MONTE CARLO APPROXIMATIONS
IN BAYESIAN DECISION THEORY

by

Jun Shao
Purdue University

Technical Report #88-39

PURDUE UNIVERSITY

DEPARTMENT OF STATISTICS

MONTE CARLO APPROXIMATIONS IN BAYESIAN DECISION THEORY

by

Jun Shao
Purdue University

Technical Report #88-39

Department of Statistics
Purdue University

August 1988
Revised January 1989

MONTE CARLO APPROXIMATIONS IN BAYESIAN DECISION THEORY

PART I

by

Jun Shao *

Purdue University

Abstract

In decision-making problems, the Bayesian action and its posterior expected loss usually can not be obtained analytically. ~~In~~ this paper ~~we study~~ a Monte Carlo method for approximating the Bayesian action and its posterior expected loss. The Monte Carlo approximation to the Bayesian action is obtained through ~~(1)~~ approximating the posterior expected loss function by using the Monte Carlo integration method and ~~(2)~~ searching the minimum of the approximated posterior expected loss function. As the Monte Carlo sample size diverges to infinity, the Monte Carlo approximations are shown to be convergent in very general situations. The rates of the convergence are also obtained under some regularity conditions on the loss function. Two accuracy measures of the Monte Carlo approximations are proposed. Some examples are presented for illustration.

Key words: Bayesian action; posterior expected loss; Monte Carlo integration; Almost sure convergence; Accuracy measures.

* The research of this author was partially supported by the Office of Naval Research Contract N00014-88-K-0170 and NSF Grant DMS-8717799, DMS-8702620 at Purdue University.

1. INTRODUCTION

Many problems in Bayesian analysis can only be solved numerically. In this article, we study the Monte Carlo method for determining the Bayesian action in a decision-making problem. Let X be a random n-vector whose distribution has a density $f(x \mid \theta)$ (with respect to a measure ν), where $\theta \in \Theta$ is an unknown parameter and Θ is a subset of $\mathbf{R}^p$. For a particular problem (e.g., the estimation of θ), a decision is to be made after observing the data $X = x$. These decisions are commonly referred to as actions in the literature. We will denote a particular action by a and the set of all possible actions under consideration by $\mathcal{A}$. Let $L(\theta, a)$ be the loss incurred when the action a is taken and θ turns out to be the *true* parameter. $L(\theta, a)$ is assumed to be a nonnegative function defined on $\Theta \times \mathcal{A}$ (measurable in θ for each a). Suppose that $\pi(\theta)$ is the density (with respect to a measure μ) of a prior distribution on Θ. The posterior expected loss of an action a, given the data x, is defined to be

$$r(a) = \int_{\Theta} L(\theta, a) p(\theta \mid x) d\mu, \tag{1.1}$$

where

$$p(\theta \mid x) = f(x \mid \theta)\pi(\theta)/m(x)$$

is the posterior of θ and

$$m(x) = \int_{\Theta} f(x \mid \theta)\pi(\theta) d\mu. \tag{1.2}$$

A Bayesian optimal action is then an action a^* which minimizes $r(a)$, or equivalently, an action a^* which minimizes

$$\rho(a) = \int_{\Theta} L(\theta, a) f(x \mid \theta)\pi(\theta) d\mu. \tag{1.3}$$

If the problem under consideration does not involve a conjugate prior-likelihood pair or the loss function is not of a special form (e.g., the squared error loss), this minimization problem can not be solved explicitly. That is, a^* does not have a closed form since the integrals (1.1)-(1.3) usually do not have explicit forms. Consequently, numerical methods for calculating the integrals (1.1)-(1.3) and the Bayesian action a^* are required. In particular, one useful method for calculating the posterior moments $\int_{\Theta} g(\theta) p(\theta \mid x) d\mu$ is the Monte Carlo method, which was studied in Kloek and van Dijk (1978), Stewart (1979, 1983) and Geweke (1986, 1988a,b). A Monte Carlo method for calculating the Bayesian action is described as follows:

Step 1. Choose a probability density $h(\theta)$ (with respect to μ) such that the support of $h(\theta)$ includes that of $\pi(\theta)$. Let P_h be the probability distribution corresponding to $h(\theta)$. Generate m independent and identically distributed (i.i.d.) random vectors $\theta_1, \theta_2, \ldots, \theta_m$ from P_h.

Step 2. Approximate $\rho(a)$ for any $a \in \mathcal{A}$ by

$$\rho_m(a, \omega) = m^{-1} \sum_{i=1}^{m} L(\theta_i, a) w(\theta_i), \tag{1.4}$$

where ω denotes a particular sequence $\{ \theta_i \}_{i=1}^{\infty}$ and

$$w(\theta) = f(x \mid \theta) \pi(\theta) / h(\theta). \tag{1.5}$$

Step 3. Find an $a_m(\omega) \in \mathcal{A}$ which minimizes $\rho_m(a, \omega)$, i.e.,

$$\rho_m(\omega) = \rho_m(a_m(\omega), \omega) = \min\{ \rho_m(a, \omega) : a \in \mathcal{A} \}. \tag{1.6}$$

$a_m = a_m(\omega)$ is the Monte Carlo approximation to the Bayesian action a^*. For the computation of $a_m(\omega)$, working with $\rho(a)$ is preferable to working with $r(a)$ since it avoids the computation of $m(x)$. A Monte Carlo approximation to $r(a_m(\omega))$, the posterior expected loss of $a_m(\omega)$, is $r_m(\omega) = \rho_m(\omega) / m_m(\omega)$, where $\rho_m(\omega)$ is defined in (1.6) and $m_m(\omega) = m^{-1} \sum_{i=1}^{m} w(\theta_i)$. $r_m(\omega)$ can also be used to estimate the minimum posterior expected loss $r(a^*)$.

The function h in Step 1 is called the *importance function.* It should be chosen so that the random generation process in Step 1 can be easily carried out. The accuracy of the Monte Carlo approximations depends on the choice of h. Generally speaking, h should mimic the posterior $p(\theta \mid x)$ and satisfy several technical conditions such as B3, (2.8) and (3.1) in Sections 2 and 3. Extensive discussions on how to choose this function can be found in Berger (1985, Section 4.9) and Geweke (1988b). In Step 2, we have reduced computation by using the same random sequence $\{ \theta_i \}_{i=1}^{m}$ for computing $\rho_m(a, \omega)$ in (1.4) for all $a \in \mathcal{A}$. Not only does this method allow us to use a large m, but it also results in certain desirable properties for a_m, which will be discussed in the next section. In Step 3, one may employ any available algorithm for solving a minimization problem.

Throughout the paper, we assume that the integrals (1.2)-(1.3) are finite and the likelihood $f(x \mid \theta)$ and the prior $\pi(\theta)$ can be evaluated easily when x and θ are known. No other assumption on $f(x \mid \theta)$ and $\pi(\theta)$ is made. Also, there is no restriction on the length of the vector x,

i.e., the total number of observations. If $\pi(\theta)$ is a hierarchical prior of the form

$$\pi(\theta) = \int_{\Lambda} \pi_1(\theta|\lambda)\pi_2(\lambda)d\tau,$$

where τ is a measure on $\Lambda \subset \mathbf{R}^s$, then $\pi(\theta)$ is not easy to evaluate (for a given θ) and the generation of θ_i's from $\pi(\theta)$ may also be difficult. However, we may draw m i.i.d. random $(p+s)$-vectors $\xi_i=(\theta_i, \lambda_i)^\tau$ from the distribution with density $h(\xi)$ (with respect to $\mu\times\tau$ on $\Theta\times\Lambda$), and compute

$$\rho_m(a, \omega) = m^{-1}\sum_{i=1}^{m} L(\theta_i, a) f(x|\theta_i)\pi_1(\theta_i|\lambda_i)\pi_2(\lambda_i)/h(\xi_i).$$

The rest of this paper is organized as follows. In Section 2, convergence of the Monte Carlo approximations $a_m(\omega)$ and $r_m(\omega)$ is studied. The rates of convergence are also obtained. In Section 3, we propose two measures which can be used to indicate the accuracy of the Monte Carlo approximations. Section 4 contains an example of application.

2. LIMITING BEHAVIOR OF THE MONTE CARLO APPROXIMATIONS

2.1. Convergence

Here, an immediate question is whether or not a_m converges to a^* as m diverges to infinity. Since $a_m=a_m(\omega)$ is a function of random quantity ω, we say that a_m converges to a^* almost surely (*a.s.*) if for almost all ω (with respect to P_h), $a_m(\omega)\to a^*$ as $m\to\infty$.

To provide an answer to the above question, let us first consider the simple case of finite action problems, i.e., $\mathcal{A}$ consists of finitely many actions $a^{(1)}, a^{(2)},..., a^{(k)}$. The most important examples are hypothesis testing and classification problems. See also Example 2 in Section 4. Without loss of generality, assume that $a^{(1)},..., a^{(i)}$ are Bayesian actions, where i is an integer $\le k$. Let $\varepsilon=2^{-1}\min_{i<l\le k}\{\rho(a^{(l)})-\rho(a^{(1)})\}$. Since k is fixed, it follows from the strong law of large numbers (SLLN) that for almost all ω, there is an integer m_ω such that when $m\ge m_\omega$,

$$\rho_m(a^{(j)}, \omega) < \rho(a^{(j)}) + \varepsilon = \rho(a^{(1)}) + \varepsilon \le \rho(a^{(l)}) - \varepsilon < \rho_m(a^{(l)}, \omega)$$

holds for any $j\le i$ and $l>i$. This proves that $a_m(\omega)=a^{(j)}$ for a $j\le i$, i.e., $a_m(\omega)$ is a Bayesian action for $m\ge m_\omega$. From the SLLN, $M_m(\omega)\to M(x)$ *a.s.* Thus, both $r(a_m(\omega))$ and $r_m(\omega)$ converge to $r(a^{(1)})$ *a.s.* If, in addition, the Bayesian action is unique ($i=1$), then $a_m(\omega)\to a^{(1)}$ *a.s.*

The above results can be extended to the case where $\mathcal{A}$ is a compact subset of $\mathbf{R}^q$. We state the following result. Its proof uses a similar argument to the above discussion and a uniform SLLN and is given in the Appendix.

Theorem 1. Suppose that $\mathcal{A}$ is a compact subset of $\mathbf{R}^q$. Let

$$\mathcal{A}_B = \{\, a : a \text{ is a Bayesian action} \,\} \tag{2.1}$$

and for a fixed ω,

$$A_\omega = \{\, a : a \text{ is a limit point of } \{\, a_m(\omega) \,\}_{m=1}^{\infty} \,\}.$$

If $L(\theta, a)$ is continuous in a for each $\theta \in \Theta$ and there is a measurable function $M(\theta)$ such that

$$\sup\{\, L(\theta, a) : a \in \mathcal{A} \,\} \le M(\theta) \quad \text{and} \quad \int_\Theta M(\theta) f(x \mid \theta) \pi(\theta) d\mu < \infty,$$

then there exists a Bayesian action and the following results hold:

$$A_\omega \subset \mathcal{A}_B \quad \text{for almost all } \omega, \tag{2.2}$$

and

$$r(a_m(\omega)) \to r(a^*) \quad a.s. \quad \text{and} \quad r_m(\omega) \to r(a^*) \quad a.s. \tag{2.3}$$

If the Bayesian action is unique, then $a_m(\omega) \to a^*$ $a.s.$

When a^* is not unique, $a_m(\omega)$ may not converge since $\{\, a_m(\omega) \,\}_{m=1}^{\infty}$ may have several limit points. However, $a_m(\omega)$ can still be used in practice as a good decision since the posterior expected loss of $a_m(\omega)$ converges to the minimum posterior expected loss. In addition, $a_m(\omega)$ is close to a Bayesian action in the sense of (2.2). Regardless of uniqueness of a^*, $r_m(\omega)$ is close to $r(a_m(\omega))$ (or $r(a^*)$) according to (2.3).

Often $\mathcal{A}$ is a convex subset of $\mathbf{R}^q$ (e.g., $\mathcal{A} = \Theta$ for an estimation problem) and $\mathcal{A}$ is not compact. In this case, let $\mathcal{A}'$ be the closure of $\mathcal{A}$ and assume that the loss function $L(\theta, a)$ is defined on $\Theta \times \mathcal{A}'$. Furthermore, let $\mathcal{A}$ be an unbounded subset of $\mathbf{R}^q$. (If $\mathcal{A}$ is bounded, then $\mathcal{A}'$ is compact and Theorem 1 applies.) To establish the convergence of a_m in this case, we need some additional regularity conditions. In the following, $\mathbf{B}(c)$ and $\mathbf{N}(c)$ denote the sets $\{\, a \in \mathcal{A}' : \| a - a^* \| = c \,\}$ and $\{\, a \in \mathcal{A}' : \| a - a^* \| \le c \,\}$ for a positive constant c, where $\| \ \|$ is the Euclidean norm on $\mathbf{R}^q$.

Condition A.

A1. $\mathcal{A}$ is convex and $L(\theta, a)$ is a convex continuous function of a for each $\theta \in \Theta$.

A2. $r(a)<\infty$ for all $a \in \mathcal{A}$ and there exists an $a^* \in \mathcal{A}$ such that

$$r(a^*) = \min\{ r(a): a \in \mathcal{A} \}.$$

A3. There is a nonempty set $\mathbf{B}(c)$ such that

$$\inf \{ r(a): a \in \mathbf{B}(c) \} > r(a^*).$$

A4. There is a measurable function $M(\theta)$ such that

$$\sup\{ L(\theta, a): a \in \mathbf{N}(c) \} \le M(\theta) \quad \text{and} \quad \int_{\Theta} M(\theta) f(x \mid \theta)\pi(\theta) d\mu < \infty.$$

Examples of convex loss functions which are commonly used in the estimation of scalar θ are the squared error loss $(\theta-a)^2$ and the absolute error loss $|\theta-a|$. See also Example 2 in Section 4. For vector $\theta=(\theta^{(1)},...,\theta^{(p)})^\tau$ and $a=(a^{(1)},...,a^{(p)})^\tau$, a commonly used loss is the weighted average of the loss functions $L_i(\theta^{(i)}, a^{(i)})$, which satisfies A1 if each L_i does.

Under A1, $r(a)$ is convex. Condition A2 simply says that a Bayesian action for the problem under consideration exists. Let $\mathcal{A}_B$ be defined as in (2.1) and

$$\mathcal{A}'_B=\{ a \in \mathcal{A}': r(a)=r(a^*) \}.$$

Then $\mathcal{A}_B \subset \mathcal{A}'_B$ and $\mathcal{A}_B = \mathcal{A}'_B$ iff $r(a)>r(a^*)$ for all a in the boundary of $\mathcal{A}$ but not in $\mathcal{A}$. Condition A3 rules out the possibility that $r(a)$ is too flat in the sense that $\mathcal{A}_B$ is unbounded. Condition A3 is also necessary for $\mathcal{A}_B$ to be bounded. A sufficient condition for A3 is that $L(\theta, a)$ is strictly convex, which also ensures the uniqueness of the Bayesian action. Condition A4 is mild since $\mathbf{N}(c)$ is compact and $L(\theta, a)$ is continuous in a.

Theorem 2. Under Condition A, the assertions of Theorem 1 hold (with $\mathcal{A}_B$ replaced by $\mathcal{A}'_B$).

The proof of Theorem 2 is given in the Appendix. The convergence of $a_m(\omega)$ in the situation where the loss function is non-convex is studied in Shao (1988).

2.2. Convergence Rates and Asymptotic Distributions

The convergence rates and asymptotic distributions of $a_m(\omega)$ and $r_m(\omega)$ are obtained under some further regularity conditions on the loss function. The variances of the asymptotic distributions can be used as accuracy measures of $a_m(\omega)$ and $r_m(\omega)$ (see Section 3). Let

$L'(\theta, a)$ and $L''(\theta, a)$ be the gradient $\partial L(\theta, a)/\partial a$ and the Hessian mat $\partial^2 L(\theta, a)/\partial a \partial a^\tau$, respectively.

Condition B.

B1. A1 and A2 hold.

B2. For each θ, $L(\theta, a)$ is a twice continuously differentiable function of a in a neighborhood $\mathbf{N}(c)$ of a^*.

B3. $\int_\Theta \|L'(\theta, a^*)\|^2 w^2(\theta) h(\theta) d\mu < \infty$, where $w(\theta)$ is defined in (1.5).

B4. There is a measurable function $v(\theta)$ such that for any (k, l),
$\sup\{ |L''_{kl}(\theta, a)| : a \in \mathbf{N}(c) \} \le v(\theta)$ and $\int_\Theta v(\theta) f(x|\theta)\pi(\theta) d\mu < \infty$,
where $L''_{kl}(\theta, a)$ is the (k, l)th element of $L''(\theta, a)$.

B5. $L''(\theta, a^*)$ is positive definite for each θ.

Condition B5 implies that $\int L''(\theta, a^*) f(x|\theta)\pi(\theta) d\mu$ is positive definite. Under Condition B, A3 and A4 are satisfied and a^* is unique. Hence $a_m(\omega)$ converges to a^* *a.s.*

Theorem 3. Under Condition B, we have

$$a_m(\omega) - a^* = O(m^{-1/2}(\log\log m)^{1/2}) \quad a.s., \tag{2.4}$$

$$r(a_m(\omega)) - r(a^*) = O(m^{-1}\log\log m) \quad a.s. \tag{2.5}$$

and

$$m^{1/2}(a_m(\omega) - a^*) \to_d N(0, \Sigma), \tag{2.6}$$

where $\to_d$ denotes convergence in distribution and

$$\Sigma = \Sigma_2^{-1}\Sigma_1\Sigma_2^{-1} \tag{2.7}$$

with

$$\Sigma_1 = \int_\Theta [L'(\theta, a^*)][L'(\theta, a^*)]^\tau w^2(\theta) h(\theta) d\mu$$

and

$$\Sigma_2 = \int_\Theta L''(\theta, a^*) f(x|\theta)\pi(\theta) d\mu.$$

If

$$\sigma_r^2 = \int_\Theta [L(\theta, a^*) - r(a^*)]^2 p^2(\theta \mid x)/h(\theta) d\mu < \infty, \tag{2.8}$$

then

$$r_m(\omega) - r(a^*) = O(m^{-1/2}(\log\log m)^{1/2}) \quad a.s., \tag{2.9}$$

$$r_m(\omega) - r(a_m(\omega)) = O(m^{-1/2}(\log\log m)^{1/2}) \quad a.s., \tag{2.10}$$

$$m^{1/2}(r_m(\omega) - r(a^*)) \to_d N(0, \sigma_r^2) \tag{2.11}$$

and

$$m^{1/2}(r_m(\omega) - r(a_m(\omega))) \to_d N(0, \sigma_r^2). \tag{2.12}$$

The proof is given in the Appendix.

An important special case is the problem of estimating θ with loss $L(\theta, a) = \| \theta - a \|^2$, $\theta \in \Theta \subset \mathbf{R}^p$, $a \in \mathcal{A} \subset \mathbf{R}^p$. The Bayesian action is the posterior mean and

$$a_m(\omega) = \frac{m^{-1}\sum_{i=1}^m \theta_i f(x \mid \theta_i)\pi(\theta_i)/h(\theta_i)}{m^{-1}\sum_{i=1}^m f(x \mid \theta_i)\pi(\theta_i)/h(\theta_i)}.$$

A straightforward calculation shows that

$$\Sigma = \int_\Theta (\theta - a^*)(\theta - a^*)^\tau p^2(\theta \mid x)/h(\theta) d\mu$$

and

$$\sigma_r^2 = \int_\Theta [\| \theta - a^* \|^2 - r(a^*)]^2 p^2(\theta \mid x)/h(\theta) d\mu.$$

When $h(\theta)$ is the posterior density $p(\theta \mid x)$, Σ reduces to the posterior covariance matrix. The posterior expected loss of $a_m(\omega)$ in this case is

$$\begin{aligned} r(a_m(\omega)) &= r(a^*) + \| a_m(\omega) - a^* \|^2 \\ &= r(a^*) + O(m^{-1}\log\log m) \quad a.s. \end{aligned}$$

Conditions B3 and (2.8) suggest that the importance function $h(\theta)$ should be selected so that the tails of $h(\theta)$ are heavier than the tails of the posterior $p(\theta \mid x)$. Note that the posterior mean and variance of $L(\theta, a^*)$ are $r(a^*)$ and $\sigma^2 = \sigma_r^2$ with $h(\theta) = p(\theta \mid x)$, respectively. From Theorem 3, the standard error due to Monte Carlo approximation is the fraction $(\gamma m)^{-1/2}$ of σ,

where $\gamma=\sigma^2/\sigma_r^2$ is called the relative numerical efficiency (Geweke, 1988b). This may lead to a guide to choosing the importance function $h(\theta)$ and the Monte Carlo approximation sample size m. See the discussions in Geweke (1988a,b).

3. MEASURES OF ACCURACY

In practice, it is also desirable to indicate the precisions of $a_m(\omega)$ and $r_m(\omega)$ (as approximations to the Bayesian action and posterior expected loss of $a_m(\omega)$, respectively) by using some measures of accuracy. A nice feature of the Monte Carlo method is that accuracy estimates of $a_m(\omega)$ and $r_m(\omega)$ can be computed. With a method of estimating the precision of $a_m(\omega)$, the Monte Carlo sample size m can be chosen by using an iterative method. That is, starting with an initial m_0, we compute a_{m_j} and p_{m_j} (an estimate of the precision of a_{m_j}) for $m_j=m_0+jt$, where t is a step size and $j=0,1,2,...$ Stop if p_{m_j} is less than a predescribed small number. This method is used in Example 2 of Section 4.

We discuss two accuracy measures of the Monte Carlo approximations and study them in an illustrative example.

3.1. Asymptotic Variance Approach

The variances of the asymptotic distributions of $a_m(\omega)$ and $r_m(\omega)$ can be used as accuracy measures. Since in general, Σ (2.7) and σ_r^2 (2.8) do not have closed forms, they have to be approximated by the Monte Carlo approximations

$$V_m^a = V_2^{-1}V_1V_2^{-1}$$

and

$$V_m^r = m^{-1}\sum_{i=1}^m [L(\theta_i, a_m(\omega)) - r_m(\omega)]^2 w^2(\theta_i)/[\mathrm{M}_m(\omega)]^2,$$

respectively, where

$$V_1 = m^{-1}\sum_{i=1}^m [L'(\theta_i, a_m(\omega))][L'(\theta_i, a_m(\omega))]^\tau w^2(\theta_i)$$

and

$$V_2 = m^{-1}\sum_{i=1}^m L''(\theta_i, a_m(\omega))w(\theta_i).$$

The estimated asymptotic variances V_m^a/m and V_m^r/m can then be used as accuracy measures of $a_m(\omega)$ and $r_m(\omega)$, respectively. The following theorem proves the almost sure convergence

of these approximations.

Theorem 4. Assume the conditions in Theorem 3. If, in addition, there is a measurable function $u(\theta)$ such that

$$\sup\{ \|L'(\theta, a)\|^2 : a \in \mathbf{N}(c) \} \le u(\theta) \quad \text{and} \quad \int_{\Theta} u(\theta) w^2(\theta) h(\theta) d\mu < \infty, \tag{3.1}$$

then

$$V_m^a \to \Sigma \quad a.s. \quad \text{and} \quad V_m^r \to \sigma_r^2 \quad a.s.$$

Proof. The results follow from the convergence of a_m and Theorem 2 of Jennrich (1969). □

3.2. Independent Samples

In some situations (e.g., $\mathcal{A}$ is a finite set or $L(\theta, a)$ is not differentiable), the asymptotic variance approach does not apply. Another method for obtaining accuracy measures of $a_m(\omega)$ and $r_m(\omega)$ is as follows: we repeat the calculation of the Bayesian action (with independent sets of random θ's) and compute the sample variance. Suppose that $m=gk$ and we generate (from P_h) g independent sets of random variables $\omega_j=(\theta_{1j},...,\theta_{kj})$, $j=1,...,g$. Calculate the $a_k(\omega_j)$, $\rho_k(\omega_j)$, $r_k(\omega_j)$ according to (1.6), $j=1,...,g$, and the sample variances

$$S_{gk}^a = (g-1)^{-1}\sum_{j=1}^{g}(a_k(\omega_j)-\bar{a}_m)(a_k(\omega_j)-\bar{a}_m)^{\tau}, \quad \bar{a}_m = g^{-1}\sum_{j=1}^{g} a_k(\omega_j) \tag{3.2}$$

and

$$S_{gk}^r = (g-1)^{-1}\sum_{j=1}^{g}(r_k(\omega_j)-\bar{r}_m)^2, \quad \bar{r}_m = g^{-1}\sum_{j=1}^{g} r_k(\omega_j).$$

S_{gk}^a/g and S_{gk}^r/g are then used as accuracy measures of $a_m(\omega)$ and $r_m(\omega)$, respectively.

The motivation of the use of independent samples for assessing the accuracy of $a_m(\omega)$ and $r_m(\omega)$ are intuitively obvious. Unlike the asymptotic variance approach, this method does not require conditions B and (3.1). But it may require more computations if the computation of $a_m(\omega)$ involves a difficult minimization problem, since it needs to solve the same minimization problem g times.

Usually k should be chosen to be large and g may be chosen to be small or moderate (for computational saving). For example, if m=10,000, we may choose g=10 (k=1,000). Unless both g and k are large, the accuracy estimates obtained by using asymptotic variance approach and independent samples may be different (see Example 1).

When independent samples are used, to reduce the computation, we may use $\bar{a}_m$ defined in (3.2) as the Monte Carlo approximation to a^*. This is specially preferred if the computation of a_m is expensive. Although a_m and $\bar{a}_m$ are generally different, the difference is usually inappreciable for large k.

3.3. An Example

We study the accuracy estimates V_m^a and S_{gk}^a in an illustrative example where the exact values of a^* and the asymptotic variance Σ can be obtained and therefore we can assess the Monte Carlo approximations.

Example 1. (Berger, 1985, Section 4.3). Let $X \sim N(\theta,1)$, where θ is a measure of some positive quantity. In this case, it is difficult to estimate θ by using the classical approach. For example, the maximum likelihood estimate of θ is $\max(x, 0)$, which is unsuitable if the observation $x \le 0$. Under the noninformative prior $\pi(\theta) = I_{(0,\infty)}(\theta)$, the posterior mean of θ is

$$\mu_x = x + \phi(x)/\Phi(x),$$

where $\Phi(x)$ is the standard normal distribution function and $\phi(x) = \Phi'(x)$. In this case, no Monte Carlo approximation for μ_x is needed. To see the accuracy of the Monte Carlo approximations, ten independent sets (each of size m=5,000) of random variables are generated from the density $h(\theta) = e^{-\theta} I_{(0,\infty)}(\theta)$ and the Monte Carlo approximations a_m and $v_m = (V_m^a/m)^{1/2}$ are calculated and reported in Table 1 for x=1.5, 0.5 and −0.5. The exact values of the posterior mean μ_x and the asymptotic standard deviation $\sigma = (\Sigma/m)^{1/2}$ are included. Both a_m and v_m are quite accurate.

For comparison, we also include $s_m = (S_{gk}^a)^{1/2}$ (g=10, k=5,000) in Table 1. In all three cases, s_m is quite different from v_m (or σ), even though k=m=5,000. This is not surprise since g=10 is small.

3.4. Concluding Remarks

(1) When the loss function is differentiable, accuracy estimates based on asymptotic variances are recommended. The importance function $h(\theta)$ should be chosen so that conditions (2.8) and (3.1) are satisfied.

(2) When the asymptotic variance approach can not be applied, use independent samples to estimate the accuracy of the Monte Carlo approximations.

(3) In some situations, it may be appropriate to focus on relative accuracy measures such as $v_m^a/|a_m|$ and $s_m^a/|\bar{a}_m|$, where $v_m^a=(V_m^a/m)^{1/2}$ and $s_m^a=(S_{gk}^a/g)^{1/2}$ are asymptotic standard deviations.

4. NUMERICAL IMPLEMENTATION

In some situations, the minimization problem in Step 3 can be solved analytically, e.g., a^* is a posterior moment or quantile. However, there are situations where $a_m(\omega)$ can only be obtained numerically. There is a large body of literature on optimization and nonlinear programming which provides many efficient algorithms for solving $\min\{\rho_m(a,\omega): a\in\mathcal{A}\}$. See Avriel (1976) and Rao (1984), which also provide many other references.

As we discussed in Section 1, the same random $\{\theta_i\}$ should be used for approximating $\rho(a)$. If storage is not a problem, we may reduce computations by storing $\{w(\theta_i)\}$ and computing $\rho_m(a,\omega)=m^{-1}\sum_{i=1}^m L(\theta_i,a)w(\theta_i)$ for all needed a, where $w(\theta)$ is defined in (1.5).

In the following we use an example to illustrate the numerical implementation of the Monte Carlo method in practical problems.

Example 2. A company has developed a new type of product and must decide the number of the new products (denoted by a) to produce based on a marketing survey. Let θ be the unknown proportion of customers who will favor the new product.

(1) *Likelihood function.* A sample of 36 customers was interviewed and 11 of them indicated that they favor the new product. Hence the likelihood function is $\binom{36}{11}\theta^{11}(1-\theta)^{25}$.

(2) *Loss function.* Using the method introduced in Becker, DeGroot and Marschak (1964), DeGroot (1970, Chapter 7) and Berger (1985, Chapter 2), we obtain the company's loss function

$$L(\theta,a)=\begin{cases}-2{,}500{,}000ca+l & a\le c\theta/2\\ 10{,}562{,}500\theta^{-1}(8a/13-c\theta/2)^2-1{,}640{,}625c^2\theta+l & c\theta/2\le a\le c\theta\\ 1{,}500{,}000c(a-2c\theta)+l & a\ge c\theta\end{cases}$$

where $c=2\times10^5$ and $I=1{,}640{,}625c^2$. Here, $\theta\in\Theta=(0,1)$ and $a\in\mathcal{A}=\{$ all the integers $\le c\ \}$.

(3) *Prior distribution.* It is felt that $0.2<\theta<0.5$ is twice as likely as $\theta\le0.2$ or $\theta\ge0.5$ and there is no other information about θ available. Thus, we use the vague prior

$$(2/9)I_{(0.2,0.5)}(\theta) + (1/21)[1-I_{(0.2,0.5)}(\theta)].$$

Neither the posterior expected loss nor its Monte Carlo approximation has a closed form in this problem. Note that $\mathcal{A}$ contains finitely many actions. However, the number of actions is so large that it is too expensive to evaluate $\rho_m(a, \omega)$ for all actions. Since $L(\theta, a)$ is a convex function of a, we used the *golden section* method (see Avriel, 1976) with 24 evaluations of $\rho_m(a, \omega)$ to find the Monte Carlo approximation to the Bayesian action a^*.

To assess the accuracy of the Monte Carlo approximations, we used independent samples (Section 3.2) since the asymptotic variance approach is not applicable to this case. The following iterative method was used for selecting a Monte Carlo sample size: Compute $\bar{a}_{m_j}$ and $\bar{r}_{m_j}$ for $m_j = (k_0+jt)g$, $j=0,1,2,...$, with step size $t=1{,}000$, $g=10$ and initial $k_0=1{,}000$ ($m_0=10{,}000$). At each iteration, the relative accuracy measures $c_m^a = s_m^a/\bar{a}_m$ and $c_m^r = s_m^r/\bar{r}_m$ were computed, where $s_m^a = (S_{gk}^r/g)^{1/2}$ and $s_m^r = (S_{gk}^a/g)^{1/2}$. Stop if both $c_{m_j}^a$ and $c_{m_j}^r$ are less than 0.001.

The importance function was chosen to be proportional to the likelihood function. The random θ_i's were generated by using the IMSL subroutine GGBTR. The computation was done (after four iterations) on a VAX 11/780 (Unix 4.3BSD) at Purdue University. The total CPU time used is about 1.96 minutes.

The selected Monte Carlo sample size is $m=40{,}000$ and the resulting Monte Carlo approximations to a^* and $r(a^*)$ are 51,223 and 459,861.09, respectively. Table 2 shows the accuracy measures and the results from four iterations.

APPENDIX

Proof of Theorem 1. The existence of a Bayesian action is guaranteed by the continuity of $r(a)$. From Theorem 2 of Jennrich (1969), for almost every ω,

$$\rho_m(a, \omega) \to \rho(a) \quad \text{uniformly in } a\in\mathcal{A}.$$

Suppose that $a'\in A_\omega$. Then by compactness of $\mathcal{A}$, $a'\in\mathcal{A}$. Consequently, there is a sequence $\{ m_j \}_{j=1}^{\infty}$ such that $a_{m_j}(\omega) \to a'$ and $\rho_{m_j}(a_{m_j}(\omega), \omega) \to \rho(a')$. Since for $a^*\in\mathcal{A}_B$,

$\rho_{m_j}(a_{m_j}(\omega), \omega) \le \rho_{m_j}(a^*) \to \rho(a^*)$, $\rho(a') \le \rho(a^*)$ and therefore $a' \in \mathcal{A}_B$. Thus, $A_\omega \subset \mathcal{A}_B$.

If a^* is the unique Bayesian action, then $a_m(\omega) \to a^*$, $\rho(a_m(\omega)) \to \rho(a^*)$ and $\rho_m(a_m(\omega), \omega) \to \rho(a^*)$. Hence (2.2) and (2.3) hold since $M_m \to M(x)$ *a.s.*

In the case where a^* is not unique, let ρ' be a limit point of $\{ \rho_m(a_m(\omega), \omega) \}_{m=1}^{\infty}$. Then there is a sequence $\{ m_j \}_{j=1}^{\infty}$ such that $\rho_{m_j}(a_{m_j}(\omega), \omega) \to \rho'$. From the compactness of $\mathcal{A}$, there is a subsequence $\{ m_l \}_{l=1}^{\infty} \subset \{ m_j \}_{j=1}^{\infty}$ such that $a_{m_l}(\omega) \to a'$. By the above proof, $\rho_{m_l}(a_{m_l}(\omega), \omega) \to \rho(a') = \rho(a^*)$. Thus, $\rho' = \rho(a^*)$ and

$$r_m(\omega) = \rho_m(a_m(\omega), \omega) \to \rho(a^*).$$

Similarly, we can show that $\rho(a_m(\omega)) \to \rho(a^*)$ and therefore (2.2) and (2.3) hold. □

Proof of Theorem 2. From the proof of Theorem 1, the result follows if for almost all ω, there exists an m_ω such that

$$a_m(\omega) \in \mathbf{N}(c) \quad \text{for } \textit{all } m \ge m_\omega. \tag{A1}$$

Let ε be such that $\varepsilon > 0$ and $\inf \{ \rho(a): a \in \mathbf{B}(c) \} > \rho(a^*) + 2\varepsilon$. From Condition A and Theorem 2 of Jennrich (1969), for almost all ω, there is an m_ω such that when $m \ge m_\omega$,

$$|\rho_m(a, \omega) - \rho(a)| < \varepsilon \quad \text{for } \textit{all } a \in \mathbf{N}(c). \tag{A2}$$

Let $b_m(\omega)$ be such that

$$\rho_m(b_m(\omega), \omega) = \min\{ \rho_m(a, \omega): a \in \mathbf{N}(c) \}. \tag{A3}$$

For a fixed $m \ge m_\omega$, suppose that $\| a_m(\omega) - a^* \| > c$. Then there exist an $a_m^*(\omega) \in \mathbf{B}(c)$ and a λ such that $0 < \lambda \le 1$ and $a_m^*(\omega) = \lambda b_m(\omega) + (1-\lambda) a_m(\omega)$. By the convexity of $L(\theta, a)$,

$$\begin{aligned} \rho_m(a_m^*(\omega), \omega) &\le \lambda \rho_m(b_m(\omega), \omega) + (1-\lambda) \rho_m(a_m(\omega), \omega) \\ &\le \rho_m(b_m(\omega), \omega) \le \rho_m(a_m^*(\omega), \omega), \end{aligned}$$

i.e., $\rho_m(b_m(\omega), \omega) = \rho_m(a_m^*(\omega), \omega)$. However, from (A2),

$$\begin{aligned} \rho_m(a^*, \omega) &< \rho(a^*) + \varepsilon < \inf \{ \rho(a): a \in \mathbf{B}(c) \} - \varepsilon \\ &\le \rho(a_m^*(\omega)) - \varepsilon < \rho_m(a_m^*(\omega), \omega) = \rho_m(b_m(\omega), \omega), \end{aligned}$$

which is contrary to (A3). Thus, (A1) holds. □

Proof of Theorem 3. By Theorem 2, we may focus on the event $\{ a_m(\omega) \in \mathbf{N}(c) \}$. Under B2-B4, $\rho'(a) = \int_\Theta L'(\theta, a) f(x \mid \theta) \pi(\theta) d\mu$ is continuous on $\mathbf{N}(c)$ and

$$\rho'(a^*) = \int_\Theta L'(\theta, a^*) f(x \mid \theta) \pi(\theta) d\mu = 0. \tag{A4}$$

Note that $\sum_{i=1}^m L'(\theta_i, a_m(\omega)) w(\theta_i) = 0$ under B2. By the mean value theorem and (A4),

$$m^{-1} \sum_{i=1}^m L'(\theta_i, a^*) w(\theta_i) = -[m^{-1} \sum_{i=1}^m L''(\theta_i, \tilde{a}_{m1}) w(\theta_i)](a_m(\omega) - a^*), \tag{A5}$$

$$\rho(a_m(\omega)) - \rho(a^*) = (a_m(\omega) - a^*)^\tau [\rho''(\tilde{a}_{m2})](a_m(\omega) - a^*) \tag{A6}$$

and

$$r_m(\omega) - \rho_m(a^*, \omega) = m^{-1} [\sum_{i=1}^m L'(\theta_i, \tilde{a}_{m3})]^\tau (a_m(\omega) - a^*), \tag{A7}$$

where $\tilde{a}_{mj}$, $j = 1,2,3$, are on the line segment between a^* and $a_m(\omega)$. From B4 and Theorem 2,

$$m^{-1} \sum_{i=1}^m L''(\theta_i, \tilde{a}_{m1}) w(\theta_i) \to \Sigma_2 \quad a.s. \tag{A8}$$

Hence we obtain (2.4) by applying the law of iterated logarithm to each component of the left hand side of (A5). Also, (2.5) follows from (2.5), (A6), the continuity of $\rho''(a)$ and $\mathrm{M}_m \to \mathrm{M}(x)$ *a.s.* From the central limit theorem, (2.6) follows from (A5), (A8) and Theorem 2.

If σ_r^2 is finite, by the law of iterated logarithm,

$$r_m(a^*, \omega) - r(a^*) = O(m^{-1/2} (\log\log m)^{1/2}) \quad a.s. \tag{A9}$$

where $r_m(a^*, \omega) = \rho_m(a^*, \omega) / \mathrm{M}_m(\omega)$. From B2-B4, (A4) and Theorem 2 of Jennrich (1969),

$$m^{-1} \sum_{i=1}^m L'(\theta_i, \tilde{a}_{m3}) \to 0 \quad a.s. \tag{A10}$$

Hence (2.9) follows from (2.4), (A7), (A9) and (A10), and (2.10) follows from (2.5) and (2.9). From the central limit theorem and $\mathrm{M}_m \to \mathrm{M}(x)$ *a.s.*,

$$m^{1/2} [r_m(a^*, \omega) - r(a^*)] \to_d N(0, \sigma_r^2). \tag{A11}$$

Then (2.11) follows from (A7), (A10), (A11) and (2.6). Finally, (2.12) follows from (2.5) and (2.11). □

REFERENCES

Avriel, M. (1976). *Nonlinear Programming: Analysis and Methods*. Prentice-Hall, New Jersey.

Becker, G. M., DeGroot, M. H. and Marschak, J. (1964). Measuring utility by a single-response sequential method. *Behavioral Sci.* **9**, 226-232.

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*, second edition. Spring-Verlag, New York.

DeGroot, M. H. (1970). *Optimal Statistical Decisions*. McGraw-Hill, New York.

Geweke, J. (1986). Exact inference in the inequality constrained normal linear regression model. *J. Applied Econometrics* **1**, 127-142.

Geweke, J. (1988a). Antithetic acceleration of Monte Carlo integration in Bayesian inference. *J. Econometrics* **38**, 73-90.

Geweke, J. (1988b). Bayesian inference in econometric models using Monte Carlo integration. *Econometrica*, to appear.

Jennrich, R. I. (1969). Asymptotic properties of non-linear least squares estimators. *Ann. Math. Statist.* **40**, 633-643.

Kloek, T. and van Dijk, H. K. (1978). Bayesian estimates of equation system parameters; an application of integration by Monte Carlo. *Econometrica* **46**, 1-19.

Rao, S. S. (1984). *Optimization: Theory and Applications*, second edition. Wiley, New York.

Shao, J. (1988). Monte Carlo approximations in Bayesian decision theory III. Technical Report #88-64C, Statistics Department, Purdue University.

Stewart, L. (1979). Multiparameter univariate Bayesian analysis. *J. Amer. Statist. Assoc.* **74**, 684-693.

Stewart, L. (1983). Bayesian analysis using Monte Carlo integration — a powerful methodology for handling some difficult problems. *The Statistician* **32**, 195-200.

Table 1
Monte Carlo approximations to μ_x and σ (Example 1)
($h(\theta) = e^{-\theta} I_{(0,\infty)}(\theta)$, m=5,000)

	$x = 1.5$ $\mu_x = 1.6388$ $\sigma = 0.015230$		$x = 0.5$ $\mu_x = 1.0091$ $\sigma = 0.009418$		$x = -0.5$ $\mu_x = 0.6412$ $\sigma = 0.006580$	
j	a_m	v_m	a_m	v_m	a_m	v_m
1	1.6540	0.015389	1.0063	0.009581	0.6310	0.006559
2	1.6307	0.014763	1.0271	0.009275	0.6602	0.006682
3	1.6149	0.015108	1.0033	0.009359	0.6409	0.006526
4	1.6339	0.015107	1.0118	0.009407	0.6443	0.006576
5	1.6325	0.015632	0.9986	0.009285	0.6421	0.006471
6	1.6573	0.015331	1.0114	0.009530	0.6378	0.006605
7	1.6915	0.015439	1.0219	0.009605	0.6382	0.006653
8	1.6595	0.015293	1.0173	0.009413	0.6483	0.006598
9	1.6186	0.015091	1.0070	0.009333	0.6430	0.006584
10	1.6555	0.015415	1.0054	0.009512	0.6330	0.006595
mean	1.6448	0.015257	1.0110	0.009430	0.6419	0.006585
	$s_m = 0.021902$		$s_m = 0.008367$		$s_m = 0.007819$	

Table 2
Results from four iterations (Example 2)

j	m	a_m	s_m^a	c_m^a	r_m	s_m^r	c_m^r
0	10,000	51,137	75.15	0.0015	460,042	211.39	0.0005
1	20,000	51,195	74.87	0.0014	460,068	220.82	0.0005
2	30,000	51,195	60.35	0.0012	460,037	169.24	0.0004
3	40,000	51,223	40.42	0.0008	459,861	122.52	0.0003

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION / AVAILABILITY OF REPORT Approval for public release, distribution unlimited.
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) Technical Report #88-39	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Purdue University 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION
6c. ADDRESS (City, State, and ZIP Code) Department of Statistics West Lafayette, IN 47907	7b. ADDRESS (City, State, and ZIP Code)
8a. NAME OF FUNDING / SPONSORING ORGANIZATION Office of Naval Research 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER N00014-88-K-0170 and NSF Grant DMS-8717799 DMS-8702620
8c. ADDRESS (City, State, and ZIP Code) Arlington, VA 22217-5000	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. / PROJECT NO. / TASK NO. / WORK UNIT ACCESSION NO

11. TITLE (Include Security Classification)
MONTE CARLO APPROXIMATIONS IN BAYESIAN DECISION THEORY

12. PERSONAL AUTHOR(S)
Jun Shao

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Technical	FROM _____ TO _____	August 1988 (Revised 1/89)	19

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Bayesian action; posterior expected loss; Monte Carlo integration; Almost sure convergence; Accuracy measures.

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

In decision-making problems, the Bayesian action and its posterior expected loss usually can not be obtained analytically. In this paper we study a Monte Carlo method for approximating the Bayesian action and its posterior expected loss. The Monto Carlo approximation to the Bayesian action is obtained through (1) approximating the posterior expected loss function by using the Monte Carlo integration method and (2) searching the minimum of the approximated posterior expected loss function. As the Monte Carlo sample size diverges to infinity, the Monte Carlo approximations are shown to be convergent in very general situations. The rates of the convergence are also obtained under some regularity conditions on the loss function. Two accuracy measures of the Monte Carlo approximations are proposed. Some examples are presented for illustration.

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT ☐ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL Jun Shao	22b. TELEPHONE (Include Area Code) (317) 494-6039 22c. OFFICE SYMBOL

DD FORM 1473, 84 MAR 83 APR edition may be used until exhausted. All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED