Trajectory Averaging for Stochastic Approximation MCMC Algorithms

Faming Liang

Department of Statistics
Texas A&M University

1. Liang, F. (2009) Trajectory Averaging for Stochastic Approximation MCMC Algorithms. Manuscript.

2. Liang, F. (2009) On the use of SAMC for Monte Carlo Integration. *Statistics & Probability Letters*, **79**, 581-587.

3. Liang, F., Liu, C. and Carroll, R.J. (2007) Stochastic approximation in Monte Carlo computation. *JASA*, **102**, 305-320.

Stochastic Approximation

Consider the vectorial stochastic approximation algorithm:

$$\theta_{k+1} = \theta_k + a_k H(\theta_k, X_{k+1}), \tag{1}$$

where a_k is called gain factor, and X_{k+1} is a stochastic disturbance distributed according to the density function $f_{\theta_k}(x)$ with $x \in \mathcal{X} \subset \mathbb{R}^d$, and d is the dimension of x.

Such an algorithm is often studied by rewriting it as an algorithm used for the search of zeros of a function $h(\theta)$,

$$\theta_{k+1} = \theta_k + a_k[h(\theta_k) + \epsilon_{k+1}], \tag{2}$$

where $h(\theta_k) = \int_{\mathcal{X}} H(\theta_k, x) f_{\theta_k}(x) dx$ corresponds to the mean effect of $H(\theta_k, X_{k+1})$, and $\epsilon_k = H(\theta_k, X_{k+1}) - h(\theta_k)$ is called the observation noise.

Optimal Convergence Rate and Newton Algorithm

It is well known that the optimal convergence rate of (2) can be achieved with

$$a_k = -F^{-1}/k,$$

where $F = \partial h(\theta^*)/\partial\theta$, and θ^* denotes the zero point of $h(\theta)$. In this case, the stochastic approximation algorithm is reduced to Newton's algorithm. Unfortunately, it is often impossible to use this algorithm, as the matrix F is generally unknown.

Trajectory Averaging

It has been shown (Ruppert, 1988; Polyak, 1990; Polyak and Juditsky, 1992) that the trajectory averaging estimator is asymptotically efficient; that is,

$$\bar{\theta}_n = \sum_{k=1}^{n} \theta_k / n$$

can converge in distribution to a normal random variable with mean θ^* and covariance matrix Σ, where Σ is the smallest possible covariance matrix in an appropriate sense.

Averaging paradoxical principle: A slow algorithm having less than optimal convergence rate must be averaged.

The trajectory averaging estimator allows $\{a_k\}$ to be relatively large, decreasing slower than $O(1/k)$.

The averaging estimator has not yet been explored for the stochastic approximation MCMC algorithm.

- Chen (1993, 2002) considered the case where the observation noise can be state dependent, their results are not directly applicable to stochastic approximation MCMC algorithms.

- The theory established by Kushner and Yin (2003) can potentially be extended to the stochastic approximation MCMC algorithm, but, as mentioned in Kushner and Yin (2003, p.375), more work needs to be done for the extension.

Results

- We show that the trajectory averaging estimator is asymptotically efficient for the stochastic approximation MCMC algorithm under mild conditions.
- The theoretical result is illustrated by a numerical example, which shows that the trajectory averaging estimator has a constant variance independent of the choice of the gain factor sequence.

Notations

- Let $\{\mathcal{K}_s, s \geq 0\}$ be a sequence of compact subsets of Θ such that

$$\bigcup_{s \geq 0} \mathcal{K}_s = \Theta, \quad \text{and} \quad \mathcal{K}_s \subset \text{int}(\mathcal{K}_{s+1}), \quad s \geq 0, \tag{3}$$

  where int(A) denotes the interior of set A.

- Let $\{a_k\}$ and $\{b_k\}$ be two monotone, nonincreasing, positive sequences.

- Let $\mathcal{X}_0$ be a subset of $\mathcal{X}$, and let $\mathbb{T} : \mathcal{X} \times \Theta \to \mathcal{X}_0 \times \mathcal{K}_0$ be a measurable function which maps a point in $\mathcal{X} \times \Theta$ to a random point in $\mathcal{X}_0 \times \mathcal{K}_0$. Let σ_k denote the number of truncations performed until iteration k.

Varying Truncation Stochastic Approximation MCMC Algorithm

The algorithm starts with a random choice of (θ_0, x_0) in the space $\mathcal{K}_0 \times \mathcal{X}_0$, and then iterates between the following steps:

- Draw sample x_{k+1} from a Markov transition kernel with the invariant distribution $f_{\theta_k}(x)$.
- Set $\theta_{k+\frac{1}{2}} = \theta_k + a_k H(\theta_k, x_{k+1})$.
- If $\|\theta_{k+\frac{1}{2}} - \theta_k\| \leq b_k$ and $\theta_{k+\frac{1}{2}} \in \mathcal{K}_{\sigma_k}$, then set $(\theta_{k+1}, x_{k+1}) = (\theta_{k+\frac{1}{2}}, x_{k+1})$ and $\sigma_{k+1} = \sigma_k$; otherwise, set $(\theta_{k+1}, x_{k+1}) = \mathbb{T}(\theta_k, x_k)$ and $\sigma_{k+1} = \sigma_k + 1$.

Why varying truncation?

- It removes the bound constraint on the solution space.
- It removes the growth rate restriction on the mean effect function $h(\theta)$.

Condition A_1: Lyapunov condition on $h(\theta)$

The function $h : \Theta \to \mathbb{R}^d$ is continuous, and there exists a continuously differentiable function $v : \Theta \to [0, \infty)$ such that

(i) There exists a single point $\theta^* \in \Theta$ such that $\langle \nabla v(\theta^*), h(\theta^*) \rangle = 0$ and $\langle \nabla v(\theta), h(\theta) \rangle < 0$ for any $\theta \in \Theta \setminus \{\theta^*\}$.

(ii) There exist M_0 and M_1 such that $M_1 > M_0 > 0$, $\theta^* \in \text{int}(\mathcal{V}_{M_0})$, and $\mathcal{V}_{M_1}$ is a compact set, where int(A) denotes the interior of the set A.

Condition A_2: stability condition on $h(\theta)$

The mean field function $h(\theta)$ is measurable and locally bounded. There exist a stable matrix F (i.e., all eigenvalues of F are with negative real parts), $\gamma > 0$, $\rho \in (\tau, 1]$, and a constant c such that

$$\|h(\theta) - F(\theta - \theta^*)\| \leq c\|\theta - \theta^*\|^{1+\rho}, \quad \forall \theta \in \{\theta : \|\theta - \theta^*\| \leq \gamma\}.$$

Condition A_3: Drift conditions on the transition kernel P_θ

For any given $\theta \in \Theta$, the transition kernel P_θ is irreducible and aperiodic. In addition, there exists a function $V : \mathcal{X}^\kappa \to [1, \infty)$ and a constant $\alpha \geq 2$ such that for any compact subset $\mathcal{K} \subset \Theta$,

(i) There exist a set $\boldsymbol{C} \subset \mathcal{X}$, an integer l, constants $0 < \lambda < 1$, b, ς, $\delta > 0$ and a probability measure ν such that

- $$\sup_{\theta \in \mathcal{K}} P_\theta^l V^\alpha(x) \leq \lambda V^\alpha(x) + bI(x \in \boldsymbol{C}), \quad \forall x \in \mathcal{X}. \tag{4}$$
- $$\sup_{\theta \in \mathcal{K}} P_\theta V^\alpha(x) \leq \varsigma V^\alpha(x), \quad \forall x \in \mathcal{X}. \tag{5}$$
- $$\sup_{\theta \in \mathcal{K}} P_\theta^l(x, A) \geq \delta \nu(A), \quad \forall x \in \boldsymbol{C}, \forall A \in \mathcal{B}. \tag{6}$$

(ii) There exists a constant c such that for all $x \in \mathcal{X}$,

- $$\sup_{\theta \in \mathcal{K}} \|H(\theta, x)\| \leq cV(x). \tag{7}$$
- $$\sup_{(\theta,\theta') \in \mathcal{K}} \|H(\theta, x) - H(\theta', x)\| \leq cV(x)\|\theta - \theta'\|. \tag{8}$$

(iii) There exists a constant c such that for all $(\theta, \theta') \in \mathcal{K} \times \mathcal{K}$,

- $$\|P_\theta g - P_{\theta'} g\|_V \leq c_2 \|g\|_V |\theta - \theta'|, \quad \forall g \in \mathcal{L}_V. \quad (9)$$
- $$\|P_\theta g - P_{\theta'} g\|_{V^\alpha} \leq c_2 \|g\|_{V^\alpha} |\theta - \theta'|, \quad \forall g \in \mathcal{L}_{V^\alpha}. \quad (10)$$

Condition A_4: Conditions on the step-sizes

The sequences $\{a_k\}$ and $\{b_k\}$ are nonincreasing, positive, and satisfy the conditions:

$$\lim_{k\to\infty} (ka_k) = \infty, \quad \frac{a_{k+1} - a_k}{a_k} = o(a_{k+1}), \quad \lim_{k\to\infty} b_k = 0, \tag{11}$$

for some $\tau \in (0, 1)$,

$$\sum_{k=1}^{\infty} \frac{a_k^{(1+\tau)/2}}{\sqrt{k}} < \infty, \tag{12}$$

and for some constant $\alpha \geq 2$ as defined in condition (A_3),

$$\sum_{i=1}^{\infty} \left\{a_i^2 + a_i b_i + (b_i^{-1} a_i)^{\alpha}\right\} < \infty. \tag{13}$$

For instance, we can set $a_k = 1/k^{\eta}$ for some $\eta \in (1/2, 1)$. In this case, (12) is satisfied for any $\tau > 1/\eta - 1$, and (13) is satisfied by setting $b_k = C/k^{\xi}$ for some constants C and $\xi \in (1 - \eta, \eta - \frac{1}{\alpha})$.

Convergence theorem of SAMCMC (Andrieu *et al.*, 2005)

Theorem 0.1 *Assume the conditions (A_1), (A_3), and (A_4) hold. Let α_σ denote the number of iterations for which the σ-th truncation occurs in the stochastic approximation MCMC simulation. Let $\mathcal{X}_0 \subset \mathcal{X}$ be such that $\sup_{x \in \mathcal{X}_0} V(x) < \infty$ and that $\mathcal{K}_0 \subset \mathcal{V}_{M_0}$, where $\mathcal{V}_{M_0}$ is defined in (A_1). Then there exists a number σ such that $\alpha_\sigma < \infty$ a.s., $\alpha_{\sigma+1} = \infty$ a.s., and $\{\theta_k\}$ has no truncation for $k \geq \alpha_\sigma$, i.e.,*

$$\theta_{k+1} = \theta_k + a_k H(\theta_k, x_{k+1}), \quad \forall\, k \geq \alpha_\sigma,$$

and

$$\theta_k \longrightarrow \theta^*, \quad a.s.$$

Noise decomposition

Lemma 0.1 *Assume the conditions A_1, A_3 and A_4 hold. If $\sup_x V(x) < \infty$ and $\alpha_\sigma < \infty$, then there exist $\mathbb{R}^d$-valued random processes $\{e_k\}_{k\geq 1}$, $\{\nu_k\}_{k\geq 1}$, and $\{\varsigma_k\}_{k\geq 1}$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that when $k > \alpha_\sigma$,*

(i) $\epsilon_k = e_k + \nu_k + \varsigma_k$.

(ii) $\{e_k\}$ *is a martingale difference sequence, and* $\frac{1}{\sqrt{n}} \sum_{k=1}^n e_k \longrightarrow N(0, Q)$ *in distribution, where* $Q = \lim_{k\to\infty} E(e_k e_k')$.

(iii) $\|\nu_k\| = O(a_k^{(1+\tau)/2})$, *where* τ *is given in condition* (A_4).

(iv) $\|\sum_{k=0}^n a_k \varsigma_k\| = O(a_n)$.

Asymptotic normality of $\bar{\theta}_k$

Theorem 0.2 *Assume the conditions* (A_1), (A_2), (A_3), *and* (A_4) *hold. If* $\sup_x V(x) < \infty$ *and* $\mathcal{K}_0 \subset \mathcal{V}_{M_0}$, *where* $\mathcal{V}_{M_0}$ *is defined in* (A_1). *Then we have*

$$\sqrt{k}(\bar{\theta}_k - \theta^*) \longrightarrow N(\mathbf{0}, \Gamma)$$

where $\Gamma = F^{-1}Q(F^{-1})'$, $F = \partial h(\theta^*)/\partial\theta$ *is negative definite,* $Q = \lim_{k\to\infty} E(e_k e_k')$, *and* e_k *is as defined in Lemma 0.1.*

Asymptotic Efficiency

Definition 0.1 *Consider the stochastic approximation algorithm (2). Let $\{Z_n\}_{n\geq 0}$, given as a function of $\{\theta_n\}_{n\geq 0}$, be a sequence of estimators of θ^*. The algorithm $\{Z_n\}_{n\geq 0}$ is said to be asymptotically efficient if*

$$\sqrt{n}(Z_n - \theta^*) \longrightarrow N\left(\mathbf{0}, F^{-1}\widetilde{Q}(F^{-1})'\right), \tag{14}$$

where $F = \partial h(y^*)/\partial y$, and $\widetilde{Q}$ is the asymptotic covariance matrix of $(1/\sqrt{n})\sum_{k=1}^{n}\epsilon_k$.

Asymptotic Efficiency of $\bar{\theta}_k$

Theorem 0.3 *Assume the conditions* (A_1), (A_2), (A_3), *and* (A_4) *hold. If* $\sup_x V(x) < \infty$ *and* $\mathcal{K}_0 \subset \mathcal{V}_{M_0}$, *where* $\mathcal{V}_{M_0}$ *is defined in* (A_1), *then* $\bar{\theta}_k$ *is asymptotically efficient.*

Basic Idea

- Partition the sample space into different subregions: $E_1, \ldots, E_m$, $\bigcup_{i=1}^{M} E_i = \mathcal{X}$, and $E_i \cap E_j = \emptyset$ for $i \neq j$.
- Let $g_i = \int_{E_i} \psi(x)dx$, and choose $\pi = (\pi_1, \ldots, \pi_m)$, $\pi_i \geq 0$, and $\sum_i \pi_i = 1$.
- Sampling from the distribution

$$p_\theta(x) \propto \sum_{i=1}^{m} \frac{\psi(x)}{e^{\theta^{(i)}}} I(x \in E_i).$$

If $\theta^{(i)} = \log(g_i/\pi_i)$ for all i, sampling from $p_\theta(x)$ will result in a random walk in the space of subregions with each subregion being sampled with probability π_i (viewing each subregion as a "single point"). Therefore, sampling from $p_\theta(x)$ can avoid the local trap problem encountered in sampling from $f(x)$.

SAMC Algorithm

(a) (Sampling) Simulate a sample x_{k+1} by a single MH update with the target distribution

$$f_{\theta_k}(x) \propto \sum_{i=1}^{m-1} \frac{\psi(x)}{e^{\theta_k^{(i)}}} I_{\{x \in E_i\}} + \psi(x) I_{\{x \in E_m\}}, \qquad (15)$$

provided that E_m is non-empty, i.e., $\omega_m > 0$. Note that the assumption $w_m > 0$ is only made for the reason of theoretical simplicity. As a practical matter, this is not necessary.

(a.1) Generate y according to a proposal distribution $q(x_k, y)$.

(a.2) Calculate the ratio

$$r = e^{\theta_k^{(J(x_k))} - \theta_k^{(J(y))}} \frac{\psi(y)q(y, x_k)}{\psi(x_t)q(x_k, y)},$$

where $J(z)$ denotes the index of the subregion that the sample z belongs to.

(a.3) Accept the proposal with probability $\min(1, r)$. If it is accepted, set $x_{k+1} = y$; otherwise, set $x_{k+1} = x_k$.

Continuation of Algorithm

(b) (Weight updating) Set

$$\theta_{k+\frac{1}{2}}^{(i)} = \theta_k^{(i)} + a_{k+1}\left(I_{\{x_{k+1}\in E_i\}} - \pi_i\right), \quad i = 1, \ldots, m-1. \tag{16}$$

(c) (Varying truncation) If $\theta_{k+\frac{1}{2}} \in \mathcal{K}_{\sigma_k}$, then set $(\theta_{k+1}, x_{k+1}) = (\theta_{k+\frac{1}{2}}, x_{k+1})$ and $\sigma_{k+1} = \sigma_k$; otherwise, set $(\theta_{k+1}, x_{k+1}) = \mathbb{T}(\theta_k, x_k)$ and $\sigma_{k+1} = \sigma_k + 1$, where σ_k and $\mathbb{T}(\cdot,\cdot)$ are as defined in Section 2.

To show that the trajectory averaging estimator is asymptotically efficient for SAMC, we assume the following conditions.

(C_1) The sample space $\mathcal{X}$ is compact, and $f(x)$ is bounded away from 0 and ∞. The proposal distribution $q(x, y)$ used in step (a.1) of the SAMC algorithm satisfies the condition: For every $x \in \mathcal{X}$, there exist ϵ_1 and ϵ_2 such that

$$\|x - y\| \leq \epsilon_1 \Longrightarrow q(x, y) \geq \epsilon_2, \tag{17}$$

where $\|z\|$ denotes the norm of the vector z.

(C_2) The sequence $\{a_k\}$ is positive and non-increasing,

$$\lim_{k\to\infty}(ka_k) = \infty, \quad \frac{a_k - a_{k+1}}{a_k} = o(a_{k+1}), \tag{18}$$

and for some $\tau \in (0, 1)$

$$\sum_{k=1}^{\infty} \frac{a_k^{(1+\tau)/2}}{\sqrt{k}} < \infty. \tag{19}$$

Theoretical Results

Theorem 0.4 *(Convergence) Assume the conditions* (C_1) *and* (C_2)*. Then there exists a number* σ *such that* $\alpha_\sigma < \infty$ *a.s.,* $\alpha_{\sigma+1} = \infty$ *a.s., and* $\{\theta_k\}$ *given by the SAMC algorithm has no truncation for* $k \geq \alpha_\sigma$*, i.e.,*

$$\theta_{k+1} = \theta_k + a_k H(\theta_k, x_{k+1}), \quad \forall\, k \geq \alpha_\sigma, \tag{20}$$

and

$$\theta_k \longrightarrow \theta^*, \quad a.s., \tag{21}$$

where $H(\theta_k, x_{k+1}) = \big(I_{\{x_{k+1} \in E_1\}} - \pi_1, \ldots, I_{\{x_{k+1} \in E_{m-1}\}} - \pi_{m-1}\big)'$,
$\theta^* = c\mathbf{1}_{m-1} + \Big(\log(\omega_1/\pi_1), \ldots, \log(\omega_{m-1}/\pi_{m-1})\Big)', c = -\log(\omega_m/\pi_m)$,
and $\mathbf{1}_{m-1}$ *denotes an* $(m-1)$*-vector of ones.*

Theorem 0.5 *(Asymptotic Efficiency) Assume the conditions* (C_1) *and* (C_2). *Then* $\bar{\theta}_k$ *is asymptotically efficient; that is,*

$$\sqrt{k}(\bar{\theta}_k - \theta^*) \longrightarrow N(\mathbf{0}, \Gamma) \quad \textit{as } k \to \infty,$$

where $\Gamma = F^{-1}Q(F^{-1})'$, $F = \partial h(\theta^*)/\partial\theta$ *is negative definite, and* $Q = \lim_{k\to\infty} E(e_k e_k')$.

Theorem 0.6 *(Weighted averaging) Assume the conditions (C_1) and (C_2). For a set of samples generated by SAMC, the random variable/vector Y generated by*

$$P(Y = y_i) = \frac{\sum_{t=1}^{n} e^{\theta_{tJ(x_t)}} I(x_t = y_i)}{\sum_{t=1}^{n} e^{\theta_{tJ(x_t)}}}, \quad i = 1, \ldots, n',$$

is asymptotically distributed as $f(\cdot)$.

This implies that for an integrable function $h(x)$, the expectation $E_f h(x)$ can be estimated by

$$\widehat{E_f h(x)} = \frac{\sum_{t=1}^{n} e^{\theta_{tJ(x_t)}} h(x_t)}{\sum_{t=1}^{n} e^{\theta_{tJ(x_t)}}}. \tag{22}$$

As $n \to \infty$, $\widehat{E_f h(x)} \to E_f h(x)$ for the same reason that the usual importance sampling estimate converges (Geweke, 1989).

Let $g_t(x, w)$ be the joint distribution of the sample (x, w) drawn at iteration t, where $w = \exp(\theta_{tJ(x)})$. The principle of IWIW (invariance with respect to importance weights, Wong and Liang, 1997; Liang, 2002) can be defined as follows:

The joint distribution $g_t(x, w)$ is said to be correctly weighted with respect to a distribution $f(x)$ if

$$\int w g_t(x, w) dw \propto f(x). \tag{23}$$

A transition rule is said to satisfy IWIW if it maintains the correctly weighted property for the joint distribution $g_t(x, w)$ whenever an initial joint distribution is correctly weighted.

Theorem 0.7 *Assume the conditions (C_1) and (C_2). Then SAMC asymptotically satisfies the IWIW principle.*

x	1	2	3	4	5	6	7	8	9	10
$f(x)$	1	100	2	1	3	3	1	200	2	1

Table 1: The unnormalized mass function of the 10-state distribution.

The sample space was partitioned according to the mass function into five subregions: $E_1 = \{8\}$, $E_2 = \{2\}$, $E_3 = \{5, 6\}$, $E_4 = \{3, 9\}$ and $E_5 = \{1, 4, 7, 10\}$. The desired sampling distribution is set to

$$\pi_i \propto \frac{1}{1+i}, \quad i = 1, \ldots, 5.$$

The transition proposal matrix was set to a stochastic matrix with each row being generated independently from the Dirichlet distribution $Dir(1, \ldots, 1)$, and the gain factor sequences we tried included

$$a_k = \frac{T_0}{\max\{k^\eta, T_0\}}, \qquad T_0 = 10, \quad \eta \in \{0.6, 0.7, 0.8, 0.9, 1.0\}.$$

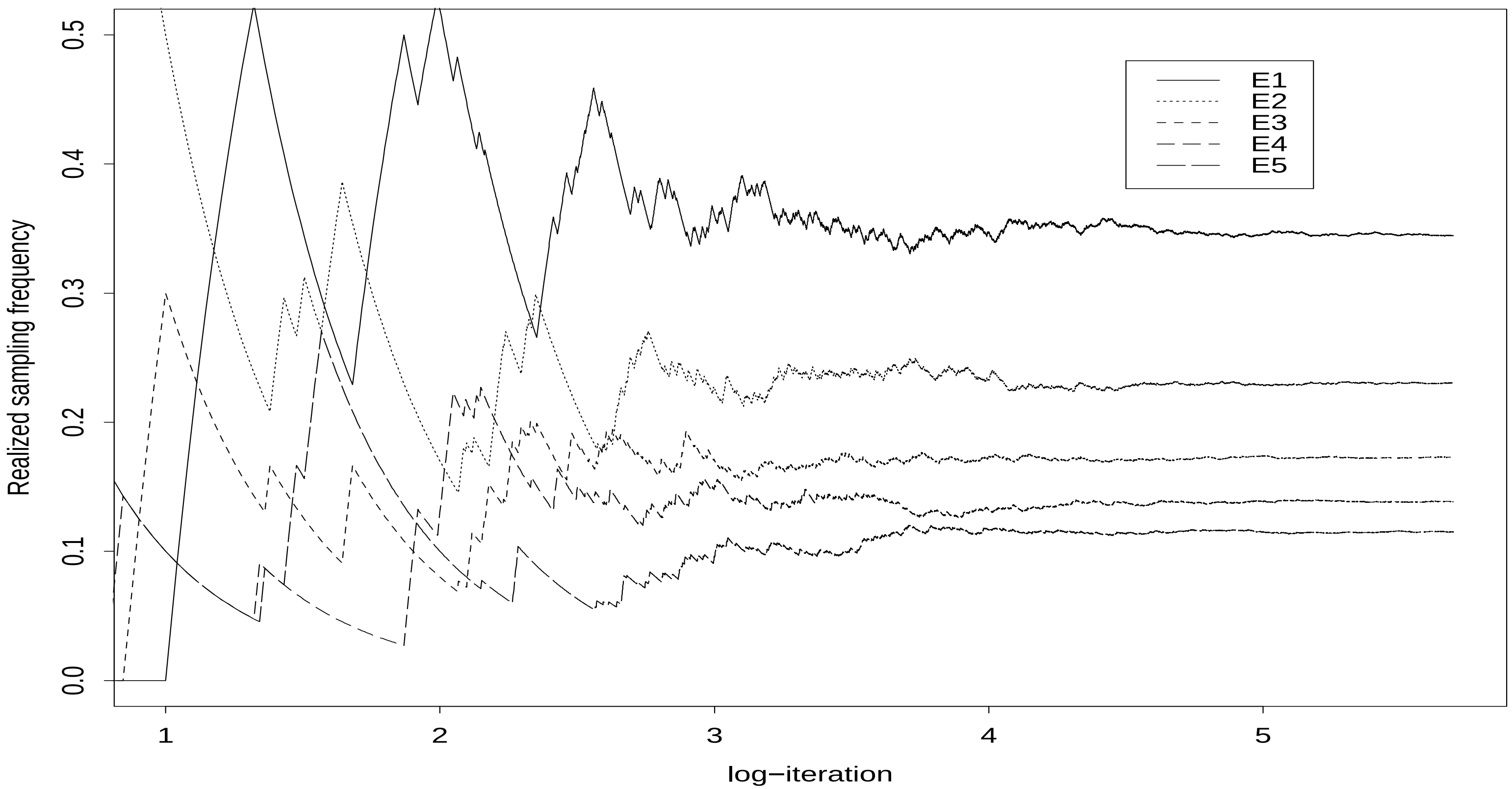

Figure 1: Progression plot of $\widehat{\pi}_k$, the realized sampling frequency of the five subregions $E_1, \ldots, E_5$, obtained in a single run of SAMC.

Table 2: Comparison of two estimators of $\omega = (200, 100, 6, 4, 4)$: the trajectory averaging estimator $\widehat{\omega}$ and the conventional SAMC estimator $\widetilde{\omega}$, where each row corresponds to one component of ω. The simulations are done with the desired distribution: $\pi_i \propto 1/(1+i)$ for $i = 1, \ldots, 5$, "bias" and "sd" are calculated based on 100 independent runs, and "rmse" is calculated as the square root of "bias^2+sd^2".

	$\eta = 0.7$			$\eta = 0.8$			$\eta = 0.9$		
	bias	sd	rmse	bias	sd	rmse	bias	sd	rmse
$\widetilde{\omega}$	-0.81	0.52	0.96	-0.15	0.33	0.36	0.03	0.21	0.21
	0.71	0.49	0.87	0.14	0.31	0.34	-0.03	0.20	0.20
	0.04	0.03	0.05	0.01	0.01	0.02	0.00	0.01	0.01
	0.02	0.02	0.03	0.00	0.01	0.01	0.00	0.01	0.01
	0.03	0.02	0.04	0.01	0.01	0.02	0.01	0.01	0.01
$\widehat{\omega}$	-0.24	0.09	0.26	-0.02	0.11	0.11	0.00	0.1	0.10
	0.19	0.09	0.21	0.00	0.10	0.10	-0.01	0.1	0.10
	0.03	0.00	0.03	0.01	0.00	0.01	0.00	0.0	0.01
	0.00	0.00	0.01	0.00	0.00	0.01	0.00	0.0	0.00
	0.02	0.00	0.02	0.01	0.00	0.01	0.02	0.0	0.02

x	1	2	3	4	5	6	7	8	9	10
$f(x)$	1	100	2	1	3	3	1	200	2	1

Table 3: The unnormalized mass function of the 10-state distribution.

Table 4: Comparison of SAMC and MH for the 10-state example, where the Bias and Standard Error (of the Bias) were calculated based on 100 independent runs.

Algorithm	Bias ($\times 10^{-3}$)	Standard Error ($\times 10^{-3}$)	CPU time (seconds)
SAMC	-0.528	1.513	0.38
MH	-3.685	4.634	0.20

This time our goal is to estimate $E_f X$, the mean of the distribution.

The sample space was partitioned according to the mass function into five subregions: $E_1 = \{8\}$, $E_2 = \{2\}$, $E_3 = \{5,6\}$, $E_4 = \{3,9\}$ and $E_5 = \{1,4,7,10\}$. The desired sampling distribution is uniform over 5 subregions.

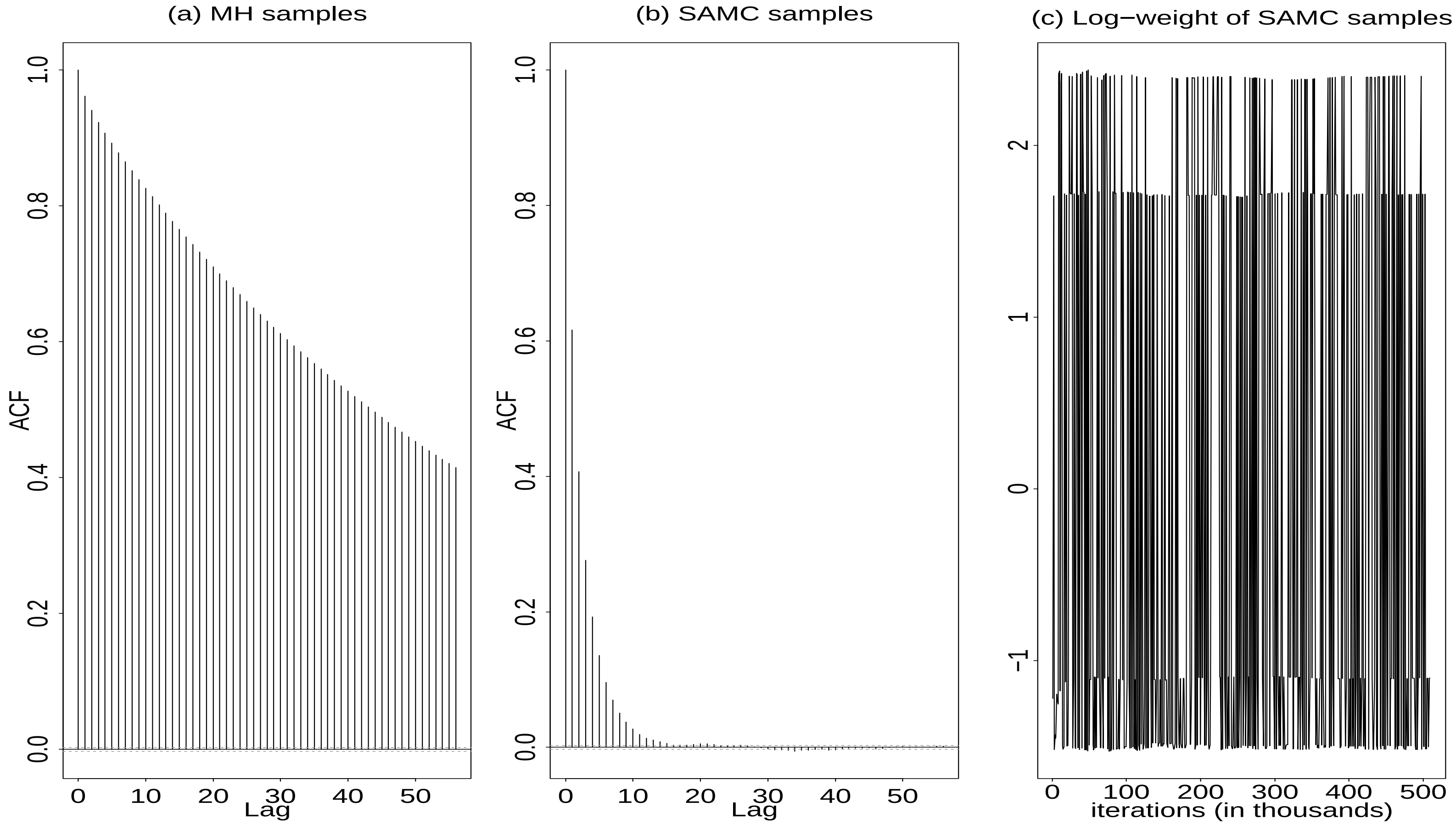

Figure 2: Computational results for the 10-state example. (a) Autocorrelation plot of the MH samples. (b) Autocorrelation plot of the SAMC samples. (c) Log-weight of the SAMC samples.