Applied Probability Trust (12 October 2013)

SUPPLEMENTARY MATERIAL FOR "WEAK CONVERGENCE RATES OF POPULATION VERSUS SINGLE-CHAIN STOCHASTIC APPROXIMATION MCMC ALGORITHMS "

QIFAN SONG,* *Texas A&M University*

MINGQI WU,** *Shell Global Solutions (US) Inc.*

FAMING LIANG,*** *Texas A&M University*

1. Proof of Theorem 2.1

Proof. Let $M = \sup_{\theta\in\Theta} \max\{\|h(\theta)\|, |v(\theta)|\}$ and $\mathcal{V}_\varepsilon = \{\theta : d(\theta,\mathcal{L}) \le \varepsilon\}$. Applying Taylor's expansion formula (Folland, 1990), we have

$$v(\theta_{t+1}) = v(\theta_t) + \gamma_{n+1} v_h(\theta_{t+1}) + R_{t+1}, \quad t \ge 0,$$

which implies that

$$\sum_{i=0}^{t} \gamma_{i+1} v_h(\theta_i) = v(\theta_{t+1}) - v(\theta_0) - \sum_{i=0}^{t} R_{i+1} \ge -2M - \sum_{i=0}^{t} R_{i+1}.$$

Since $\sum_{i=0}^{t} R_{i+1}$ converges (owing to Lemma A.2), $\sum_{i=0}^{t} \gamma_{i+1} v_h(\theta_i)$ also converges. Furthermore,

$$v(\theta_t) = v(\theta_0) + \sum_{i=0}^{t-1} \gamma_{i+1} v_h(\theta_i) + \sum_{i=0}^{t-1} R_{i+1}, \quad t \ge 0,$$

$\{v(\theta_t)\}_{t\ge 0}$ also converges. On the other hand, conditions (A_1) and (A_2) imply $\underline{\lim}_{t\to\infty} d(\theta_t,\mathcal{L}) = 0$. Otherwise, there exists $\varepsilon > 0$ and n_0 such that $d(\theta_t,\mathcal{L}) \ge$

* Postal address: Department of Statistics, Texas A&M University, College Station, TX 77840, US.
** Postal address: Shell Technology Center Houston, 3333 Highway 6 South, Houston, TX 77082, US.
*** Postal address: Department of Statistics, Texas A&M University, College Station, TX 77840, US.
Email: fliang@stat.tamu.edu

ε, $t \geq n_0$; as $\sum_{t=1}^{\infty} \gamma_t = \infty$ and $p = \sup\{v_h(\theta) : \theta \in \mathcal{V}_\varepsilon^c\} < 0$, it is obtained that $\sum_{t=n_0}^{\infty} \gamma_{t+1} v_h(\theta_t) \leq p \sum_{t=1}^{\infty} \gamma_{t+1} = -\infty$.

Suppose that $\overline{\lim}_{t\to\infty} d(\theta_t, \mathcal{L}) > 0$. Then, there exists $\varepsilon > 0$ such that $\overline{\lim}_{t\to\infty} d(\theta_t, \mathcal{L}) \geq 2\varepsilon$. Let $t_0 = \inf\{t \geq 0 : d(\theta_t, \mathcal{L}) \geq 2\varepsilon\}$, while $t_k' = \inf\{t \geq t_k : d(\theta_t, \mathcal{L}) \leq \varepsilon\}$ and $t_{k+1} = \inf\{t \geq t_k' : d(\theta_t, \mathcal{L}) \geq 2\varepsilon\}$, $k \geq 0$. Obviously, $t_k < t_{k'} < t_{k+1}$, $k \geq 0$, and

$$d(\theta_{t_k}, \mathcal{L}) \geq 2\varepsilon,\ d(\theta_{t_k'}, \mathcal{L}) \leq \varepsilon,\ \text{and}\ d(\theta_t, \mathcal{L}) \geq \varepsilon,\ t_k \leq t < t_k',\ k \geq 0.$$

Let $q = \sup\{v_h(\theta) : \theta \in \mathcal{V}_\varepsilon^c\}$. Then

$$q \sum_{k=0}^{\infty} \sum_{i=t_k}^{t_k'-1} \gamma_{i+1} \geq \sum_{k=0}^{\infty} \sum_{i=t_k}^{t_k'-1} \gamma_{i+1} v_h(\theta_i) \geq \sum_{t=0}^{\infty} \gamma_{t+1} v_h(\theta_t) > -\infty.$$

Therefore, $\sum_{k=0}^{\infty} \sum_{i=t_k}^{t_k'-1} \gamma_{i+1} < \infty$, and consequently, $\lim_{k\to\infty} \sum_{i=t_k}^{t_k'-1} \gamma_{i+1} = 0$. Since $\sum_{t=1}^{\infty} \gamma_t \xi_t$ converges (owing to Lemma A.2), we have

$$\varepsilon \leq \|\theta_{t_k'} - \theta_{t_k}\| \leq M \sum_{i=t_k}^{t_k'-1} \gamma_{i+1} + \left\| \sum_{i=t_k}^{t_k'-1} \gamma_{i+1} \xi_{i+1} \right\| \longrightarrow 0,$$

as $k \to \infty$. This contradicts with our assumption $\varepsilon > 0$. Hence, $\overline{\lim}_{t\to\infty} d(\theta_t, \mathcal{L}) > 0$ does not hold. Therefore, $\lim_{t\to\infty} d(\theta_t, \mathcal{L}) = 0$ almost surely.

2. Proofs of Theorems for Pop-SAMC

In order to study the convergence of the Pop-SAMC algorithm, we introduce an equivalent variation of the Pop-SAMC algorithm. Without loss of generality, we assume that $E_1, \ldots, E_{m_0}$ are nonempty subregions, and $E_{m_0+1}, \ldots, E_m$ are all empty.

1. (Population sampling) The sampling step is the same as described in Section 3.2 of the main text.

2'. (Weight updating) Set

$$\theta_{t+1} = \theta_t + \gamma_{t+1} \tilde{\boldsymbol{H}}(\theta_t, \boldsymbol{x}_{t+1}), \tag{1}$$

where $\tilde{\boldsymbol{H}}(\theta_t, \boldsymbol{x}_{t+1}) = \sum_{i=1}^{\kappa} \tilde{H}(\theta_t, x_{t+1}^{(i)})/\kappa$, and $\tilde{H}(\theta_t, x_{t+1}^{(i)}) = \boldsymbol{z}_{t+1} - \boldsymbol{\pi} - (I(x_{t+1}^{(i)} \in E_{m_0}) - \pi_{m_0})\mathbf{1}$. where $\boldsymbol{z}_{t+1}$ and $\boldsymbol{\pi}$ are as specified in the SAMC algorithm, and $\mathbf{1}$ denotes a vector of 1s.

The difference of this variational Pop-SAMC algorithm is that it adds a constant vector $-\gamma_{t+1} \sum_{i=1}^{\kappa} (I(x_{t+1}^{(i)} \in E_{m_0}) - \pi_{m_0})\mathbf{1}/\kappa$ to the estimate of $\boldsymbol{\theta}$ of the original algorithm and thus keeps $\theta^{(m_0)}$ unchanged, say $\theta_t^{(m_0)} \equiv 0$. Hence, below we only need to prove that Theorem 3.1 and Theorem 3.2 are true for this variational Pop-SAMC algorithm.

2.1. Proof of Theorem 3.1

Since $E_{m_0+1}, \ldots, E_m$ are empty, $\theta_{m_0+1}, \ldots, \theta_m$ are auxiliary variable, which do not affect the updating of $(\theta_i)_{i=1}^{m_0-1}$ and sampling step at all. Therefore, we can view the algorithm as of it is only update $\theta = (\theta_1, \ldots, \theta_{m_0-1})^T$ with function $(\tilde{\boldsymbol{H}}^{(1)}, \ldots, \tilde{\boldsymbol{H}}^{(m_0-1)})^T$. Once we prove that for $i = 1, \ldots, m_0 - 1$,

$$\theta_t^{(i)} \to \log\left(\int_{E_i} \psi(x)dx\right) - \log(\pi_i + \nu) - \log\left(\int_{E_{m_0}} \psi(x)dx\right) + \log(\pi_{m_0} + \nu),$$

almost surely, then it is trivial to see that $\theta_t^{(i)} \to -\infty$ for $i > m_0$. (Because $\sum_{j=1}^{t} I(x_j^{(k)} \in E_{m_0})/t \to \pi_{m_0} + \nu$, and $\theta_t^{(i)} = -t\pi_i - \sum_{k=1}^{\kappa}\sum_{j=1}^{t} I(x_j^{(k)} \in E_{m_0})/\kappa + t\pi_{m_0}$ for any $i > m_0$.)

To prove the convergence of $\theta_t^{(i)}$ for $i < m_0$, it follows from Theorem 2.1 that we only need to verify that Pop-SAMC satisfies the conditions (A_1), (A_3) and (A_4). This is done as follows.

(A_1) This condition can be verified as in Liang *et al.* (2007). Since a part of the proof will be used in proving Theorem 3.2, we re-produce the proof below. Since the invariant distribution of the kernel $\boldsymbol{P}_{\theta_t}(\boldsymbol{x}, \boldsymbol{y})$ is $f_{\theta_t}(\boldsymbol{x})$, for any fixed value of θ, we have

$$\begin{aligned} E(\tilde{\boldsymbol{H}}^{(i)}(\theta, \boldsymbol{x})) &= \frac{\int_{E_i} \psi(x)dx/e^{\theta_i} - \int_{E_{m_0}} \psi(x)dx/e^{\theta_{m_0}}}{\sum_{k=1}^{m}[\int_{E_k} \psi(x)dx/e^{\theta_k}]} - \pi_i + \pi_{m_0} \\ &= \frac{S_i - S_{m_0}}{S} - \pi_i + \pi_{m_0}, \quad i = 1, \ldots, m_0 - 1, \end{aligned} \tag{2}$$

where $\tilde{\boldsymbol{H}}^{(i)}(\theta, \boldsymbol{x})$ denotes the ith component of $\tilde{\boldsymbol{H}}(\theta, \boldsymbol{x})$, $S_i = \int_{E_i} \psi(x)dx/e^{\theta_i}$ and $S = \sum_{k=1}^{m_0} S_k$. Thus,

$$h(\theta) = \int_{\mathcal{X}} H(\theta, \boldsymbol{x})f(d\boldsymbol{x}) = \left(\frac{S_1}{S} - \pi_1, \ldots, \frac{S_{m_0-1}}{S} - \pi_{m_0-1}\right)^T - \frac{S_{m_0}}{S} + \pi_{m_0}.$$

It follows from (2) that $h(\theta)$ is a continuous function of θ. Let

$$v(\theta) = \frac{1}{2}\sum_{k=1}^{m_0}\left(\frac{S_k}{S} - \pi_k\right)^2,$$

which, as shown below, has continuous partial derivatives of the first order. Solving the system of equations formed by (2), we have

$$\mathcal{L} = \left\{(\theta_1, \ldots, \theta_{m_0-1}) : \theta_i = \mathrm{C} + \log\left(\int_{E_i} \psi(x)dx\right) - \log(\pi_i + \nu), \theta \in \Theta\right\},$$

where constant $\mathrm{C} = \log(\pi_{m_0} + \nu) - \log\int_{E_{m_0}} \psi(x)dx$. It is obvious that $\mathcal{L}$ is nonempty and $v(\theta) = 0$ for every $\theta \in \mathcal{L}$.

To verify the conditions related to $\nabla v(\theta)$, we have the following calculations:

$$\begin{aligned} \frac{\partial S}{\partial \theta_i} &= \frac{\partial S_i}{\partial \theta_i} = -S_i, & \frac{\partial S_i}{\partial \theta_j} &= \frac{\partial S_j}{\partial \theta_i} = 0, \\ \frac{\partial(\frac{S_i}{S})}{\partial \theta_i} &= -\frac{S_i}{S}(1 - \frac{S_i}{S}), & \frac{\partial(\frac{S_i}{S})}{\partial \theta_j} &= \frac{\partial(\frac{S_j}{S})}{\partial \theta_i} = \frac{S_i S_j}{S^2}, \end{aligned} \tag{3}$$

for $i, j = 1, \ldots, m_0 - 1$ and $i \neq j$.

$$\begin{aligned} \frac{\partial v(\theta)}{\partial \theta_i} &= \frac{1}{2}\sum_{k=1}^{m_0} \frac{\partial(\frac{S_k}{S} - \pi_k)^2}{\partial \theta_i} \\ &= \sum_{j=1}^{m_0}(\frac{S_j}{S} - \pi_j)\frac{S_i S_j}{S^2} - (\frac{S_i}{S} - \pi_i)\frac{S_i}{S} \\ &= \mu_{\eta^*}\frac{S_i}{S} - (\frac{S_i}{S} - \pi_i)\frac{S_i}{S}, \end{aligned} \tag{4}$$

for $i = 1, \ldots, m_0 - 1$, where $\mu_{\eta^*} = \sum_{j=1}^{m_0} (\frac{S_j}{S} - \pi_j)\frac{S_j}{S}$. Thus,

$$
\begin{aligned}
v_h(\theta) &= \langle \nabla v(\theta), h(\theta) \rangle \\
&= \mu_{\eta^*} \sum_{i=1}^{m_0-1} (\frac{S_i}{S} - \pi_i)\frac{S_i}{S} - \sum_{i=1}^{m_0-1} (\frac{S_i}{S} - \pi_i)^2 \frac{S_i}{S} \\
&\quad - \sum_{i=1}^{m_0-1} \left(\mu_{\eta^*}\frac{S_i}{S} - (\frac{S_i}{S} - \pi_i)\frac{S_i}{S} \right) \left(\frac{S_{m_0}}{S} - \pi_{m_0} \right) \qquad (5) \\
&= - \{ \sum_{i=1}^{m_0} (\frac{S_i}{S} - \pi_i)^2 \frac{S_i}{S} - \mu_{\eta^*}^2 \} \\
&= - \sigma_{\eta^*}^2 \le 0,
\end{aligned}
$$

where $\sigma_{\eta^*}^2$ denotes the variance of the discrete distribution defined in the following table,

State (η^*)	$\frac{S_1}{S} - \pi_1$	$\cdots$	$\frac{S_{m_0}}{S} - \pi_m$
Prob.	$\frac{S_1}{S}$	$\cdots$	$\frac{S_{m_0}}{S}$

If $\theta \in \mathcal{L}$, $v_h(\theta) = 0$. Otherwise, $v_h(\theta) < 0$ and for any compact set $\mathcal{K} \subset \mathcal{L}^c$, $\sup_{\theta \in \mathcal{L}} v_h(\theta) < 0$.

(A_3) Let $\boldsymbol{x}_{t+1} = (x_{t+1}^{(1)}, \ldots, x_{t+1}^{(\kappa)})$, which is a sample produced by κ independent Markov chains on the product space $\mathbb{X} = \mathcal{X} \times \cdots \times \mathcal{X}$ with the transition kernel

$$\boldsymbol{P}_{\theta_t}(\boldsymbol{x}, \boldsymbol{y}) = P_{\theta_t}(x^{(1)}, y^{(1)}) P_{\theta_t}(x^{(2)}, y^{(2)}) \cdots P_{\theta_t}(x^{(\kappa)}, y^{(\kappa)}),$$

where $P_{\theta_t}(x, y)$ denotes a one-step MH kernel at a given value of θ_t. Under the assumptions that both Θ and $\mathcal{X}$ are compact and the proposal distribution is local positive, it has been shown in Liang *et al.* (2007) that $P_\theta(x, y)$ satisfies the drift condition (A_3). In what follows, we will show that $\boldsymbol{P}_\theta(\boldsymbol{x}, \boldsymbol{y})$ also satisfies (A_3), given that $P_\theta(x, y)$ satisfies (A_3).

To simplify notations, in what follows we will drop the subscript t, denoting x_t by x, $\boldsymbol{x}_t$ by $\boldsymbol{x}$, and $\theta_t = (\theta_{t1}, \ldots, \theta_{tm})$ by $\theta = (\theta_1, \ldots, \theta_m)$. Roberts and Tweedie (1996) (Theorem 2) show that if the target distribution is bounded away from 0 and ∞ on every compact set of $\mathcal{X}$, then the MH chain with a proposal distribution satisfying the local positive condition is irreducible and aperiodic, and every nonempty compact set is small. It follows from this result that $P_\theta(x,y)$ is irreducible and aperiodic, and thus $\boldsymbol{P}_\theta(x,y)$ is also irreducible and aperiodic.

If $\mathcal{X}$ is compact, and furthermore $f(x)$ is bounded away from 0 and ∞, by equation (20) of the main text, $f_\theta(x)$ is uniformly bounded away from 0 and ∞ since Θ is compact. By Roberts and Tweedie's arguments, these imply that $\mathcal{X}$ is a small set and the minorization condition uniformly holds on $\mathcal{X}$ for all kernel $P_\theta(x,y)$, $\theta \in \Theta$; i.e., there exist a constant δ and a probability measure $\nu'(\cdot)$ such that

$$P_\theta(x,A) \geq \delta'\nu'(A), \quad \forall x \in \mathcal{X},\ \forall A \in \mathcal{B}_{\mathcal{X}}.$$

Therefore,

$$\boldsymbol{P}_\theta(\boldsymbol{x},\boldsymbol{A}) \geq \delta\nu(\boldsymbol{A}), \quad \forall \boldsymbol{x} \in \mathbb{X},\ \forall \boldsymbol{A} \in \mathcal{B}_{\mathbb{X}},$$

where $\boldsymbol{A} = A_1 \times A_2 \times \ldots \times A_\kappa$, $\delta = (\delta')^\kappa$, and $\nu(\boldsymbol{A}) = \nu'(A_1) \times \nu'(A_2) \times \ldots \times \nu'(A_\kappa)$. Hence, ($A_3$-i) is satisfied.

For Pop-SAMC, we have $\boldsymbol{H}(\theta,\boldsymbol{x}) = \sum_{i=1}^{\kappa} H(\theta, x^{(i)})/\kappa$. Since each component of $\boldsymbol{H}(\theta,\boldsymbol{x})$ takes a value between 0 and 1, there exists a constant $c_1 = \sqrt{m}$ such that for any $\theta \in \Theta$ and all $\boldsymbol{x} \in \mathbb{X}$,

$$\|\boldsymbol{H}(\theta,\boldsymbol{x})\| \leq c_1. \tag{6}$$

Also, $\boldsymbol{H}(\theta,\boldsymbol{x})$ does not depend on θ for a given sample $\boldsymbol{x}$. Hence, $\boldsymbol{H}(\theta,\boldsymbol{x}) - \boldsymbol{H}(\theta',\boldsymbol{x}) = 0$ for all $(\theta,\theta') \in \Theta \times \Theta$, and the following condition holds,

$$\|\boldsymbol{H}(\theta,\boldsymbol{x}) - \boldsymbol{H}(\theta',\boldsymbol{x})\| \leq c_1\|\theta - \theta'\|, \tag{7}$$

for all $(\theta,\theta') \in \Theta \times \Theta$. Equations (6) and (7) imply that (A_3-ii) is satisfied.

In Liang *et al.* (2007), it has been shown for the single-chain MH kernel that there exists a constant c_2 such that

$$|P_\theta(x,A) - P_{\theta'}(x,A)| \leq c_2\|\theta - \theta'\|, \tag{8}$$

for any measurable set $A \subset \mathcal{X}$. Therefore, there exists a constant c_3 such that

$$
\begin{aligned}
&|\boldsymbol{P}_\theta(\boldsymbol{x},\boldsymbol{A}) - \boldsymbol{P}_{\theta'}(\boldsymbol{x},\boldsymbol{A})| \\
=&\Big| \int_{A_1} \cdots \int_{A_\kappa} \big[P_\theta(x^{(1)},y^{(1)})P_\theta(x^{(2)},y^{(2)})\cdots P_\theta(x^{(\kappa)},y^{(\kappa)}) \\
&- P_{\theta'}(x^{(1)},y^{(1)})P_{\theta'}(x^{(2)},y^{(2)})\cdots P_{\theta'}(x^{(\kappa)},y^{(\kappa)})\big] dy^{(1)}\cdots dy^{(\kappa)}\Big| \\
\le& \sum_{i=1}^{\kappa} \int_{\mathcal{X}} \cdots \int_{\mathcal{X}} \int_{A_i} \int_{\mathcal{X}} \cdots \int_{\mathcal{X}} P_{\theta'}(x^{(1)},y^{(1)})\cdots P_{\theta'}(x^{(i-1)},y^{(i-1)}) \\
&\times \left| P_\theta(x^{(i)},y^{(i)}) - P_{\theta'}(x^{(i)},y^{(i)}) \right| \\
&\times P_\theta(x^{(i+1)},y^{(i+1)})\cdots P_\theta(x^{(\kappa)},y^{(\kappa)}) dy^{(1)}\cdots dy^{(\kappa)} \\
&\le c_3\|\theta-\theta'\|,
\end{aligned}
$$

which implies A_3-(iii) is satisfied.

(A_4) This condition is automatically satisfied by the choice of $\{\gamma_t\}$.

2.2. Proof of Theorem 3.2.

Following from Theorem 2.2 and Theorem 3.1, this theorem can be proved by verifying that SAMC and Pop-SAMC satisfy (A_2). To verify (A_2), we first show that $h(\theta)$ has bounded first and second derivatives. Continuing the calculation in (3), we have

$$
\frac{\partial^2(\frac{S_i}{S})}{\partial(\theta^{(i)})^2} = \frac{S_i}{S}(1-\frac{S_i}{S})(1-\frac{2S_i}{S}), \quad \frac{\partial^2(\frac{S_i}{S})}{\partial\theta^{(j)}\partial\theta^{(i)}} = -\frac{S_iS_j}{S^2}(1-\frac{2S_i}{S}), \tag{9}
$$

where S and S_i are as defined in (2). This implies that the first and second derivatives of $h(\theta)$ are uniformly bounded by noting the inequality $0 < \frac{S_i}{S} < 1$. Hence, $h(\theta)$ is differentiable and its derivative is Lipschitz continuous.

Let $F = \partial h(\theta)/\partial\theta$. From (3) and (9), we have $F = (\mathbf{1}\mathbf{1}^T + I)F_0$, where

$$
F_0 = \begin{pmatrix}
-\frac{S_1}{S}(1-\frac{S_1}{S}) & \frac{S_1S_2}{S^2} & \cdots & \frac{S_1S_{m_0-1}}{S^2} \\
\frac{S_2S_1}{S^2} & -\frac{S_2}{S}(1-\frac{S_2}{S}) & \cdots & \frac{S_2S_{m_0-1}}{S^2} \\
\vdots & \ddots & \vdots & \vdots \\
\frac{S_{m_0-1}S_1}{S^2} & \cdots & \cdots & -\frac{S_{m_0-1}}{S}(1-\frac{S_{m_0-1}}{S})
\end{pmatrix}.
$$

Thus, for any nonzero vector $\boldsymbol{z} = (z_1, \ldots, z_{m_0-1})^T$,

$$\boldsymbol{z}^T F_0 \boldsymbol{z} = -\Big[\sum_{i=1}^{m_0} z_i^2 \frac{S_i}{S} - \left(\sum_{i=1}^{m_0} z_i \frac{S_i}{S}\right)^2\Big] = -\operatorname{var}(Z) < 0, \qquad (10)$$

where $z_{m_0} = 0$, and $\operatorname{var}(Z)$ denotes the variance of the discrete distribution defined by the following table (note that $\operatorname{var}(Z)$ is strict positive here):

State (Z)	z_1	$\cdots$	z_{m_0}
Prob.	$\frac{S_1}{S}$	$\cdots$	$\frac{S_{m_0}}{S}$

Thus, the matrix F_0 is negative definite, $\mathbf{1}\mathbf{1}^T + I$ is positive definite, by Duan and Patton (1998), F is stable. Applying Taylor expansion to $h(\theta)$ at a point θ_*, we have

$$\|h(\theta) - F(\theta - \theta_*)\| \le c\|\theta - \theta_*\|^2,$$

for some value $c > 0$. Therefore, (A_2) is satisfied by both SAMC and Pop-SAMC.

References

Duan, G.-R. and Patton, R.J. (1998). A Note on Hurwitz Stability of Matrices. *Automatica*, **34**, 509-511.

Folland, G.B. (1990). Remainder estimates in Taylor's theorem. *American Mathematical Monthly*, **97**, 233-235.

Roberts, G.O., and Tweedie, R.L. (1996). Geometric Convergence and Central Limit Theorems for Multidimensional Hastings and metropolis Algorithms. *Biometrika*, **83**, 95-110.