A Novel Formulation for Approximate Bayesian Computation Based on Signed Roots of Log-Density Ratios

Samer A Kharroubi and Alan Brennan

Centre for Bayesian Statistics in Health Economics, Department of Probability and Statistics, University of Sheffield, Sheffield, S3 7RH, UK

February 16, 2005

Summary

In this paper we develop some novel theory for approximate Bayesian computation based on the asymptotic theory of signed root log-density ratios. Some new accurate approximations for Bartlett corrections, posterior expectations and predictive densities are derived. These approximations are modifications of formulae based on signed root log-likelihood ratios obtained in Sweeting and Kharroubi (2003) but are designed to absorb the prior into the likelihood function. This is an attractive property for various practical implications of interest, including the calculation of the expected value of sample information from decision theory. A case study decision model from the field of health economics is used to show that the expected value of sample information approximated results are similar to the standard nested Monte Carlo sampling method, but are achieved with up to 100-fold computation time reductions. Further applied research is recommended for more complex decision models. This new approximation formulation has wider potential benefits in many fields of Bayesian approximation.

Keywords: Signed root log-density ratio, approximate Bayesian inference, Bayesian Bartlett correction, higher-order asymptotics, Laplace approximation, predictive distribution, expected value of sample information.
1 Introduction

In this paper we develop some novel theory for approximate Bayesian computation based on the asymptotic theory of signed root log-density ratios. This theory provides fourth-order correct formulae for various posterior quantities of interest, including Bayesian Bartlett corrections, posterior expectations and predictive densities.

The existing theory on the use of the signed root log-density ratios goes back to Lawley (1956) and most of the related work has been done for the frequentist inference. Since then, they have been pursued by many authors, in both frequentist and Bayesian inference. Efron (1985) was among the first authors to investigate Bayesian computation based on the signed root log-density ratio and proved the asymptotic normality of the posterior distribution of the signed root log-density ratio. DiCiccio and Stern (1993) derive approximate Bayesian Bartlett corrections. Sweeting (1996) provides third- and fourth-order correct formulae for univariate posterior distribution functions, Bayesian Bartlett corrections and highest likelihood/posterior density regions, posterior expectations and predictive densities. Much of this work was based on the signed root log-likelihood ratio. Most recently, Sweeting and Kharroubi (2003) present some new accurate approximations for posterior expectations and Bartlett corrections. These approximations are modifications of formulae based on signed root log-likelihood ratios obtained in Sweeting (1996), designed to address practical application problems in the multiparameter case (i.e. computation associated with inversion of signed root log-likelihood ratios, and secondly, the inconvenient form of the posterior expectation formula when computing predictive densities).

In this paper we follow on from the work of Sweeting and Kharroubi (2003) to develop some alternative formulae for Bayesian Bartlett corrections, posterior expectations and predictive densities based on the asymptotic theory of signed root log-density ratios. Our new formulation changes the earlier approach of Sweeting and Kharroubi (2003) in one fundamental way. It is designed to absorb the prior into the likelihood function, constructing formulae that focus on the posterior which will be an attractive property for various practical implications of interest, including assessing the sensitivity of posterior expectations to the prior distribution and the influence of one or more outliers. Just such an important context - the calculation of expected value of sample information (EVSI) in decision theory - led us to the development of these new formulae.

In general, calculation of EVSI in decision analytic models demands computationally intensive nested Monte-Carlo simulations, sometimes with a Markov-Chain Monte
Carlo (MCMC) process embedded in each simulation. This often precludes successful
calculation of EVSI in a reasonable timeframe. In such a context, application of the the-
ory developed here can potentially provide substantial efficiency and enable previously
infeasible computation.

The organisation of the paper is as follows. We begin in Section 2 by reviewing some
of the methods and results in Sweeting and Kharroubi (2003). This is necessary in order
to ease the development in the subsequent sections and to facilitate comparison with
the new results in the paper. In Section 3 we develop alternative approximations to
posterior Bartlett corrections, expectations and predictive densities that are based on
the asymptotic theory of signed root log-density ratios. Section 4 sets out the theory
of expected value of sample information, the existing approaches to its computation,
and shows how the new approximations, particularly for the posterior expectation of a
multiparameter function, result in amended equations for EVSI. In Section 5 we apply
the new method to a simple, previously used in Brennan et al (2002a, b), decision analytic
model from the field of health economics. We compare EVSI results using existing
Monte-Carlo methods with those using the new formulae and report on the efficiency
gain in this one case study. To this end, some concluding remarks are given in Section
6, including a few more directions for future research.

2 Approximations based on signed root log-likelihood ratio

2.1 Preliminaries

We begin by reviewing some basic quantities associated with signed root log-likelihood
ratios in the case of a scalar parameter $\theta$. Suppose that $(\pi_n)$ is any sequence of densities
of the form

$$
\pi_n(\theta) \propto L_n(\theta) \lambda_n(\theta) \{1 + \eta_n(\theta - \hat{\theta}_n)\} ,
$$

(1)

where $(\lambda_n)$ is a suitable sequence of carrier densities, $(\eta_n)$ is an $O(n^{-1})$ sequence inde-
dependent of $\theta$ and $\hat{\theta}_n = \arg\max\{L_n(\theta)\}$. If $\pi_n(\theta)$ is the posterior density of $\theta$, then $\eta_n = 0$
and $L_n$ could be taken to be the likelihood and $\lambda_n$ the prior. We include the sequence
$(\eta_n)$ in (1) since it turns out that, under suitable regularity conditions, the marginal
posterior density of a single parameter in the multiparameter case is of the form (1).

Write $l_n(\theta) = \log L_n(\theta)$ and let $w_n(\theta) = 2\{l_n(\hat{\theta}_n) - l_n(\theta)\}$ be the log-likelihood ratio.
The transformation
\[ r_n(\theta) = \text{sign}(\theta - \hat{\theta}_n) \{ w_n(\theta) \}^{1/2} \]

is then an increasing function of \( \theta \) and is referred to as the \textit{signed root log-likelihood ratio} (SRLLR), or directed likelihood ratio. Since

\[ L_n(\theta) \propto \exp \{ l_n(\hat{\theta}_n) - l_n(\theta) \} = e^{-r^2/2}, \]

where \( r = r_n(\theta) \), the likelihood function \( L_n(\theta) \) is of standard normal form when expressed in terms of \( r \). It is shown in Barndorff-Nielsen and Cox (1989) that, in the case of independent and identically distributed (i.i.d.) observations, the sampling distribution of \( R_n \equiv r_n(\theta) \) is asymptotically standard normal.

Now define \( j_n(\theta) = -l_n''(\theta) \), \( J_n = j_n(\hat{\theta}_n) \), \( z_n(\theta) = J_n^{1/2}(\theta - \hat{\theta}_n) \), \( g_n(r) = dz/dr \) and \( h_n(r) = \lambda_n(\theta)/\lambda_n(\hat{\theta}_n) \), where \( \theta = \theta_n(r) \) is the inverse of \( r_n(\theta) \). Then the posterior density of \( R_n \) satisfies

\[ f_n(r) \propto \phi(r)q_n(r)(1 + \epsilon_n r), \]

where \( \phi(\cdot) \) is the standard normal density, \( g_n(r) = g_n(r)h_n(r) \) and \( \epsilon_n \) is an \( O(n^{-3/2}) \) sequence independent of \( \theta \). It is shown in Sweeting (1995) that, under suitable assumptions on the likelihood function, both \( g_n \) and \( h_n \), and hence \( q_n \), are of the form

\[ 1 + a_n r + b_n r^2 + c_n r^3 \]  

(2)

to \( O(n^{-2}) \), with \( a_n = O(n^{-1/2}) \), \( b_n = O(n^{-1}) \) and \( c_n = O(n^{-3/2}) \). Then, from Sweeting (1995), we obtain

\[ f_n(r) \doteq s_n^{-1} \phi(r)q_n(r)(1 + \epsilon_n r), \]

(3)

where \( s_n = 1 + b_n \) and the symbol \( \doteq \) indicates equality up to \( O(n^{-2}) \). Furthermore \( s_n^{-2} W_n \sim \chi^2_1 \) to \( O(n^{-2}) \), where \( W_n = R_n^2 \) is the loglikelihood ratio; that is, \( s_n^2 \) is a \textit{Bayesian Bartlett correction} to \( W_n \). Approximations to the posterior distribution function of \( R_n \) are discussed in Sweeting (1995).

### 2.2 Bartlett corrections and posterior expectations for a scalar parameter

Define

\[ Q_n(z) = \frac{z}{r} q_n(r). \]

Suppressing \( n \) from now on, it is shown in Sweeting and Kharroubi (2003) that \( z/r \), and hence \( Q(z) \), are of the form

\[ 1 + Az + Bz^2 + Cz^3, \]

(4)
where \( A = O(n^{-1/2}), B = O(n^{-1}), C = O(n^{-3/2}) \) and the coefficient of \( z^2 \) is

\[
B = d_2 + b, \quad (5)
\]

where \( d_2 \) is the coefficient of \( z^2 \) in the expansion of \( z/r \). It follows that \( S \equiv 1 + B \equiv (1 + b)(1 + d_2) \) and so a formula for the Bayesian Bartlett correction is \( s^2 \), where

\[
s = 1 + b \doteq \frac{1 + B}{1 + d_2}. \quad (6)
\]

It is shown in Sweeting and Kharrouri (2003) that \( s \) may be obtained without expansion of \( l(\theta) \). Note that \( Q(z) = G(z)H(z) \), where \( G(z) = -zJ^{1/2}/l'(\theta) \) and \( H(z) = \lambda(\theta)/\lambda(\hat{\theta}) \), where \( \theta = \theta(z) = \hat{\theta} + J^{-1/2}z \) is the inverse of \( z(\theta) \). Define \( \theta^- \) by \( z(\theta^-) = -1 \) and \( \theta^+ \) by \( z(\theta^+) = 1 \); that is, \( \theta^\pm = \hat{\theta} \pm J^{-1/2} \). Since \( B \) is the coefficient of \( z^2 \) in the expansion of \( Q(z) \), the numerator in (6) is

\[
S = 1 + B \doteq \frac{1}{2} \left\{ Q(-1) + Q(1) \right\} = \frac{1}{2} J^{1/2} \{ \lambda(\hat{\theta}) \}^{-1} \tau, \quad (7)
\]

where \( \tau = \{ \lambda(\theta^-)/l'(\theta^-) \} + \{ -\lambda(\theta^+)/l'(\theta^+) \} \). Similarly, since \( d_2 \) is the coefficient of \( z^2 \) in the expansion of \( z/r \), the denominator in (6) takes the form \( 1 + d_2 \doteq \frac{1}{2} \tau \omega \), where \( \omega = \{ -r(\theta^-) \}^{-1} + \{ r(\theta^+) \}^{-1} \). This implies that

\[
s = J^{1/2} \{ \omega \lambda(\hat{\theta}) \}^{-1} \tau. \quad (8)
\]

Approximation (6) can also be used to obtain a formula for the posterior expectation of a general function \( v(\theta) \) by applying it to both the numerator and denominator in

\[
E\{ v(\theta) | X \} = v(\hat{\theta}) \frac{s^*}{s} \quad (9)
\]

(Sweeting, 1996), where \( s^* = 1 + b^* \) and \( b^* \) is the coefficient of \( r^2 \) in the expansion \( q^*(r) = \{ v(\theta)/v(\hat{\theta}) \} q(r) \). The fourth-order approximation (6) is already available for the denominator. A similar approximation can be applied to the numerator as follows. By setting \( \lambda(\theta) \) to be \( \lambda(\theta)v(\theta) \) in \( H(z) \), it follows from (4) and (5) that

\[
Q^*(z) \equiv \frac{v(\theta)}{v(\hat{\theta})} Q(z) = 1 + A^*z + B^*z^2 + C^*z^3,
\]

where \( A^* = O(n^{-1/2}), B^* = O(n^{-1}), C^* = O(n^{-3/2}) \) and the coefficient of \( z^2 \) is \( B^* = d_2 + b^* \). Therefore \( S^* \equiv 1 + B^* \doteq (1 + b^*)(1 + d_2) \) so that a fourth-order approximation to the numerator of (9) is

\[
s^* = 1 + b^* \doteq \frac{1 + B^*}{1 + d_2}. \quad (10)
\]
Substituting (6) and (10) in (9) produces the approximation
\[ E\{v(\theta) | X\} = v(\hat{\theta}) S^* \frac{S}{S}. \tag{11} \]

In a similar way to (7), we can write
\[ S^* = 1 + B^* = 1 + \frac{1}{2} \{Q^*(-1) + Q^*(1)\} = \frac{1}{2} J^{1/2} \{\lambda(\hat{\theta})v(\hat{\theta})\}^{-1} \tau^*, \tag{12} \]
where \( \tau^* = \{\lambda(\theta^-)v(\theta^-)/l'(\theta^-)\} + \{-\lambda(\theta^+)v(\theta^+)/l'(\theta^+)\} \).

Finally, substituting (7) and (12) in (11) we obtain
\[ E\{v(\theta) | X\} = \frac{\tau^*}{\tau} = \alpha^- v(\theta^-) + \alpha^+ v(\theta^+), \tag{13} \]
where \( \alpha^- = \tau^{-1} \{\lambda(\theta^-)/l'(\theta^-)\} \) and \( \alpha^+ = \tau^{-1} \{-\lambda(\theta^+)/l'(\theta^+)\} = 1 - \alpha^- \). We note that \( \alpha^\pm = \frac{1}{2} + O(n^{-1/2}) \).

Formula (13) is closely related to the approximation of Sweeting (1996). Firstly, the values \( \theta^- , \theta^+ \) and the corresponding weights \( \alpha^- , \alpha^+ \) are independent of \( v(\theta) \) and hence need only be computed once when many expectations are required. Secondly, when assessing the sensitivity of a posterior expectation to a change of prior, the weights \( \alpha^- , \alpha^+ \) are all one needs to recompute. However, (13) has the significant advantage that, since \( \theta^\pm = \hat{\theta} \pm J^{-1/2} \), these values are easily computed, whereas it is usually necessary to use Newton’s method to obtain \( \theta^- \) and \( \theta^+ \) as defined in Sweeting (1996). On the other hand, formula (13) is not invariant under reparameterisation and so the problem arises of how to choose an appropriate parameterisation. For further details, see Sweeting and Kharroubi (2003).

Formula (13) can be used to compute a Bayesian predictive density and gives the formula
\[ p(y | X) = \alpha^- p(y | \theta^- , X) + \alpha^+ p(y | \theta^+ , X). \tag{14} \]
Notice that, since it is a nonnegatively weighted average of the two predictive densities \( p(y | \theta^- , X) \) and \( p(y | \theta^+ , X) \), Formula (14) is a valid density function. It also has the advantage that no recomputation of the weights \( \alpha^- , \alpha^+ \) or the arguments \( \theta^- , \theta^+ \) is required as \( y \) is varied.

### 2.3 The multiparameter case

In this section we indicate how the theory in Sections 2.1 and 2.2 generalizes to the multi-parameter case, giving rise to multiparameter formulae for Bayesian Bartlett corrections, posterior expectations and predictive densities.
As in Section 2.1, we study the behaviour of densities of the form
\[
\pi_n(\theta) \propto L_n(\theta) \lambda_n(\theta) \{1 + \eta_n^T(\theta - \hat{\theta}_n)\},
\]
where now \( \theta \in \mathbb{R}^d \). Let \( l_n(\theta) = \log L_n(\theta) \), \( j_n(\theta) = -d^2 l_n/d\theta^2 \), the matrix of second-order partial derivatives of \(-l_n\), and \( J_n = j_n(\hat{\theta}_n) \). Let \( \theta^1, \ldots, \theta^d \) denote the components of \( \theta \) and write \( \theta_i = (\theta^1, \ldots, \theta^i) \), the vector of the first \( i \) components, and \( \theta^{(i)} = (\theta^i, \ldots, \theta^d) \), the vector of the last \( d - i + 1 \) components. In addition to \( \hat{\theta}_n = (\hat{\theta}_n^1, \ldots, \hat{\theta}_n^d) = \text{argmax} \{L_n(\theta)\} \), define \( \hat{\theta}_n^{(i+1)}(\theta_i) \) to be the maximizer of \( L_n \) conditional on \( \theta_i \). For \( j > i \), \( \hat{\theta}_n^{(i)}(\theta_i) \) denotes the \( j \)th component of \( (\theta_i, \hat{\theta}_n^{(i+1)}(\theta_i)) \). For a function \( g(\theta) \), when \( i < d \) we use \( g(\theta_i) \) to denote \( g(\theta_i, \hat{\theta}_n^{(i+1)}(\theta_i)) \).

Now define, for \( i = 1, \ldots, d \), the loglikelihood ratios
\[
w_n^i = w_n^i(\theta_i) = 2\{l_n(\theta_{i-1}) - l_n(\theta_i)\}
\]
and the signed root transformation
\[
r_n^i = r_n^i(\theta_i) = \text{sign}\{\theta^i - \hat{\theta}_n^i(\theta_{i-1})\}\{w_n^i\}^{1/2}.
\]
Note that \( w_n \equiv \sum_i w_n^i = 2\{l_n(\hat{\theta}_n) - l_n(\theta)\} \).

Write \( W_n = w_n(\theta) \) and \( R_n = (r_n^1(\theta_1), \ldots, r_n^d(\theta_d)) \). Suppressing \( n \) from now on, as in Sweeting (1996) the density \( f(r) \) of \( R \) satisfies
\[
f(r) \propto \phi(r) \prod_{i=1}^d q^i(r_i)(1 + c^Tr) ,
\]
where now \( \phi(\cdot) \) is the \( d \)-dimensional standard normal density, \( c \) is an \( O(n^{-3/2}) \) sequence independent of \( \theta \) and
\[
q^i(r_i) = \left\{-r^i/l_i(\theta_i)\right\}\left\{\nu_i(\theta_i)/\nu_{i-1}(\theta_{i-1})\right\},
\]
where \( l_i(\theta) = \partial(\theta)/\partial \theta^i \), \( \nu_i(\theta) = \lambda(\theta) |j^{(i+1)}(\theta)|^{-1/2} \) and \( j^{(i)} \) is the submatrix of \( j \) corresponding to \( \theta^{(i)} \) (setting \( |j^{(d+1)}(\theta)| = 1 \)). We assume that \( q^i \), considered as a function of \( r^i \) for fixed \( r_{i-1} \), is of the form (2), so that
\[
q^i(r_i) = 1 + a^i(r_{i-1})r^i + b^i(r_{i-1})(r^i)^2 + c^i(r_{i-1})(r^i)^3 ,
\]
where \( a^i(r_{i-1}) = O(n^{-1/2}) \), \( b^i(r_{i-1}) = O(n^{-1}) \) and \( c^i(r_{i-1}) = O(n^{-3/2}) \). To \( O(n^{-2}) \), the constant of proportionality in (17) is \( \prod_i s^i \), where \( s^i = 1 + b^i(0) \). In addition, the
quantity $s^2$, where $s = \sum_i s^i/d$, is a multiparameter Bayesian Bartlett correction; that is,
\[ s^{-2}W \sim \chi^2_d \] (19)
to $O(n^{-2})$.

Define
\[ z^i = z^i(\theta_i) = \{ k^i(\theta_{i-1}) \}^{1/2} \{ \theta^i - \hat{\theta}^i(\theta_{i-1}) \} \] (20)
where $k^i(\theta_i) = -\partial^2 l(\theta_i)/\partial \theta^i \partial \theta^i$. As is shown in Sweeting and Kharroubi (2003), for $i = 1, \ldots, d$, $r^i = r^i(\theta_i)$ defined in (16) has an expansion in ascending powers of $z^i$ of the form $r^i \equiv \sum_{j=1}^4 c^i_j(z^j)$, where $c^i_j = O(n^{-1/2})$.

Now define, for $i = 1, \ldots, d$,
\[ Q^i(z_i) = \frac{z^i_{(0)}}{r^i} q^i(r_i) , \]
where $r^i = r^i(\theta_i)$ and, for $i = 1, \ldots, d$, $\theta^i$ is implicitly defined by the relation (20). As in the univariate case, regarded as a function of $z^i$ with $z_{i-1}$ fixed, each $z^i/r^i$, and hence $Q^i$ are of the form
\[
1 + A^i(z_{i-1}) z^i + B^i(z_{i-1})(z^i)^2 + C^i(z_{i-1})(z^i)^3 ,
\]
where $A^i(z_{i-1}) = O(n^{-1/2}), B^i(z_{i-1}) = O(n^{-1}), C^i(z_{i-1}) = O(n^{-3/2})$ and $B^i(z_{i-1}) = d^i_2(z_{i-1}) + b^i(r_{i-1})$ where $d^i_2(z_{i-1})$ is the coefficient of $(z^i)^2$ in the expansion of $z^i/r^i$.

Exactly as in the single parameter case, it follows that, for $i = 1, \ldots, d$,
\[ s^i = 1 + b^i(0) = \frac{1 + B^i(0)}{1 + d^i_2(0)} , \] (21)
and $s^i$ may be obtained without expansion. The numerator of (21) can be written
\[ S^i \equiv 1 + B^i(0) \equiv \frac{1}{2} \{ Q^i(-e_i) + Q^i(e_i) \} , \]
where $e_i$ is the $i$-dimensional vector $(0, \cdots, 0, 1)$, while from (18) we have
\[ Q^i(z_i) = \{ -z^i/l_i(\theta_i) \} \{ \nu_i(\theta_i)/\nu_{i-1}(\theta_{i-1}) \} . \]
Now let $\theta_i^{\pm}$ be the solutions to the equations $z^i(\hat{\theta}_{i-1}, \theta^i) = \pm 1$, and write $\theta_i^\pm = (\hat{\theta}_{i-1}, \theta_i^{\pm})$. The numerator of (21) is then
\[ S^i \equiv \frac{1}{2} |J(i)|^{1/2} \{ \lambda(\hat{\theta}) \}^{-1} \tau^i , \]
where \( \tau^i = \{ \nu_i(\theta^-_i)/l_i(\theta^-_i) \} + \{-\nu_i(\theta^+_i)/l_i(\theta^+_i) \} \). Similarly, the denominator of (21) can be written as \( 1 + d^2_0(0) = \frac{1}{2} \omega^j \), where \( \omega^j = \{-\tau^i(\theta^-_i)\}^{-1} + \{\tau^i(\theta^+_i)\}^{-1} \). These results imply that, for \( i = 1, \ldots, d \),

\[
s^i = |J^{(i)}|^{1/2} \left\{ \omega^i \lambda(\hat{\theta}) \right\}^{-1} \tau^i .
\]  

(22)

These formulae lead to simple computation of Bayesian Bartlett corrections.

As in Section 2.2, formula (21) can be used to compute an approximation to the posterior expectation of a general function \( v(\theta) \). This leads to the formula

\[
E\{v(\theta)|X\} \doteq v(\hat{\theta}) \prod_{i=1}^d S^{s^i}/S^r ,
\]  

(23)

where \( S^{s^i} = \frac{1}{2}|J^{(i)}|^{1/2} \{ \lambda(\hat{\theta})v(\hat{\theta}) \}^{-1} \tau^{s^i} \) and \( \tau^{s^i} = \{ \nu_i(\theta^-_i)v(\theta^-_i)/l_i(\theta^-_i) \} + \{-\nu_i(\theta^+_i)v(\theta^+_i)/l_i(\theta^+_i) \} \). Let \( \alpha^-_i = (\tau^-)^{-1}\{ \nu_i(\theta^-_i)/l_i(\theta^-_i) \} \) and \( \alpha^+_i = (\tau^+)^{-1}\{-\nu_i(\theta^+_i)/l_i(\theta^+_i) \} \). Then, from (23), we see that

\[
E\{v(\theta)|X\} \doteq \hat{v} \prod_{i=1}^d \left\{ \frac{\alpha^-_i v^-_i + \alpha^+_i v^+_i}{\hat{v}} \right\} ,
\]  

(24)

where \( v^-_i = v(\theta^-_i), v^+_i = v(\theta^+_i) \) and \( \hat{v} = v(\hat{\theta}) \). Since \( S^{s^i}/S^i = 1 + O(n^{-1}) \), we can deduce the alternative summation form

\[
E\{v(\theta)|X\} \doteq \hat{v} + \sum_{i=1}^d \left( \alpha^-_i v^-_i + \alpha^+_i v^+_i - \hat{v} \right) .
\]  

(25)

Formula (25) exhibits the posterior expectation of \( v(\theta) \) as \( \hat{v} \) plus a correction term. As in the single parameter case, the weights do not depend on the function \( v \) at all.

Finally, Formula (25) can be used to compute predictive densities, leading to the formula

\[
p(y|X) \doteq p(y|\hat{\theta}, X) + \sum_{i=1}^d \left\{ \alpha^-_i p(y|\theta^-_i, X) + \alpha^+_i p(y|\theta^+_i, X) - p(y|\hat{\theta}, X) \right\} .
\]

3 Alternative approximations based on the signed root log-density ratio

3.1 Bartlett corrections and posterior expectations for a scalar parameter

In this section we obtain variants of formulae (8) and (13) for Bartlett corrections and posterior expectations that are based on the signed root log-density ratio as opposed to
the SRLLR. We begin by studying the behaviour of densities of the form

\[ \pi_n(\theta) \propto L_n(\theta) \lambda_n(\theta) \{1 + \eta_n(\theta - \hat{\theta}_n)\} , \]  

where now \( L_n(\theta) = L_n(\theta) \lambda_0(\theta), \quad \lambda_n(\theta) = \lambda_n(\theta)/\lambda_0(\theta), \)

\( \lambda_0 \) is any suitable function, \( (\eta_n) \) is an \( O(n^{-1}) \) sequence independent of \( \theta \) and \( \hat{\theta}_n = \text{argmax}\{L_n(\theta)\lambda_0(\theta)\} \). If \( \lambda_0(\theta) = 1 \), then \( \pi_n(\theta) \) is exactly of the form (1). Alternatively, \( \lambda_0(\theta) \) might be chosen to be a reference prior for \( \theta \) and \( \lambda_n(\theta) \) a particular proper form. Kass, Tierney and Kadane (1989) discuss these and other ideas in a general context. Alternative choices of \( \lambda_0(\theta) \) and \( \lambda_n(\theta) \) are discussed in Sweeting (1995). Here we follow on from the work of Sweeting to regard \( L_n(\theta) \) as a density with respect to \( \lambda_n(\theta)d\theta; \) that is, we take \( \lambda_0(\theta) = \lambda_n(\theta) \) and so \( L_n(\theta) = \lambda_n(\theta)L(\theta) \) and \( \lambda_n(\theta) = 1. \)

Write \( l_n(\theta) = \log L_n(\theta) \) and let \( w_n(\theta) = 2\{l_n(\hat{\theta}_n) - l_n(\theta)\} \) be the log-density ratio. In a similar way to Section 2.1, the transformation

\[ r_n(\theta) = \text{sign}(\theta - \hat{\theta}_n)\{w_n(\theta)\}^{1/2} \]  

is an increasing function of \( \theta \) and is referred to as the \textit{signed root log-density ratio} (SRLDR). Since

\[ L_n(\theta) \propto \exp\{l_n(\hat{\theta}_n) - l_n(\theta)\} = e^{-r^2/2}, \]

where \( r = r_n(\theta) \), the posterior density function \( L_n(\theta) \) is of standard normal form when expressed in terms of \( r \).

Now define \( j_n(\theta) = -l''_n(\theta), \) \( J_n = j_n(\hat{\theta}_n) \) and \( z_n(\theta) = J_n^{1/2}(\theta - \hat{\theta}_n). \) Suppressing \( n \) for notational convenience, Note that \( Q(z) = G(z)H(z), \) where \( G(z) = -zJ^{1/2}/l'(\theta) \) and \( H(z) = 1. \) Define \( \theta^- \) by \( z(\theta^-) = -1 \) and \( \theta^+ \) by \( z(\theta^+) = 1; \) that is, \( \theta^\pm = \hat{\theta} \pm J^{-1/2}. \) From the results of Section 2.2, we find that the numerator in (6) is

\[ S = 1 + B = \frac{1}{2} \{Q(-1) + Q(1)\} = \frac{1}{2} J^{1/2} \tau, \]  

where now \( \tau = \{1/l'(\theta^-)\} + \{-1/l'(\theta^+)\}. \) Similarly, the denominator in (6) takes the form \( 1 + d_2 \geq \frac{1}{2} \omega, \) where \( \omega = \{-r(\theta^-)\}^{-1} + \{r(\theta^+)\}^{-1}. \) It follows that the expression

\[ s = J^{1/2} \tau / \omega \]  

provides an alternative Bayesian Bartlett correction to (8).

Next we deduce an alternative formula to (13) based on the SRLDR for the posterior expectation of a general real-valued function \( v(\theta). \) Consider formula (11). The
fourth-order approximation (28) is already available for the denominator. A similar approximation can be applied to the numerator as follows. By setting \( H(z) \) to be \( v(\theta) \), it follows from the results of Section 2.2 that

\[
S^* = 1 + B^* = \frac{1}{2} \{ Q^*(-1) + Q^*(1) \} = \frac{1}{2} J^{1/2} \{ v(\hat{\theta}) \}^{-1} \tau^*,
\]

where \( \tau^* = \{ v(\theta^-)/l'(\theta^-) \} + \{ -v(\theta^+)/l'(\theta^+) \} \). Substituting (28) and (30) in (11) we finally obtain

\[
E \{ v(\theta)|X \} = \alpha^- v(\theta^-) + \alpha^+ v(\theta^+),
\]

where \( \alpha^- = \tau^- \{ 1/l'(\theta^-) \} \) and \( \alpha^+ = \tau^- \{ -1/l'(\theta^+) \} = 1 - \alpha^- \). Note that \( \alpha^\pm = \frac{1}{2} + O(n^{-1/2}) \).

As is the case with formula (13), the weights \( \alpha^- \) and \( \alpha^+ \) in (31) are independent of the function \( v(\theta) \) and therefore need only be calculated once when many expectations are required. The additional advantage of (31) is that the prior was absorbed into the likelihood function in \( r(\theta) \) which would be important for assessing the sensitivity of the posterior expectations to alternative specifications for \( \lambda(\theta) \) and also for assessing the influence of one or more outliers. Just such an important context - where formula (31) is employed in the calculation of expected value of sample information in decision theory - led us to the development of these new formulae. We discuss this point further in Section 4. On the other hand, unlike (13) which requires recomputing just \( \alpha^- \) and \( \alpha^+ \) for a change of prior, formula (31) requires in addition \( \hat{\theta}, J \) and the computation of \( l'(\theta) \), all of which would normally be available for the routine maximization of \( l(\theta) \).

Formula (31) can be used to compute approximate predictive densities, leading to the formula

\[
p(y|X) = \alpha^- p(y|\theta^-, X) + \alpha^+ p(y|\theta^+, X). \tag{32}
\]

### 3.2 The multiparameter case

In this section we consider the multivariate generalization of the results in Section 3.1. In a similar way to Section 3.1, we study the behaviour of densities of the form

\[
\pi_n(\theta) \propto L_n(\theta)\lambda_n(\theta)\{1 + \eta_n^T(\theta - \hat{\theta}_n)\}, \tag{33}
\]

where now \( \theta \in \mathcal{R}^d \),

\[
L_n(\theta) = \lambda_n(\theta) L_n(\theta), \quad \lambda_n(\theta) = 1,
\]

\( \eta_n = O(n^{-1}) \) and \( \hat{\theta}_n = \arg\max \{ L_n(\theta)\lambda_n(\theta) \} \). Let \( l_n(\theta) = \log L_n(\theta), j_n(\theta) = -d^2 l_n/d\theta^2 \), the matrix of second-order partial derivatives of \( -l_n \), and \( J_n = j_n(\hat{\theta}_n) \). Now define, for
From the results of Section 2.3, we find that the numerator in (21) can be written
\[ w'_{ni} = w^i_n(\theta_i) = 2\{l_n(\theta_{i-1}) - l_n(\theta_i)\} \]
and the signed root transformation
\[ r^i_n = r^i_n(\theta_i) = \text{sign}\{\theta^i - \hat{\theta}^i_n(\theta_{i-1})\}\{w^i_n\}^{1/2}. \]  
(34)
Note that \( w_n = \sum_i w^i_n = 2\{l_n(\hat{\theta}_n) - l_n(\theta)\} \). Suppressing the dependence on \( n \), define
\[ z^i = z^i(\theta_i) = \{k^i(\theta_{i-1})\}^{1/2}\{\theta^i - \hat{\theta}^i(\theta_{i-1})\} \]
(35)
where \( k^i(\theta_i) = -\partial^2 l(\theta_i)/\partial \theta^2 \). Let
\[ Q^i(z_i) = \{-z^i/l_i(\theta_i)\}\{\nu_i(\theta_i)/\nu_{i-1}(\theta_{i-1})\}. \]
where \( l_i(\theta) = \partial l(\theta)/\partial \theta^i \), \( \nu_i(\theta) = |J^{(i+1)}(\theta)|^{-1/2} \) and \( J^{(i)} \) is the submatrix of \( j \) corresponding to \( \theta^{(i)} \). Now let \( \theta^- \), \( \theta^+ \) be the solutions to the equations \( z^i(\hat{\theta}_{i-1}, \theta^i) = -1 \), \( z^i(\hat{\theta}_{i-1}, \theta^i) = 1 \) respectively; that is \( \theta^- = \hat{\theta} - (k^i)^{-1/2} \) and \( \theta^+ = \hat{\theta} + (k^i)^{-1/2} \), where \( k^i \) is the reciprocal of the first entry in \( \{J^{(i)}\}^{-1} \). Write \( \hat{\theta}^- = (\hat{\theta}_{i-1}, \theta^-) \) and \( \hat{\theta}^+ = (\hat{\theta}_{i-1}, \theta^+) \).
From the results of Section 2.3, we find that the numerator in (21) can be written
\[ S^i \equiv 1 + B^i(0) \equiv \frac{1}{2}\{Q^i(-e_i) + Q^i(e_i)\} = \frac{1}{2}|J^{(i)}|^{1/2}r^i, \]
where \( e_i \) is the \( i \)-dimensional vector \((0, \cdots, 0, 1)\) and \( r^i = \{\nu_i(\theta_i^-)/l_i(\theta_i^-)\} + \{-\nu_i(\theta_i^+)/l_i(\theta_i^+)\} \).
Similarly, the denominator of (21) can be written as \( 1 + d_2(0) = \frac{1}{2}\omega^i \), where \( \omega^i = \{-r^i(\theta_i^-)\}^{-1} + \{r^i(\theta_i^+)\}^{-1} \). These results imply that, for \( i = 1, \ldots, d, \)
\[ s^i = |J^{(i)}|^{1/2}r^i/\omega^i. \]  
(36)
As is the case in Sweeting and Kharroubi (2003), these formulae lead to simple computation of Bayesian Bartlett corrections. The quantity \( s^2 \), where \( s = \sum_is^i/d \), is a multiparameter Bayesian Bartlett correction satisfying (19).

We turn now to posterior expectations. Exactly as in Section 2.3, we find that, for \( i = 1, \ldots, d, \)
\[ S^{*i} \equiv \frac{1}{2}|J^{(i)}|^{1/2}\{v(\hat{\theta})\}^{-1}\tau^{*i}, \]
where \( \tau^{*i} = \{\nu_i(\theta_i^-)v(\theta_i^-)/l_i(\theta_i^-)\} + \{-\nu_i(\theta_i^+)/v(\theta_i^+)/l_i(\theta_i^+)\} \). Let \( \alpha^- = \tau^{*i}^{-1}\{\nu_i(\theta_i^-)/l_i(\theta_i^-)\} \) and \( \alpha^+ = \tau^{*i}^{-1}\{-\nu_i(\theta_i^+)/l_i(\theta_i^+)\} \). Then, from (23), we have
\[ E\{v(\theta)|X\} \doteq \hat{v} \prod_{i=1}^d \left\{ \frac{\alpha^-v^- + \alpha^+v^+}{\hat{v}} \right\}, \]  
(37)
where \( v_i^- = v(\theta_i^-), v_i^+ = v(\theta_i^+) \) and \( \hat{v} = v(\hat{\theta}) \).

It also follows from the derivation of \( S_i \) that each of the quantities in parenthesis in (37) is \( 1 + O(n^{-1}) \), from which we obtain the alternative summation form

\[
E\{v(\theta)|X\} \doteq \hat{v} + \sum_{i=1}^{d} (\alpha_i^- v_i^- + \alpha_i^+ v_i^+ - \hat{v}) .
\]  

As in Section 2.3. Formula (38) exhibits the posterior expectation of \( v(\theta) \) as \( \hat{v} \) plus a correction term and the weights do not depend on the function \( v \) at all.

Finally, formula (38) gives rise to the formula

\[
p(y|X) = p(y|\hat{\theta}, X) + \sum_{i=1}^{d} \left\{ \alpha_i^- p(y|\theta_i^-, X) + \alpha_i^+ p(y|\theta_i^+, X) - p(y|\hat{\theta}, X) \right\}
\]

for the predictive density.

## 4 Expected Value of Sample Information

EVSI was first developed as part of decision theory (Raiffa et al., 1967). The central concept is the expected value of each decision option. EVSI effectively quantifies the value to be gained by obtaining sample information before making a large investment decision (e.g. test drilling for oil before setting up a major oil platform). In health economics, the use of EVSI methods is an active area of methodological development, with authors investigating and promoting its use as a tool for quantifying the societal value of expensive medical research projects and for determination of optimum sample sizes and allocation rates in randomized clinical trails (Chilcott et al., 2003) and (Claxton, 1996, 1999, 2000).

We begin with a mathematical description of EVSI using the context of health economics (Claxton 1996, Brennan et al., 2002a and Ades et al., 2004). We assume a decision model with unknown parameters \( \theta \), with a choice to be made between a fixed number of treatments \( t = 1, 2, \ldots, T \). Each treatment delivers health or `utility’ gains measured in quality adjusted life years (QALYs), with an associated treatment cost. We adopt a net benefit approach to cost-effectiveness analysis (Stinnett et al., 1998), whereby health gains are monetarised by multiplying by the decision maker’s willingness to pay per additional QALY (\( \lambda \)). \( NB(t, \theta) \) is the net benefit of treatment \( t \) if the parameters take the value \( \theta \). The net benefit for treatment \( t \) is thus

\[
NB(t, \theta) = \lambda U(t, \theta) - C(t, \theta),
\]  

(39)
where the functions $C$ and $U$ give the costs and the QALYs under treatment $t$.

The optimal decision given the current information is the decision that yields the highest expected net benefit

$$\max_t E_{\theta} NB(t, \theta)$$

Assume that a particular study is being considered with a specific sample size vector $n$. The new study will provide new data $X_{\theta[i]}$ relating to parameters of interest $\theta[i]$, where $\theta[i]$ is a vector of $i$ components subset of $\theta$. Then the above expression can be further written as

$$\max_t E_{\theta[i]}(\theta[i], X_{\theta[i]}) NB(t, \theta[i], \theta[i]),$$

where the expectation is taken over the prior density of $\theta[i]$, the complement of $\theta[i]$, and the posterior density of $\theta[i]$ given $X_{\theta[i]}$. As we don’t know yet what the proposed collection of data $X_{\theta[i]}$ will be, we must average over the distribution of $X_{\theta[i]}$ to calculate the expected value of a decision made after data $X_{\theta[i]}$ have been collected

$$E_{X_{\theta[i]}} \max_t E_{\theta[i]}(\theta[i], X_{\theta[i]}) NB(t, \theta[i], \theta[i]).$$

Finally, the expected value of sample information is the difference between the expected value of a decision made after data $X_{\theta[i]}$ have been collected and expected value of a decision made now

$$E_{X_{\theta[i]}} \max_t E_{\theta} NB(t, \theta) - \max_t E_{\theta} NB(t, \theta)$$

where $\theta = (\theta[i], \theta[i])$. To calculate EVSI, we need to be able to evaluate the first term of (40). This term contains an inner expectation of net benefit over $\theta[i]$ and an outer expectation of net benefit over $X_{\theta[i]}$. Current methods of evaluating these expectations (Brennan et al., 2002b and Ades et al., 2004) involve a two level Monte-Carlo sampling algorithm. The $X_{\theta[i]}$ are sampled first, and then for each individual sampled $X_{\theta[i]}$, both the posterior density of $\theta[i]$ given $X_{\theta[i]}$ and then the evaluation of the inner expectation using Monte Carlo sampling are required. This nested evaluation of Monte-Carlo estimates with embedded Bayesian updating produces significant computation time. Previous authors restrict their work to situations in which the likelihood for the proposed data $X_{\theta[i]}$ is conjugate with priors, so that means or other parameters for posterior distributions are available in closed form and to scalar priors with no correlations. If this conjugate form applies, then analytic computation of posterior density means that the overall computation is simplified, but it still requires large numbers of runs of the decision model. Generally, in the absence of conjugate distributions, the Bayesian updating could be
undertaken for each simulated dataset using Markov Chain Monte Carlo methods (e.g. repeated application in WinBUGS), creating a level of computation, which might be impractical in all but the simplest of decision models.

Here we show how the use of formula (38) to the inner expectation of the first term of (40) provides a new and valuable approach that transforms the efficiency of the EVSI computations as follows. First note that, if \( v(\theta) = NB(t, \theta) \), then the inner expectation of the first term of (40) is

\[
NB(t, \hat{\theta}) + \sum_{i=1}^{d} \left( \alpha_i^- NB(t, \theta_i^-) + \alpha_i^+ NB(t, \theta_i^+) - NB(t, \hat{\theta}) \right), \tag{41}
\]

and so the approximated EVSI is given by

\[
E_{X_{\theta[j]}} \max_t \left\{ NB(t, \hat{\theta}) + \sum_{i=1}^{d} \left( \alpha_i^- NB(t, \theta_i^-) + \alpha_i^+ NB(t, \theta_i^+) - NB(t, \hat{\theta}) \right) \right\} - \max_t E_\theta NB(t, \theta). \tag{42}
\]

In the case study below, we are just interested in the first term of (41) in order to expose the main ideas. This implies that

\[
E_{X_{\theta[j]}} \max_t NB(t, \hat{\theta}) - \max_t E_\theta NB(t, \theta) \tag{43}
\]

provides the required approximation to the EVSI calculations.

5 Hypothetical Cost-Effectiveness model

To explore the feasibility of using formula (43) for the EVSI approximation, we utilized a previously developed hypothetical cost-effectiveness model. EVSI calculations were previously undertaken and reported using the nested Monte-Carlo algorithm approach (Brennan et al., 2002b). In this section we apply the new EVSI approximation formula and compare the results.

The cost-effectiveness model compares two strategies: treatment with drug T0 versus treatment with drug T1. Table 1 shows the nineteen uncertain model parameters, with prior mean values shown for T0 (column a), T1 (column b) and hence the incremental analysis (column c). Costs include cost of drug and cost of hospitalisations (the product of the percentage of patients admitted to hospital, days in hospital, and cost per day). For example, cost of strategy T0 = $1000 + 10% x 5.20 x $400 = $1208. Health benefits are measured as QALYs gained and come from two sources: responders receive a utility improvement for a specified duration, and some patients have side effects with a utility
decrement for a specified duration i.e. QALY for strategy T0 = 70% responders x 0.3 x 3 years + 25% side effects x -0.1 x 0.5 years = 0.6175. The willingness to pay i.e. threshold cost per QALY is set at $10000. Thus, the net benefit of T0 is = $10000 x 0.6175 - $1208 = $4967.

In this simple, hypothetical model, the uncertain model parameters are characterised with independent normal distributions. Standard deviations for the model parameters are shown in columns (d) and (e). Each parameter can be informed by collection of further data on individual patients. It is assumed in this simple model that the patient level variance is known for each parameter and this is given in columns (f) and (g). Note, by assuming independent normal distributions with known patient level variance, we are able to use the conjugate assumption to calculate posterior densities analytically.

Given the existing data, the basic model results show $5405 expected net for T1 compared with $4967 for T0 (difference = $437.80), which means that our baseline decision given current information should be to adopt strategy T1. Probabilistic sensitivity analysis (Briggs et al, 1999) shows that T1 provides greater net benefits than T0 on only 54.5% of 1000 Monte Carlo samples. This suggests that obtaining more data on the uncertain parameters might help us with our decision.

EVSI results were previously published, using the nested Monte Carlo sampling algorithm. 1000 simulations for the outer expectation were used and for each outer sampled data 1000 simulations were also used to evaluate the inner expectation in Formula (41). In total therefore, there were 1,000,000 model runs for each EVSI value calculated. Five different sample sizes (n= 10, 25, 50, 100 and 200) were investigated on five different proposed data collection exercises. The five proposed data collection exercises were based on subgroups of parameters as follows:

a) a proposed randomized controlled clinical trial measuring only response rate parameters (parameters 5, 14)

b) an observational study on utility only (parameters 6, 15),

c) a trail combined with utility data collection (parameters 5, 6, 14, 15),

d) an observational study of the duration of response to therapy (parameters 7, 16)

e) a trial combined with utility study alongside an observational study on duration of response (parameters 5, 6, 7, 14, 15, 16).
<table>
<thead>
<tr>
<th>Parameters</th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
<th>e</th>
<th>f</th>
<th>g</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost of Drug</td>
<td>$1000</td>
<td>$1500</td>
<td>$500</td>
<td>$1</td>
<td>$1</td>
<td>$500</td>
<td>$500</td>
</tr>
<tr>
<td>% Admissions</td>
<td>10%</td>
<td>8%</td>
<td>-2%</td>
<td>2%</td>
<td>2%</td>
<td>25%</td>
<td>25%</td>
</tr>
<tr>
<td>Days in Hospital</td>
<td>5.20</td>
<td>6.10</td>
<td>0.90</td>
<td>1.00</td>
<td>1.00</td>
<td>4.00</td>
<td>4.00</td>
</tr>
<tr>
<td>Cost per day</td>
<td>$400</td>
<td>$400</td>
<td>...</td>
<td>$200</td>
<td>$200</td>
<td>$200</td>
<td>$200</td>
</tr>
<tr>
<td>% Responding</td>
<td>70%</td>
<td>80%</td>
<td>10%</td>
<td>10%</td>
<td>10%</td>
<td>20%</td>
<td>20%</td>
</tr>
<tr>
<td>Utility Change if respond</td>
<td>0.30</td>
<td>0.30</td>
<td>...</td>
<td>0.10</td>
<td>0.05</td>
<td>0.20</td>
<td>0.20</td>
</tr>
<tr>
<td>Duration of response (years)</td>
<td>3.0</td>
<td>3.0</td>
<td>...</td>
<td>0.5</td>
<td>1.0</td>
<td>1.0</td>
<td>2.0</td>
</tr>
<tr>
<td>% Side effects</td>
<td>25%</td>
<td>20%</td>
<td>-5%</td>
<td>10%</td>
<td>5%</td>
<td>20%</td>
<td>20%</td>
</tr>
<tr>
<td>Change in utility if side effect</td>
<td>-0.10</td>
<td>-0.10</td>
<td>-0.00</td>
<td>0.02</td>
<td>0.02</td>
<td>0.10</td>
<td>0.10</td>
</tr>
<tr>
<td>Duration of side effect (years)</td>
<td>0.50</td>
<td>0.50</td>
<td>...</td>
<td>0.20</td>
<td>0.20</td>
<td>0.80</td>
<td>0.80</td>
</tr>
</tbody>
</table>

**Central Estimate Results**

- Total Cost: $1208 $1695 $487
- Total QALY: 0.6175 0.7100 0.0925
- Cost per QALY: $1956 $2388 $5267

**Net Benefit**

(Threshold cost per QALY = $10000) $4967 $5405 $438
Table 2 and Figure 1 show the EVSI results using the nested Monte Carlo sampling algorithm. EVSI for a trial shows the lowest expected values. As one would expect, EVSI is greater for higher sample sizes but with diminishing returns as n increases. The lower bound of EVSI for a sample size of zero is clearly zero (there is no value to be obtained from no data). The upper bound is the expected value of perfect information, known as partial EVPI (i.e. the value of a study with infinite sample size for a particular subgroup of parameters). In this example model, the value of a trial (a) is substantially lower than that for a utility study (b) or a duration study (c). Data collection on further combinations of parameters (d and e) provide greater value but again with diminishing returns because they are bounded above by the overall EVPI (i.e. from a study which tells us the exact true value of each currently uncertain model parameter).

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Parameter</th>
<th>10</th>
<th>25</th>
<th>50</th>
<th>100</th>
<th>200</th>
</tr>
</thead>
<tbody>
<tr>
<td>% Response trial</td>
<td>297</td>
<td>331</td>
<td>330</td>
<td>355</td>
<td>357</td>
<td></td>
</tr>
<tr>
<td>Utility trial</td>
<td>605</td>
<td>669</td>
<td>716</td>
<td>736</td>
<td>769</td>
<td></td>
</tr>
<tr>
<td>Duration of response</td>
<td>690</td>
<td>815</td>
<td>856</td>
<td>855</td>
<td>860</td>
<td></td>
</tr>
<tr>
<td>% Response trial/Utility trial</td>
<td>776</td>
<td>853</td>
<td>878</td>
<td>877</td>
<td>907</td>
<td></td>
</tr>
<tr>
<td>Duration of response</td>
<td>1100</td>
<td>1217</td>
<td>1277</td>
<td>1310</td>
<td>1321</td>
<td></td>
</tr>
</tbody>
</table>

Table 2. EVSI results:- Monte Carlo Simulations (1000 x 1000 iterations).

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Parameter</th>
<th>10</th>
<th>25</th>
<th>50</th>
<th>100</th>
<th>200</th>
</tr>
</thead>
<tbody>
<tr>
<td>% Response trial</td>
<td>243</td>
<td>279</td>
<td>298</td>
<td>304</td>
<td>309</td>
<td></td>
</tr>
<tr>
<td>Utility trial</td>
<td>558</td>
<td>659</td>
<td>706</td>
<td>730</td>
<td>742</td>
<td></td>
</tr>
<tr>
<td>Duration of response</td>
<td>693</td>
<td>773</td>
<td>790</td>
<td>839</td>
<td>844</td>
<td></td>
</tr>
<tr>
<td>% Response trial/Utility trial</td>
<td>689</td>
<td>789</td>
<td>833</td>
<td>860</td>
<td>877</td>
<td></td>
</tr>
<tr>
<td>Duration of response</td>
<td>1045</td>
<td>1171</td>
<td>1232</td>
<td>1267</td>
<td>1286</td>
<td></td>
</tr>
</tbody>
</table>

Table 3. EVSI results:- Approximation formula (43) (150000 iterations).
Figure 1. EVSI results:– Monte Carlo Simulations (1000 x 1000 iterations).

Figure 2. EVSI results:– Approximation formula (43) (150000 iterations).
We now apply the new EVSI approximation Formula (43). The analysis was undertaken based on 150,000 iterations. Table 3 and Figure 2 set out the approximated EVSI results, showing are very similar order of magnitude to those produced via the nested Monte Carlo method. We index the overall EVPI at 100. The differences between the two methods shown on this indexed scale are shown in Table 4. In all cases the EVSI approximation is within 7 percentage indexed points of the calculation by the nested Monte-Carlo approach, and for all but one evaluated point the 2 estimates are within 5 percentage points. It must also be remembered that the nested Monte-Carlo results themselves are estimates and are subject to the error associated with evaluating expectations of functions using just 1000 Monte Carlo samples. For practical purposes, this level of alignment in estimates means that the approximation based on Formula (43) provides an adequate estimate of EVSI for this decision model.

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>10</th>
<th>25</th>
<th>50</th>
<th>100</th>
<th>200</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameter</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Response trial</td>
<td>4.1</td>
<td>3.9</td>
<td>2.4</td>
<td>3.9</td>
<td>3.6</td>
</tr>
<tr>
<td>Utility trial</td>
<td>3.6</td>
<td>0.8</td>
<td>0.8</td>
<td>0.5</td>
<td>2.0</td>
</tr>
<tr>
<td>Duration of response</td>
<td>-0.2</td>
<td>3.2</td>
<td>5.0</td>
<td>1.2</td>
<td>1.2</td>
</tr>
<tr>
<td>% Response trial/ Utility trial</td>
<td>6.6</td>
<td>4.8</td>
<td>3.4</td>
<td>1.3</td>
<td>2.3</td>
</tr>
<tr>
<td>% Response trial/ Utility trial/ Duration of response</td>
<td>4.2</td>
<td>3.5</td>
<td>3.4</td>
<td>3.3</td>
<td>2.6</td>
</tr>
</tbody>
</table>

Table 4. Differences between EVSI approximation and nested Monte Carlo algorithm results (Indexed to overall EVPI = 100).

Computation time is significantly reduced. The nested Monte Carlo approach with 1000 x 1000 iterations takes 15 minutes per evaluated point using the interpreted language R. To produce the 25 points in Figure 1 therefore takes 6.25 hours. This is for a very simple decision model. Formula (43) with 150000 iterations using R took a maximum of 18 seconds for the evaluated points, a computation time of 7.5 minutes to produce Figure 2. In fact, the computation time reductions were seen to be up to 100 times shorter using the approximation method.

6 Discussion

In this paper we have developed some new fourth-order correct formulae for various posterior quantities of interest, including Bayesian Bartlett corrections, posterior expec-
tations and predictive densities based on the asymptotic theory of signed root log-density ratios. As is the case with the Sweeting and Kharroubi (2003), these new formulae require little more than standard likelihood or posterior maximization computer output for their implementation, since they do not require computation of log-likelihood derivatives beyond second order. The formulae are all correct to the same asymptotic order as those given in previous approximations.

The presented approximations are designed to absorb the prior into the likelihood function, constructing formulae that focus on the posterior. The practical value of new formulation lies in reducing the number of computations and the time required when applied to the calculation of the expected value of sample information from decision theory. The illustrative model demonstrates the feasibility of the new approach. It shows that the expected value of sample information approximated results are similar to the standard nested Monte Carlo sampling method, but are achieved with up to 100-fold computation time reductions.

The model presented in this paper is very simple and is only intended to illustrate the potential of the new approach given here. Although it is relatively straightforward for our model, the calculation of EVSI can often be completely impractical for computationally expensive models and so the trade-off between the standard approach and the new approximation will be even more favourable. In our simple model, the posterior mode was calculated analytically, using the conjugate nature of the independent normal distribution, and only the first term of formula (41) was needed to provide adequate estimates in a much shorter time. Further research is required on using the approximation approach to evaluate EVSI in more complex decision models and for more complex distributions. The practical difficulties need to be explored when the posterior mode is not available analytically and so Newton or some other technical methods like MCEM algorithm (see Wei and Tanner, 1990) are then needed. Also important is to explore the circumstances when formula (43) is inadequate. It would be worthwhile to explore further the value of the 'correction' term of formula (41) (i.e. formula (42)) in different circumstances. The development of code to implement the approach on a generic set of EVSI problems might also be of value.

In summary, we have developed a new formulation for Bayesian approximation. This new theory enables swifter computation of several important posterior quantities of interest, in particular posterior expectations, Bayesian Bartlett corrections, and predictive densities. We have applied the posterior expectation formula within the calculation of
expected value of sample information, providing a new approximation Formula (43) for EVSI. This methodology can transform the efficiency of EVSI calculation. Tested against a previously published hypothetical decision model, the approximation produces EVSI results of a similar order of magnitude but with computation time reduced by be up to 100 times. More research in complex decision models is required. We hope the new formulation presented here can be of further practical value in many fields of Bayesian approximation.

References


