MONTE CARLO APPROXIMATIONS IN BAYESIAN DECISION THEORY PART III: LIMITING BEHAVIOR OF MONTE CARLO APPROXIMATIONS *

by

Jun Shao
Purdue University

Technical Report #88-65C

Department of Statistics
Purdue University

December 1988

* The research of this author was partially supported by the Office of Naval Research Contract N00014-88-K-1070 and NSF Grant DMS-8717799, DMS-8702620 at Purdue University.

★

MONTE CARLO APPROXIMATIONS IN BAYESIAN DECISION THEORY PART III: LIMITING BEHAVIOR OF MONTE CARLO APPROXIMATIONS

Jun Shao *

Purdue University

ABSTRACT

Monte Carlo approximation is a useful method in obtaining a numerical approximation to a Bayesian action (an action which minimizes the posterior expected loss). We study the behavior of the Monte Carlo approximation when the Monte Carlo sample size is large. Convergence and convergence rate of the Monte Carlo approximation are established under some weak conditions on the loss function.

Keywords: almost sure convergence; convergence rate; loss function; posterior expected loss.

* 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

Monte Carlo integration (Hammersley and Handscomb, 1964) is a very useful method for numerical calculation in Bayesian analysis when the Bayesian solution (action) of the problem can not be obtained analytically. Unlike other numerical methods, the use of Monte Carlo method does not require restrictive conditions such as the dimension of the parameter space is low (say one or two) and the total number of sample observations is large. Applications of this method in Bayesian analysis can be found in Stewart and Johnson (1972), Kloek and van Dijk (1978), Stewart (1979), van Dijk and Kloek (1980), Zellner and Rossi (1984), Bauwens and Richard (1985), Geweke (1988), and Berger and Deely (1988).

Let θ be a parameter of interest, $\theta \in \Theta \subset \mathbf{R}^k$, $l_x(\theta)$ be the likelihood function based on the observed data x (an n-vector), and $\Pi(\theta)$ be a prior distribution. The posterior distribution is then

$$P_x(\theta) = \int_{S(\theta)} l_x(\theta) d\Pi(\theta) / \mathrm{M}_x,$$

where $S(\theta) = (-\infty, \theta^{(1)}] \times (-\infty, \theta^{(2)}] \times \cdots \times (-\infty, \theta^{(k)}]$, $\theta^{(j)}$ is the jth component of θ and $\mathrm{M}_x = \int l_x(\theta) d\Pi(\theta)$. Let $\mathcal{A}$ denote the collection of all possible actions we may take for a problem under consideration (e.g., $\mathcal{A} = \Theta$ in the problem of estimating θ). $\mathcal{A}$ is assumed to be a subset of $\mathbf{R}^p$. Let $L(\theta, a) \geq 0$ be the loss incurred when the action a is taken and θ is the true parameter. A Bayesian solution of the problem is an action a^* which minimizes the posterior expected loss

$$r(a) = \int L(\theta, a) dP_x(\theta).$$

Since M_x is fixed for given x, a^* is a solution of

$$\rho(a^*) = \min_{a \in \mathcal{A}} \rho(a),$$

where

$$\rho(a) = \int L(\theta, a) l_x(\theta) d\Pi(\theta).$$

The solution a^* is referred to as Bayesian action in the literature. Note that a^* may not be unique. Only in special cases a^* can be obtained analytically.

The numerical approximation to a^* using Monte Carlo method is obtained as follows. Select a distribution $H(\theta)$ such that the Radon-Nikodym derivative $\frac{d\Pi}{dH}(\theta)$ exists and it is easy to generate a random θ from H. Let $\{ \theta_i, i=1,...,m \}$ be m independent and identically distributed (i.i.d.) random k-vectors generated from $H(\theta)$. Approximate $\rho(a)$ by

$$\rho_m(a) = \frac{1}{m}\sum_{i=1}^{m} L(\theta_i, a) w(\theta_i), \tag{1.1}$$

where

$$w(\theta) = l_x(\theta)\frac{d\Pi}{dH}(\theta).$$

The Monte Carlo approximation to a^* is an action a_m satisfying

$$\rho_m(a_m) = \min_{a \in \mathcal{A}} \rho_m(a). \tag{1.2}$$

This approach is motivated by the fact that for any fixed a,

$$\lim_{m\to\infty}\rho_m(a) = \int L(\theta, a) w(\theta) dH(\theta) = \rho(a)$$

for almost all $\theta_1, \theta_2, \ldots$ (with respect to the probability distribution H), according to the strong law of large numbers (SLLN).

A theoretical justification of the use of this Monte Carlo method is the convergence of the approximation a_m to a Bayesian action a^* in some sense. In some simple cases, such as $L(\theta, a)$ is the squared error loss (in an estimation problem) or the action space $\mathcal{A}$ is a compact subset of $\mathbf{R}^p$, the convergence of a_m is a direct consequence of the SLLN or uniform SLLN. Shao (1988) proved the almost sure convergence of a_m in the situation where $\mathcal{A}$ is non-compact but the loss function is convex in a. There are some important examples of convex loss functions. Also, convex loss usually ensures the uniqueness of the Bayesian action. However, reasonable loss functions derived through utility analyses are often not convex but bounded and concave for large errors (see Berger, 1985, Chapter 2). For a convex loss, large errors are penalized much too severely. In addition, a bounded loss function usually provides a robust Bayesian solution to the problem (see Section 2).

In this note we study the limiting behavior of the Monte Carlo approximation a_m for general unbounded $\mathcal{A}$ and non-convex loss functions. The convergence of a_m is studied in Section 3 for two large classes of loss functions introduced in Section 2. The rate of convergence and the asymptotic distribution of a_m (which provides an accuracy measure for the Monte Carlo approximation) are obtained in Section 4.

Throughout the paper we assume that x is a fixed data vector, $0<\text{M}_x<\infty$, $\mathcal{A}$ is a closed subset of $\mathbf{R}^p$, $\rho(a)$ is finite for any $a \in \mathcal{A}$ and a Bayesian action a^* exists, and H is a selected distribution for generating random θ_i. Discussions for the selection of the distribution H can be found in Berger (1985, Section 4.9) and Geweke (1988).

2. Preliminaries

We first consider the modes of convergence of a_m as $m \to \infty$. Let ω denote a particular sequence $(\theta_1, \theta_2, \ldots)$ and $a_m(\omega)$ denote the corresponding a_m for fixed ω. Since a_m is random, we may consider the almost sure convergence: for almost all ω (with respect to H),

$$a_m(\omega) \to a^*. \tag{2.1}$$

However, unless there is a unique Bayesian action, (2.1) usually does not hold. For practical uses, $a_m(\omega)$ might be considered as a good approximation as long as $a_m(\omega)$ is close to a Bayesian action a^*. That is, if

$$\mathcal{A}^* = \{\ a^* : a^* \text{ is a Bayesian action } \}$$

and

$$\mathcal{A}_\omega = \{\ a : a \text{ is a limit point of } \{\ a_m(\omega), m=1,2,\ldots\ \}\ \},$$

then for almost all ω,

$$\mathcal{A}_\omega \subset \mathcal{A}^*. \tag{2.2}$$

Another way is to consider the posterior expected losses $r(a_m)$ and $r(a^*)$, since the posterior expected loss is used to judge the performance of an action. Note that $r(a^*)$ is uniquely defined although a^* may not. Denote $r(a^*)$ and $\rho(a^*)$ by r^* and ρ^*, respectively. It is desired to show that for almost all ω,

$$r(a_m(\omega)) \to r^*,$$

which is equivalent to (since $\rho(a) = \text{M}_x r(a)$)

$$\rho(a_m(\omega)) \to \rho^*. \tag{2.3}$$

Usually $\rho(a)$ is continuous in a. Then (2.3) is weaker than (2.1).

The following result relates (2.2) to (2.3) and the boundedness of $a_m(\omega)$:

$$\| a_m(\omega) \| \leq C_\omega \quad \text{for all } m, \tag{2.4}$$

where $C_\omega > 0$ is a constant for each ω, $\| a \| = (a^\tau a)^{1/2}$ and a^τ is the transpose of a.

Lemma 1. Let ω be fixed. Suppose that $\rho(a)$ is continuous and that

$$\liminf_{\| a \| \to \infty} \rho(a) > \rho^*. \tag{2.5}$$

Then (2.3) is equivalent to (2.2) and (2.4).

Proof. Suppose that (2.3) holds. If (2.4) does not hold, then there is a subsequence $\{ a_{m_j}(\omega), j=1,2,... \}$ such that

$$\lim_{j \to \infty} \| a_{m_j}(\omega) \| = \infty.$$

From condition (2.5), this implies

$$\lim_{j \to \infty} \rho(a_{m_j}(\omega)) > \rho^*,$$

which contradicts (2.3). Hence (2.4) holds. Let $c \in \mathcal{A}_\omega$. Then there is a subsequence $\{ a_{m_l}(\omega), l=1,2,... \}$ such that $\lim_{l \to \infty} a_{m_l}(\omega) = c$. From the continuity of ρ,

$$\lim_{l \to \infty} \rho(a_{m_l}(\omega)) = \rho(c).$$

From (2.3), $\rho(c) = \rho^*$. Hence $c \in \mathcal{A}^*$ and therefore (2.2) holds.

Suppose now (2.2) and (2.4) hold. Let η be any limit point of $\{ \rho(a_m(\omega)), m=1,2,... \}$. Then there is a subsequence $\{ m_j \}$ such that

$$\lim_{j \to \infty} \rho(a_{m_j}(\omega)) = \eta.$$

From (2.2) and (2.4), there is a subsequence $\{ m_l \} \subset \{ m_j \}$ such that

$$\lim_{l \to \infty} a_{m_l}(\omega) = a^* \in \mathcal{A}^*.$$

From the continuity of ρ, $\eta = \rho^*$. This proves (2.3). □

Berger (1985) pointed out that reasonable loss functions are usually bounded. Consider loss functions satisfying the following condition:

Condition (L1).

(1) $L(\theta, a)$ is continuous in a and $\sup_{\theta,a} L(\theta, a) < \infty$.

(2) For any constant $C>0$, $\lim_{\|a\| \to \infty} L(\theta, a) = A$ uniformly for all θ satisfying $\|\theta\| \leq C$, where A is a fixed constant.

A simple example of a loss function satisfying (L1) is

$$L(\theta, a) = \min(\|\theta - a\|^2, A),$$

where A is a constant. Note that A usually is an upper bound of the loss function. Hence for a reasonable loss function satisfying (L1), it is usually true that $A > r^*$.

The following result shows condition (L1) implies the conditions in Lemma 1.

Lemma 2. Assume (L1). Then $\rho(a)$ is continuous and (2.5) holds if $A > r^*$.

Proof. It is obvious that the continuity and boundedness of L imply the continuity of ρ. For (2.5), it suffices to show that

$$\lim_{\|a\| \to \infty} r(a) = A. \tag{2.6}$$

For any $\varepsilon>0$, since $M_x = \int l_x(\theta) d\Pi(\theta)$ is finite, there exists a constant $C>0$ such that

$$\int_{\|\theta\|>C} l_x(\theta) d\Pi(\theta) < \varepsilon.$$

For this $C>0$, under condition (L1), there exists a $K>0$ such that when $\|a\| > K$,

$$\int_{\|\theta\| \leq C} |L(\theta, a) - A| l_x(\theta) d\Pi(\theta) < \varepsilon.$$

Hence for $\|a\| > K$,

$$\int |L(\theta, a) - A| l_x(\theta) d\Pi(\theta) < (A+B+1)\varepsilon,$$

where $B = \sup_{\theta,a} L(\theta, a)$. This proves (2.6) since ε is arbitrary. □

It can be shown that when the loss function satisfies (L1) and the Bayesian action a^* is unique, the Bayesian action is robust in the sense that for any sequence of posteriors $\{ G_n \}$ converging weakly to P_x, we have $a_n \to a^*$, where a_n is a Bayesian action corresponding to G_n. We will not discuss this issue here.

An unbounded loss function usually satisfies $\lim_{\|a\|\to\infty} L(\theta, a) = \infty$ for fixed θ. We consider the loss functions satisfying the following condition.

Condition (L2).

(1) $L(\theta, a)$ is continuous in a and for any $C>0$, there is a function $B_C(\theta)$ such that $\sup_{\|a\|\le C} L(\theta, a) \le B_C(\theta)$ and $\int B_C(\theta) l_x(\theta) d\Pi(\theta) < \infty$.

(2) There is a constant C_0 such that $\int_{\|\theta\|\le C_0} dP_x(\theta) > 0$ and $\lim_{\|a\|\to\infty} L(\theta, a) = \infty$ holds uniformly for θ satisfying $\|\theta\| \le C_0$.

From the dominated convergence theorem, the first condition in (L2) implies the continuity of $\rho(a)$. The second condition in (L2) implies (2.5), as the following lemma shows.

Lemma 3. Assume the second condition in (L2). Then

$$\lim_{\|a\|\to\infty} r(a) = \infty$$

and therefore (2.5) holds.

Proof. For any $K > 0$, there is a $K_1 > 0$ such that for any a with $\|a\| > K_1$,

$$\inf_{\|\theta\|\le C_0} L(\theta, a) > K.$$

Then

$$r(a) \ge \int_{\|\theta\|\le C_0} L(\theta, a) dP_x(\theta) \ge K \int_{\|\theta\|\le C_0} dP_x(\theta).$$

The result follows since K is arbitrary and $\int_{\|\theta\|\le C_0} dP_x(\theta) > 0$. □

We also need the following technical lemma for the proof of the main results.

Lemma 4. Let $g(\theta, a)$ be a function on $\mathbf{R}^k \times \mathbf{R}^p$ and F be a distribution function on $\mathbf{R}^k$. Suppose that for fixed a, g is measurable and for fixed θ, g is continuous. Suppose also that for any $C > 0$, there is a function $B_C(\theta)$ such that $\sup_{\|a\|\le C} |g(\theta, a)| \le B_C(\theta)$ and $\int B_C(\theta) dF(\theta) < \infty$. Let $\theta_1, \ldots, \theta_m$ be i.i.d. samples from F. Then for almost all $\theta_1, \theta_2, \ldots,$

$$\sup_{\|a\| \le C} \left| \frac{1}{m} \sum_{i=1}^{m} g(\theta_i, a) - \int g(\theta, a) dF(\theta) \right| \to 0 \quad \text{for all positive rational } C. \tag{2.7}$$

Proof. For any fixed C, from Theorem 2 of Jennrich (1969),

$$\sup_{\|a\| \le C} \left| \frac{1}{m} \sum_{i=1}^{m} g(\theta_i, a) - \int g(\theta, a) dF(\theta) \right| \to 0$$

holds for almost all $\theta_1, \theta_2, \ldots$ Then (2.7) follows from the fact that the set of all rational numbers is countable.

3. Convergence of Monte Carlo approximations

For loss functions satisfying either condition (L1) or (L2), the convergence of a_m (in the sense of (2.2) and (2.3)) is established in the following theorems.

Theorem 1. Assume that $L(\theta, a)$ satisfies condition (L1) with $A > r^*$. Then (2.2) and (2.3) hold for almost all ω (with respect to H).

Proof. From Lemma 2, the conditions of Lemma 1 are satisfied. Using Lemma 1, we only need to show (2.2) and (2.4).

Consider (2.4) first. Note that $\int w(\theta) dH(\theta) = \int l_x(\theta) d\Pi(\theta) = M_x < \infty$. From the SLLN, for almost all ω,

$$\frac{1}{m} \sum_{i=1}^{m} w(\theta_i) I_{(\|\theta_i\| > C)} \to \int_{\|\theta\| > C} w(\theta) dH(\theta) \quad \text{for all positive rational } C, \tag{3.1}$$

where I_S is the indicator function of the set S. Let a^* be a Bayesian action. From (1.1) and the SLLN, for almost all ω,

$$\rho_m(a^*) \to \rho^*. \tag{3.2}$$

Let ω be fixed such that (3.1) and (3.2) hold. Suppose that (2.4) does not hold. Then there is a subsequence of $a_m(\omega)$ diverging to infinity. Without loss of generality, assume that $\|a_m(\omega)\| \to \infty$. For any $\varepsilon > 0$, there is a rational $C > 0$ such that

$$\int_{\|\theta\| > C} w(\theta) dH(\theta) < \varepsilon.$$

From condition (L1), for this $C > 0$, there is an $N_\omega > 0$ such that for all $\|\theta\| \le C$ and $m > N_\omega$,

$$|L(\theta, a_m(\omega)) - A| < \varepsilon.$$

Then

$$\frac{1}{m}\sum_{i=1}^{m} |L(\theta_i, a_m(\omega)) - A| w(\theta_i) \le \frac{\varepsilon}{m}\sum_{i=1}^{m} w(\theta_i) + \frac{(A+B)}{m}\sum_{i=1}^{m} w(\theta_i) I_{(\|\theta_i\| > C)},$$

where $B = \sup_{\theta,a} L(\theta, a)$. From (3.1),

$$\limsup_{m\to\infty} \frac{1}{m}\sum_{i=1}^{m} |L(\theta_i, a_m(\omega)) - A| w(\theta_i) \le (A+B+\mathrm{M}_x)\varepsilon$$

and therefore

$$\rho_m(a_m(\omega)) \to \mathrm{M}_x A. \tag{3.3}$$

But from (1.2),

$$\rho_m(a_m(\omega)) \le \rho_m(a^*)$$

for all m. Hence from (3.2),

$$\limsup_{m\to\infty} \rho_m(a_m(\omega)) \le \rho^*.$$

This contradicts (3.3) since $\mathrm{M}_x A > \rho^*$. Hence (2.4) holds for almost all ω.

From Lemma 4, for almost all ω,

$$\sup_{\|a\| \le C} |\rho_m(a) - \rho(a)| \to 0 \quad \text{for all positive rational } C. \tag{3.4}$$

Let ω be fixed such that (2.4) and (3.4) hold and (3.2) holds for an $a^* \in \mathcal{A}^*$. Then

$$|\rho_m(a_m(\omega)) - \rho(a_m(\omega))| \to 0. \tag{3.5}$$

Let $a_1 \in \mathcal{A}_\omega$. Then $\|a_1\| \le C_\omega$ and there is a subsequence $\{ m_j \}$ such that

$$\lim_{j\to\infty} a_{m_j}(\omega) = a_1.$$

From the continuity of ρ and (3.5), we have

$$\lim_{j\to\infty} \rho_{m_j}(a_{m_j}(\omega)) = \rho(a_1).$$

But

$$\rho_{m_j}(a_{m_j}(\omega)) \le \rho_{m_j}(a^*),$$

which converges to ρ^* by (3.2). Hence $\rho(a_1) = \rho^*$ and $a_1 \in \mathcal{A}^*$. This proves (2.2). □

Theorem 2. Assume that $L(\theta, a)$ satisfies condition (L2). Then (2.2) and (2.3) hold for almost all ω.

Proof. From Lemma 3, the conditions of Lemma 1 are satisfied and therefore we only need to show (2.2) and (2.4). Note that for almost all ω,

$$\lim_{m\to\infty}\frac{1}{m}\sum_{i=1}^{m} w(\theta_i)I_{(\|\theta_i\|\le C_0)} = \int_{\|\theta\|\le C_0} w(\theta)dH(\theta). \tag{3.6}$$

For a fixed ω such that (3.2) and (3.6) hold, we have

$$\begin{aligned}
\liminf_{m\to\infty}\rho_m(a_m(\omega)) &\ge \liminf_{m\to\infty}\frac{1}{m}\sum_{i=1}^{m} L(\theta_i, a_m(\omega))w(\theta_i)I_{(\|\theta_i\|\le C_0)} \\
&\ge \liminf_{m\to\infty}[\inf_{\|\theta\|\le C_0} L(\theta, a_m(\omega))\frac{1}{m}\sum_{i=1}^{m} w(\theta_i)I_{(\|\theta_i\|\le C_0)}] \\
&\ge \liminf_{m\to\infty}\inf_{\|\theta\|\le C_0} L(\theta, a_m(\omega))\int_{\|\theta\|\le C_0} w(\theta)dH(\theta)
\end{aligned}$$

and

$$\limsup_{m\to\infty}\rho_m(a_m(\omega)) \le \lim_{m\to\infty}\rho_m(a^*) = \rho^*.$$

From condition (L2), $\lim_{\|a\|\to\infty}\inf_{\|\theta\|\le C_0} L(\theta, a) = \infty$. Hence $\{ a_m(\omega) \}$ can not have any subsequence diverging to infinity. Therefore (2.4) holds for almost all ω.

From Lemma 4, (3.4) holds under condition (L2). Then the proof of (2.2) is the same as that in the proof of Theorem 1. This completes the proof of Theorem 2. □

From (2.2), if the Bayesian action is unique, then a_m converges to a^* in the ordinary sense.

Corollary 1. Assume the conditions of Theorem 1 or Theorem 2. If the Bayesian action is unique, then (2.1) holds for almost all ω.

In estimation problems, it is desired to indicate the accuracy of the Bayes estimate a^*. The posterior expected loss of a^*, of course, can be used as an accuracy measure. Let $r_m = m^{-1}\sum_{i=1}^{m} L(\theta_i, a_m)w(\theta_i)$. r_m can be used to approximate the posterior expected loss r^*. The following result is a direct consequence of (3.5) and Theorem 1 (or Theorem 2).

Corollary 2. Assume the conditions of Theorem 1 or Theorem 2. Then for almost all ω,

$$r_m(\omega) \to r^*.$$

4. Convergence rate and limiting distribution

We study the convergence rates of $a_m(\omega)$ and $r(a_m(\omega))$ for differentiable loss functions. Let $\nabla L(\theta, a) = \partial L(\theta, a)/\partial a$, $\nabla^2 L(\theta, a) = \partial^2 L(\theta, a)/\partial a \partial a^\tau$ and $g_{uv}(\theta, a)$ be the (u,v)th element of $\nabla^2 L(\theta, a)$, $1 \le u, v \le p$.

Condition (L3).

(1) For almost all θ (with respect to H), $\nabla^2 L(\theta, a)$ exists for any a and is continuous in a.

(2) $\int \|\nabla L(\theta, a^*)\|^2 w^2(\theta) dH(\theta) < \infty$ for any $a^* \in \mathcal{A}^*$.

(3) For any $C > 0$, there is a function $B_C(\theta)$ such that $\int B_C(\theta) l_x(\theta) d\Pi(\theta) < \infty$ and $\sup_{\|a\| \le C} |g_{uv}(\theta, a)| \le B_C(\theta)$ for all $1 \le u, v \le p$.

(4) For any $a^* \in \mathcal{A}^*$, $\int \nabla^2 L(\theta, a^*) l_x(\theta) d\Pi(\theta)$ is positive definite.

Under condition (L3), $\rho(a)$ is second order continuously differentiable and $\nabla\rho(a^*) = 0$ for any $a^* \in \mathcal{A}^*$, where $\nabla\rho$ is the gradient of ρ.

If the Bayesian action is unique, $a_m(\omega)$ converges to a^* for almost all ω (Corollary 1) and the convergence rates of $a_m(\omega)$ and $r(a_m(\omega))$ can be obtained by using standard techniques. Shao (1988) obtained the convergence rates when the loss is convex. The result is extended to general non-convex loss situations (see Theorem 4 below). When the loss function is not convex, the Bayesian action may not be unique. If there are more than one Bayesian actions, $a_m(\omega)$ may not converge in the ordinary sense and it is also much more difficult to obtain the convergence rate of $r(a_m(\omega))$ (although $r(a_m(\omega))$ converges according to Theorems 1 and 2). In Theorem 3, without assuming the uniqueness of the Bayesian action, we establish a convergence rate for $r(a_m(\omega))$ in some situations.

Theorem 3. Assume (L3) and either (L1) or (L2). Assume also that for almost all ω,

$$\sum_{i=1}^m \nabla L(\theta_i, a^*) w(\theta_i) = O(m^{1/2}(\log\log m)^{1/2}) \quad \text{for all } a^* \in \partial\mathcal{A}^*, \tag{4.1}$$

where $\partial\mathcal{A}^*$ is the boundary of $\mathcal{A}^*$. Then for almost all ω,

$$r(a_m(\omega)) - r^* = o(m^{-1/2}(\log\log m)^{1/2}).$$

Remark. Note that for any $a^* \in \partial\mathcal{A}^*$, $\sum_{i=1}^m \nabla L(\theta_i, a^*)w(\theta_i)$ is a sum of i.i.d. random variables and $\nabla\rho(a^*) = \int \nabla L(\theta, a^*)w(\theta)dH(\theta) = 0$. Hence from condition (L3) and the law of iterated logarithm, for almost all ω,

$$\sum_{i=1}^m \nabla L(\theta_i, a^*)w(\theta_i) = O(m^{1/2}(\log\log m)^{1/2}).$$

Thus, condition (4.1) is clearly satisfied if $\partial\mathcal{A}^*$ is a countable set. An important example of countable $\partial\mathcal{A}^*$ is that $\mathcal{A} \subset \mathbf{R}$ and $\mathcal{A}^* = \cup_{i\in\Lambda}[\alpha_i, \beta_i]$, where $\alpha_i \le \beta_i$ are constants and Λ is a countable index set. Another example is that $\mathcal{A}^*$ is a countable set ($\partial\mathcal{A}^* \subset \mathcal{A}^*$ since $\mathcal{A}^*$ is closed).

Proof. Let $Q(a) = \int \nabla^2 L(\theta, a)w(\theta)dH(\theta)$. Note that under either (L1) or (L2), (2.2) and (2.4) hold (Theorems 1 and 2). Also, from condition (L3) and Lemma 4,

$$\sup_{\|a\|\le C} \left| \frac{1}{m}\sum_{i=1}^m \nabla^2 L(\theta_i, a)w(\theta_i) - Q(a) \right| \to 0 \quad \text{for all positive rational } C. \tag{4.2}$$

Let ω be fixed such that (2.2), (2.4) and (4.1)-(4.2) hold, and

$$z_m(\omega) = \frac{m^{1/2}[\rho(a_m(\omega)) - \rho^*]}{(\log\log m)^{1/2}}.$$

It suffices to show that for any subsequence $\{ m_l \}$, there is a subsequence $\{ m_j \} \subset \{ m_l \}$ such that

$$\lim_{j\to\infty} z_{m_j}(\omega) = 0.$$

Let $\{ m_l \}$ be a given subsequence. From (2.4), there is a subsequence $\{ m_j \} \subset \{ m_l \}$ such that

$$\lim_{j\to\infty} a_{m_j}(\omega) = a^* \in \mathcal{A}^*. \tag{4.3}$$

Case 1. $a^* \in \mathcal{A}^* - \partial\mathcal{A}^*$. Since a^* is an interior point of $\mathcal{A}^*$, there is a constant $\delta > 0$ such that $\rho(a) = \rho^*$ or all a satisfying $\|a - a^*\| < \delta$. Then from (4.3),

$$z_{m_j}(\omega) = 0 \quad \text{for sufficiently large } j.$$

Case 2. $a^* \in \partial\mathcal{A}^*$. Note that $\sum_{i=1}^{m_j} \nabla L(\theta_i, a_{m_j}(\omega))w(\theta_i) = m_j \nabla\rho_{m_j}(a_{m_j}(\omega)) = 0$. Then from the mean value theorem,

$$\sum_{i=1}^{m_j} \nabla L(\theta_i, a^*) w(\theta_i) = [\sum_{i=1}^{m_j} \nabla^2 L(\theta_i, \xi_j(\omega)) w(\theta_i)][a^* - a_{m_j}(\omega)] \qquad (4.4)$$

and

$$\rho(a_{m_j}(\omega)) - \rho^* = [\nabla \rho(\zeta_j(\omega))]^\tau [a_{m_j}(\omega) - a^*], \qquad (4.5)$$

where $\xi_j(\omega)$ and $\zeta_j(\omega)$ are on the line segment between a^* and $a_{m_j}(\omega)$. From (2.4) and (4.2),

$$\lim_{j\to\infty} \frac{1}{m_j} \sum_{i=1}^{m_j} \nabla^2 L(\theta_i, \xi_j(\omega)) w(\theta_i) = Q(a^*), \qquad (4.6)$$

which is positive definite under (L3). Since $a^* \in \partial\mathcal{A}^*$, (4.1) holds and

$$\lim_{j\to\infty} \nabla \rho(\zeta_j(\omega)) = \nabla \rho(a^*) = 0.$$

Hence $\lim_{j\to\infty} z_{m_j}(\omega) = 0$ follows from (4.1) and (4.3)-(4.6). This completes the proof since $r(a) = \rho(a)/M_x$. □

Theorem 4. Assume the same conditions as in Theorem 3 and there is a unique Bayesian action a^*. Then for almost all ω,

$$a_m(\omega) - a^* = O(m^{-1/2}(\log\log m)^{1/2}), \qquad (4.7)$$

and

$$r(a_m(\omega)) - r^* = O(m^{-1}\log\log m). \qquad (4.8)$$

Proof. By the same argument as in the proof of Theorem 3, we can show that for almost all ω,

$$\frac{1}{m} \sum_{i=1}^{m} \nabla^2 L(\theta_i, b_m) w(\theta_i) \to \int \nabla^2 L(\theta, a^*) w(\theta) dH(\theta) > 0, \qquad (4.9)$$

where $\{ b_m \}$ is any sequence satisfying $\| b_m - a^* \| \le \| a_m(\omega) - a^* \|$. From the mean value theorem and the fact that $\nabla \rho(a^*) = 0$ and $\sum_{i=1}^{m} \nabla L(\theta_i, a_m(\omega)) w(\theta_i) = 0$, we have

$$\sum_{i=1}^{m} \nabla L(\theta_i, a^*) w(\theta_i) = [\sum_{i=1}^{m} \nabla^2 L(\theta_i, \xi_m(\omega))][a^* - a_m(\omega)] \qquad (4.10)$$

and

$$\rho(a_m(\omega)) - \rho^* = [a_m(\omega) - a^*]^\tau \nabla^2 \rho(\zeta_m(\omega))[a_m(\omega) - a^*], \qquad (4.11)$$

where $\xi_m(\omega)$ and $\zeta_m(\omega)$ are on the line segment between a^* and $a_m(\omega)$. Then (4.7) follows from (4.9)-(4.10) and the law of iterated logarithm and (4.8) follows from (4.7), (4.9) and

(4.11). □

From (4.7)-(4.8), $r(a_m(\omega))$ converges much faster than $a_m(\omega)$. In the following we obtain a limiting distribution of a_m, which provides an accuracy measure for a_m.

Theorem 5. Assume the same conditions as in Theorem 4. Then

$$m^{1/2}(a_m - a^*) \to N(0, D) \quad \text{in distribution},$$

where $N(0, D)$ is the p dimensional normal distribution with

$$D = U^{-1}VU^{-1},$$
$$V = \int [\nabla L(\theta, a^*)][\nabla L(\theta, a^*)]^{\tau} w^2(\theta) dH(\theta),$$
$$U = \int \nabla^2 L(\theta, a^*) w(\theta) dH(\theta).$$

Proof. The result follows from (4.9)-(4.10) and the central limit theorem. □

A Monte Carlo approximation to D is

$$D_m(\omega) = U_m^{-1}(\omega) V_m(\omega) U_m^{-1}(\omega)$$

with

$$V_m(\omega) = \frac{1}{m} \sum_{i=1}^{m} [\nabla L(\theta_i, a_m(\omega))][\nabla L(\theta_i, a_m(\omega))]^{\tau} w^2(\theta_i)$$

and

$$U_m(\omega) = \frac{1}{m} \sum_{i=1}^{m} \nabla^2 L(\theta_i, a_m(\omega)) w(\theta_i).$$

Using the same argument as in the above proofs, we can show the following result.

Theorem 6. Assume the same conditions as in Theorem 4. If for any $C > 0$, there is a function $B_C(\theta)$ such that $\int B_C(\theta) w^2(\theta) dH(\theta) < \infty$ and $\sup_{\|a\| \le C} \|\nabla L(\theta, a)\|^2 \le B_C(\theta)$, then for almost all ω,

$$\lim_{m \to \infty} D_m(\omega) = D.$$

References

Bauwens, W. and Richard, J. F. (1985). A 1-1 plot-t random variable generator with application to Monte Carlo integration. *J. Econometrics* **29**, 19-46.

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

Berger, J. O. and Deely, J. (1988). A Bayesian approach to ranking and selection of related means with alternatives to analysis-of-variance methodology. *J. Amer. Statist. Assco.* **83**, 364-373.

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

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

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.

Shao, J. (1988). Monte Carlo approximations in Bayesian decision theory. Technical Report, Department of Statistics, Purdue University.

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

Stewart, L. and Johnson, J. D. (1972). Determining optimum burn-in and replacement times using Bayesian decision theory. *IEEE Transactions on Reliability R-21* 170-175.

van Dijk, H. K. and Kloek, T. (1980). Further experience in Bayesian analysis using Monte Carlo integration. *J. Econometrics* **14**, 307-328.

Zellner, A. and Rossi, P. (1984). Bayesian analysis of dichotomous quantal response models. *J. Econometrics* **25**, 365-394.

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION	Unclassified
1b. RESTRICTIVE MARKINGS	
2a. SECURITY CLASSIFICATION AUTHORITY	
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE	
3. DISTRIBUTION / AVAILABILITY OF REPORT	Approved for public release, distribution unlimited.
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	Technical Report #88-65C
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	NSF DMS-8717799, DMS-8702620 N00014-88-K-1070
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 PART III: LIMITING BEHAVIOR OF MONTE CARLO APPROXIMATIONS

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 ______	December 1988	17

16. SUPPLEMENTARY NOTATION

17. COSATI CODES

FIELD	GROUP	SUB-GROUP

18. SUBJECT TERMS *(Continue on reverse if necessary and identify by block number)*

Almost sure convergence; convergence rate; loss function; posterior expected loss.

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

Monte Carlo approximation is a useful method in obtaining a numerical approximation to a Bayesian action (an action which minimizes the posterior expected loss). We study the behavior of the Monte Carlo approximation when the Monte Carlo sample size is large. Convergence and convergence rate of the Monte Carlo approximation are established under some weak conditions on the loss function.

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT	21. ABSTRACT SECURITY CLASSIFICATION
☐ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS	Unclassified

22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE *(Include Area Code)*	22c. OFFICE SYMBOL
Jun Shao	317-494-6039	