2} + \frac{\theta}{4}, \frac{1}{4}(1-\theta), \frac{1}{4}(1-\theta), \frac{\theta}{4}\right)$. Wei and Tanner (1990) apply this model to data $y = (14, 0, 1, 5)$ on 20 animals. Under a uniform prior for θ, the posterior density of θ is

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

For these data we find that $\hat{\theta} = 0.9034$ and $J = 115.042$.

We begin by comparing the importance sampling weight function $h(\theta)$ based on $r(\theta)$ with the weight function $h_1(\theta)$ based on a straight (truncated) $N(\hat{\theta}, J^{-1})$ approximation to the posterior distribution of θ. Figure 1 exhibits the two weight functions, plotted along with the likelihood function. As can be seen, the function h is stable over the main range of the likelihood function, whereas the use of h_1 would lead to a very poor importance sampler with a huge variance. Of course the latter can be improved by a judicious choice of parameterization, but the point is that such a choice is unnecessary for the importance sampling scheme proposed here.

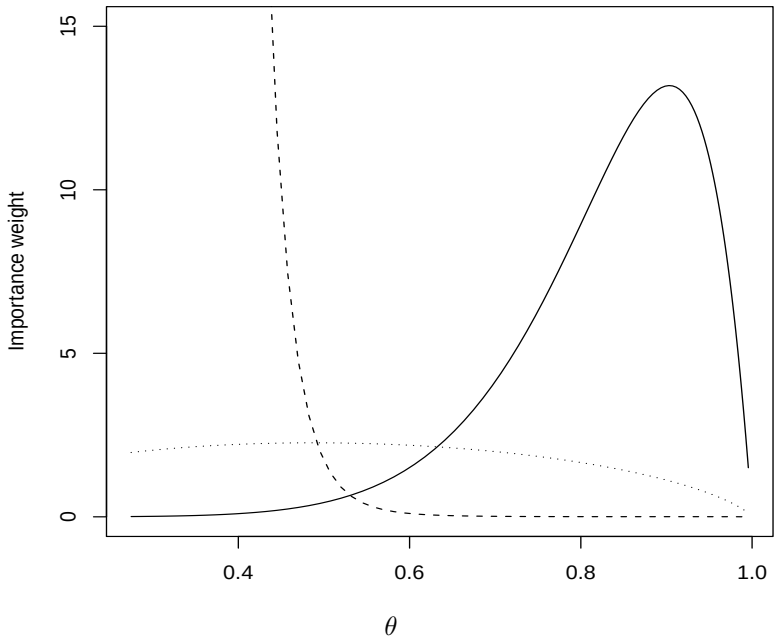

Figure 1: Genetic linkage data: importance weight functions based on normality of θ (dashed line) and normality of $r(\theta)$ (dotted line). The solid line is the likelihood function.

In order to implement the importance sampling algorithm, the non-linear equation $r(\theta) = r$ was solved using the method given in Appendix B in terms of the log-odds parameterization $\phi = \log\{\theta/(1-\theta)\}$ in order to avoid difficulties when $|r|$ is large. The algorithm was run three independent times, each with $m = 100$, yielding the values in

column 2 of Table 1 for the normalizing constant, together with the exact value obtained by numerical integration. For a more precise estimate, the algorithm may be run with a larger value of m. As can be seen from Table 1, when $m = 10000$ the resulting estimate of c^{-1} is extremely accurate.

m	c^{-1}	μ	s	$3s/\sigma$
100	$2.5386e-05$	0.8178	0.0154	0.4951
100	$2.6254e-05$	0.8521	0.0116	0.3720
100	$2.2332e-05$	0.8064	0.0164	0.5276
10000	$2.4074e-05$	0.8313	0.0013	0.0420
Exact	$2.4056e-05$	0.8311		

Table 1: Exact and approximate posterior computations for the genetic linkage model

We next examine the accuracy of formula (8) with $v(\theta) = \theta$. Given the output from the simulation algorithm, computation of (8) is straightforward and the resulting posterior expectation estimates are listed in column 3 of Table 1, again with the exact value obtained by numerical integration. Whilst these estimates exhibit some variability when $m = 100$, the estimate when $m = 10000$ is accurate to three decimal places. The entries in column 4 of Table 1 are the estimated Monte Carlo standard errors s of (8). In column 5 we give the relative errors $3s/\sigma$, where $\sigma = 0.0932$ is the exact posterior standard deviation of $v(\theta)$. These relative errors measure the size of the possible Monte Carlo error in relation to the spread of the posterior distribution. We observe that these errors are quite high when $m = 100$ but acceptably small when $m = 10000$.

We now apply the method of antithetic variates to this example. Table 2 lists the computations of interest based on three independent runs of the simulation algorithm, each with $m = 50$. As can be seen from column 2 of Table 2, the estimates of c^{-1} show a marked improvement over those in Table 1. The posterior expectation estimates show a slight improvement. We return to these results in §3 when we use a control variate.

m	c^{-1}	μ	s	$3s/\sigma$
50	$2.4113e-05$	0.8493	0.0100	0.3215
50	$2.3994e-05$	0.8135	0.0149	0.4796
50	$2.4041e-05$	0.8329	0.0108	0.3489

Table 2: Approximate posterior computations for the genetic linkage model using antithetic variates

It is not hard to check that the conditions of Lemma 1 apply in this example. We do not anticipate that the user would actually check these conditions in more complex examples, however, but rather that the algorithm would be run, any tail problem identified and the algorithm possibly modified (see §4).

EXAMPLE 2. *Censored regression*

We next consider the censored failure data given by Crawford (1970). These data arise from temperature accelerated life tests on electrical insulation in $n = 40$ motorettes. Ten motorettes were tested at each of four temperatures in degrees Centigrade (150°, 170°, 190° and 220°), resulting in a total of $l = 17$ failed units and $n - l = 23$ unfailed (*i.e.* censored) units.

As in Schmee and Hahn (1979), we fit a model of the form

$$y_i = \beta_0 + \beta_1 v_i + \sigma\epsilon_i, \quad i = 1, \ldots, n\,,$$

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

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

where Φ is the standard normal distribution function. Here we find that $\hat\theta = (\hat\beta_0, \hat\beta_1, \hat\sigma) = (-6.0193, 4.3112, 0.2592)$.

For the purpose of illustration, we examine the accuracy of (8) with $\lambda(\theta) \propto \sigma^{-1}$ and $v(\theta) = \beta_0 + 2\beta_1 + \sigma$. The signed root based simulation algorithm was run three times, each with $m = 100$, yielding the results in Table 3. The exact values were obtained using the data augmentation scheme in Tanner and Wong (1987). To examine the precision of the estimates of μ, the estimated standard errors (9) are given in column 4 along with the relative error estimates $3s/\sigma$ in the final column, where the posterior standard deviation $\sigma = 0.1075$. Even with m as low as 100, we obtain very reasonable estimates of c^{-1} and μ. The values in the final column, however, indicate that the algorithm should be run with a larger value of m.

m	c^{-1}	μ	s	$3s/\sigma$
100	$6.6451e-10$	2.8821	0.0118	0.3280
100	$6.6442e-10$	2.9012	0.0141	0.3935
100	$6.7416e-10$	2.9131	0.0139	0.3880
Exact	$6.7187e-10$	2.9048	0.0007	0.0203

Table 3: Exact and approximate posterior computations for the motorette data

We next apply the method of antithetic variates. Based on three independent runs of the simulation algorithm, each with $m = 50$, Table 4 lists the approximate posterior quantities of interest. We observe that the posterior expectation estimates have greater precision than those in Table 3. The method of antithetic variates here provides a small but worthwhile reduction in variability.

m	c^{-1}	μ	s	$3s/\sigma$
50	$6.7566e-10$	2.9103	0.0075	0.2106
50	$6.7237e-10$	2.9005	0.0061	0.1703
50	$6.8092e-10$	2.9044	0.0055	0.1542

Table 4: Approximate posterior computations for the motorette data using antithetic variates

2.3 Marginal densities

In this section 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.

First note that we can always use the independently sampled values of θ to produce an estimate of the posterior density of any parametric function by forming a weighted kernel density estimate, using standard computer software. However, in cases where the importance sample is not large the resulting density estimates may not be very smooth. In Appendix C we show how to produce a simulation-consistent smooth estimator of the marginal posterior density $p(\theta_i|y)$ of the first i components of θ from the sampled values $R_{[1]}, \ldots, R_{[m]}$ of R. This is given by

$$\hat{p}(\theta_i|y) = \frac{(2\pi)^{(d-i)/2}}{\hat{c}m} \sum_{j=1}^{m} \frac{L(\theta_{[j]i})L(\theta_i, \bar{\theta}_{[j]}^{(i+1)})\lambda(\theta_i, \bar{\theta}_{[j]}^{(i+1)})}{L(\theta_{[j]})} \prod_{k=i+1}^{d} \frac{-R_{[j]}^{k}}{l_k(\theta_{[j]k})}, \tag{12}$$

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

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

In particular, formula (12) provides a smooth functional form for the marginal posterior density of θ^1. Thus, if a specific parameter is of interest then it would be useful to include this as the first component of θ. Notice that if we set θ_i to be the observed sample values obtained via the simulation algorithm, then formula (12) requires no additional maximization procedure for its implementation.

EXAMPLE 2. *Censored regression (continued)*
Given the ordering of the parameters, it is straightforward to calculate the marginal posterior density estimate for β_0. Having applied the simulation algorithm three independent times, each with $m = 100$, to this data set in §2.2, we have three independent samples from $g(\beta_0, \beta_1, \sigma)$. The resulting β_0 marginal density estimates (dashed lines) are shown in Figure 2. The irregularity in the tails here is due to there being few observed sample values in the tails and it may be necessary to perform some additional conditional maximization in order to produce an overall smooth curve. This has been done in Figure 3. Finally, Figure 4 shows the normalized densities, revealing that the

main source of variation here is associated with the normalizing constant. Antithetic variates may also be applied in an obvious way, which results in a slight improvement in this example.

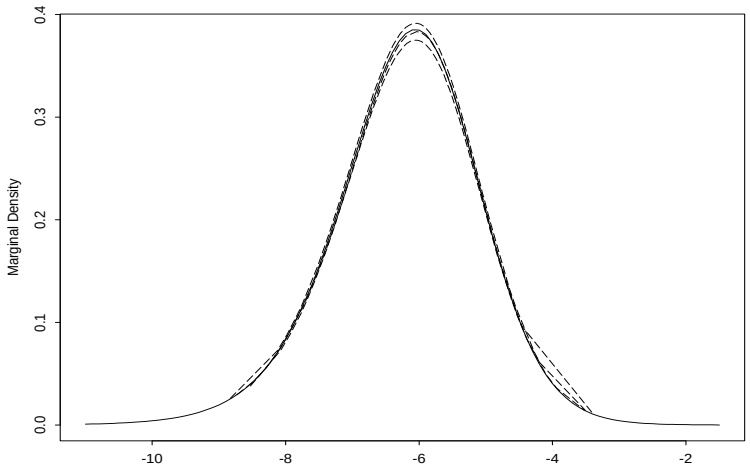

Figure 2: Censored regression data: marginal posterior density of β_0. Importance sampling with $m = 100$ (dashed lines) using sampled values; data augmentation (solid line).

3 Control variates based on signed roots

In this section we show how to construct control variates based on signed roots in order to achieve further variance reduction in a signed root based importance sampling scheme. We begin in §3.1 by reviewing some asymptotic approximations based on signed roots. These approximations will form the basis of the control variates constructed in §3.2 and §3.3. Some of the more technical details are relegated to Appendices D and E.

3.1 Approximations based on signed roots

We begin by reviewing some results in Sweeting (1995, 1996) and Sweeting and Kharroubi (2003) associated with signed root log-likelihood ratios. The reader is referred to these papers for detailed derivations.

Define $\nu_i(\theta) = \lambda(\theta)|j^{(i+1)}(\theta)|^{-1/2}$, where, for $1 < i \leq d$, $j^{(i)}(\theta)$ is the submatrix of $j(\theta)$ corresponding to $\theta^{(i)}$, with $|j^{(d+1)}(\theta)|$ set to one. For each $i = 1, \ldots, d$ define the function

$$q^i(r_i) = \{-r^i/l_i(\theta_i)\}\{\nu_i(\theta_i)/\nu_{i-1}(\theta_{i-1})\}\,,$$

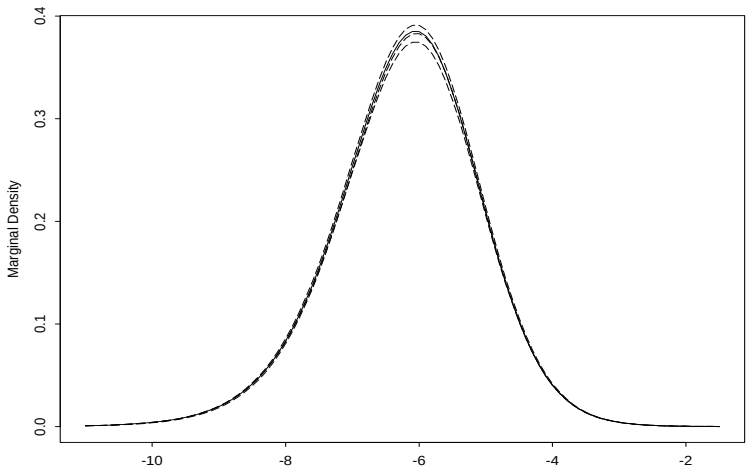

Figure 3: Censored regression data: marginal posterior density of β_0. Importance sampling with $m = 100$ (dashed lines) with additional values in the tails; data augmentation (solid line).

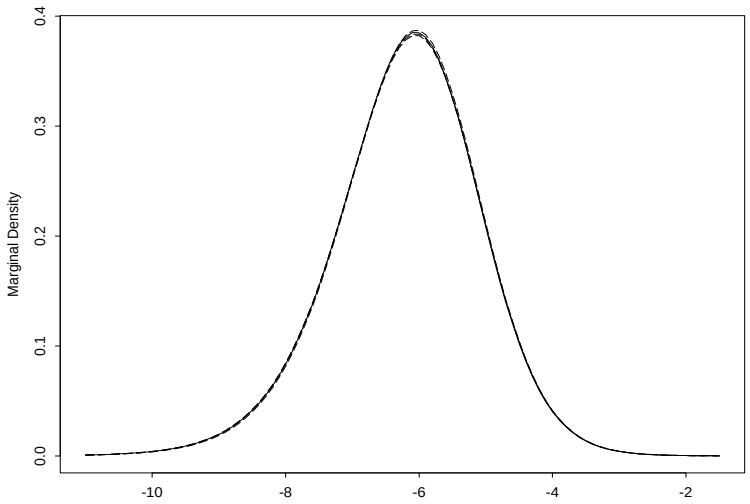

Figure 4: Censored regression data: marginal posterior density of β_0. normalized densities from Figure 3 (dashed lines); data augmentation (solid line).

where we set $\nu_0(\theta_0) = \lambda(\hat{\theta})|J|^{-1/2}$. It is shown in Sweeting (1996) that the exact posterior density $f(r)$ of $r(\theta)$ can be expressed as

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

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

Notice that the $\nu_i(\theta)$ terms for $1 < i < d$ in (14) cancel. However, it is necessary to include these terms in order that each $q^i(r_i)$ term has the asymptotic form

$$q^i(r_i) = 1 + a^i(r_{i-1})r^i + b^i(r_{i-1})(r^i)^2 + c^i(r_{i-1})(r^i)^3 + O(n^{-2})\,, \tag{15}$$

where $a^i(r_{i-1}) = O(n^{-1/2})$, $b^i(r_{i-1}) = O(n^{-1})$ and $c^i(r_{i-1}) = O(n^{-3/2})$ and n indexes the amount of information in the data (the sample size when y comprises n i.i.d. observations); see Sweeting (1995). Here q^i is considered as a function of r^i for fixed r_{i-1}.

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

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

The following result gives 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.

THEOREM 1.

$$c = (2\pi)^{d/2}|J|^{-1/2}L(\hat{\theta})\lambda(\hat{\theta})k \tag{16}$$

$$\mu = v(\hat{\theta})k^*/k \tag{17}$$

$$p(\theta_i|y) = c^{-1}(2\pi)^{(d-i)/2}L(\theta_i)\nu_i(\theta_i)k_i \tag{18}$$

Formulae (16), (17) and (18) in Theorem 1 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, $\bar{t}, \bar{t}^*$ and $\bar{t}_i$ respectively, based on results in Sweeting and Kharroubi (2003), are presented in Appendix E. These approximations are asymptotically accurate to $O(n^{-2})$ and lead to the approximations

$$c_{\text{asy}} = (2\pi)^{d/2}|J|^{-1/2}L(\hat{\theta})\lambda(\hat{\theta})\bar{t} \tag{19}$$

$$\mu_{\text{asy}} = v(\hat{\theta})\bar{t}^*/\bar{t} \tag{20}$$

$$p_{\text{asy}}(\theta_i|y) = c_{\text{asy}}^{-1}(2\pi)^{(d-i)/2}L(\theta_i)\nu_i(\theta_i)\bar{t}_i \tag{21}$$

Finally, for the case $d = 1$ consider the posterior distribution function $F(r)$ of $r(\theta)$, which may be written $F(r) = k^{-1}k(r)$, where $k(r) = \int_{-\infty}^{r} q(s)\phi(s)\,ds$. The usual $O(n^{-3/2})$ approximation to $F(r)$ (see, for example, Sweeting, 1995) is

$$\Phi(r) - \phi(r)\left\{\frac{q(r)-1}{r}\right\}.$$

It will turn out to be more convenient here, however, to use the $O(n^{-3/2})$ equivalent expression

$$F_{\text{asy}}(r) = \Phi(r) - \phi(r)(a + br)\,, \tag{22}$$

where the constants a and b are defined in Appendix E.

3.2 Control variates

The use of control variates requires there to be a closely related integral whose value is known. Evans and Swartz (1995a, 2000) used a first-order normal approximation to the posterior density multiplied by a Laplace approximation as a control variate in an importance sampling scheme. Here we base our control variates on formulas (19), (20) and (22) for the constant of proportionality, posterior expectation of $v(\theta)$ and distribution function of $r(\theta)$ respectively, with the definitions of $\bar{t}, \bar{t}^*$ and $\bar{t}_i$ given in Appendix E.

Consider first the constant of proportionality, c, given by formula (16). We break the integral for the constant k in (14) into two terms, so that

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

where

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

integrate the first term analytically and estimate the second term by importance sampling.

Since q has the asymptotic form (15), a sensible choice for u is

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

for a suitable choice of coefficients a^i and b^i. A particular choice based on an asymptotic analysis is given in Appendix E. With this choice the first integral in (23) is simply $\bar{t}$, so that $k = \bar{t} + C$. Therefore, from (16) we see that

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

Thus $1 + \bar{t}^{-1}C$ may be regarded as a multiplicative correction term to the asymptotic approximation (19).

Consider next the posterior expectation $\mu = E\{v(\theta)|y\}$. In a similar way, we break the integral defining k^* into two terms to give

$$k^* = \int u^*(r)\phi(r)\,dr + C^*\,,$$

where now

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

and

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

Again, a particular choice of constants based on asymptotic considerations is given in Appendix E. With this choice the first integral is t^*, giving

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

Formula (27) exhibits the posterior expectation of $v(\theta)$ as approximation (20) multiplied by a correction term.

Finally, we estimate C and C^* by importance sampling. Transforming back to the $\theta-$parameterization and using the m values $\theta_{[1]}, \ldots, \theta_{[m]}$ from $g(\theta)$ produced by the simulation algorithm in §2.2, we find that C and C^* are consistently estimated by

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

and

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

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

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

and

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

as our Monte Carlo estimators of the normalizing constant and posterior expectation respectively. By straightforward Taylor expansion, an approximation to the standard error of (31) is found to be

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

We next derive an approximate expression for the distribution function $F(r)$ in the case $d = 1$. Again, we break the integral $k(r) = \int_{-\infty}^{r} q(s)\phi(s)\,ds$ into two terms

$$\int_{-\infty}^{r} u(s)\phi(s)\,ds + C(r)\,,$$

where $C(r) = \int_{-\infty}^{r}\{q(s) - u(s)\}\phi(s)\,ds$ and $u(r) = 1 - ar - br^2$. Then we obtain

$$F(r) = k^{-1}\bar{t}\{F_{\text{asy}}(r) + \bar{t}^{-1}C(r)\} = (1 + \bar{t}^{-1}C)^{-1}\{F_{\text{asy}}(r) + \bar{t}^{-1}C(r)\}\,,$$

where $F_{\text{asy}}(r)$ is given by (22). Finally, we estimate $C(r)$ using the m sampled values $\theta_{[1]}, \ldots, \theta_{[m]}$ from $g(\theta)$ by

$$\hat{C}(r) = \frac{1}{m}\sum_{j=1}^{m} I_r(\theta_{[j]})\left\{\frac{-J^{1/2}\lambda(\theta_{[j]})R_{[j]}}{\lambda(\hat{\theta})l'(\theta_{[j]})} - u(R_{[j]})\right\}\,, \tag{33}$$

where

$$I_r(\theta) = \begin{cases} 1 & \theta \le \theta(r) \\ 0 & \theta > \theta(r) \end{cases}\,.$$

Thus a simulation-consistent estimator of $F(r)$ is given by

$$\hat{F}(r) = (1 + \bar{t}^{-1}\hat{C})^{-1}\{F_{\text{asy}}(r) + \bar{t}^{-1}\hat{C}(r)\}\,. \tag{34}$$

We can further apply the method of antithetic variates in the same way as was done in §2.2. Suppose that $\theta_{[1]}, \ldots, \theta_{[m]} \sim g(\theta)$ based on R and $\tilde{\theta}_{[1]}, \ldots, \tilde{\theta}_{[m]} \sim g(\tilde{\theta})$ based on $\tilde{R} = -R$. Then, from (30), (31) and (32), the estimators of the normalizing constant, posterior expectation and standard error using antithetic and control variates are

$$\hat{c} = c_{\text{asy}}(1 + (2\bar{t})^{-1}(\hat{C} + \tilde{C})\} \tag{35}$$

$$\hat{\mu} = \mu_{\text{asy}}\left\{\frac{1 + (2\bar{t}^*)^{-1}(\hat{C}^* + \tilde{C}^*)}{1 + (2\bar{t})^{-1}(\hat{C} + \tilde{C})}\right\} \tag{36}$$

and

$$s = \mu_{\text{asy}}[\hat{\text{var}}\{(2t^*)^{-1}(\hat{C}^* + \tilde{C}^*) - (2t)^{-1}(\hat{C} + \tilde{C})\}]^{1/2} \tag{37}$$

respectively, where $\tilde{C}$ and $\tilde{C}^*$ are (28) and (29) with θ and R replaced by $\tilde{\theta}$ and $\tilde{R}$ respectively. Finally, it follows from (34) that an approximation to the distribution function in the case $d = 1$ is

$$\hat{F}(r) = \{1 + (2\bar{t})^{-1}(\hat{C} + \tilde{C})\}^{-1}\{F_{\text{asy}}(r) + (2\bar{t})^{-1}(\hat{C}(r) + \tilde{C}(r))\}\,,$$

where $\tilde{C}(r)$ is (33) with θ and R replaced by $\tilde{\theta}$ and $\tilde{R}$ respectively.

EXAMPLE 1. *Genetic Linkage (continued)*
The three independent runs of the simulation algorithm from §2.2 with $m = 100$ and

m	c^{-1}	μ	s	$3s/\sigma$
100	$2.3825e-05$	0.8334	0.00476	0.1532
100	$2.4006e-05$	0.8316	0.00405	0.1302
100	$2.4226e-05$	0.8292	0.00445	0.1433
10000	$2.4044e-05$	0.8312	0.00044	0.0142
Exact	$2.4056e-05$	0.8311		

Table 5: Approximate posterior computations for the genetic linkage model using control variates

the one with $m = 10000$ produce the results in Table 5. In comparison with Table 1, the reduction in sampling variability achieved by (30) is clearly apparent and we obtain excellent estimates even with m as low as 100.

Figure 5 shows the posterior distribution estimates. The asymptotic approximation (22) (short dash line) is plotted along with the true posterior distribution (solid line) calculated via numerical integration and the estimated distribution (34) (long dash line) when $m = 100$. The dotted line is the straight importance sampling estimator, also when $m = 100$. The estimated distribution (34) is indistinguishable from the exact distribution here.

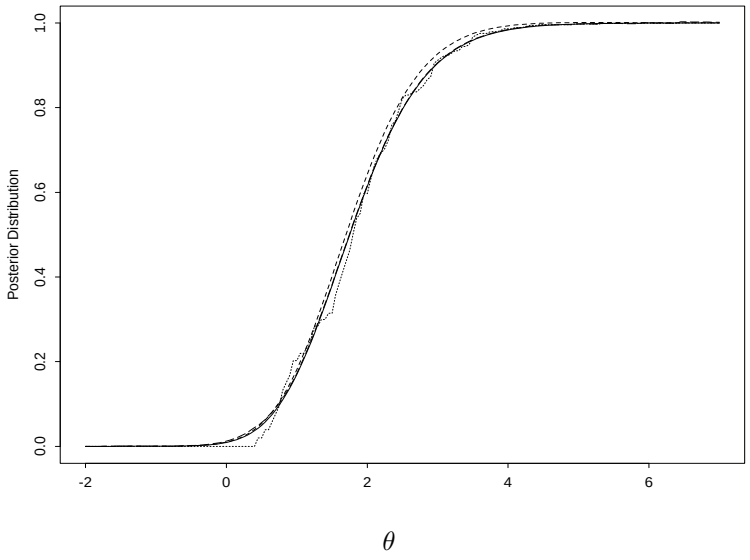

Figure 5: Genetic linkage data: posterior distribution function of θ. Control variates (long dash) and importance sampling (dotted), both with $m = 100$; asymptotic formula (short dash); exact (solid).

Finally, we apply the method of antithetic variates to this example. Table 6 lists the posterior results based on three independent runs of the simulation algorithm, each with $m = 50$ pairs $(R, \tilde{R})$. The improvements using antithetic and control variates can be seen from a comparison of the entries of Tables 1, 2, 5 and 6.

m	c^{-1}	μ	s	$3s/\sigma$
50	$2.4060e-05$	0.8318	0.00372	0.1196
50	$2.4057e-05$	0.8334	0.00246	0.0793
50	$2.4048e-05$	0.8306	0.00303	0.0973

Table 6: Approximate posterior computations for the genetic linkage model using control and antithetic variates

EXAMPLE 2. *Censored regression (continued)*
The three independent samples ($m = 100$) obtained in §2.2 yield the results in Table 7. The entries in column 3 are the posterior expectation estimates of $v(\theta) = \beta_0 + 2\beta_1 + \sigma$ obtained by applying (31). These results, which should be compared with those in Table 3, show a marked reduction in sampling variability, which can also be seen from the standard and relative error estimates.

We now apply antithetic variates along with control variates. Clearly, versions of formulae (35), (36) and (37) hold with the definitions for the multiparameter case. Based on the three independent samples ($m = 50$ pairs $(R, \tilde{R})$) obtained in §2.2, Table 8 lists the results of interest and shows a further useful decrease in sampling variability.

m	c^{-1}	μ	s	$3s/\sigma$
100	$6.7241e-10$	2.9024	0.00335	0.0934
100	$6.7114e-10$	2.9079	0.00425	0.1185
100	$6.6938e-10$	2.9047	0.00420	0.1170
Exact	$6.7187e-10$	2.9048	0.0007	0.0203

Table 7: Approximate posterior computations for the motorette data using control variates

m	c^{-1}	μ	s	$3s/\sigma$
50	$6.7306e-10$	2.9042	0.0023	0.0635
50	$6.6896e-10$	2.9043	0.0029	0.0751
50	$6.7555e-10$	2.9027	0.0011	0.0311

Table 8: Approximate posterior computations for the motorette data using control and antithetic variates

3.3 Marginal densities

In this section we consider estimation of the marginal density $p(\theta_i)$ of the first i components of θ using control variates based on formula (21). As usual, we start by splitting the integral k_i in (18) into two parts,

$$k_i = \int u_i(r)\phi(r^{(i+1)})\, dr^{(i+1)} + C_i\,,$$

where

$$C_i = \int \left\{ \prod_{k=i+1}^{d} q^k(r_k) - u_i(r) \right\} \phi(r^{(i+1)})\, dr^{(i+1)} \tag{38}$$

and

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

with a^i and b^i as defined in Appendix E. Then the first integral is equal to $\bar{t}_i$ defined in Appendix E, and it follows from (18) and (47) that

$$p(\theta_i|y) = c^{-1}(2\pi)^{(d-i)/2} L(\theta_i)\nu_i(\theta_i)(\bar{t}_i + C_i)\,. \tag{39}$$

To estimate the first integral in (38), we use the device of §2.3, which gives

$$\hat{C}_i = \frac{1}{L(\theta_i)\nu_i(\theta_i)} \frac{1}{m} \sum_{j=1}^{m} \frac{L(\theta_{[j]i}) L(\theta_i, \bar{\theta}_{[j]}^{(i+1)}) \lambda(\theta_i, \bar{\theta}_{[j]}^{(i+1)})}{L(\theta_{[j]})} \prod_{u=i+1}^{d} \frac{-R_{[j]}^u}{l_u(\theta_{[j]u})} - \frac{1}{m} \sum_{j=1}^{m} u_i(R_{[j]})\,,$$

where $\bar{\theta}_{[j]}^{(i+1)}$ is given by (13). From (39) and (30) we finally obtain

$$\hat{p}(\theta_i|y) = p_{\text{asy}}(\theta_i|y)\{1 + (\bar{t}_i)^{-1}\hat{C}_i\}\{1 + \bar{t}^{-1}\hat{C}\}^{-1} \tag{40}$$

as a simulation-consistent estimator of $p(\theta_i|y)$. Again, formula (39) exhibits the marginal posterior density of θ_i as the asymptotic approximation (47) multiplied by a correction term. As is the case with the marginal density formula in §2.3, by setting θ_i to be the observed sample values obtained via the simulation algorithm, no additional computational effort is needed to obtain $\hat{C}_i$ and the implementation of formula (40) is quite straightforward.

EXAMPLE 2. *Censored regression (continued)*
We first examine the β_0 marginal. Setting β_0 to be each of the observed sample values ($m = 100$) obtained in §2.2, along with some additional values in the tails, an estimate of $p(\beta_0|y)$ is obtained by applying (40) with $i = 1$. The resulting marginal estimates, represented by the dashed lines in Figure 6, are plotted along with the estimated marginal (solid line) calculated via data augmentation. These graphs, which should be compared with those of Figure 3, show a decrease in sampling variability. An additional small decrease in sampling variability can also be achieved by applying antithetic variates.

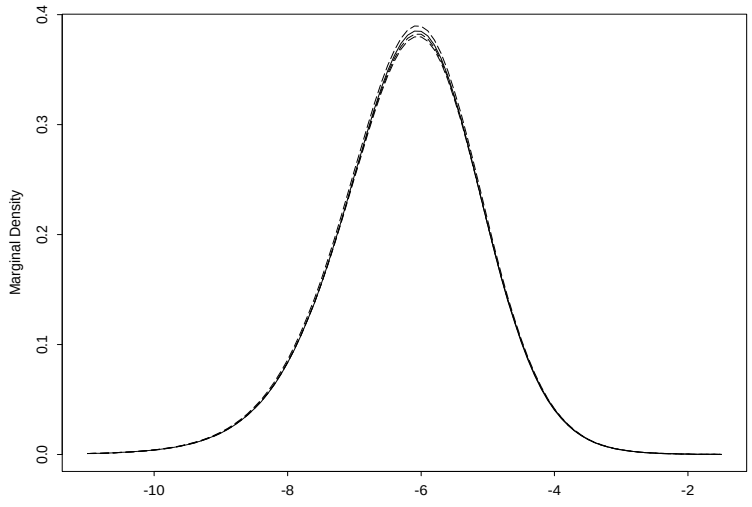

Figure 6: Censored regression data: marginal posterior density of β_0. Control variates with $m = 100$ (dashed lines); data augmentation (solid line).

As a further illustration, we include a power transformation in the model, indexed by the unknown parameter ξ, and obtain the marginal posterior density of ξ. Let y be a failure time and consider the family of Box-Cox transformations

$$y(\xi) = \begin{cases} (y^\xi - 1)/\xi & \xi \neq 0, \\ \log y & \xi = 0 \end{cases} .$$

Suppose that, given ξ,

$$y_i(\xi) = \alpha + \beta(v_i - \bar{v}) + \sigma\epsilon_i \,,$$

where $v_i, \bar{v}$ and ϵ_i are as before. It is convenient to rewrite this model as

$$y_i(\xi) = \nu(\xi) + \nu^\xi \phi(v_i - \bar{v}) + (\nu^\xi \gamma)\epsilon_i \,,$$

where

$$\nu = \begin{cases} (1 + \xi\alpha)^{1/\xi} & \xi \neq 0, \\ e^\alpha & \xi = 0 \end{cases} ,$$

$\phi = \nu^{-\xi}\beta$ and $\gamma = \nu^{-\xi}\sigma$ (Sweeting, 1984). The resulting parameterization also agrees with the Cox and Reid (1987) approximate orthogonal parameterization for this problem. Finally define $\eta = \log \nu$, $\psi = \log \gamma$.

Write $\tilde{y}_i = \nu^{-1} y_i$, $\tilde{c}_i = \nu^{-1} c_i$, $g = (\prod_{i=1}^m y_i)^{1/m}$ and $m_i = \phi(v_i - \bar{v})$. Reordering the

original data as before, the log-likelihood function of $\theta = (\xi, \eta, \phi, \psi)$ is

$$l(\theta) = l\xi(\log g - \eta) - l\psi - \frac{1}{2}\sum_{i=1}^{l}\left\{\frac{\tilde{y}(\xi) - m_i}{e^{\psi}}\right\}^2 + \sum_{i=l+1}^{n}\log\left\{1 - \Phi\left(\frac{\tilde{c}(\xi) - m_i}{e^{\psi}}\right)\right\},$$

from which $\hat{\theta}$ is found to be $(-0.2519, 7.9017, 11.3760, -0.3411)$. We will use a relatively diffuse prior for θ. Taking a uniform prior for θ would be equivalent to the local approximation recommended in Sweeting (1984), using asymptotic approximations. However, a uniform prior cannot be used for exact inference here since the posterior distribution would be improper. This occurs because the likelihood function does not tend to zero as $\xi\eta \to -\infty$. To avoid this impropriety we use the prior proportional to $\exp(\xi\eta)$.

The signed root importance sampler was run three times with $m = 100$ and the resulting estimated ξ marginals (dashed lines) obtained by applying (40) are shown in Figure 7 along with the exact marginal (solid line), which was obtained in this case by taking a very large sample ($m = 20,000$). Again, the method is seen to behave well and gives excellent marginal approximations with very few Monte Carlo draws.

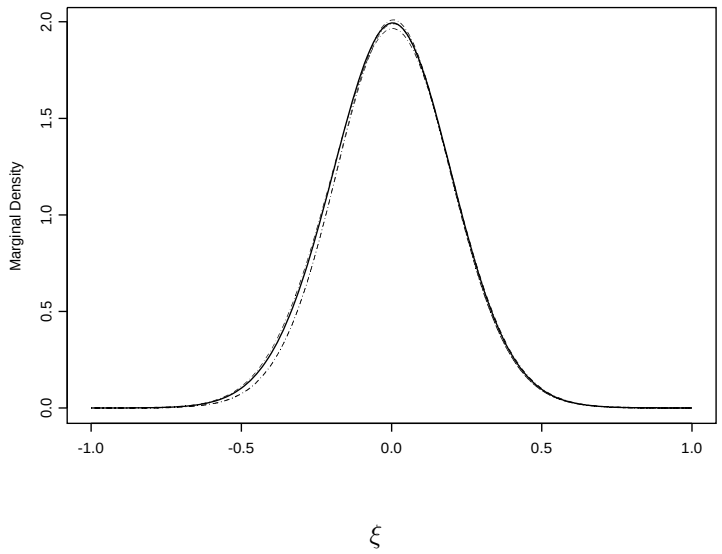

Figure 7: Censored regression data: marginal posterior density of ξ. Control variates with $m = 100$ (dashed lines) and $m = 20,000$ (solid line).

4 Discussion

In this paper we have explored a hybrid method for Bayesian computation involving importance sampling and asymptotic analysis and have provided some numerical il-

lustrations. The proposed signed root based importance sampling algorithm has been shown to be useful for Bayesian computation in moderately parameterized models in which there is a unique maximum likelihood estimate. The results of §3 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.

In all the examples we have studied importance sampling based on signed roots has worked well due to the good tail behaviour of the importance function g. However, it is possible to generalize Lemmas 1 and 2 by taking the variance of the normal importance function to be greater than one. This provides a modified importance sampler that will apply to a wider class of problems.

The main computational burden associated with the importance sampling scheme described in this paper is the solution of the equation $r(\theta) = r$ at each sampled value r of R. As already discussed, alternative methods of solution to those proposed in Appendix B can be devised. Local linear approximations to conditional maxima can be used during an iterative cycle, as discussed in Sweeting (1996). A judicious reparameterization of the form $\phi^i = \phi^i(\theta_i)$, $i = 1, \ldots d$, for which the parameter space is $\mathcal{R}^d$, would help in the Newton scheme described here. Note, however, that the partial invariance property referred to in §2.1 implies that the performance of the importance sampler does not depend on which parameterization is used. Another possibility would be to employ exponential tilting in order to avoid conditional maximization altogether. In fact, even in cases where the likelihood function does not possess a local maximum, a suitable exponential tilt may be used to create one (Sweeting and Kharroubi, 2005). As discussed in Appendix B, an alternative approach to determining initial values would be to carry out an initial charting of the inverse of the function $r(\theta)$ at suitable values of $(r^1, \ldots, r^d)$. Although the inversion of $r(\theta)$ requires some work, the effort should be worthwhile as we anticipate good approximations from a relatively small number of sampled values, especially when using control variates, as demonstrated in the paper. Furthermore, of course, the algorithm produces an independent weighted sample from the posterior distribution, dispensing with the need for convergence diagnostics as required for Markov dependent simulations.

Appendix A: Tail behaviour

The following lemma indicates that, under weak conditions, the tail behaviour of g makes it suitable as an importance function. For the sake of simplicity we first state the result for the scalar parameter case. Here var_g denotes variance under the density $g(\theta)$.

LEMMA 1. *Let v be any continuous function of θ with $v(\hat{\theta}) \neq 0$. Let $\epsilon > 0$ and define the interval $B \equiv B(\epsilon) = \{\theta : |\theta - \hat{\theta}| \leq \epsilon\}$. Suppose that there exist positive constants $c_1 \equiv c_1(\epsilon)$ and $c_2 \equiv c_2(\epsilon)$ with $c_1 < 1/2$ such that*

$$\{v(\theta)\lambda(\theta)\}^{-1} \left| \frac{d}{d\theta}\{L(\theta)\}^{-c_1} \right| > c_2 \tag{41}$$

in the complement B^c of B. Then $\mathrm{var}_g\{v(\theta)h(\theta)\} < \infty$.

Setting $v(\theta) = 1$ gives a condition for the importance sampling estimator of c to converge at rate $m^{-1/2}$. Condition (41) is a fairly weak condition. For example, if $L(\theta) \propto \theta^{-s}$ for $\theta > K$ and $\lambda(\theta) = v(\theta) = 1$, then (41) holds for $\theta > K$ when $s > 2$. We note that the criterion (41) is invariant, so it could be checked in any parameterization.

The corresponding multiparameter result is as follows.

LEMMA 2. *Let v be any continuous function of θ with $v(\hat{\theta}) \neq 0$. Let $\epsilon > 0$ and define the region $B \equiv B(\epsilon) = \{\theta : \|\theta^i - \hat{\theta}^i(\theta_{i-1})\| \leq \epsilon,\ i = 1, \ldots, d\}$. Suppose that there exist positive constants $c_1 \equiv c_1(\epsilon)$ and $c_2 \equiv c_2(\epsilon)$ with $c_1 < 1/2$ such that*

$$\frac{\prod_{i=1}^{d} |l_i(\theta_i)|}{v(\theta)\lambda(\theta)\{L(\theta)\}^{c_1}} > c_2 \tag{42}$$

in in the complement B^c of B. Then $\mathrm{var}_g\{v(\theta)h(\theta)\} < \infty$.

Proof of Lemmas 1 and 2.

We prove the more general result in Lemma 2. Since $r^i(\theta_i)/l_i(\theta_i) \to \{-k^i(\theta_{i-1})\}^{-1/2}$ as $\theta^i \to \hat{\theta}^i(\theta_{i-1})$, $v(\hat{\theta}) \neq 0$ and v is continuous, there exists $\epsilon > 0$ such that $v(\theta)h(\theta)$ is bounded away from zero in the region $0 < \|\theta^i - \hat{\theta}^i(\theta_{i-1})\| \leq \epsilon, i = 1, \ldots, d$. It follows that $\int_B \{v(\theta)h(\theta)\}^2 g(\theta)d\theta$ is finite.

Condition (42) implies that, within B^c,

$$\begin{aligned} \{v(\theta)h(\theta)\}^2 &= \{v(\theta)\lambda(\theta)\}^2 \left\{ \prod_{i=1}^{d} \frac{\{r^i(\theta_i)\}^2}{l_i(\theta_i)} \right\}^2 \\ &< c_2^{-2} \left[\prod_{i=1}^{d} \{r^i(\theta_i)\}^2 \right] \{L(\theta)\}^{-2c_1} = c_3 \left[\prod_{i=1}^{d} \{r^i(\theta_i)\}^2 \right] e^{c_1 \|r(\theta)\|^2} \end{aligned}$$

since $L(\theta) = L(\hat{\theta})e^{-\|r\|^2/2}$. Therefore

$$\int_{B^c} \{v(\theta)h(\theta)\}^2 g(\theta)d\theta < c_4 \int \left\{ \prod_{i=1}^{d} (r^i)^2 \right\} e^{-(\frac{1}{2} - c_1)\|r\|^2} dr\,,$$

which is finite since $c_1 < 1/2$ and so $E_g\{v(\theta)h(\theta)\}^2 < \infty$ as required.

Appendix B: Inversion of the signed root

We begin by discussing the scalar parameter case, for which we use the update formula from Sweeting (1996) given by $z' = z - [r\{r(\theta) - r\}]/z$, where $z \equiv z(\theta) = J^{1/2}(\theta - \hat{\theta})$. Define $z^+ = z(\theta^+)$ and $z^- = z(\theta^-)$, where θ^+, θ^- are the solutions of the equations $r(\theta^+) = 1$ and $r(\theta^-) = -1$ respectively; see also §3.1. Then the iteration may be initialized at $z = r + ar^2 + br^3$ with $a = (z^+ + z^-)/2$ and $b = (z^+ - z^-)/2 - 1$. A little algebra shows that this choice of constants matches the initial values of z with the exact values when $r = -1, 0$ and 1. The initialization suggested here is a sensible one because in the i.i.d. case the asymptotic error in this cubic approximation to z is $O(n^{-3/2})$, where n is sample size (Sweeting, 1996).

In the multiparameter case, the equations $r^i(\theta_i) = r^i$ for each sampled value r of R may be solved sequentially using the update formulae $z'^i = z^i - [r^i\{r^i(\theta_i) - r^i\}]/z^i$, where $z^i = z^i(\theta_i) = \left\{k^i(\theta_{i-1})\right\}^{1/2}\{\theta^i - \hat{\theta}^i(\theta_{i-1})\}$; see Sweeting (1996). In a similar way to the scalar parameter case, this process can be initialized at $z^i = r^i + a^i(r^i)^2 + b^i(r^i)^3$ for a suitable choice of a^i and b^i. Define $z^{i+} = z^i(\hat{\theta}_{i-1}, \theta^{i+})$ and $z^{i-} = z^i(\hat{\theta}_{i-1}, \theta^{i-})$, where θ^{i+} and θ^{i-} are the solutions to the equations $r^i(\hat{\theta}_{i-1}, \theta^{i+}) = \sqrt{d}$ and $r^i(\hat{\theta}_{i-1}, \theta^{i-}) = -\sqrt{d}$ respectively. We note that any constant could be used in place of $\sqrt{d}$ here; this choice is for consistency with §3.1. Now define $a^i = (z^{i+} + z^{i-})/(2d)$ and $b^i = \{(z^{i+} - z^{i-})/(2\sqrt{d}) - 1\}/d$. It may be verified that the initial values of z^i are the exact values when (i) $r^i = 0$ and (ii) $r_{i-1} = 0$ and $r^i = \sqrt{d}$ or $r^i = -\sqrt{d}$.

As previously noted, it is important to have an efficient method of solving $r(\theta) = r$ in the multiparameter case, especially since conditional maximization is involved at each step of the iteration. Some further possibilities are discussed here. An alternative approach would be to create either an initial 'cloud' or a rectangular grid of θ values. In the former case the points could be generated from a normal distribution for θ with covariance matrix cJ^{-1}, $c > 1$, and the corresponding values of $r(\theta)$ calculated. The initial value of θ in the equation $r(\theta) = r$ would then be the nearest neighbour within the cloud of $r-$values and a non-sequential Newton-Raphson scheme employed. In the latter case, again the value of $r(\theta)$ would be calculated for each value in the grid, the initial value for each θ^i, $1 \leq i \leq d$, would be the closest value in the grid, and the equation solved in a sequential manner.

In any approach to the solution of $r(\theta) = r$ it is a good idea to use a two-stage procedure, where at the first stage the conditional maximizations are replaced by first-order linear approximations, with full conditional maximization only used at the second pass. This idea was discussed in the author's reply to the discussion in Sweeting (1996) and has been successfully implemented. Additional stability may be achieved by an initial reparameterization of the form $\phi = J^{1/2}(\theta - \hat{\theta})$, where $J^{1/2}$ is any square root of J.

Appendix C: Proof of equation (12)

In order to ease the notation we write $\alpha = \theta_i$, $\beta = \theta^{(i+1)}$ and drop the conditioning on y. Consider the one-to-one transformation (α, γ) of (α, β), where $\gamma = \beta - \hat{\beta}(\alpha)$. Then the densities of (α, γ) corresponding to the densities p and g of (α, β) are $\bar{p}(\alpha, \gamma) = p(\alpha, \gamma + \hat{\beta}(\alpha))$ and $\bar{g}(\alpha, \gamma) = g(\alpha, \gamma + \hat{\beta}(\alpha))$ respectively. Therefore

$$\begin{aligned} p(\alpha) &= \int \bar{p}(\alpha, \gamma)\, d\gamma = \int\int \bar{p}(\alpha, \gamma)\bar{g}(\alpha')\, d\alpha' d\gamma \\ &= \int \frac{\bar{p}(\alpha, \gamma)}{\bar{g}(\gamma|\alpha')} \bar{g}(\alpha', \gamma)\, d\alpha' d\gamma \\ &= \int \frac{p(\alpha, \beta' + \hat{\beta}(\alpha) - \hat{\beta}(\alpha'))}{g(\beta'|\alpha')} g(\alpha', \beta')\, d\alpha' d\beta' \end{aligned}$$

on applying the change of variables $\gamma = \beta' - \hat{\beta}(\alpha')$. A consistent estimator of $p(\alpha)$ is therefore

$$\hat{p}(\alpha) = \frac{1}{m} \sum_{j=1}^{m} \frac{p(\alpha, \beta^*_{[j]})}{g(\beta_{[j]}|\alpha_{[j]})}\,, \tag{43}$$

where $\beta^*_{[j]} = \beta_{[j]} + \hat{\beta}(\alpha) - \hat{\beta}(\alpha_{[j]})$. But, transforming from the normal distribution of R_i, we find that the density of the first i components of θ is

$$g(\theta_i) = (2\pi)^{-i/2} \frac{L(\theta_i)}{L(\hat{\theta})} \prod_{k=1}^{i} \frac{-l_k(\theta_k)}{r^k(\theta_k)}$$

so that, from (4), the conditional density of $\theta^{(i+1)}$ given θ_i is

$$g(\theta^{(i+1)}|\theta_i) = (2\pi)^{-(d-i)/2} \frac{L(\theta)}{L(\theta_i)} \prod_{k=i+1}^{d} \frac{-l_k(\theta_k)}{r^k(\theta_k)}\,.$$

Substituting this expression and (1) into (43) gives formula (12).

Appendix D: Proof of Theorem 1

From (14) and the usual multivariate transformation formula, the exact posterior density of θ can be expressed as

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

from which formula (16) for the constant of proportionality c follows, since $c = \int L(\theta)\lambda(\theta)\, d\theta$.

Now let $v(\theta)$ be a smooth function of θ. By applying formula (16) with $\lambda(\theta)$ set to $v(\theta)\lambda(\theta)$ we obtain

$$\int v(\theta) L(\theta)\lambda(\theta)\, d\theta = (2\pi)^{d/2} |J|^{-1/2} v(\hat{\theta}) L(\hat{\theta})\lambda(\hat{\theta}) k^*\,.$$

Formula (17) for the posterior expectation of $v(\theta)$ now follows on dividing by the formula (16) for c.

Finally, from (14) we see that the exact marginal posterior density of the first i components of $r(\theta)$ is

$$f(r_i|y) = k^{-1}(2\pi)^{-i/2}\left\{\prod_{j=1}^{i}\frac{-r^j}{l_j(\theta_j)}\right\}\left\{\frac{\nu_i(\theta_i)L(\theta_i)}{\lambda(\hat{\theta})L(\hat{\theta})}\right\}|J|^{1/2}k_i\,.$$

Formula (18) for the marginal posterior density $p(\theta_i|y)$ of θ_i follows on transforming this density to that of θ_i and using formula (16).

Appendix E: Asymptotic formulae for control variates

We start by obtaining asymptotic approximations $\bar{t}, \bar{t}^*$ and $\bar{t}_i$ respectively to the constants k, k^* and k_i appearing in expressions (16), (17) and (18) in §3.1.

Suppose that t^i is any $O(n^{-2})$ approximation to $1 + db^i(0)$, where $b^i(0)$ is the coefficient of $(r^i)^2$ in equation (15) when $r_{i-1} = 0$, and define $\bar{t} = d^{-1}\sum_i t^i$. Then it is shown in Sweeting and Kharroubi (2003) that $k = \bar{t} + O(n^{-2})$. An $O(n^{-2})$ approximation to c is therefore given by

$$(2\pi)^{d/2}|J|^{-1/2}L(\hat{\theta})\lambda(\hat{\theta})\bar{t}\,. \tag{44}$$

Similarly, suppose that t^{*i} is any $O(n^{-2})$ approximation to $1 + db^{*i}(0)$, where $b^{*i}(0)$ is the coefficient of $(r^i)^2$ in the expansion of $q^{*i}(r_i)$ when $r_{i-1} = 0$, and define $\bar{t}^* = d^{-1}\sum_i t^{*i}$. Then $k^* = \bar{t} + O(n^{-2})$ to $O(n^{-2})$, and an $O(n^{-2})$ approximation to μ is therefore

$$v(\hat{\theta})\frac{\bar{t}^*}{\bar{t}}\,. \tag{45}$$

We now use a result from Sweeting and Kharroubi (2003) to choose suitable values for t^i and t^{i*} in equations (44) and (45). As in Appendix B, θ^{i+} and θ^{i-} are chosen to satisfy the equations $r^i(\hat{\theta}_{i-1}, \theta^{i+}) = \sqrt{d}$ and $r^i(\hat{\theta}_{i-1}, \theta^{i-}) = -\sqrt{d}$ respectively. Further write $\theta_i^+ = (\hat{\theta}_{i-1}, \theta^{i+})$ and $\theta_i^- = (\hat{\theta}_{i-1}, \theta^{i-})$. Define $\tau^i = \{\nu_i(\theta_i^-)/l_i(\theta_i^-)\} + \{-\nu_i(\theta_i^+)/l_i(\theta_i^+)\}$, let e_i be the i-dimensional vector $(0, \cdots, 0, 1)$ and define $J^{(i)} = j^{(i)}(\hat{\theta})$. Then it is shown in Sweeting and Kharroubi (2003) that

$$t^i \equiv \frac{1}{2}\{q^i(\sqrt{d}e_i) + q^i(-\sqrt{d}e_i)\} = \frac{1}{2}d^{1/2}|J^{(i)}|^{1/2}\{\lambda(\hat{\theta})\}^{-1}\tau^i \tag{46}$$

is an $O(n^{-2})$ approximation to $1 + db^i(0)$. Substitution of (46) into (44) now gives

$$c_{\text{asy}} = \frac{1}{2}d^{-1/2}(2\pi)^{d/2}|J|^{-1/2}L(\hat{\theta})\sum_{j=1}^{d}|J^{(j)}|^{1/2}\tau^j$$

as an $O(n^{-2})$ approximation to c.

Similarly, t^{*i}, the expression corresponding to (46) with $\lambda(\theta)$ set to $v(\theta)\lambda(\theta)$, is an $O(n^{-2})$ approximation to $1 + db^{i*}(0)$. Now define $\alpha_i^- = (\tau^i)^{-1}\{\nu_i(\theta_i^-)/l_i(\theta_i^-)\}$, $\alpha_i^+ = (\tau^i)^{-1}\{-\nu_i(\theta_i^+)/l_i(\theta_i^+)\}$ and $\gamma^i = |J^{(i)}|^{1/2}\tau^i / \sum_{k=1}^d |J^{(k)}|^{1/2}\tau^k$. Then, after some algebra, it follows from (45) that

$$\mu_{\text{asy}} = \sum_{i=1}^{d} \gamma^i \{\alpha_i^- v(\theta_i^-) + \alpha_i^+ v(\theta_i^+)\}$$

is an $O(n^{-2})$ approximation to μ. This formula, which is a variant of a formula given by Sweeting (1996), was derived by Sweeting and Kharroubi (2003).

Now define $\bar{t}_i = (i + \sum_{k=i+1}^d t^k)/d$, where t^i is given by equation (46). Then an $O(n^{-3/2})$ form of Laplace approximation of $p(\theta_i|y)$ that is amenable to the application of control variates is given by

$$p_{\text{asy}}(\theta_i|y) = c_{\text{asy}}^{-1}(2\pi)^{(d-i)/2} L(\theta_i)\nu_i(\theta_i)\bar{t}_i \,. \tag{47}$$

Next the constants a and b for equation (22) may be taken as $a = \frac{1}{2}\{q(1) - q(-1)\} = (\alpha^+ - \alpha^-)t$, $b = t - 1$, where t is given by equation (46) in the case $d = 1$.

Finally, we consider the choice of coefficients in expressions (24) and (26). In expression (24) we choose $a^i = \frac{1}{2}d^{-1/2}\{q^i(d^{1/2}e_i) - q^i(-d^{1/2}e_i)\} = d^{-1/2}(\alpha_i^+ - \alpha_i^-)t^i$ and $b^i = d^{-1}(t^i - 1)$, where t^i is given by (46). Note that in the case of n i.i.d. observations, from (15) $u(r)$ is an $O(n^{-3/2})$ approximation to $q(r)$. Similarly, in expression (26) we choose $a^{*i} = \frac{1}{2}d^{-1/2}\{q^{*i}(d^{1/2}e_i) - q^{*i}(-d^{1/2}e_i)\} = \{dv(\hat{\theta})\}^{-1}\{\alpha_i^+ v(\theta_i^+) - \alpha_i^- v(\theta_i^-)\}t^i$ and $b^{*i} = d^{-1}(t^{*i} - 1)$, where $t^* = \{v(\hat{\theta})\}^{-1}\mu_{\text{asy}}t$.

References

Cox, D. R. and Reid, N. (1987). Parameter orthogonality and approximate conditional inference (with discussion). *J. R. Statist. Soc.*, B, **49**, 1–39.

Crawford, D. E. (1970). Analysis of incomplete life test data on motorettes. *Insulation Circuits*, **16**, 43–48.

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

DiCiccio, T. J. and Field, C. A. (1991). An accurate method for approximate conditional and Bayesian inference about linear regression models from censored data. *Biometrika*, **78**, 903–910.

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

Evans, M. and Swartz, T. (1995a). Methods for approximating integrals in statistics with special emphasis on Bayesian integration problems. *Stat. Sci.*, **10**, 254–272.

Evans, M. and Swartz, T. (1995b). Bayesian integration using multivariate Student importance sampling. In *Computing Science and Statistics*, **27** (ed. M. M. Meyer and J. L. Rosenberger) 456–461. Interface Foundation of North America, Inc., Fairfax Station, VA.

Evans, M. and Swartz, T. (2000). *Approximating Integrals via Monte Carlo and Deterministic Methods*. Oxford: Oxford University Press.

Glynn, P. W. and Iglehart, D. L. (1989). Importance sampling for stochastic simulations. *Management Science*, **35**, 1367–1392.

Hammersley, J. M. and Handscomb, D. C. (1964). *Monte Carlo Methods*. Methuen, London.

Hesterberg, T. (1990). Importance sampling for Bayesian estimation. *Computing and Graphics in Statistics* (A. Buja and P. A. Tukey, eds.) 63–75. Springer, New York.

Rao, C. R. (1973). *Linear Statistical Inference and Applications*. New York: Wiley.

Ripley, B. (1987). *Stochastic Simulation*. New York: Wiley.

Schmee, J. and Hahn, G. J. (1979). A simple method for regression analysis with censored data. *Technometrics*, **21**, 417– 432.

Sweeting, T. J. (1984). On the choice of prior distribution for the Box-Cox transformed linear model. *Biometrika*, **71**, 127–134.

Sweeting, T. J. (1992). Discussion of paper by D. A. Pierce and D. Peters. *J. R. Statist. Soc.*, B, **54**, 732–733.

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

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

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

Sweeting, T. J. and Kharroubi, S. A. (2005). Application of a predictive distribution formula to Bayesian computation for incomplete data models. *Statistics and Computing*, **15**, 167–178.

Tanner, M.A. and Wong, W.H. (1987). The calculation of posterior distributions by data augmentation. *J. Amer. Statist. Assoc.*, **82**, 528–540.

Van Dijk, H. K., Kloek, T. and Louter, A. S. (1986). An algorithm for the computation of posterior moments and densities using simple importance sampling. *The Statistician*, **35**, 83– 90.

Wei, G. C. G. and Tanner, M. A. (1990). A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithm. *J. Amer. Statist. Assoc.*, **85**, 699–704.

Wolpert, R. L. (1991). Monte Carlo integration in Bayesian statistical analysis. *Statistical Multiple Integration* (N. Flournoy and R. K. Tsutakawa, eds.) 101–116. Amer. Math. Soc., Providence, RI.

Acknowledgments

We are grateful to the Associate Editor and a referee for their many helpful comments that led to a much improved version of this paper.