Central limit theorems for adaptive MCMC: some old and new results

Yves Atchadé
(Based partly on joint work with Gersende Fort)

Department of Statistics
University of Michigan

March 17, 2009

Introduction

Markov Chain Monte Carlo (MCMC) is a popular tool for Monte Carlo simulations.

- $(\mathcal{X}, \mathcal{B}, \lambda)$ a meas. space. Want to sample from $\pi(x)\lambda(dx)$.
- MCMC: a recipe to construct ergodic Markov chains $\{X_n,\ n \geq 0\}$ with state space $(\mathcal{X}, \mathcal{B})$ and with invariant distribution π.

Introduction

- Central limit theorems play an important role in MCMC.
  1. Quantify uncertainty in MCMC estimates.
  2. Comparing algorithms.
  3. Serving as a stopping rule.
- There are many CLT for Markov chains (See e.g. G. Jones' review).

Introduction

- Adaptive MCMC: you update the transition kernel (TK) of the sampler as it runs.
- Can be used to improve performances in a variety of situations. Can be useful for MCMC softwares.
- What do we know about central limit theorems for adaptive MCMC?

Introduction

Outline

- CLT for fixed-target AMCMC
- CLT for the Equi-Energy sampler

Fixed-target AMCMC

A large class of adaptive MCMC algorithms can be described as follows.

- Given π on $(\mathcal{X}, \mathcal{B}, \lambda)$, let $\{P_\theta,\ \theta \in \Theta\}$ a family of TK. P_θ is inv. wrt π for all $\theta \in \Theta$.
- Define some optimality criterion.

Definition

An adaptive Markov chain is a random process $\{(X_n, \theta_n), \mathcal{F}_n,\ n \geq 0\}$ such that

$$X_{n+1} | \mathcal{F}_n \sim P_{\theta_n}(X_n, \cdot), \quad \text{and } \theta_{n+1} \in \mathcal{F}_{n+1}.$$

Fixed-target AMCMC

A large class of adaptive MCMC algorithms can be described as follows.

- Given π on $(\mathcal{X}, \mathcal{B}, \lambda)$, let $\{P_\theta,\ \theta \in \Theta\}$ a family of TK. P_θ is inv. wrt π for all $\theta \in \Theta$.
- Define some optimality criterion.

Definition

An adaptive Markov chain is a random process $\{(X_n, \theta_n), \mathcal{F}_n,\ n \geq 0\}$ such that

$$X_{n+1} | \mathcal{F}_n \sim P_{\theta_n}(X_n, \cdot), \quad \text{and } \theta_{n+1} \in \mathcal{F}_{n+1}.$$

Fixed-target AMCMC

Many possibilities for the optimality criterion.

- Maximize the square-jump distance as in the example (Prasarica & Gelman (2005), Andrieu & Robert (2001)).
- Target a given acceptance rate Andrieu & Robert (2001)), (Atchade & Rosenthal (2005)).
- Moment matching (Haario et al. (2001)).
- Minimize the Kulback-Leibler between the proposal and the target (Andrieu & Moulines (2005), Holden et al. (2009), Giordani & Kohn (2008)).

Fixed-target AMCMC

Many possibilities for the optimality criterion.

- Maximize the square-jump distance as in the example (Prasarica & Gelman (2005), Andrieu & Robert (2001)).
- Target a given acceptance rate Andrieu & Robert (2001)), (Atchade & Rosenthal (2005)).
- Moment matching (Haario et al. (2001)).
- Minimize the Kulback-Leibler between the proposal and the target (Andrieu & Moulines (2005), Holden et al. (2009), Giordani & Kohn (2008)).

Fixed-target AMCMC

Many possibilities for the optimality criterion.

- Maximize the square-jump distance as in the example (Prasarica & Gelman (2005), Andrieu & Robert (2001)).
- Target a given acceptance rate Andrieu & Robert (2001)), (Atchade & Rosenthal (2005)).
- Moment matching (Haario et al. (2001)).
- Minimize the Kulback-Leibler between the proposal and the target (Andrieu & Moulines (2005), Holden et al. (2009), Giordani & Kohn (2008)).

Fixed-target AMCMC

Many possibilities for the optimality criterion.

- Maximize the square-jump distance as in the example (Prasarica & Gelman (2005), Andrieu & Robert (2001)).
- Target a given acceptance rate Andrieu & Robert (2001)), (Atchade & Rosenthal (2005)).
- Moment matching (Haario et al. (2001)).
- Minimize the Kulback-Leibler between the proposal and the target (Andrieu & Moulines (2005), Holden et al. (2009), Giordani & Kohn (2008)).

Fixed-target AMCMC

Let $\{(X_n, \theta_n), \mathcal{F}_n, \; n \geq 0\}$ the adaptive chain.

$$X_{n+1} | \mathcal{F}_n \sim P_{\theta_n}(X_n, \cdot), \quad \text{and } \theta_{n+1} = \theta_n + \gamma_n H(\theta_n, X_{n+1}).$$

Objectives: We want conditions under which $n^{-1/2} \sum_{k=1}^{n} \bar{f}(X_k) \Rightarrow Z$, $\bar{f} = f - \pi(f)$.

Definition

$$D_\beta(\theta, \theta') := \sup_{x \in \mathcal{X}} \sup_{|f|_{V^\beta} \leq 1} \frac{|P_\theta f(x) - P_{\theta'} f(x)|}{V^\beta(x)}.$$

Fixed-target AMCMC

Let $\{(X_n, \theta_n), \mathcal{F}_n,\ n \geq 0\}$ the adaptive chain.

$$X_{n+1}|\mathcal{F}_n \sim P_{\theta_n}(X_n, \cdot), \quad \text{and } \theta_{n+1} = \theta_n + \gamma_n H(\theta_n, X_{n+1}).$$

Objectives: We want conditions under which $n^{-1/2} \sum_{k=1}^{n} \bar{f}(X_k) \Rightarrow Z$, $\bar{f} = f - \pi(f)$.

Definition

$$D_\beta(\theta, \theta') := \sup_{x \in \mathcal{X}} \sup_{|f|_{V^\beta} \leq 1} \frac{|P_\theta f(x) - P_{\theta'} f(x)|}{V^\beta(x)}.$$

Fixed-target AMCMC

Theorem (Andrieu-Moulines (2005))

Assume

- (A1) *P_θ is invariant wrt π, ϕ-irreducible, aperiodic and there exist a 1-small set C, $V : \mathcal{X} \to [1, +\infty)$, $\lambda \in (0, 1)$ and constants b such that for any $\theta \in \Theta$, $P_\theta V \leq \lambda V + b\mathbf{1}_C$.*
- (A2) $D_\beta(\theta, \theta') \leq C|\theta - \theta'|, \quad \theta, \theta' \in \Theta.$
- (A3) $\gamma_n = O(n^{-1})$ *and* $\sup_{\theta \in \Theta} |H(\theta, \cdot)|_{V^\alpha} < \infty.$
- (A4) $|\theta_n - \theta_\star| \xrightarrow{Prob} 0.$

Then for $|f|_{V^\beta} < \infty$ $(2\beta + \alpha < 1)$, $n^{-1/2} \sum_{k=1}^n \bar{f}(X_k) \Rightarrow N(0, \sigma^2(f))$ *where*

$$\sigma^2(f) = \pi(|\bar{f}|^2) + 2\sum_{j \geq 1} \pi\left(\bar{f} P^j_{\theta_\star} \bar{f}\right).$$

Fixed-target AMCMC

Theorem (Andrieu-Moulines (2005))

Assume

(A1) P_θ *is invariant wrt* π, ϕ*-irreducible, aperiodic and there exist a* 1*-small set* C, $V : \mathcal{X} \to [1, +\infty)$, $\lambda \in (0, 1)$ *and constants b such that for any* $\theta \in \Theta$, $P_\theta V \leq \lambda V + b\mathbf{1}_C$.

(A2) $D_\beta(\theta, \theta') \leq C|\theta - \theta'|, \quad \theta, \theta' \in \Theta.$

(A3) $\gamma_n = O(n^{-1})$ *and* $\sup_{\theta\in\Theta} |H(\theta, \cdot)|_{V^\alpha} < \infty$.

(A4) $|\theta_n - \theta_\star| \overset{Prob}{\to} 0.$

Then for $|f|_{V^\beta} < \infty$ $(2\beta + \alpha < 1)$, $n^{-1/2} \sum_{k=1}^n \bar{f}(X_k) \Rightarrow N(0, \sigma^2(f))$ *where*

$$\sigma^2(f) = \pi(|\bar{f}|^2) + 2\sum_{j\geq 1} \pi\left(\bar{f} P^j_{\theta_\star} \bar{f}\right).$$

Fixed-target AMCMC

Theorem (Andrieu-Moulines (2005))

Assume

(A1) P_θ *is invariant wrt* π, ϕ*-irreducible, aperiodic and there exist a* 1*-small set* C, $V : \mathcal{X} \to [1, +\infty)$, $\lambda \in (0,1)$ *and constants b such that for any* $\theta \in \Theta$, $P_\theta V \leq \lambda V + b\mathbf{1}_C$.

(A2) $D_\beta(\theta, \theta') \leq C|\theta - \theta'|, \quad \theta, \theta' \in \Theta$.

(A3) $\gamma_n = O(n^{-1})$ *and* $\sup_{\theta\in\Theta} |H(\theta, \cdot)|_{V^\alpha} < \infty$.

(A4) $|\theta_n - \theta_\star| \overset{Prob}{\to} 0$.

Then for $|f|_{V^\beta} < \infty$ *(*$2\beta + \alpha < 1$*),* $n^{-1/2} \sum_{k=1}^n \bar{f}(X_k) \Rightarrow N(0, \sigma^2(f))$ *where*

$$\sigma^2(f) = \pi(|\bar{f}|^2) + 2\sum_{j\geq 1} \pi\left(\bar{f} P^j_{\theta_\star} \bar{f}\right).$$

Fixed-target AMCMC

Theorem (Andrieu-Moulines (2005))

Assume

(A1) P_θ *is invariant wrt* π, ϕ*-irreducible, aperiodic and there exist a* 1*-small set* C, $V : \mathcal{X} \to [1, +\infty)$, $\lambda \in (0, 1)$ *and constants* b *such that for any* $\theta \in \Theta$, $P_\theta V \leq \lambda V + b\mathbf{1}_C$.

(A2) $D_\beta(\theta, \theta') \leq C|\theta - \theta'|, \quad \theta, \theta' \in \Theta.$

(A3) $\gamma_n = O(n^{-1})$ *and* $\sup_{\theta\in\Theta} |H(\theta, \cdot)|_{V^\alpha} < \infty$.

(A4) $|\theta_n - \theta_\star| \overset{Prob}{\to} 0.$

Then for $|f|_{V^\beta} < \infty$ $(2\beta + \alpha < 1)$, $n^{-1/2} \sum_{k=1}^{n} \bar{f}(X_k) \Rightarrow N(0, \sigma^2(f))$ *where*

$$\sigma^2(f) = \pi(|\bar{f}|^2) + 2\sum_{j\geq 1} \pi\left(\bar{f} P^j_{\theta_\star} \bar{f}\right).$$

Fixed-target AMCMC

Theorem (Andrieu-Moulines (2005))

Assume

- (A1) P_θ *is invariant wrt* π, ϕ*-irreducible, aperiodic and there exist a* 1*-small set* C, $V : \mathcal{X} \to [1, +\infty)$, $\lambda \in (0,1)$ *and constants* b *such that for any* $\theta \in \Theta$, $P_\theta V \leq \lambda V + b\mathbf{1}_C$.
- (A2) $D_\beta(\theta, \theta') \leq C|\theta - \theta'|, \quad \theta, \theta' \in \Theta.$
- (A3) $\gamma_n = O(n^{-1})$ *and* $\sup_{\theta\in\Theta} |H(\theta, \cdot)|_{V^\alpha} < \infty.$
- (A4) $|\theta_n - \theta_\star| \overset{Prob}{\to} 0.$

Then for $|f|_{V^\beta} < \infty$ $(2\beta + \alpha < 1)$, $n^{-1/2}\sum_{k=1}^n \bar{f}(X_k) \Rightarrow N(0, \sigma^2(f))$ *where*

$$\sigma^2(f) = \pi(|\bar{f}|^2) + 2\sum_{j\geq 1} \pi\left(\bar{f} P^j_{\theta_\star} \bar{f}\right).$$

Fixed-target AMCMC

There are many MCMC algorithms that are not geometrically ergodic.

1. If

$$\int e^{s|x|}\pi(dx) = \infty, \qquad \text{for all } s > 0$$

the Random Walk Metropolis algorithm cannot be geometrically ergodic. (Jarner & Tweedie (2001)).

2. If $\liminf_{|x|\to\infty} \frac{|\nabla \log \pi(x)|}{|x|} = \infty$ or $\liminf_{|x|\to\infty} |\nabla \log \pi(x)| = 0$, then the Metropolis Adjusted Langevin algorithm cannot be geometrically ergodic. (Roberts & Tweedie (1996)).

Fixed-target AMCMC

There are many MCMC algorithms that are not geometrically ergodic.

1. If
$$\int e^{s|x|}\pi(dx) = \infty, \qquad \text{for all } s > 0$$
the Random Walk Metropolis algorithm cannot be geometrically ergodic. (Jarner & Tweedie (2001)).

2. If $\liminf_{|x|\to\infty} \frac{|\nabla \log \pi(x)|}{|x|} = \infty$ or $\liminf_{|x|\to\infty} |\nabla \log \pi(x)| = 0$, then the Metropolis Adjusted Langevin algorithm cannot be geometrically ergodic. (Roberts & Tweedie (1996)).

Fixed-target AMCMC

Theorem (Atchade-Fort (2009))

Assume

(B1) P_θ *is invariant wrt* π, ϕ*-irreducible, aperiodic and there exist a* 1*-small set* C, $V : \mathcal{X} \to [1, +\infty)$, $\alpha \in (0, 1/2)$ *and constants* b, c *such that for any* $\theta \in \Theta$
$P_\theta V \leq V - cV^{1-\alpha}(x) + b\mathbf{1}_C$.

(B2) $D_\beta(\theta, \theta') \leq C|\theta - \theta'|, \quad \theta, \theta' \in \Theta$.

(B3) $\gamma_n = O(n^{-1})$ *and* $\sup_{(x,\theta) \in \mathcal{X} \times \Theta} |H(\theta, x)| < \infty$.

(B4) $|\theta_n - \theta_\star| \stackrel{Prob}{\to} 0$.

Then for $|f| \leq C$, $n^{-1/2} \sum_{k=1}^{n} \bar{f}(X_k) \Rightarrow N(0, \sigma^2(f))$ *where*

$$\sigma^2(f) = \pi(|\bar{f}|^2) + 2 \sum_{j \geq 1} \pi\left(\bar{f} P_{\theta_\star}^j \bar{f}\right).$$

Fixed-target AMCMC

Theorem (Atchade-Fort (2009))

Assume

(B1) P_θ *is invariant wrt* π, ϕ*-irreducible, aperiodic and there exist a* 1*-small set* C, $V : \mathcal{X} \to [1, +\infty)$, $\alpha \in (0, 1/2)$ *and constants* b, c *such that for any* $\theta \in \Theta$
$P_\theta V \leq V - cV^{1-\alpha}(x) + b\mathbf{1}_C$.

(B2) $D_\beta(\theta, \theta') \leq C|\theta - \theta'|, \quad \theta, \theta' \in \Theta$.

(B3) $\gamma_n = O(n^{-1})$ *and* $\sup_{(x,\theta)\in\mathcal{X}\times\Theta} |H(\theta, x)| < \infty$.

(B4) $|\theta_n - \theta_\star| \stackrel{Prob}{\to} 0$.

Then for $|f| \leq C$, $n^{-1/2} \sum_{k=1}^n \bar{f}(X_k) \Rightarrow N(0, \sigma^2(f))$ *where*

$$\sigma^2(f) = \pi(|\bar{f}|^2) + 2\sum_{j\geq 1} \pi\left(\bar{f} P^j_{\theta_\star} \bar{f}\right).$$

Fixed-target AMCMC

Theorem (Atchade-Fort (2009))

Assume

(B1) P_θ *is invariant wrt* π, ϕ*-irreducible, aperiodic and there exist a* 1*-small set* C, $V : \mathcal{X} \to [1, +\infty)$, $\alpha \in (0, 1/2)$ *and constants* b, c *such that for any* $\theta \in \Theta$
$P_\theta V \le V - cV^{1-\alpha}(x) + b\mathbf{1}_C$.

(B2) $D_\beta(\theta, \theta') \le C|\theta - \theta'|, \quad \theta, \theta' \in \Theta$.

(B3) $\gamma_n = O(n^{-1})$ *and* $\sup_{(x,\theta)\in\mathcal{X}\times\Theta} |H(\theta, x)| < \infty$.

(B4) $|\theta_n - \theta_\star| \overset{Prob}{\to} 0$.

Then for $|f| \le C$, $n^{-1/2}\sum_{k=1}^n \bar{f}(X_k) \Rightarrow N(0, \sigma^2(f))$ *where*

$$\sigma^2(f) = \pi(|\bar{f}|^2) + 2\sum_{j\ge1} \pi\left(\bar{f} P_{\theta_\star}^j \bar{f}\right).$$

Fixed-target AMCMC

Theorem (Atchade-Fort (2009))

Assume

(B1) P_θ *is invariant wrt* π, ϕ*-irreducible, aperiodic and there exist a* 1*-small set* C, $V : \mathcal{X} \to [1, +\infty)$, $\alpha \in (0, 1/2)$ *and constants* b, c *such that for any* $\theta \in \Theta$
$P_\theta V \leq V - cV^{1-\alpha}(x) + b\mathbf{1}_C$.

(B2) $D_\beta(\theta, \theta') \leq C|\theta - \theta'|, \quad \theta, \theta' \in \Theta$.

(B3) $\gamma_n = O(n^{-1})$ *and* $\sup_{(x,\theta)\in\mathcal{X}\times\Theta} |H(\theta, x)| < \infty$.

(B4) $|\theta_n - \theta_\star| \stackrel{Prob}{\to} 0$.

Then for $|f| \leq C$, $n^{-1/2} \sum_{k=1}^{n} \bar{f}(X_k) \Rightarrow N(0, \sigma^2(f))$ *where*

$$\sigma^2(f) = \pi(|\bar{f}|^2) + 2\sum_{j\geq 1} \pi\left(\bar{f} P_{\theta_\star}^j \bar{f}\right).$$

Fixed-target AMCMC

Theorem (Atchade-Fort (2009))

Assume

(B1) P_θ *is invariant wrt* π, ϕ*-irreducible, aperiodic and there exist a* 1*-small set* C, $V : \mathcal{X} \to [1, +\infty)$, $\alpha \in (0, 1/2)$ *and constants* b, c *such that for any* $\theta \in \Theta$
$P_\theta V \leq V - cV^{1-\alpha}(x) + b\mathbf{1}_C$.

(B2) $D_\beta(\theta, \theta') \leq C|\theta - \theta'|, \quad \theta, \theta' \in \Theta$.

(B3) $\gamma_n = O(n^{-1})$ *and* $\sup_{(x,\theta)\in\mathcal{X}\times\Theta} |H(\theta, x)| < \infty$.

(B4) $|\theta_n - \theta_\star| \overset{Prob}{\to} 0$.

Then for $|f| \leq C$, $n^{-1/2} \sum_{k=1}^n \bar{f}(X_k) \Rightarrow N(0, \sigma^2(f))$ *where*

$$\sigma^2(f) = \pi(|\bar{f}|^2) + 2 \sum_{j\geq 1} \pi\left(\bar{f} P_{\theta_\star}^j \bar{f}\right).$$

Fixed-target AMCMC

Practical consequence:
In a MCMC simulation problem where uniform drift condition of the form

$$P_\theta V(x) \leq V(x) - \phi \circ V(x) + b\mathbf{1}_C(x)$$

is available, we recommend adaptive MCMC.

- In such cases, AMCMC algorithms are very stable.
- When good ideas are available on how to adapt, AMCMC perform better than plain MCMC.

Fixed-target AMCMC

Practical consequence:
In a MCMC simulation problem where uniform drift condition of the form

$$P_\theta V(x) \leq V(x) - \phi \circ V(x) + b\mathbf{1}_C(x)$$

is available, we recommend adaptive MCMC.

- In such cases, AMCMC algorithms are very stable.
- When good ideas are available on how to adapt, AMCMC perform better than plain MCMC.

Fixed-target AMCMC

A quick comment on the proof:

- Similar to the proof of Andrieu & Moulines (2005).
- Based on Poisson equation.

Fixed-target AMCMC

A quick comment on the proof:

- Similar to the proof of Andrieu & Moulines (2005).
- Based on Poisson equation.

Fixed-target AMCMC

- Let $\{Z_n\}$ be a Markov chain with t.k. P and inv. dist. π. Let f s.t. $\pi(f) = 0$. Notation: $Pf(x) := \int P(x, dy)f(y)$.
- Suppose that

$$g = \sum_{k=0}^{\infty} P^k f,$$

exists. Then

$$g(x) - Pg(x) = f(x).$$

- We can use g to approximate partial sum of Markov chains by martingales.

Fixed-target AMCMC

In the non-geometric case, solutions to the Poisson equation are hard to work with.

- If $PV \leq V - \chi + b\mathbf{1}_C \Rightarrow f \leq \chi$ implies $|g| \leq V$.
- If $P_\theta V \leq V - V^{1-\alpha} + b\mathbf{1}_C \Rightarrow$ for $|f| \leq V^\beta$, $|g| \leq V^{\beta+\alpha}$.

$$g_\theta - g_{\theta'} = \sum_{j\geq 0} P_\theta^j \circ (P_\theta - P_{\theta'}) \circ \sum_{j\geq 0} P_{\theta'}^j f(x).$$

So that

$$|g_\theta - g_{\theta'}| \leq V^{\beta+2\alpha}.$$

Fixed-target AMCMC

In the non-geometric case, solutions to the Poisson equation are hard to work with.

- If $PV \leq V - \chi + b\mathbf{1}_C \Rightarrow f \leq \chi$ implies $|g| \leq V$.
- If $P_\theta V \leq V - V^{1-\alpha} + b\mathbf{1}_C \Rightarrow$ for $|f| \leq V^\beta$, $|g| \leq V^{\beta+\alpha}$.

$$g_\theta - g_{\theta'} = \sum_{j\geq 0} P_\theta^j \circ (P_\theta - P_{\theta'}) \circ \sum_{j\geq 0} P_{\theta'}^j f(x).$$

So that

$$|g_\theta - g_{\theta'}| \leq V^{\beta+2\alpha}.$$

Fixed-target AMCMC

In the non-geometric case, solutions to the Poisson equation are hard to work with.

- If $PV \leq V - \chi + b\mathbf{1}_C \Rightarrow f \leq \chi$ implies $|g| \leq V$.
- If $P_\theta V \leq V - V^{1-\alpha} + b\mathbf{1}_C \Rightarrow$ for $|f| \leq V^\beta$, $|g| \leq V^{\beta+\alpha}$.

$$g_\theta - g_{\theta'} = \sum_{j\geq 0} P_\theta^j \circ (P_\theta - P_{\theta'}) \circ \sum_{j\geq 0} P_{\theta'}^j f(x).$$

So that

$$|g_\theta - g_{\theta'}| \leq V^{\beta+2\alpha}.$$

Fixed-target AMCMC

The solution is to work with $H(x, y) = g(y) - Pg(x)$.

Theorem (Maxwell-Woodroofe (2000))

Define $V_n(f) = \left\| \sum_{k=0}^{n-1} P^k f \right\|_{L^2(\pi)}$. *If*

$$\sum_{k \geq 0} n^{-3/2} V_n(f) < \infty,$$

Then $H_n(x, y) = \sum_{j=0}^{n-1} P^j f(y) - \sum_{j=0}^{n} P^j f(x)$ *converges to a limit* H $(H = g(y) - Pg(x))$ *in* $L^2(\pi \times P)$.

Kipnis-Varadhan (1986) has a similar result for reversible chains.

Fixed-target AMCMC

Proposition

Assume $P_\theta V \leq V - V^{1-\alpha} + b\mathbf{1}_C$. *For* $f \in L_{V^\beta}$, $\beta \in [0, 1-2\alpha)$, *define* $H_\theta(x,y) = g_\theta(y) - P_\theta g_\theta(x)$.

$$\sup_{x,y} \frac{|H_\theta(x,y) - H_{\theta'}(x,y)|}{V^{\beta+\alpha\kappa}(x) + V^{\beta+\alpha\kappa}(y)} \leq C|\theta - \theta'|.$$

for any $\kappa > 1$.

Equi-Energy sampler

1. Let us now consider another class of adaptive MCMC where the transition kernels do not all have the same invariant dist.
2. There are actually many such algorithms (Equi-Energy sampler, Wang-Landau, SAMC...).
3. We focus on the Equi-Energy sampler (EE).

Equi-Energy sampler

1. Let us now consider another class of adaptive MCMC where the transition kernels do not all have the same invariant dist.
2. There are actually many such algorithms (Equi-Energy sampler, Wang-Landau, SAMC...).
3. We focus on the Equi-Energy sampler (EE).

Equi-Energy sampler

$(\mathcal{X}, \mathcal{B})$ a meas. space. π a prob. meas. and P a TK inv. wrt to π. Θ now denotes the space of all prob. meas. on $(\mathcal{X}, \mathcal{B})$

- the Equi-Energy sampler is a very clever MCMC algorithm by Kou & Zhou & Wong (2005).
- Can be looked at as an adaptive MCMC with infinite dimensional parameter space.

Equi-Energy sampler

$(\mathcal{X}, \mathcal{B})$ a meas. space. π a prob. meas. and P a TK inv. wrt to π. Θ now denotes the space of all prob. meas. on $(\mathcal{X}, \mathcal{B})$

- the Equi-Energy sampler is a very clever MCMC algorithm by Kou & Zhou & Wong (2005).
- Can be looked at as an adaptive MCMC with infinite dimensional parameter space.

Equi-Energy sampler

Let $\bar{\pi}$ a prob. meas. that is equivalent to π and $\omega(x) = \pi(dx)/\bar{\pi}(dx)$. Let P with inv. dist. π. Define

$$\alpha(x, y) = \min\left(1, \frac{\omega(y)}{\omega(x)}\right).$$

For $\varepsilon \in (0, 1)$ and $\theta \in \Theta$ a prob. meas. define

$$P_\theta(x, A) = (1-\varepsilon)P(x, A) + \varepsilon \int \theta(dz) \left[\alpha(x, z)\mathbf{1}_A(z) + (1 - \alpha(x, z))\,\mathbf{1}_A(x)\right].$$

Key insight of Equi-Energy:

(i) $P_{\bar{\pi}}(x, A)$ is invariant with respect to π.

(ii) $P_{\bar{\pi}}$ will have a very good convergence rate if $\bar{\pi}$ is close to π.

Equi-Energy sampler

Let $\bar{\pi}$ a prob. meas. that is equivalent to π and $\omega(x) = \pi(dx)/\bar{\pi}(dx)$. Let P with inv. dist. π. Define

$$\alpha(x, y) = \min\left(1, \frac{\omega(y)}{\omega(x)}\right).$$

For $\varepsilon \in (0, 1)$ and $\theta \in \Theta$ a prob. meas. define

$$P_\theta(x, A) = (1-\varepsilon)P(x, A) + \varepsilon \int \theta(dz) \left[\alpha(x, z)\mathbf{1}_A(z) + (1 - \alpha(x, z))\, \mathbf{1}_A(x)\right].$$

Key insight of Equi-Energy:

(i) $P_{\bar{\pi}}(x, A)$ is invariant with respect to π.

(ii) $P_{\bar{\pi}}$ will have a very good convergence rate if $\bar{\pi}$ is close to π.

Equi-Energy sampler

Let $\bar{\pi}$ a prob. meas. that is equivalent to π and $\omega(x) = \pi(dx)/\bar{\pi}(dx)$. Let P with inv. dist. π. Define

$$\alpha(x, y) = \min\left(1, \frac{\omega(y)}{\omega(x)}\right).$$

For $\varepsilon \in (0, 1)$ and $\theta \in \Theta$ a prob. meas. define

$$P_\theta(x, A) = (1-\varepsilon)P(x, A) + \varepsilon \int \theta(dz) \left[\alpha(x, z)\mathbf{1}_A(z) + (1 - \alpha(x, z))\, \mathbf{1}_A(x)\right].$$

Key insight of Equi-Energy:

(i) $P_{\bar{\pi}}(x, A)$ is invariant with respect to π.

(ii) $P_{\bar{\pi}}$ will have a very good convergence rate if $\bar{\pi}$ is close to π.

Equi-Energy sampler

Let $\bar{\pi}$ a prob. meas. that is equivalent to π and $\omega(x) = \pi(dx)/\bar{\pi}(dx)$. Let P with inv. dist. π. Define

$$\alpha(x, y) = \min\left(1, \frac{\omega(y)}{\omega(x)}\right).$$

For $\varepsilon \in (0, 1)$ and $\theta \in \Theta$ a prob. meas. define

$$P_\theta(x, A) = (1-\varepsilon)P(x, A) + \varepsilon \int \theta(dz) \left[\alpha(x, z)\mathbf{1}_A(z) + (1 - \alpha(x, z))\, \mathbf{1}_A(x)\right].$$

Key insight of Equi-Energy:

(i) $P_{\bar{\pi}}(x, A)$ is invariant with respect to π.

(ii) $P_{\bar{\pi}}$ will have a very good convergence rate if $\bar{\pi}$ is close to π.

Equi-Energy sampler

Implementation: take $\bar{\pi} = \pi^\gamma$, $\gamma \in (0,1)$. Let $\bar{P}$ with inv. dist. $\bar{\pi}$. Run the joint process $\{(\bar{X}_n, X_n, \theta_n),\ n \geq 0\}$ as follows.

Algorithm

Given $\sigma\{(\bar{X}_k, X_k, \theta_k),\ k \leq n\}$

- $$\bar{X}_{n+1} \sim \bar{P}(\bar{X}_n, \cdot), \qquad X_{n+1} \sim P_{\theta_n}(X_n, \cdot),$$
- $$\theta_{n+1} = (n+1)^{-1} \sum_{k=1}^{n+1} \delta_{\bar{X}_k}$$

As $n \to \infty$, P_{θ_n} should converge to $P_{\bar{\pi}}$.

Equi-Energy sampler

Implementation: take $\bar{\pi} = \pi^{\gamma}$, $\gamma \in (0,1)$. Let $\bar{P}$ with inv. dist. $\bar{\pi}$. Run the joint process $\{(\bar{X}_n, X_n, \theta_n),\ n \geq 0\}$ as follows.

Algorithm

Given $\sigma\{(\bar{X}_k, X_k, \theta_k),\ k \leq n\}$

- $$\bar{X}_{n+1} \sim \bar{P}(\bar{X}_n, \cdot), \qquad X_{n+1} \sim P_{\theta_n}(X_n, \cdot),$$
- $$\theta_{n+1} = (n+1)^{-1} \sum_{k=1}^{n+1} \delta_{\bar{X}_k}$$

As $n \to \infty$, P_{θ_n} should converge to $P_{\bar{\pi}}$.

Equi-Energy sampler

Implementation: take $\bar{\pi} = \pi^{\gamma}$, $\gamma \in (0,1)$. Let $\bar{P}$ with inv. dist. $\bar{\pi}$. Run the joint process $\{(\bar{X}_n, X_n, \theta_n),\ n \geq 0\}$ as follows.

Algorithm

Given $\sigma\{(\bar{X}_k, X_k, \theta_k),\ k \leq n\}$

- $$\bar{X}_{n+1} \sim \bar{P}(\bar{X}_n, \cdot), \qquad X_{n+1} \sim P_{\theta_n}(X_n, \cdot),$$
- $$\theta_{n+1} = (n+1)^{-1} \sum_{k=1}^{n+1} \delta_{\bar{X}_k}$$

As $n \to \infty$, P_{θ_n} should converge to $P_{\bar{\pi}}$.

Equi-Energy sampler

- The EE has been studied in Andrieu & Jasra & Del Moral & Doucet (2007) (convergence of the marginal and law of large numbers) and Atchade (2009) (law of large numbers and CLT).
- I would to discuss the CLT. That is condition under which $n^{-1/2}\sum_{k=1}^{n}\bar{f}(X_k) \Rightarrow Z$.

Equi-Energy sampler

- The EE has been studied in Andrieu & Jasra & Del Moral & Doucet (2007) (convergence of the marginal and law of large numbers) and Atchade (2009) (law of large numbers and CLT).
- I would to discuss the CLT. That is condition under which $n^{-1/2}\sum_{k=1}^{n}\bar{f}(X_k) \Rightarrow Z$.

Equi-Energy sampler

Fix f with $\pi(f)=0$. Let U that satisfies the Poisson equation $U - P_{\bar{\pi}}U = f$. Define $S_n = \sum_{k=1}^n f(X_k)$

$$\begin{aligned}
S_n &= \sum_{k=1}^n U(X_k) - P_{\bar{\pi}}U(X_k) \\
&= \left(\sum_{k=1}^n U(X_k) - P_{\theta_{k-1}}U(X_{k-1})\right) + \left(P_{\theta_0}U(X_0) - P_{\theta_n}U(X_n)\right) \\
&\quad + \sum_{k=1}^n P_{\theta_k}U(X_k) - P_{\bar{\pi}}U(X_k) \\
&= M_n + \varepsilon \sum_{k=1}^n \frac{1}{k}\sum_{l=0}^{k-1} H_{X_k}(\bar{X}_l) + o_P\left(\sqrt{n}\right), \\
&= M_n + 2\varepsilon \frac{1}{\sqrt{n}} \sum_{k=1}^n \frac{1}{\sqrt{k}} \sum_{l=0}^{k-1} H_{X_k}(\bar{X}_l) + o_P\left(\sqrt{n}\right),
\end{aligned}$$

for some function $H_x : \mathcal{X} \to \mathbb{R}$.

Equi-Energy sampler

Fix f with $\pi(f) = 0$. Let U that satisfies the Poisson equation $U - P_{\bar{\pi}} U = f$. Define $S_n = \sum_{k=1}^{n} f(X_k)$

$$
\begin{aligned}
S_n &= \sum_{k=1}^{n} U(X_k) - P_{\bar{\pi}} U(X_k) \\
&= \left(\sum_{k=1}^{n} U(X_k) - P_{\theta_{k-1}} U(X_{k-1}) \right) + \left(P_{\theta_0} U(X_0) - P_{\theta_n} U(X_n) \right) \\
&\quad + \sum_{k=1}^{n} P_{\theta_k} U(X_k) - P_{\bar{\pi}} U(X_k) \\
&= M_n + \varepsilon \sum_{k=1}^{n} \frac{1}{k} \sum_{l=0}^{k-1} H_{X_k}(\bar{X}_l) + o_P\left(\sqrt{n}\right), \\
&= M_n + 2\varepsilon \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \frac{1}{\sqrt{k}} \sum_{l=0}^{k-1} H_{X_k}(\bar{X}_l) + o_P\left(\sqrt{n}\right),
\end{aligned}
$$

for some function $H_x : \mathcal{X} \to \mathbb{R}$.

Equi-Energy sampler

Fix f with $\pi(f) = 0$. Let U that satisfies the Poisson equation $U - P_{\bar{\pi}}U = f$. Define $S_n = \sum_{k=1}^{n} f(X_k)$

$$
\begin{aligned}
S_n &= \sum_{k=1}^{n} U(X_k) - P_{\bar{\pi}}U(X_k) \\
&= \left(\sum_{k=1}^{n} U(X_k) - P_{\theta_{k-1}}U(X_{k-1})\right) + \left(P_{\theta_0}U(X_0) - P_{\theta_n}U(X_n)\right) \\
&\quad + \sum_{k=1}^{n} P_{\theta_k}U(X_k) - P_{\bar{\pi}}U(X_k) \\
&= M_n + \varepsilon \sum_{k=1}^{n} \frac{1}{k} \sum_{l=0}^{k-1} H_{X_k}(\bar{X}_l) + o_P\left(\sqrt{n}\right), \\
&= M_n + 2\varepsilon \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \frac{1}{\sqrt{k}} \sum_{l=0}^{k-1} H_{X_k}(\bar{X}_l) + o_P\left(\sqrt{n}\right),
\end{aligned}
$$

for some function $H_x : \mathcal{X} \to \mathbb{R}$.

Equi-Energy sampler

Fix f with $\pi(f) = 0$. Let U that satisfies the Poisson equation $U - P_{\bar{\pi}} U = f$. Define $S_n = \sum_{k=1}^n f(X_k)$

$$
\begin{aligned}
S_n &= \sum_{k=1}^n U(X_k) - P_{\bar{\pi}} U(X_k) \\
&= \left(\sum_{k=1}^n U(X_k) - P_{\theta_{k-1}} U(X_{k-1})\right) + \left(P_{\theta_0} U(X_0) - P_{\theta_n} U(X_n)\right) \\
&\quad + \sum_{k=1}^n P_{\theta_k} U(X_k) - P_{\bar{\pi}} U(X_k) \\
&= M_n + \varepsilon \sum_{k=1}^n \frac{1}{k} \sum_{l=0}^{k-1} H_{X_k}(\bar{X}_l) + o_P\left(\sqrt{n}\right), \\
&= M_n + 2\varepsilon \frac{1}{\sqrt{n}} \sum_{k=1}^n \frac{1}{\sqrt{k}} \sum_{l=0}^{k-1} H_{X_k}(\bar{X}_l) + o_P\left(\sqrt{n}\right),
\end{aligned}
$$

for some function $H_x : \mathcal{X} \to \mathbb{R}$.

Equi-Energy sampler

Fix f with $\pi(f) = 0$. Let U that satisfies the Poisson equation $U - P_{\bar{\pi}} U = f$. Define $S_n = \sum_{k=1}^n f(X_k)$

$$
\begin{aligned}
S_n &= \sum_{k=1}^{n} U(X_k) - P_{\bar{\pi}} U(X_k) \\
&= \left(\sum_{k=1}^{n} U(X_k) - P_{\theta_{k-1}} U(X_{k-1}) \right) + \left(P_{\theta_0} U(X_0) - P_{\theta_n} U(X_n) \right) \\
&\quad + \sum_{k=1}^{n} P_{\theta_k} U(X_k) - P_{\bar{\pi}} U(X_k) \\
&= M_n + \varepsilon \sum_{k=1}^{n} \frac{1}{k} \sum_{l=0}^{k-1} H_{X_k}(\bar{X}_l) + o_P\left(\sqrt{n}\right), \\
&= M_n + 2\varepsilon \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \frac{1}{\sqrt{k}} \sum_{l=0}^{k-1} H_{X_k}(\bar{X}_l) + o_P\left(\sqrt{n}\right),
\end{aligned}
$$

for some function $H_x : \mathcal{X} \to \mathbb{R}$.

Equi-Energy sampler

Suppose that $\mathcal{X}$ is finite. Then

$$\left(\frac{1}{\sqrt{n}}\sum_{l=1}^{n} H_x(\bar{X}_l)\right)_x \Rightarrow G,$$

where G is a Gaussian r.v with zero mean and covariance

$$\Gamma(x,y) = \int \left[U_x(z)U_y(z) - \left(\bar{P}U_x(z)\right)\left(\bar{P}U_y(z)\right)\right]\bar{\pi}(dz),$$

Then

$$\frac{1}{\sqrt{n}}S_n = \frac{1}{\sqrt{n}}M_n + 2\varepsilon\frac{1}{n}\sum_{k=1}^{n} G(X_k) + o_P\left(1\right).$$

Equi-Energy sampler

Theorem

Suppose that $\mathcal{X}$ is finite and $\bar{P}$ and P are ergodic. Let $f : \mathcal{X} \to \mathbb{R}$ bounded.

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n} f(X_k) \Rightarrow Z + 2\varepsilon \sum_{x \in \mathcal{X}} \pi(x) G(x) \quad \text{as} \quad n \to \infty, \tag{1}$$

where Z and $\sum_{x \in \mathcal{X}} \pi(x) G(x)$ are independent random variables and $Z \sim N(0, \sigma_\star^2(f))$, with $\sigma_\star^2(f) := \pi(f^2) + 2 \sum_{k=1}^{\infty} \int_{\mathcal{X}} \pi(dx) f(x) P_{\bar{\pi}}^k f(x)$

Equi-Energy sampler

If $\mathcal{X}$ is not finite, we need empirical process theory for Markov chains (uniform CLT for Markov chains). Suppose that:

1. $\mathcal{X}$ is compact $\bar{P}$ and P are uniformly geometrically ergodic.
2. Suppose that a uniform CLT over the class $\{H_x,\ x \in \mathcal{X}\}$ hold for the Markov chain $\{\bar{X}_n,\ n \geq 0\}$.

Equi-Energy sampler

If $\mathcal{X}$ is not finite, we need empirical process theory for Markov chains (uniform CLT for Markov chains). Suppose that:

1. $\mathcal{X}$ is compact $\bar{P}$ and P are uniformly geometrically ergodic.
2. Suppose that a uniform CLT over the class $\{H_x,\ x \in \mathcal{X}\}$ hold for the Markov chain $\{\bar{X}_n,\ n \geq 0\}$.

Equi-Energy sampler

If $\mathcal{X}$ is not finite, we need empirical process theory for Markov chains (uniform CLT for Markov chains). Suppose that:

1. $\mathcal{X}$ is compact $\bar{P}$ and P are uniformly geometrically ergodic.
2. Suppose that a uniform CLT over the class $\{H_x,\ x \in \mathcal{X}\}$ hold for the Markov chain $\{\bar{X}_n,\ n \geq 0\}$.

Equi-Energy sampler

Theorem

Assume the above. Let $f : \mathcal{X} \to \mathbb{R}$ *a bounded meas. function.*

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n} f(X_k) \Rightarrow Z + 2\varepsilon \int \pi(dx) G(x) \quad \text{as} \quad n \to \infty, \tag{2}$$

where Z *is as above and* G *is a zero-mean Gaussian process on* $\mathcal{X}$ *with covariance function*

$$\Gamma(x, y) = \int \left[U_x(z) U_y(z) - \left(\bar{P} U_x(z) \right) \left(\bar{P} U_y(z) \right) \right] \bar{\pi}(dz),$$

Equi-Energy sampler

Practical implications:

1. In the EE sampler (and more generally for adaptive MCMC with varying invariant distributions) you pay the price of the adaptation.
2. For small problems, the cost of the adaptation (the term $2\varepsilon \int \pi(dx)G(x)$) can make the algorithm inefficient compared with simpler MCMC sampler.
3. To minimize the cost of the adaptation, ε should be kept small.

Equi-Energy sampler

Practical implications:

1. In the EE sampler (and more generally for adaptive MCMC with varying invariant distributions) you pay the price of the adaptation.
2. For small problems, the cost of the adaptation (the term $2\varepsilon \int \pi(dx)G(x)$) can make the algorithm inefficient compared with simpler MCMC sampler.
3. To minimize the cost of the adaptation, ε should be kept small.

Equi-Energy sampler

Practical implications:

1. In the EE sampler (and more generally for adaptive MCMC with varying invariant distributions) you pay the price of the adaptation.
2. For small problems, the cost of the adaptation (the term $2\varepsilon \int \pi(dx)G(x)$) can make the algorithm inefficient compared with simpler MCMC sampler.
3. To minimize the cost of the adaptation, ε should be kept small.