Approximations of geometrically ergodic Markov chains

Daniel Rudolf

University Jena / TU Chemnitz

March 2016, Luminy
(joint work with N. Schweizer)

Markov chain Monte Carlo

Approximate

$$\pi(f) = \int_G f(x)\,\pi(x)\mathrm{d}x.$$

($G = \mathbb{R}^d$ and $\mathrm{d}x$ is the Lebesgue measure.)

Markov chain Monte Carlo:

Construct a Markov chain $(X_n)_{n\in\mathbb{N}}$ with limit distribution π and

$$\frac{1}{n}\sum_{j=1}^{n} f(X_j) \underset{n\to\infty}{\longrightarrow} \pi(f).$$

Typically Metropolis-Hastings algorithm is used.

Metropolis-Hastings (MH) algorithm

$\pi_u(x)$ is the unnormalized density, i.e. $\pi(x) = c\pi_u(x)$.

MH algorithm M with transition from x_n to x_{n+1}:

1. Draw x' from proposal density $q(x_n, \cdot)$;
2. Set
$$x_{n+1} = \begin{cases} x' & \text{with probab.} \quad a(x_n, x') \\ x_n & \text{otherwise} \end{cases}$$
with
$$a(x_n, x') = \min\left\{1, \frac{\pi_u(x')q(x', x_n)}{\pi_u(x_n)q(x_n, x')}\right\}.$$

Latent variables

Assume $\pi_u(x)$ cannot be computed but

$$\pi_u(x) = \int_T \widehat{\pi}_u(x,t)\,\theta_x(\mathrm{d}t).$$

(see Andrieu, Roberts 2009, Andrieu, Vihola 2015)

Substitute $\pi_u(x)$ in the MH algorithm by an unbiased approximation,

$$\pi_{u,N}(x) = \frac{1}{N}\sum_{i=1}^{N} \widehat{\pi}_u(x, T_i^x),$$

with i.i.d. sample $T_1^x, \dots, T_N^x$ and $T_i^x \sim \theta_x$.

Monte Carlo within Metropolis (MCWM)

MCWM algorithm M_N with transition from $\widetilde{x}_n$ to $\widetilde{x}_{n+1}$:

1. Draw x' from proposal density $q(\widetilde{x}_n, \cdot)$;
2. Compute independently $\pi_{u,N}(x')$ and $\pi_{u,N}(\widetilde{x}_n)$.
3. Set
$$\widetilde{x}_{n+1} = \begin{cases} x' & \text{with probab.} \quad a_N(\widetilde{x}_n, x') \\ \widetilde{x}_n & \text{otherwise} \end{cases}$$
with
$$a_N(\widetilde{x}_n, x') = \min\left\{1, \frac{\pi_{u,N}(x')q(x', \widetilde{x}_n)}{\pi_{u,N}(\widetilde{x}_n)q(\widetilde{x}_n, x')}\right\}.$$

(Beaumont 2003, Andrieu, Roberts 2009 and Medina-Aguayo et al. 2015)

(Korattikara et al. 2014, Alquier et al. 2014, Bardenet et al. 2015, Pillai, Smith 2015)

Abstract problem

Setting:

- $(X_i)_{i\in\mathbb{N}_0}$, $(\widetilde{X}_i)_{i\in\mathbb{N}_0}$ Markov chains with transition kernels P, $\widetilde{P}$
- distribution of X_n and $\widetilde{X}_n$ denoted by p_n and $\widetilde{p}_n$, assume $p_0 = \widetilde{p}_0$
- $\widetilde{P}$ is an approximation or perturbation of P

Problem:

"What is the difference of p_n and $\widetilde{p}_n$?"

Quantitative bounds?

Total variation and V-norm

Assume π and $\widetilde{\pi}$ are distributions on G.

Total variation

$$\|\pi - \widetilde{\pi}\|_{\text{tv}} := 2 \sup_{A \subseteq G} |\pi(A) - \widetilde{\pi}(A)| = \sup_{|f| \leq 1} \left| \int_G f(y)(\pi(\mathrm{d}y) - \widetilde{\pi}(\mathrm{d}y)) \right|.$$

V-norm

$$\|\pi - \widetilde{\pi}\|_V = \sup_{|f| \leq V} \left| \int_G f(y)(\pi(\mathrm{d}y) - \widetilde{\pi}(\mathrm{d}y)) \right|$$

for a measurable function $V \colon G \to [1, \infty)$.

Ergodicity assumption

Unperturbed Markov chain $(X_i)_{i\in\mathbb{N}}$ is geometrically ergodic:

$$
\begin{aligned}
& P \text{ is geometrically ergodic} \\
\iff \quad & P \text{ is } V\text{-uniformly ergodic} \\
\iff \quad & \exists C \in (0,\infty) \quad \exists \varrho \in [0,1) \quad \text{s.t.} \\
& \qquad \|P^n(x,\cdot) - \pi\|_V \leq C V(x) \varrho^n
\end{aligned}
$$

Lyapunov assumption

V is a Lyapunov function of perturbed Markov chain $(\widetilde{X}_i)_{i\in\mathbb{N}}$:

$\exists L \in (0,\infty) \quad \exists \delta \in [0,1) \quad$ s.t.

$$\widetilde{P}V(x) := \int_G V(y)\widetilde{P}(x,\mathrm{d}y) \leq \delta V(x) + L.$$

Abstract result

Theorem (quantitative upper bound)

Define

$$\gamma_{\mathrm{tv}} = \sup_{x\in G} \frac{\left\| P(x,\cdot) - \widetilde{P}(x,\cdot) \right\|_{\mathrm{tv}}}{V(x)}, \qquad \gamma_V = \sup_{x\in G} \frac{\left\| P(x,\cdot) - \widetilde{P}(x,\cdot) \right\|_V}{V(x)},$$

and

$$\kappa = \max\left\{ \int_G V(x)\, \widetilde{p}_0(\mathrm{d}x), \frac{L}{1-\delta} \right\}.$$

Then, for any $r \in (0,1]$,

$$\left\| p_n - \widetilde{p}_n \right\|_{\mathrm{tv}} \le \gamma_{\mathrm{tv}}^{1-r} \gamma_V^r \, \frac{C^r \kappa}{(1-\varrho) r}.$$

Notes and remarks

- We have

$$\gamma_{\text{tv}} \leq \min\{2, \gamma_V\},$$

$$PV(x) \leq V(x) + L \implies \gamma_V \leq L + 2.$$

- If π and $\widetilde{\pi}$ are stationary distributions of P and $\widetilde{P}$, then

$$\|\pi - \widetilde{\pi}\|_{\text{tv}} \leq \gamma_{\text{tv}}^{1-r} \gamma_V^r \, \frac{C^r L}{(1-\delta)(1-\varrho) r}.$$

(Essentially follows by letting $n \to \infty$)

Perturbation of Markov chains literature

- Ferré, Hervé, Ledoux, *Regular Perturbation of V-geometrically ergodic Markov chains,* 2013.
- Kartashov, *Strong Stable Markov Chains*, 1996.
- Mitrophanov, *Sensitivity and convergence of uniformly ergodic Markov chains,* 2005.
- Pillai, Smith, *Ergodicity of Approximate MCMC Chains with Applications to Large Data Sets,* 2015.
- Rudolf, Schweizer, *Perturbation theory for Markov chains via Wasserstein distance*, 2015.
- Stuart, Shardlow, *A Perturbation theory for ergodic Markov chains and application to numerical approximations,* 2000.

Application to MCWM

MCWM algorithm M_N for large N should be close to MH algorithm M.

Question:

$$\left\| m_n - m_{n,N} \right\|_{\text{tv}} \leq ?$$

($m_{n,N}$ and m_n distributions of M_N and M after n steps)

For the Theorem we need

- M is V-uniformly ergodic.
- V Lyapunov function of M_N, i.e. for some $\delta \in [0,1)$ and $L \in (0,\infty)$

$$M_N V(x) \leq \delta V(x) + L.$$

- Estimate of γ_{tv} and/or γ_V.

Standing assumption

M is V-uniformly ergodic and for some $\alpha \in [0,1)$, $R \in (0,\infty)$ holds

$$MV(x) \leq \alpha V(x) + R.$$

Define

$$K_1 = \sup_{x\in G} \mathbb{E} \left| \frac{\pi_u(x)}{\widehat{\pi}_u(x, T_1^x)} \right|^2,$$

$$K_2 = \sup_{x\in G} \mathbb{E} \left| \frac{\widehat{\pi}_u(x, T_1^x)}{\pi_u(x)} - 1 \right|^2.$$

Recall $\pi_{u,N}(x)$ unbiased estimate of $\pi_u(x)$ given by

$$\pi_{u,N}(x) = \frac{1}{N} \sum_{j=1}^{N} \widehat{\pi}_u(x, T_i^x) \quad \text{with i.i.d. sample} \quad T_1^x, \dots, T_N^x \sim \theta_x$$

Consequence I (K_1 involved)

With arguments from Medina-Aguayo et al. 2015 follows

$$N > \frac{K_2(K_1\alpha+3)^3}{(1-\alpha)^3} \quad \Longrightarrow \quad \begin{cases} \exists \delta \in [0,1),\ L \in [R,\infty) \text{ s.t.} \\ M_N V(x) \leq \delta V(x) + L \end{cases}$$

and

$$N > 4400 K_2 \quad \Longrightarrow \quad \gamma_{\text{tv}} \leq \sup_{x\in G} \|M_N(x,\cdot) - M(x,\cdot)\|_{\text{tv}} \leq \frac{6K_2^{1/3}}{N^{1/3}}.$$

Corollary I (MCWM quantitative upper bound)

$$\left\| m_n - m_{n,N} \right\|_{\text{tv}} \leq \frac{2K_2^{1/3} \log\left(\frac{N}{216K_2}\right)}{N^{1/3}} \cdot \frac{C(L+2)\kappa}{(1-\varrho)}.$$

Consequence II (K_1 not involved)

Additionally assume for proposal q that

$$\int_G V(y)q(x,y)\mathrm{d}y \leq K_3 V(x).$$

With arguments from Medina-Aguayo et al. 2015 follows

$$N > 64K_2 \quad \Longrightarrow \quad \gamma_V \leq \frac{6(1+K_3)K_2^{1/3}}{N^{1/3}}$$

and

$$N > \frac{216K_2(K_3+1)^3}{(1-\alpha)^3} \quad \Longrightarrow \quad \begin{cases} \exists \delta \in [0,1), \text{ s.t.} \\ \quad M_N V(x) \leq \delta V(x) + R. \end{cases}$$

Consequence II (K_1 not involved)

Corollary II (MCWM quantitative upper bound)

Additionally assume for proposal q that

$$\int_G V(y)q(x,y)\mathrm{d}y \leq K_3 V(x).$$

Then

$$\left\| m_n - m_{n,N} \right\|_{\mathrm{tv}} \leq \frac{6(1+K_3)K_2^{1/3}}{N^{1/3}} \cdot \frac{C\kappa}{(1-\varrho)}.$$

Work in progress: Integration error

Approximate

$$\pi(f) = \int_G f(x)\,\pi(x)\mathrm{d}x.$$

- $(X_i)_{i\in\mathbb{N}_0}$, $(\widetilde{X}_i)_{i\in\mathbb{N}_0}$ Markov chains with transition kernels P, $\widetilde{P}$
- π stationary distribution of P
- $\widetilde{P}$ is an approximation or perturbation of P

Question:

$$\mathbb{E}\left|\frac{1}{n}\sum_{j=1}^{n} f(\widetilde{X}_j) - \pi(f)\right| \leq ?$$

Theorem (Integration error)

- $\exists \varrho \in [0,1)$ s.t. $\|P^n(x,\cdot) - \pi\|_V \leq V(x)\varrho^n$.
- $\exists \delta \in [0,1)$ $\exists L \in (0,\infty)$ s.t. $\widetilde{P}V(x) \leq \delta V(x) + L$.
- Define
$$|f|_V = \sup_{x \in G} \frac{|f(x)|}{V(x)},$$

Then

$$\mathbb{E}\left|\frac{1}{n}\sum_{j=1}^{n} f(\widetilde{X}_j) - \pi(f)\right| \leq \frac{\kappa\gamma_V}{1-\varrho}\,|f|_V + \mathbb{E}\left|\frac{1}{n}\sum_{j=1}^{n} f(X_j) - \pi(f)\right|.$$

References

- Alquier, Friel, Everitt, Boland, *Noisy Monte Carlo: Convergence of Markov chains with approximate transition kernels,* 2014.
- Andrieu, Roberts, *The pseudo-marginal approach for efficient Monte Carlo computations*, 2009.
- Andrieu, Vihola, *Convergence properties of pseudo-marginal Markov chain Monte Carlo algorithms*, 2015
- Bardenet, Doucet, Holmes, *On Markov chain Monte Carlo methods for tall data*, 2015.
- Beaumont. *Estimation of population growth or decline in generally monitored populations*, 2003.
- Korattikara, Chen, Welling *Austerity in MCMC Land: Cutting the Metropolis-Hastings Budget*, 2013.
- Medina-Aguayo, Lee, Roberts, *Stability of noisy Metropolis-Hastings,* 2015.