Current developments in MCMC II: Comparison of Markov chains via weak Poincaré inequalities

Andi Q. Wang

University of Warwick

Joint with: Christophe Andrieu, Anthony Lee, Sam Power.

Dynstoch Satellie Spring School

May 2024

Overview

Overview

Recap

This morning we looked at random walk Metropolis–Hastings (RWM) on $\mathbb{R}^d$:
$q(X_{i-1}, \cdot) = \mathcal{N}(X_{i-1}, \Sigma)$.

Recap

This morning we looked at random walk Metropolis–Hastings (RWM) on $\mathbb{R}^d$:
$q(X_{i-1}, \cdot) = \mathcal{N}(X_{i-1}, \Sigma)$.

We showed under certain conditions it possessed a positive spectral gap γ_d, and that when tuned properly, $\gamma_d \sim O(1/d)$.

Recap

This morning we looked at random walk Metropolis–Hastings (RWM) on $\mathbb{R}^d$: $q(X_{i-1}, \cdot) = \mathcal{N}(X_{i-1}, \Sigma)$.

We showed under certain conditions it possessed a positive spectral gap γ_d, and that when tuned properly, $\gamma_d \sim O(1/d)$. Many more sophisticated MCMC methods can be seen as extensions of RWM, and some of them are known to be subgeometric (no spectral gap).

Recap

This morning we looked at random walk Metropolis–Hastings (RWM) on $\mathbb{R}^d$:
$q(X_{i-1}, \cdot) = \mathcal{N}(X_{i-1}, \Sigma)$.

We showed under certain conditions it possessed a positive spectral gap γ_d, and that when tuned properly, $\gamma_d \sim O(1/d)$. Many more sophisticated MCMC methods can be seen as extensions of RWM, and some of them are known to be subgeometric (no spectral gap).

Can our results be extended into these settings?

L^2 convergence

More precisely we looked at convergence in $L^2(\pi)$:

L^2 convergence

More precisely we looked at convergence in $L^2(\pi)$:

$$L^2(\pi) = \{f : \mathcal{X} \to \mathbb{R} \mid \|f\|_2^2 = \int |f|^2 \, \mathrm{d}\pi < \infty\}, \quad \langle f, g \rangle = \int fg \, \mathrm{d}\pi.$$

L^2 convergence

More precisely we looked at convergence in $L^2(\pi)$:

$$L^2(\pi) = \{f : \mathcal{X} \to \mathbb{R} \mid \|f\|_2^2 = \int |f|^2 \, \mathrm{d}\pi < \infty\}, \quad \langle f, g \rangle = \int fg \, \mathrm{d}\pi.$$

Given π-invariant Markov kernel P, it follows that $P : L^2(\pi) \to L^2(\pi)$ given by

$$Pf(x) = \mathbb{E}_x[f(X_1)]$$

is a bounded linear mapping.

L^2 convergence

More precisely we looked at convergence in $L^2(\pi)$:

$$L^2(\pi) = \{f : \mathcal{X} \to \mathbb{R} \mid \|f\|_2^2 = \int |f|^2 \, \mathrm{d}\pi < \infty\}, \quad \langle f, g \rangle = \int fg \, \mathrm{d}\pi.$$

Given π-invariant Markov kernel P, it follows that $P : L^2(\pi) \to L^2(\pi)$ given by

$$Pf(x) = \mathbb{E}_x[f(X_1)]$$

is a bounded linear mapping.

Interested to study for $f \in L^2(\pi)$, how fast do we have

$$\|P^2 f - \pi(f)\|_2 \to 0.$$

Exponential convergence and spectral gap

In many cases the convergence is exponential: for any $f \in L^2(\pi)$,

$$\|Pf - \pi(f)\|_2 \leq \|f\| \cdot (1 - \gamma)^n,$$

where for reversible P, the best possible rate is given by the spectral gap $0 \leq \gamma \leq 1$.

Exponential convergence and spectral gap

In many cases the convergence is exponential: for any $f \in L^2(\pi)$,

$$\|Pf - \pi(f)\|_2 \leq \|f\| \cdot (1-\gamma)^n,$$

where for reversible P, the best possible rate is given by the spectral gap $0 \leq \gamma \leq 1$.

Theorem ([Andrieu, Lee, Power, Wang (2022a)], *c.f.* [Roberts, Gelman, Gilks (1997)])

When the target π has a strongly concave and L-smooth potential, the spectral gap of RWM with proposal variance of d^{-1} on $\mathbb{R}^d$ scales like

$$\gamma_d = O(d^{-1}).$$

Exponential convergence and spectral gap

In many cases the convergence is exponential: for any $f \in L^2(\pi)$,

$$\|Pf - \pi(f)\|_2 \leq \|f\| \cdot (1-\gamma)^n,$$

where for reversible P, the best possible rate is given by the spectral gap $0 \leq \gamma \leq 1$.

Theorem ([Andrieu, Lee, Power, Wang (2022a)], *c.f.* [Roberts, Gelman, Gilks (1997)])

When the target π has a strongly concave and L-smooth potential, the spectral gap of RWM with proposal variance of d^{-1} on $\mathbb{R}^d$ scales like

$$\gamma_d = O(d^{-1}).$$

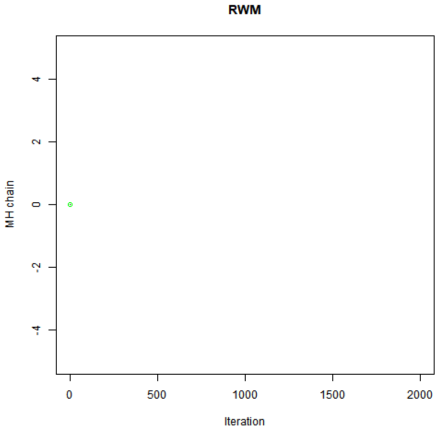
This is nice when applicable, but many chains actually converge at a subgeometric rate and have 0 spectral gap.

Metropolis–Hastings

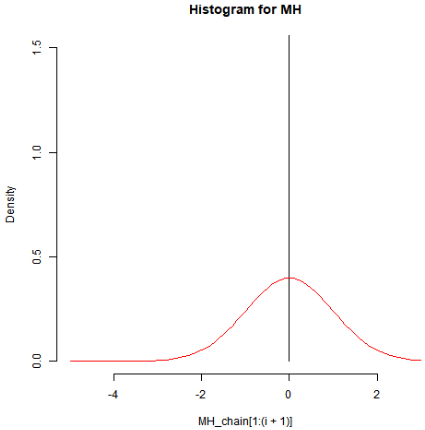
Algorithm 1 (Marginal) Metropolis–Hastings (MH)

initialise: $X_0 = x_0, i = 0$
while $i < N$ **do**
  $i \leftarrow i + 1$
  simulate $Y_i \sim q(X_{i-1}, \cdot)$
  $\alpha(X_{i-1}, Y_i) = 1 \wedge \frac{q(Y_i, X_{i-1})\pi(Y_i)}{q(X_{i-1}, Y_i)\pi(X_{i-1})}$
  with probability $\alpha(X_{i-1}, Y_i)$
    $X_i \leftarrow Y_i$
  else
    $X_i \leftarrow X_{i-1}$
return $(X_i)_{i=1,\dots,n}$

MH example



RWM
MH chain
4
2
0
-2
-4
0
500
1000
1500
2000
Iteration



Histogram for MH
Density
1.5
1.0
0.5
0.0
-4
-2
0
2
MH_chain[1:(i + 1)]

MCMC on modern datasets

Metropolis–Hastings can be implemented provided we can evaluate $\pi(x)$ pointwise (up to a constant).

MCMC on modern datasets

Metropolis–Hastings can be implemented provided we can evaluate $\pi(x)$ pointwise (up to a constant).

But recall:

$$\pi(x) \propto \nu(x)\ell_y(x).$$

MCMC on modern datasets

Metropolis–Hastings can be implemented provided we can evaluate $\pi(x)$ pointwise (up to a constant).

But recall:

$$\pi(x) \propto \nu(x)\ell_y(x).$$

For modern datasets this $\ell_y(x)$ is often intractable! E.g.

MCMC on modern datasets

Metropolis–Hastings can be implemented provided we can evaluate $\pi(x)$ pointwise (up to a constant).

But recall:

$$\pi(x) \propto \nu(x)\ell_y(x).$$

For modern datasets this $\ell_y(x)$ is often intractable! E.g.

$$\ell_y(x) = \prod_{i=1}^{N} f_x(y_i), \qquad \text{or } \ell_y(x) = \int\int\cdots\int g(x,z,y)\,\mathrm{d}z_1\,\mathrm{d}z_2\ldots\mathrm{d}z_N,$$

corresponding to 'big data' or latent variable models.

Pseudo-marginal MH [Andrieu and Roberts (2009)]

One proposed solution: we may be able to construct a nonnegative unbiased estimator (up to a constant C) $\hat{\ell}_y(x)$ of the intractable likelihood $\ell_y(x)$,

Pseudo-marginal MH [Andrieu and Roberts (2009)]

One proposed solution: we may be able to construct a nonnegative unbiased estimator (up to a constant C) $\hat{\ell}_y(x)$ of the intractable likelihood $\ell_y(x)$,

$$\mathbb{E}[\hat{\ell}_y(x)] = C \cdot \ell_y(x), \quad \hat{\ell}_y(x) \geq 0 \text{ a.s.}$$

Pseudo-marginal MH [Andrieu and Roberts (2009)]

One proposed solution: we may be able to construct a nonnegative unbiased estimator (up to a constant C) $\hat{\ell}_y(x)$ of the intractable likelihood $\ell_y(x)$,

$$\mathbb{E}[\hat{\ell}_y(x)] = C \cdot \ell_y(x), \quad \hat{\ell}_y(x) \geq 0 \text{ a.s.}$$

and then in Metropolis–Hastings, naively substitute in

$$\hat{\pi}(x) = \nu(x)\hat{\ell}_y(x)$$

wherever we needed $\pi(x)$.

Pseudo-marginal MH [Andrieu and Roberts (2009)]

One proposed solution: we may be able to construct a nonnegative unbiased estimator (up to a constant C) $\hat{\ell}_y(x)$ of the intractable likelihood $\ell_y(x)$,

$$\mathbb{E}[\hat{\ell}_y(x)] = C \cdot \ell_y(x), \quad \hat{\ell}_y(x) \geq 0 \text{ a.s.}$$

and then in Metropolis–Hastings, naively substitute in

$$\hat{\pi}(x) = \nu(x)\hat{\ell}_y(x)$$

wherever we needed $\pi(x)$.

This is the pseudo-marginal MH algorithm proposed in [Andrieu and Roberts (2009)].

Pseudo-marginal MH [Andrieu and Roberts (2009)]

One proposed solution: we may be able to construct a nonnegative unbiased estimator (up to a constant C) $\hat{\ell}_y(x)$ of the intractable likelihood $\ell_y(x)$,

$$\mathbb{E}[\hat{\ell}_y(x)] = C \cdot \ell_y(x), \quad \hat{\ell}_y(x) \geq 0 \text{ a.s.}$$

and then in Metropolis–Hastings, naively substitute in

$$\hat{\pi}(x) = \nu(x)\hat{\ell}_y(x)$$

wherever we needed $\pi(x)$.

This is the pseudo-marginal MH algorithm proposed in [Andrieu and Roberts (2009)].

(One subtlety: need to 'carry around' the previous estimator $\hat{\pi}(X_{n-1})$ in the next step rather than generating it again.)

Pseudo-marginal MH [Andrieu and Roberts (2009)]

Algorithm 2 Pseudo-marginal Metropolis–Hastings

initialise: $X_0 = x_0, i = 0$
while $i < N$ **do**
  $i \leftarrow i + 1$
  simulate $Y_i \sim q(X_{i-1}, \cdot)$
  $\alpha(X_{i-1}, Y_i) = 1 \wedge \frac{q(Y_i, X_{i-1})\hat{\pi}(Y_i)}{q(X_{i-1}, Y_i)\hat{\pi}(X_{i-1})}$
  with probability $\alpha(X_{i-1}, Y_i)$
    $X_i \leftarrow Y_i$
  else
    $X_i \leftarrow X_{i-1}$
return $(X_i)_{i=1,\ldots,n}$

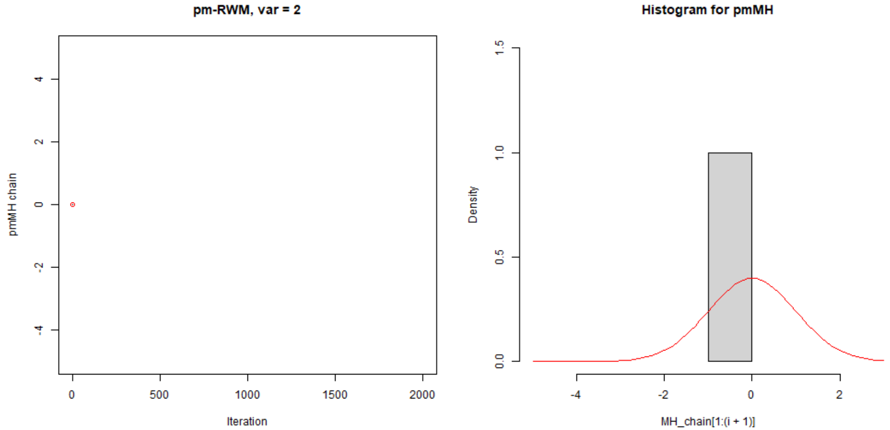
Pseudo-marginal MH [Andrieu and Roberts (2009)]

Algorithm 3 Pseudo-marginal Metropolis–Hastings

initialise: $X_0 = x_0, i = 0$
while $i < N$ **do**
  $i \leftarrow i + 1$
  simulate $Y_i \sim q(X_{i-1}, \cdot)$
  $\alpha(X_{i-1}, Y_i) = 1 \wedge \frac{q(Y_i, X_{i-1})\hat{\pi}(Y_i)}{q(X_{i-1}, Y_i)\hat{\pi}(X_{i-1})}$
  with probability $\alpha(X_{i-1}, Y_i)$
    $X_i \leftarrow Y_i$
  else
    $X_i \leftarrow X_{i-1}$
return $(X_i)_{i=1,\dots,n}$

A miracle: with an unbiased estimator, this is still a valid algorithm which targets π.

Pseudo-marginal RWM example



Two examples of pseudo-marginal MH

We have replaced $\pi(x)$ with a stochastic estimator $\hat{\pi}(x)$ (which in practice could have a large variance), so we expect the performance of pseudo-marginal MH to be worse than the original marginal MH.

Two examples of pseudo-marginal MH

We have replaced $\pi(x)$ with a stochastic estimator $\hat{\pi}(x)$ (which in practice could have a large variance), so we expect the performance of pseudo-marginal MH to be worse than the original marginal MH.

Natural question: how much worse?

Two examples of pseudo-marginal MH

We have replaced $\pi(x)$ with a stochastic estimator $\hat{\pi}(x)$ (which in practice could have a large variance), so we expect the performance of pseudo-marginal MH to be worse than the original marginal MH.

Natural question: how much worse?

Will consider two particular examples.

- Approximate Bayesian Computation (ABC) [Marin et. al. (2012)] for Simulation-Based Inference (SBI);

Two examples of pseudo-marginal MH

We have replaced $\pi(x)$ with a stochastic estimator $\hat{\pi}(x)$ (which in practice could have a large variance), so we expect the performance of pseudo-marginal MH to be worse than the original marginal MH.

Natural question: how much worse?

Will consider two particular examples.

- Approximate Bayesian Computation (ABC) [Marin et. al. (2012)] for Simulation-Based Inference (SBI);
- Particle Marginal Metropolis–Hastings (PMMH) [Andrieu et. al. (2010)] for inference with time series in State Space Models (SMMs), also known has Hidden Markov Models (HMMs).

Example 1: Approximate Bayesian Computation (ABC)

Recall we had posterior

$$\pi(x) \propto \nu(x)\ell_y(x).$$

Example 1: Approximate Bayesian Computation (ABC)

Recall we had posterior

$$\pi(x) \propto \nu(x)\ell_y(x).$$

Suppose $\ell_y(x) = f_x(y)$ is intractable, but we can simulate from $f_x(\cdot)$.

Example 1: Approximate Bayesian Computation (ABC)

Recall we had posterior

$$\pi(x) \propto \nu(x)\ell_y(x).$$

Suppose $\ell_y(x) = f_x(y)$ is intractable, but we can simulate from $f_x(\cdot)$.

So instead we use an approximate likelihood:

$$\pi_{\mathrm{ABC}}(x) \propto \nu(x)\ell_{\mathrm{ABC}}(x),$$
$$\ell_{\mathrm{ABC}}(x) := \mathbb{P}_x(|Z - y| < \epsilon), \quad Z \sim f_x.$$

Example 1: Approximate Bayesian Computation (ABC)

Recall we had posterior

$$\pi(x) \propto \nu(x)\ell_y(x).$$

Suppose $\ell_y(x) = f_x(y)$ is intractable, but we can simulate from $f_x(\cdot)$.

So instead we use an approximate likelihood:

$$\pi_{\mathrm{ABC}}(x) \propto \nu(x)\ell_{\mathrm{ABC}}(x),$$
$$\ell_{\mathrm{ABC}}(x) := \mathbb{P}_x(|Z - y| < \epsilon), \quad Z \sim f_x.$$

This is further approximated using an unbiased estimator obtained by drawing M sets of artificial data $Z_1, \ldots, Z_M \overset{\text{i.i.d.}}{\sim} f_x$.

Example 1: Approximate Bayesian Computation (ABC)

Recall we had posterior

$$\pi(x) \propto \nu(x)\ell_y(x).$$

Suppose $\ell_y(x) = f_x(y)$ is intractable, but we can simulate from $f_x(\cdot)$.

So instead we use an approximate likelihood:

$$\pi_{\mathrm{ABC}}(x) \propto \nu(x)\ell_{\mathrm{ABC}}(x),$$
$$\ell_{\mathrm{ABC}}(x) := \mathbb{P}_x(|Z - y| < \epsilon), \quad Z \sim f_x.$$

This is further approximated using an unbiased estimator obtained by drawing M sets of artificial data $Z_1, \ldots, Z_M \overset{\text{i.i.d.}}{\sim} f_x$.

Natural question for practitioner: what is the convergence rate and effect of varying M?

Example 1: Approximate Bayesian Computation (ABC)

Recall we had posterior

$$\pi(x) \propto \nu(x)\ell_y(x).$$

Suppose $\ell_y(x) = f_x(y)$ is intractable, but we can simulate from $f_x(\cdot)$.

So instead we use an approximate likelihood:

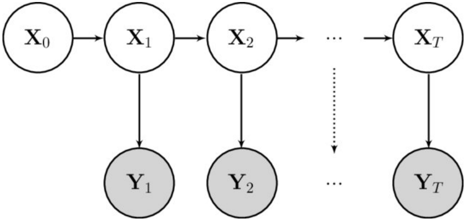
$$\pi_{\mathrm{ABC}}(x) \propto \nu(x)\ell_{\mathrm{ABC}}(x),$$
$$\ell_{\mathrm{ABC}}(x) := \mathbb{P}_x(|Z - y| < \epsilon), \quad Z \sim f_x.$$

This is further approximated using an unbiased estimator obtained by drawing M sets of artificial data $Z_1, \ldots, Z_M \overset{\text{i.i.d.}}{\sim} f_x$.

Natural question for practitioner: what is the convergence rate and effect of varying M?
N.B. the chain is almost always subgeometric [Lee and Łatuszyński (2014)].

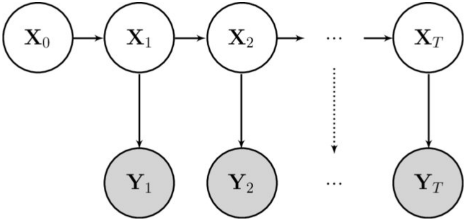
Example 2: Inference in state space models via PMMH

In state space models, we assume there is a latent Markov chain $(\mathbf{X}_n)$ which drives the observation process $(\mathbf{Y}_n)$. (Image: Wiki.)



Example 2: Inference in state space models via PMMH

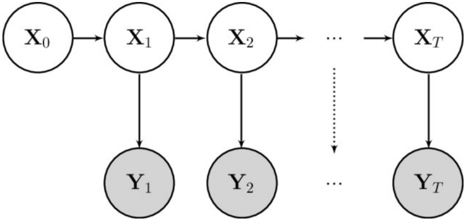
In state space models, we assume there is a latent Markov chain $(\mathbf{X}_n)$ which drives the observation process $(\mathbf{Y}_n)$. (Image: Wiki.)



The likelihood for static parameters x is typically intractable.

Example 2: Inference in state space models via PMMH

In state space models, we assume there is a latent Markov chain $(\mathbf{X}_n)$ which drives the observation process $(\mathbf{Y}_n)$. (Image: Wiki.)

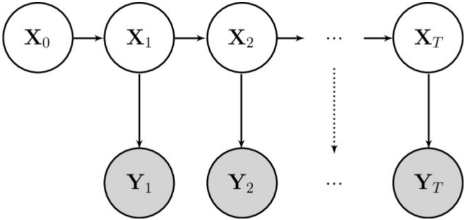


The likelihood for static parameters x is typically intractable.

However we can use a particle filter to get an unbiased estimator of $\ell_y(x)$; Particle Marginal Metropolis–Hastings (PMMH) [Andrieu et. al. (2010)].

Example 2: Inference in state space models via PMMH

In state space models, we assume there is a latent Markov chain $(\mathbf{X}_n)$ which drives the observation process $(\mathbf{Y}_n)$. (Image: Wiki.)



The likelihood for static parameters x is typically intractable.

However we can use a particle filter to get an unbiased estimator of $\ell_y(x)$; Particle Marginal Metropolis–Hastings (PMMH) [Andrieu et. al. (2010)].

Natural questions: (subgeometric) convergence rate, how to tune the particle filter?

Convergence of pseudo-marginal MH

All pseudo-marginal examples replace $\pi(x)$ with an unbiased estimator $\hat{\pi}(x)$.

Convergence of pseudo-marginal MH

All pseudo-marginal examples replace $\pi(x)$ with an unbiased estimator $\hat{\pi}(x)$.

Since there is an extra layer of randomization, we expect pseudo-marginal MH to perform worse than the original marginal MH targeting π.

Convergence of pseudo-marginal MH

All pseudo-marginal examples replace $\pi(x)$ with an unbiased estimator $\hat{\pi}(x)$.

Since there is an extra layer of randomization, we expect pseudo-marginal MH to perform worse than the original marginal MH targeting π.

This raises the natural question: how do we quantify the degradation in performance when using pseudo-marginal MH compared to marginal MH?

Convergence of pseudo-marginal MH

All pseudo-marginal examples replace $\pi(x)$ with an unbiased estimator $\hat{\pi}(x)$.

Since there is an extra layer of randomization, we expect pseudo-marginal MH to perform worse than the original marginal MH targeting π.

This raises the natural question: how do we quantify the degradation in performance when using pseudo-marginal MH compared to marginal MH?

A first answer was given in [Andrieu and Vihola (2015)]: model $\hat{\pi}(x) = W_x \cdot \pi(x)$, with $W_x \sim Q_x$ nonnegative and $\mathbb{E}[W_x] = 1$.

Convergence of pseudo-marginal MH

All pseudo-marginal examples replace $\pi(x)$ with an unbiased estimator $\hat{\pi}(x)$.

Since there is an extra layer of randomization, we expect pseudo-marginal MH to perform worse than the original marginal MH targeting π.

This raises the natural question: how do we quantify the degradation in performance when using pseudo-marginal MH compared to marginal MH?

A first answer was given in [Andrieu and Vihola (2015)]: model $\hat{\pi}(x) = W_x \cdot \pi(x)$, with $W_x \sim Q_x$ nonnegative and $\mathbb{E}[W_x] = 1$.

Theorem ([Andrieu and Vihola (2015)])

If the marginal MH is geometric, and the W_x are uniformly bounded, then the pseudo-marginal chain is geometric. If the W_x have unbounded support, then the chain is subgeometric.

Summary

I will describe a self-contained, general comparison technique for Markov chains.

Summary

I will describe a self-contained, general comparison technique for Markov chains.

For pseudo-marginal MCMC, we are able to precisely and transparently relate the degradation in performance compared to the marginal chain with the tail probabilities of the perturbations.

Summary

I will describe a self-contained, general comparison technique for Markov chains.

For pseudo-marginal MCMC, we are able to precisely and transparently relate the degradation in performance compared to the marginal chain with the tail probabilities of the perturbations.

For ABC, we can bound the subgeometric convergence rate of the chain depending on the regularity of the prior and describe the effect of varying M.

Summary

I will describe a self-contained, general comparison technique for Markov chains.

For pseudo-marginal MCMC, we are able to precisely and transparently relate the degradation in performance compared to the marginal chain with the tail probabilities of the perturbations.

For ABC, we can bound the subgeometric convergence rate of the chain depending on the regularity of the prior and describe the effect of varying M.

For PMMH, in the limiting case of lognormal weights, we bound the subgeometric convergence rate and give precise tuning advice to minimize the asymptotic variance.

Summary

I will describe a self-contained, general comparison technique for Markov chains.

For pseudo-marginal MCMC, we are able to precisely and transparently relate the degradation in performance compared to the marginal chain with the tail probabilities of the perturbations.

For ABC, we can bound the subgeometric convergence rate of the chain depending on the regularity of the prior and describe the effect of varying M.

For PMMH, in the limiting case of lognormal weights, we bound the subgeometric convergence rate and give precise tuning advice to minimize the asymptotic variance.

Overview

Standard Poincaré inequalities

We work on $\mathrm{L}^2(\mu) = \{f : \mathcal{X} \to \mathbb{R} : \|f\|_2^2 < \infty\}$, $\langle f, g \rangle := \int fg \, \mathrm{d}\mu$,
$\mathrm{L}_0^2(\mu) := \{f \in \mathrm{L}^2(\mu) : \mu(f) = 0\}$.

Standard Poincaré inequalities

We work on $\mathrm{L}^2(\mu) = \{f : \mathcal{X} \to \mathbb{R} : \|f\|_2^2 < \infty\}$, $\langle f, g \rangle := \int fg \, \mathrm{d}\mu$,
$\mathrm{L}_0^2(\mu) := \{f \in \mathrm{L}^2(\mu) : \mu(f) = 0\}$.

For a μ-invariant Markov transition kernel P with $\mathrm{L}^2(\mu)$-adjoint P^*, define the Dirichlet form $\mathcal{E}(P^*P, f)$, for $f \in \mathrm{L}_0^2(\mu)$:

$$\mathcal{E}(P^*P, f) := \langle (I - P^*P)f, f \rangle.$$

Standard Poincaré inequalities

We work on $\mathrm{L}^2(\mu) = \{f : \mathcal{X} \to \mathbb{R} : \|f\|_2^2 < \infty\}$, $\quad \langle f, g \rangle := \int fg \, \mathrm{d}\mu$,
$\mathrm{L}_0^2(\mu) := \{f \in \mathrm{L}^2(\mu) : \mu(f) = 0\}$.

For a μ-invariant Markov transition kernel P with $\mathrm{L}^2(\mu)$-adjoint P^*, define the Dirichlet form $\mathcal{E}(P^*P, f)$, for $f \in \mathrm{L}_0^2(\mu)$:

$$\mathcal{E}(P^*P, f) := \langle (I - P^*P)f, f \rangle.$$

Standard Poincaré inequality (SPI)

A SPI holds if there exists a constant $C_{\mathrm{P}} > 0$ such that for all $f \in \mathrm{L}_0^2(\mu)$,

$$C_{\mathrm{P}} \|f\|_2^2 \leq \mathcal{E}(P^*P, f).$$

Geometric convergence / spectral gap

$$C_{\mathrm{P}} \|f\|_2^2 \leq \mathcal{E}(P^* P, f).$$

Geometric convergence / spectral gap

$$C_{\mathrm{P}} \|f\|_2^2 \leq \mathcal{E}(P^* P, f).$$

Theorem (Geometric convergence)

Under a standard Poincaré inequality, we have for all $f \in \mathrm{L}_0^2(\mu)$, $n \in \mathbb{N}_0$,

$$\|P^n f\|_2^2 \leq (1 - C_{\mathrm{P}})^n \|f\|_2^2.$$

Geometric convergence / spectral gap

$$C_{\mathrm{P}}\|f\|_2^2 \leq \mathcal{E}(P^*P, f).$$

Theorem (Geometric convergence)

Under a standard Poincaré inequality, we have for all $f \in \mathrm{L}_0^2(\mu)$, $n \in \mathbb{N}_0$,

$$\|P^n f\|_2^2 \leq (1 - C_{\mathrm{P}})^n \|f\|_2^2.$$

Proof. Rewriting the SPI, see $\mathcal{E}(P^*P, f)$ behaves like a discrete derivative:

Geometric convergence / spectral gap

$$C_{\mathrm{P}} \|f\|_2^2 \le \mathcal{E}(P^*P, f).$$

Theorem (Geometric convergence)

Under a standard Poincaré inequality, we have for all $f \in \mathrm{L}_0^2(\mu)$, $n \in \mathbb{N}_0$,

$$\|P^n f\|_2^2 \le (1 - C_{\mathrm{P}})^n \|f\|_2^2.$$

Proof. Rewriting the SPI, see $\mathcal{E}(P^*P, f)$ behaves like a discrete derivative:

$$\begin{aligned} C_{\mathrm{P}} \|f\|_2^2 \le \mathcal{E}(P^*P, f) &= \|f\|_2^2 - \langle P^*Pf, f\rangle \\ &= \|f\|_2^2 - \|Pf\|_2^2 \\ \Rightarrow \|Pf\|_2^2 \le (1 - C_{\mathrm{P}}) \|f\|_2^2. \end{aligned}$$

Geometric convergence / spectral gap

$$C_{\mathrm{P}}\|f\|_2^2 \leq \mathcal{E}(P^*P, f).$$

Theorem (Geometric convergence)

Under a standard Poincaré inequality, we have for all $f \in \mathrm{L}_0^2(\mu)$, $n \in \mathbb{N}_0$,

$$\|P^n f\|_2^2 \leq (1 - C_{\mathrm{P}})^n \|f\|_2^2.$$

Proof. Rewriting the SPI, see $\mathcal{E}(P^*P, f)$ behaves like a discrete derivative:

$$\begin{aligned} C_{\mathrm{P}}\|f\|_2^2 \leq \mathcal{E}(P^*P, f) &= \|f\|_2^2 - \langle P^*Pf, f\rangle \\ &= \|f\|_2^2 - \|Pf\|_2^2 \\ \Rightarrow \|Pf\|_2^2 &\leq (1 - C_{\mathrm{P}})\|f\|_2^2. \end{aligned}$$

The rest is by induction. □

SPI to weak Poincaré inequality

$$C_{\mathrm{P}}\|f\|_2^2 \le \mathcal{E}(P^*P, f).$$

SPI to weak Poincaré inequality

$$C_{\mathrm{P}} \|f\|_2^2 \leq \mathcal{E}(P^* P, f).$$

We now generalize this to allow for subgeometric rates of convergence:

SPI to weak Poincaré inequality

$$C_{\mathrm{P}} \|f\|_2^2 \le \mathcal{E}(P^* P, f).$$

We now generalize this to allow for subgeometric rates of convergence:

Require $\beta : (0, \infty) \to [0, \infty)$ decreasing with $\beta(s) \downarrow 0$ as $s \to \infty$ and $\Phi : \mathrm{L}^2(\mu) \to [0, \infty]$ given by $\Phi(f) = \|f\|_{\mathrm{osc}}^2 = (\mathrm{ess}_\mu \sup f - \mathrm{ess}_\mu \inf f)^2$.

SPI to weak Poincaré inequality

$$C_{\mathrm{P}} \|f\|_2^2 \leq \mathcal{E}(P^* P, f).$$

We now generalize this to allow for subgeometric rates of convergence:

Require $\beta : (0, \infty) \to [0, \infty)$ decreasing with $\beta(s) \downarrow 0$ as $s \to \infty$ and $\Phi : \mathrm{L}^2(\mu) \to [0, \infty]$ given by $\Phi(f) = \|f\|_{\mathrm{osc}}^2 = (\mathrm{ess}_\mu \sup f - \mathrm{ess}_\mu \inf f)^2$.

Weak Poincaré inequality (WPI) (*c.f.* [Röckner and Wang (2001)])

A WPI holds if:

$$C_{\mathrm{P}}\|f\|_2^2 \leq \mathcal{E}(P^*P, f).$$

We now generalize this to allow for subgeometric rates of convergence:

Require $\beta : (0, \infty) \to [0, \infty)$ decreasing with $\beta(s) \downarrow 0$ as $s \to \infty$ and $\Phi : \mathrm{L}^2(\mu) \to [0, \infty]$ given by $\Phi(f) = \|f\|_{\mathrm{osc}}^2 = (\mathrm{ess}_\mu \sup f - \mathrm{ess}_\mu \inf f)^2$.

Weak Poincaré inequality (WPI) (*c.f.* [Röckner and Wang (2001)])

A WPI holds if: for some such β, Φ, $\forall s > 0$, $f \in \mathrm{L}_0^2(\mu)$,

$$\|f\|_2^2 \leq s\,\mathcal{E}(P^*P, f) + \beta(s)\Phi(f).$$

SPI to weak Poincaré inequality

$$C_{\mathrm{P}} \|f\|_2^2 \le \mathcal{E}(P^* P, f).$$

We now generalize this to allow for subgeometric rates of convergence:

Require $\beta : (0, \infty) \to [0, \infty)$ decreasing with $\beta(s) \downarrow 0$ as $s \to \infty$ and $\Phi : \mathrm{L}^2(\mu) \to [0, \infty]$ given by $\Phi(f) = \|f\|_{\mathrm{osc}}^2 = (\mathrm{ess}_\mu \sup f - \mathrm{ess}_\mu \inf f)^2$.

Weak Poincaré inequality (WPI) (*c.f.* [Röckner and Wang (2001)])

A WPI holds if: for some such β, Φ, $\forall s > 0$, $f \in \mathrm{L}_0^2(\mu)$,

$$\|f\|_2^2 \le s\,\mathcal{E}(P^* P, f) + \beta(s)\Phi(f).$$

E.g. $\beta(s) = c_0 s^{-c_1}$.

Subgeometric convergence

$$\|f\|_2^2 \leq s\,\mathcal{E}(P^*P, f) + \beta(s)\Phi(f), \quad \forall s > 0, f \in \mathrm{L}_0^2(\mu).$$

Subgeometric convergence

$$\|f\|_2^2 \le s\,\mathcal{E}(P^*P, f) + \beta(s)\Phi(f), \quad \forall s > 0, f \in \mathrm{L}^2_0(\mu).$$

Define

$$\begin{aligned} K(u) &:= u\beta(1/u), \qquad u \ge 0, \\ K^*(v) &:= \sup_{u \ge 0}\{uv - K(u)\}, \quad v \ge 0, \\ F(x) &:= \int_x^1 \frac{\mathrm{d}v}{K^*(v)}, \qquad 0 < x \le 1. \end{aligned}$$

Subgeometric convergence

$$\|f\|_2^2 \le s\,\mathcal{E}(P^*P, f) + \beta(s)\Phi(f), \quad \forall s > 0, f \in \mathrm{L}_0^2(\mu).$$

Define

$$K(u) := u\beta(1/u), \qquad u \ge 0,$$
$$K^*(v) := \sup_{u\ge 0}\{uv - K(u)\}, \quad v \ge 0,$$
$$F(x) := \int_x^1 \frac{\mathrm{d}v}{K^*(v)}, \qquad 0 < x \le 1.$$

Theorem ([Andrieu, Lee, Power, Wang (2022)])

Under a weak Poincaré inequality, we have, $\forall n \in \mathbb{N}_0$, $f \in \mathrm{L}_0^2(\mu)$,

$$\|P^n f\|_2^2 \le \Phi(f)F^{-1}(n).$$

Proof of convergence bound (I)

Fix $f \in \mathrm{L}_0^2(\mu)$. Have that

$$\|f\|_2^2 \le s\mathcal{E}(P^*P, f) + \beta(s)\Phi(f), \quad \forall s > 0$$

Proof of convergence bound (I)

Fix $f \in \mathrm{L}^2_0(\mu)$. Have that

$$\|f\|_2^2 \leq s\mathcal{E}(P^*P, f) + \beta(s)\Phi(f), \quad \forall s > 0$$

$$\Rightarrow \frac{\mathcal{E}(P^*P, f)}{\Phi(f)} \geq \frac{\|f\|_2^2}{s\Phi(f)} - \frac{\beta(s)}{s}.$$

Proof of convergence bound (I)

Fix $f \in \mathrm{L}^2_0(\mu)$. Have that

$$\|f\|_2^2 \le s\mathcal{E}(P^*P, f) + \beta(s)\Phi(f), \quad \forall s > 0$$

$$\Rightarrow \frac{\mathcal{E}(P^*P, f)}{\Phi(f)} \ge \frac{\|f\|_2^2}{s\Phi(f)} - \frac{\beta(s)}{s}.$$

Set $u := 1/s$, $K(u) := u\beta(1/u)$.

Proof of convergence bound (I)

Fix $f \in \mathrm{L}^2_0(\mu)$. Have that

$$\|f\|_2^2 \le s\mathcal{E}(P^*P, f) + \beta(s)\Phi(f), \quad \forall s > 0$$

$$\Rightarrow \frac{\mathcal{E}(P^*P, f)}{\Phi(f)} \ge \frac{\|f\|_2^2}{s\Phi(f)} - \frac{\beta(s)}{s}.$$

Set $u := 1/s$, $K(u) := u\beta(1/u)$.

$$\frac{\mathcal{E}(P^*P, f)}{\Phi(f)} \ge u \cdot \frac{\|f\|_2^2}{\Phi(f)} - K(u), \quad \forall u > 0.$$

Proof of convergence bound (I)

Fix $f \in \mathrm{L}^2_0(\mu)$. Have that

$$\|f\|_2^2 \le s\mathcal{E}(P^*P, f) + \beta(s)\Phi(f), \quad \forall s > 0$$

$$\Rightarrow \frac{\mathcal{E}(P^*P, f)}{\Phi(f)} \ge \frac{\|f\|_2^2}{s\Phi(f)} - \frac{\beta(s)}{s}.$$

Set $u := 1/s$, $K(u) := u\beta(1/u)$.

$$\frac{\mathcal{E}(P^*P, f)}{\Phi(f)} \ge u \cdot \frac{\|f\|_2^2}{\Phi(f)} - K(u), \quad \forall u > 0.$$

$$\frac{\mathcal{E}(P^*P, f)}{\Phi(f)} \ge \sup_{u>0} \left\{ u \cdot \frac{\|f\|_2^2}{\Phi(f)} - K(u) \right\}$$

Proof of convergence bound (I)

Fix $f \in \mathrm{L}^2_0(\mu)$. Have that

$$\|f\|_2^2 \leq s\mathcal{E}(P^*P, f) + \beta(s)\Phi(f), \quad \forall s > 0$$

$$\Rightarrow \frac{\mathcal{E}(P^*P, f)}{\Phi(f)} \geq \frac{\|f\|_2^2}{s\Phi(f)} - \frac{\beta(s)}{s}.$$

Set $u := 1/s$, $K(u) := u\beta(1/u)$.

$$\frac{\mathcal{E}(P^*P, f)}{\Phi(f)} \geq u \cdot \frac{\|f\|_2^2}{\Phi(f)} - K(u), \quad \forall u > 0.$$

$$\frac{\mathcal{E}(P^*P, f)}{\Phi(f)} \geq \sup_{u>0} \left\{ u \cdot \frac{\|f\|_2^2}{\Phi(f)} - K(u) \right\} =: K^* \left(\frac{\|f\|_2^2}{\Phi(f)} \right).$$

Proof of convergence bound (I)

Fix $f \in \mathrm{L}^2_0(\mu)$. Have that

$$\|f\|_2^2 \leq s\mathcal{E}(P^*P, f) + \beta(s)\Phi(f), \quad \forall s > 0$$

$$\Rightarrow \frac{\mathcal{E}(P^*P, f)}{\Phi(f)} \geq \frac{\|f\|_2^2}{s\Phi(f)} - \frac{\beta(s)}{s}.$$

Set $u := 1/s$, $K(u) := u\beta(1/u)$.

$$\frac{\mathcal{E}(P^*P, f)}{\Phi(f)} \geq u \cdot \frac{\|f\|_2^2}{\Phi(f)} - K(u), \quad \forall u > 0.$$

$$\frac{\mathcal{E}(P^*P, f)}{\Phi(f)} \geq \sup_{u>0} \left\{ u \cdot \frac{\|f\|_2^2}{\Phi(f)} - K(u) \right\} =: K^* \left(\frac{\|f\|_2^2}{\Phi(f)} \right).$$

Call this final inequality optimized WPI (oWPI).

Proof of convergence bound (II)

Now define

$$F(x) := \int_x^1 \frac{\mathrm{d}v}{K^*(v)}, \quad x \in (0, a], \qquad h_n := \frac{\|P^n f\|_2^2}{\Phi(f)}.$$

Proof of convergence bound (II)

Now define

$$F(x) := \int_x^1 \frac{\mathrm{d}v}{K^*(v)}, \quad x \in (0, a], \qquad h_n := \frac{\|P^n f\|_2^2}{\Phi(f)}.$$

Want to bound convergence of $h_n \to 0$.

Proof of convergence bound (II)

Now define

$$F(x) := \int_x^1 \frac{\mathrm{d}v}{K^*(v)}, \quad x \in (0, a], \qquad h_n := \frac{\|P^n f\|_2^2}{\Phi(f)}.$$

Want to bound convergence of $h_n \to 0$.

$$F(h_n) - F(h_{n-1}) = \int_{h_n}^{h_{n-1}} \frac{\mathrm{d}v}{K^*(v)}$$

Proof of convergence bound (II)

Now define

$$F(x) := \int_x^1 \frac{\mathrm{d}v}{K^*(v)}, \quad x \in (0, a], \qquad h_n := \frac{\|P^n f\|_2^2}{\Phi(f)}.$$

Want to bound convergence of $h_n \to 0$.

$$\begin{aligned} F(h_n) - F(h_{n-1}) &= \int_{h_n}^{h_{n-1}} \frac{\mathrm{d}v}{K^*(v)} \\ &\geq (h_{n-1} - h_n)/K^*(h_{n-1}) \end{aligned}$$

Proof of convergence bound (II)

Now define

$$F(x) := \int_x^1 \frac{\mathrm{d}v}{K^*(v)}, \quad x \in (0, a], \qquad h_n := \frac{\|P^n f\|_2^2}{\Phi(f)}.$$

Want to bound convergence of $h_n \to 0$.

$$\begin{aligned} F(h_n) - F(h_{n-1}) &= \int_{h_n}^{h_{n-1}} \frac{\mathrm{d}v}{K^*(v)} \\ &\geq (h_{n-1} - h_n)/K^*(h_{n-1}) \\ &= \frac{\mathcal{E}(P^*P, P^{n-1}f)/\Phi(f)}{K^*(h_{n-1})} \end{aligned}$$

Proof of convergence bound (II)

Now define

$$F(x) := \int_x^1 \frac{\mathrm{d}v}{K^*(v)}, \quad x \in (0, a], \qquad h_n := \frac{\|P^n f\|_2^2}{\Phi(f)}.$$

Want to bound convergence of $h_n \to 0$.

$$\begin{aligned} F(h_n) - F(h_{n-1}) &= \int_{h_n}^{h_{n-1}} \frac{\mathrm{d}v}{K^*(v)} \\ &\geq (h_{n-1} - h_n)/K^*(h_{n-1}) \\ &= \frac{\mathcal{E}(P^*P, P^{n-1}f)/\Phi(f)}{K^*(h_{n-1})} \\ &\geq K^*(h_{n-1})/K^*(h_{n-1}) = 1. \qquad \text{(oWPI)} \end{aligned}$$

Proof of convergence bound (II)

Now define

$$F(x) := \int_x^1 \frac{\mathrm{d}v}{K^*(v)}, \quad x \in (0, a], \qquad h_n := \frac{\|P^n f\|_2^2}{\Phi(f)}.$$

Want to bound convergence of $h_n \to 0$.

$$\begin{aligned} F(h_n) - F(h_{n-1}) &= \int_{h_n}^{h_{n-1}} \frac{\mathrm{d}v}{K^*(v)} \\ &\geq (h_{n-1} - h_n)/K^*(h_{n-1}) \\ &= \frac{\mathcal{E}(P^*P, P^{n-1}f)/\Phi(f)}{K^*(h_{n-1})} \\ &\geq K^*(h_{n-1})/K^*(h_{n-1}) = 1. \qquad \text{(oWPI)} \\ &\Rightarrow F(h_n) - F(h_0) \geq n. \end{aligned}$$

Proof of convergence bound (II)

Now define

$$F(x) := \int_x^1 \frac{\mathrm{d}v}{K^*(v)}, \quad x \in (0, a], \qquad h_n := \frac{\|P^n f\|_2^2}{\Phi(f)}.$$

Want to bound convergence of $h_n \to 0$.

$$\begin{aligned} F(h_n) - F(h_{n-1}) &= \int_{h_n}^{h_{n-1}} \frac{\mathrm{d}v}{K^*(v)} \\ &\geq (h_{n-1} - h_n)/K^*(h_{n-1}) \\ &= \frac{\mathcal{E}(P^*P, P^{n-1}f)/\Phi(f)}{K^*(h_{n-1})} \\ &\geq K^*(h_{n-1})/K^*(h_{n-1}) = 1. \qquad \text{(oWPI)} \\ &\Rightarrow F(h_n) - F(h_0) \geq n. \end{aligned}$$

So we invert this to obtain

$$\|P^n f\|_2^2 \leq \Phi(f) F^{-1}(n). \quad \square$$

Examples

Theorem ([Andrieu, Lee, Power, Wang (2022)])

Under a weak Poincaré inequality, we have, $\forall n \in \mathbb{N}_0$, $f \in \mathrm{L}_0^2(\mu)$,

$$\|P^n f\|_2^2 \leq \Phi(f) F^{-1}(n).$$

If $\beta(s) = c_0 s^{-c_1}$, we can bound

$$F^{-1}(n) \leq C n^{-c_1}.$$

Examples

Theorem ([Andrieu, Lee, Power, Wang (2022)])

Under a weak Poincaré inequality, we have, $\forall n \in \mathbb{N}_0$, $f \in \mathrm{L}_0^2(\mu)$,

$$\|P^n f\|_2^2 \leq \Phi(f) F^{-1}(n).$$

If $\beta(s) = c_0 s^{-c_1}$, we can bound

$$F^{-1}(n) \leq C n^{-c_1}.$$

If $\beta(s) = \eta_0 \exp(-\eta_1 s^{\eta_2})$, we can bound

$$F^{-1}(n) \leq C' \exp\left(-(Cn)^{\eta_2/(1+\eta_2)}\right).$$

Examples

Theorem ([Andrieu, Lee, Power, Wang (2022)])

Under a weak Poincaré inequality, we have, $\forall n \in \mathbb{N}_0$, $f \in \mathrm{L}^2_0(\mu)$,

$$\|P^n f\|_2^2 \le \Phi(f) F^{-1}(n).$$

If $\beta(s) = c_0 s^{-c_1}$, we can bound

$$F^{-1}(n) \le C n^{-c_1}.$$

If $\beta(s) = \eta_0 \exp(-\eta_1 s^{\eta_2})$, we can bound

$$F^{-1}(n) \le C' \exp\left(-(Cn)^{\eta_2/(1+\eta_2)}\right).$$

Intuition: the faster β decays, the faster the rate of convergence.

Overview

Application to pseudo-marginal MCMC

Recall our goal was to precisely characterise the degradation in convergence when using a pseudo-marginal chain with $\hat{\pi}(x) = W_x \cdot \pi(x)$ compared to the marginal MH chain with $\pi(x)$.

Application to pseudo-marginal MCMC

Recall our goal was to precisely characterise the degradation in convergence when using a pseudo-marginal chain with $\hat{\pi}(x) = W_x \cdot \pi(x)$ compared to the marginal MH chain with $\pi(x)$.

Let $\tilde{P}$ denote the pseudo-marginal RWM chain and let P denote the marginal RWM chain.
(Both constructed on an augmented space containing the weights w.)

Application to pseudo-marginal MCMC

Recall our goal was to precisely characterise the degradation in convergence when using a pseudo-marginal chain with $\hat{\pi}(x) = W_x \cdot \pi(x)$ compared to the marginal MH chain with $\pi(x)$.

Let $\tilde{P}$ denote the pseudo-marginal RWM chain and let P denote the marginal RWM chain.
(Both constructed on an augmented space containing the weights w.)

Have $W_x \sim Q_x$ nonnegative with $\mathbb{E}[W_x] = 1$, and set $\tilde{\pi}_x(\mathrm{d}w) := Q_x(\mathrm{d}w)w$.

Application to pseudo-marginal MCMC

Recall our goal was to precisely characterise the degradation in convergence when using a pseudo-marginal chain with $\hat{\pi}(x) = W_x \cdot \pi(x)$ compared to the marginal MH chain with $\pi(x)$.

Let $\tilde{P}$ denote the pseudo-marginal RWM chain and let P denote the marginal RWM chain.
(Both constructed on an augmented space containing the weights w.)

Have $W_x \sim Q_x$ nonnegative with $\mathbb{E}[W_x] = 1$, and set $\tilde{\pi}_x(\mathrm{d}w) := Q_x(\mathrm{d}w)w$.

Theorem ([Andrieu, Lee, Power, Wang (2022)])

We have for any $s > 0$, $f \in \mathrm{L}^2_0(\mu)$ bounded,

$$\mathcal{E}(P, f) \leq s\,\mathcal{E}(\tilde{P}, f) + \beta(s)\|f\|^2_{\mathrm{osc}},$$

where

$$\beta(s) := \int_{\mathcal{X}} \tilde{\pi}_x(w \geq s)\pi(\mathrm{d}x), \quad s > 0.$$

Application 1: Approximate Bayesian Computation (ABC)

We run an MCMC chain targeting the ABC posterior, true likelihood $\ell_y(x) = f_x(y)$,

$$\pi_{\mathrm{ABC}}(x) \propto \nu(x)\ell_{\mathrm{ABC}}(x),$$
$$\ell_{\mathrm{ABC}}(x) := \mathbb{P}_x(|Z - y| < \epsilon), \quad Z \sim f_x.$$

Application 1: Approximate Bayesian Computation (ABC)

We run an MCMC chain targeting the ABC posterior, true likelihood $\ell_y(x) = f_x(y)$,

$$\pi_{\mathrm{ABC}}(x) \propto \nu(x)\ell_{\mathrm{ABC}}(x),$$
$$\ell_{\mathrm{ABC}}(x) := \mathbb{P}_x(|Z - y| < \epsilon), \quad Z \sim f_x.$$

This is further approximated using a pseudo-marginal approach: $\mathcal{W}_M := \frac{1}{M}\sum_{j=1}^{M} W_j$,

$$W_j := 1\{\left|Z_j - y\right| < \epsilon\}/\ell_{\mathrm{ABC}}(x), \quad Z_j \overset{\mathrm{iid}}{\sim} f_x.$$

Application 1: Approximate Bayesian Computation (ABC)

We run an MCMC chain targeting the ABC posterior, true likelihood $\ell_y(x) = f_x(y)$,

$$\pi_{\mathrm{ABC}}(x) \propto \nu(x)\ell_{\mathrm{ABC}}(x),$$
$$\ell_{\mathrm{ABC}}(x) := \mathbb{P}_x(|Z - y| < \epsilon), \quad Z \sim f_x.$$

This is further approximated using a pseudo-marginal approach: $\mathcal{W}_M := \frac{1}{M}\sum_{j=1}^{M} W_j$,

$$W_j := 1\{\left|Z_j - y\right| < \epsilon\}/\ell_{\mathrm{ABC}}(x), \quad Z_j \overset{\mathrm{iid}}{\sim} f_x.$$

Theorem ([Andrieu, Lee, Power, Wang (2022)])

Suppose that $\int_{\mathcal{X}} \nu(x)\ell_{\mathrm{ABC}}^{-(p-1)}(x)\,\mathrm{d}x < \infty$. Then there is $C_{M,p} > 0$ such that for all $s > 0$,

$$\beta(s) = \int_{\mathcal{X}} \pi(\mathrm{d}x)\tilde{\pi}_x(\mathcal{W}_M \geq s) \leq C_{M,p}s^{-p},$$

and as $M \to \infty$, $C_{M,p} = 1 + O(1/M)$. The convergence rate for the chain is $O(n^{-p})$.

Application 2: Particle Marginal Metropolis–Hastings (PMMH)

PMMH is a well-established algorithm to perform MCMC, e.g. for state space models [Andrieu et. al. (2010)].

Application 2: Particle Marginal Metropolis–Hastings (PMMH)

PMMH is a well-established algorithm to perform MCMC, e.g. for state space models [Andrieu et. al. (2010)].

We can study the limiting case when the weights are log-normal$(-\sigma^2/2, \sigma^2)$, where $\sigma^2 = \sigma_0^2/N$, N number of particles [Bérard et. al. (2014)].

Application 2: Particle Marginal Metropolis–Hastings (PMMH)

PMMH is a well-established algorithm to perform MCMC, e.g. for state space models [Andrieu et. al. (2010)].

We can study the limiting case when the weights are log-normal$(-\sigma^2/2, \sigma^2)$, where $\sigma^2 = \sigma_0^2/N$, N number of particles [Bérard et. al. (2014)].

Theorem ([Andrieu, Lee, Power, Wang (2022)])

Assume the marginal chain satisfies a SPI. Then we have the convergence bound, $\forall n \in \mathbb{N}$,

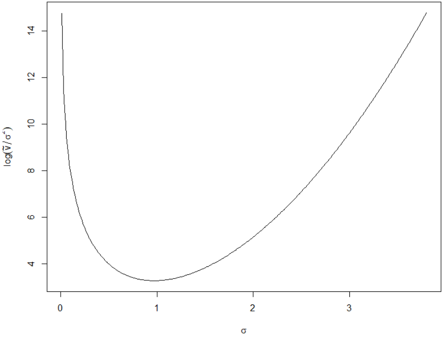
$$F_{\mathrm{PMMH}}^{-1}(n) \leq \frac{2}{C_{\mathrm{P}}} \exp\left\{-\frac{1}{2\sigma^2} \mathrm{W}^2\left(\frac{C_{\mathrm{P}}\sigma^2}{2\exp(\sigma^2/2)} \cdot n\right)\right\},$$

W *is the Lambert function (inverse of $x \mapsto x\mathrm{e}^x$).*

Application to Particle Marginal Metropolis–Hastings (PMMH) II

To minimize the asymptotic variance taking into account the computational cost, (*c.f.* our results for convergence bounds and mixing times), tune algorithm so that

$$\sigma \approx 0.973.$$



Summary

I introduced the framework of [Andrieu, Lee, Power, Wang (2022)]: weak Poincaré inequalities for Markov chains, as a new tool to compare convergence of Markov chains, which covers subgeometric convergence.

Summary

I introduced the framework of [Andrieu, Lee, Power, Wang (2022)]: weak Poincaré inequalities for Markov chains, as a new tool to compare convergence of Markov chains, which covers subgeometric convergence.

This technique has been particularly fruitful for analyzing pseudo-marginal MCMC methods, significantly extending [Andrieu and Vihola (2015)],

Summary

I introduced the framework of [Andrieu, Lee, Power, Wang (2022)]: weak Poincaré inequalities for Markov chains, as a new tool to compare convergence of Markov chains, which covers subgeometric convergence.

This technique has been particularly fruitful for analyzing pseudo-marginal MCMC methods, significantly extending [Andrieu and Vihola (2015)], but the general approach could be applied much more broadly.

Summary

I introduced the framework of [Andrieu, Lee, Power, Wang (2022)]: weak Poincaré inequalities for Markov chains, as a new tool to compare convergence of Markov chains, which covers subgeometric convergence.

This technique has been particularly fruitful for analyzing pseudo-marginal MCMC methods, significantly extending [Andrieu and Vihola (2015)], but the general approach could be applied much more broadly.

Combined with our results for this morning for RWM, this enables us to give complete convergence bounds for psuedo-marginal RWM.

Summary

I introduced the framework of [Andrieu, Lee, Power, Wang (2022)]: weak Poincaré inequalities for Markov chains, as a new tool to compare convergence of Markov chains, which covers subgeometric convergence.

This technique has been particularly fruitful for analyzing pseudo-marginal MCMC methods, significantly extending [Andrieu and Vihola (2015)], but the general approach could be applied much more broadly.

Combined with our results for this morning for RWM, this enables us to give complete convergence bounds for psuedo-marginal RWM.

Further fundamental theory in our follow-up report [Andrieu, Lee, Power, Wang (2022b)].

Later today: comparisons for slice sampling.

Overview

Thanks for listening! I

- Andrieu, C., Dobson, P., Wang, A. Q. (2021). Subgeometric hypocoercivity for piecewise-deterministic Markov process Monte Carlo methods. *Elec. J. Probab.*, 26, 1-26.
- Andrieu, C., Doucet, A., Holenstein, R. (2010). Particle Markov chain Monte Carlo methods. *J. Roy. Statist. Soc. Ser. B: Stat. Methodol.*, 72(3), 269-342.
- **Andrieu, C., Lee, A., Power, S., Wang, A. Q. (2022)**. Comparison of Markov chains via weak Poincaré inequalities with application to pseudo-marginal MCMC. *Ann. Statist.*, 50(6), 3592-3618.
- **Andrieu, C., Lee, A., Power, S., Wang, A. Q. (2022).** Explicit convergence bounds for Metropolis Markov chains: isoperimetry, spectral gaps and profiles. `https://doi.org/10.48550/arxiv.2211.08959`.
- **Andrieu, C., Lee, A., Power, S., Wang, A. Q. (2022)**. Poincaré inequalities for Markov chains: a meeting with Cheeger, Lyapunov and Metropolis. *Technical report*. `https://doi.org/10.48550/arxiv.2208.05239`.
- Andrieu, C., Roberts, G. O. (2009). The pseudo-marginal approach for efficient Monte Carlo computations. *Ann. Statist.*, 37(2), 697-725.
- Andrieu, C., Vihola, M. (2015). Convergence properties of pseudo-marginal Markov chain Monte Carlo algorithms. *Ann. Appl. Probab.*, 25(2), 1030-1077.
- Baxendale, P. H. (2005). Renewal theory and computable convergence rates for geometrically ergodic Markov chains. *Ann. Appl. Probab.*, 15(1B), 700-738.

Thanks for listening! II

Bérard, J., Del Moral, P., Doucet, A. (2014). A lognormal central limit theorem for particle approximations of normalizing constants. *Electron. J. Probab.*, 19, 1-28.

Cai, D., Adams, R.P. (2022). Multi-fidelity Monte Carlo: a pseudo-marginal approach. https://arxiv.org/abs/2210.01534.

Diaconis, P., Saloff-Coste, L. (1993). Comparison Theorems for Reversible Markov Chains. *Ann. Appl. Probab.*, 3(3), 696-730.

Garegnani (2021). Sampling Methods for Bayesian Inference Involving Convergent Noisy Approximations of Forward Maps. https://arxiv.org/abs/2111.03491.

Jarner, S. F., Hansen, E. (2000). Geometric ergodicity of Metropolis algorithms. *Stoc. Proc. Appl.*, 85(2), 341-361.

Lee, A., Łatuszyński, K. (2014). Variance bounding and geometric ergodicity of Markov chain Monte Carlo kernels for approximate Bayesian computation. *Biometrika*, 101(3), 655-671.

Marin, J. M., Pudlo, P., Robert, C. P., Ryder, R. J. (2012). Approximate Bayesian computational methods. *Statist. Comput.*, 22(6), 1167-1180.

Roberts, G. O., Gelman, A., Gilks, W. R. (1997). Weak Convergence and Optimal Scaling of random walk Metropolis algorithms. *Ann. Appl. Probab.*, 7(1), 110-120.

Roberts, G., Tweedie, R. L. (1996). Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. *Biometrika*, 83(1), 95-110.

Thanks for listening! III

Röckner, M., Wang, F.-Y. (2001). Weak Poincare Inequalities and L^2 Convergence Rates of Markov Semigroups. *J. Funct. Anal.*, 185, 564-603.

Stuart, A. (2010). Inverse problems: A Bayesian perspective. *Acta Numerica*, 19, 451-559.

Statistical modelling

Suppose we have some (potentially vast) dataset $y = \{y_1, y_2, \ldots, y_N\}$.

Statistical modelling

Suppose we have some (potentially vast) dataset $y = \{y_1, y_2, \ldots, y_N\}$.

Posit a model (density function) $f_x(y)$ which generated y, which depends upon parameters $x \in \mathcal{X} = \mathbb{R}^d$.

Statistical modelling

Suppose we have some (potentially vast) dataset $y = \{y_1, y_2, \dots, y_N\}$.

Posit a model (density function) $f_x(y)$ which generated y, which depends upon parameters $x \in \mathcal{X} = \mathbb{R}^d$.

Seek learn or infer values of the parameter x which are commensurate with the observed dataset y.

The Bayesian approach

Encode prior beliefs into a prior distribution $\nu(x)$, and define likelihood $\ell_y(x) := f_x(y)$.

The Bayesian approach

Encode prior beliefs into a prior distribution $\nu(x)$, and define likelihood $\ell_y(x) := f_x(y)$.

Given our observations, our posterior distribution is

$$\pi(x) = \pi(x|y) = \frac{\nu(x)\ell_y(x)}{\int \nu(z)\ell_y(z)\,\mathrm{d}z} \propto \nu(x)\ell_y(x).$$

The Bayesian approach

Encode prior beliefs into a prior distribution $\nu(x)$, and define likelihood $\ell_y(x) := f_x(y)$.

Given our observations, our posterior distribution is

$$\pi(x) = \pi(x|y) = \frac{\nu(x)\ell_y(x)}{\int \nu(z)\ell_y(z)\,\mathrm{d}z} \propto \nu(x)\ell_y(x).$$

We are then interested in quantities of the form

$$I = \pi(f) = \int_{\mathcal{X}} f(x)\pi(x)\,\mathrm{d}x,$$

e.g. $f(x) = \|x\|^p$ (posterior moments), $f(x) = 1_A(x)$ (credible sets / posterior tail probabilities), etc.

Sampling

So we wish to evaluate integrals

$$I = \pi(f) = \int_{\mathcal{X}} f(x)\pi(x)\,\mathrm{d}x,$$

where π is a probability density function (our posterior distribution).

Sampling

So we wish to evaluate integrals

$$I = \pi(f) = \int_{\mathcal{X}} f(x)\pi(x)\,\mathrm{d}x,$$

where π is a probability density function (our posterior distribution).

Direct integration infeasible in high-dimensions (curse of dimensionality), furthermore only have access to π up to a normalizing constant!

Sampling

So we wish to evaluate integrals

$$I = \pi(f) = \int_{\mathcal{X}} f(x)\pi(x)\,\mathrm{d}x,$$

where π is a probability density function (our posterior distribution).

Direct integration infeasible in high-dimensions (curse of dimensionality), furthermore only have access to π up to a normalizing constant!

So instead, approximate I by sampling $X_1, X_2, \ldots, X_n \sim \pi$ and consider

$$I_n := \frac{1}{n}\sum_{i=1}^{n} f(X_i) \approx I = \int_{\mathcal{X}} f(x)\pi(x)\,\mathrm{d}x.$$

Sampling

So we wish to evaluate integrals

$$I = \pi(f) = \int_{\mathcal{X}} f(x)\pi(x)\,\mathrm{d}x,$$

where π is a probability density function (our posterior distribution).

Direct integration infeasible in high-dimensions (curse of dimensionality), furthermore only have access to π up to a normalizing constant!

So instead, approximate I by sampling $X_1, X_2, \dots, X_n \sim \pi$ and consider

$$I_n := \frac{1}{n}\sum_{i=1}^{n} f(X_i) \approx I = \int_{\mathcal{X}} f(x)\pi(x)\,\mathrm{d}x.$$

Will not be discussing optimization-based approaches such as Variational Inference, INLA, ...

Monte Carlo

So instead, approximate I by sampling $X_1, X_2, \ldots, X_n \sim \pi$.

Monte Carlo

So instead, approximate I by sampling $X_1, X_2, \ldots, X_n \sim \pi$.

Exact sampling hard (e.g. rejection sampling also suffers from a curse of dimensionality)

Monte Carlo

So instead, approximate I by sampling $X_1, X_2, \ldots, X_n \sim \pi$.

Exact sampling hard (e.g. rejection sampling also suffers from a curse of dimensionality)
so instead: build an ergodic Markov chain X which possesses π as its stationary distribution.

Monte Carlo

So instead, approximate I by sampling $X_1, X_2, \ldots, X_n \sim \pi$.

Exact sampling hard (e.g. rejection sampling also suffers from a curse of dimensionality) so instead: build an ergodic Markov chain X which possesses π as its stationary distribution.

We simulate a π-reversible ergodic Markov chain,

$$X_1, X_2, \ldots$$

where $X_n \to \pi$ in distribution and considering

$$I_n := \frac{1}{n} \sum_{i=1}^{n} f(X_i) \approx I = \int_{\mathcal{X}} f(x)\pi(x)\,\mathrm{d}x.$$

Random walk Metropolis–Hastings

We will focus on random walk Metropolis–Hastings (RWM): $q(X_{i-1}, \cdot) = \mathcal{N}(X_{i-1}, \Sigma)$.

Random walk Metropolis–Hastings

We will focus on random walk Metropolis–Hastings (RWM): $q(X_{i-1}, \cdot) = \mathcal{N}(X_{i-1}, \Sigma)$.

What is the criteria for an MCMC chain to be 'good'?

Random walk Metropolis–Hastings

We will focus on random walk Metropolis–Hastings (RWM): $q(X_{i-1}, \cdot) = \mathcal{N}(X_{i-1}, \Sigma)$.

What is the criteria for an MCMC chain to be 'good'?

Classically, MCMC is good if it converges fast to equilibrium and mixes well.

Random walk Metropolis–Hastings

We will focus on random walk Metropolis–Hastings (RWM): $q(X_{i-1}, \cdot) = \mathcal{N}(X_{i-1}, \Sigma)$.

What is the criteria for an MCMC chain to be 'good'?

Classically, MCMC is good if it converges fast to equilibrium and mixes well.

One measure of the former is to look at rates of convergence:

Random walk Metropolis–Hastings

We will focus on random walk Metropolis–Hastings (RWM): $q(X_{i-1}, \cdot) = \mathcal{N}(X_{i-1}, \Sigma)$.

What is the criteria for an MCMC chain to be 'good'?

Classically, MCMC is good if it converges fast to equilibrium and mixes well.

One measure of the former is to look at rates of convergence:

Theorem ([Roberts and Tweedie (1996), Jarner and Hansen (2000)])

RWM converges to equilibrium exponentially fast if and only if π has an exponential moment (e.g. $\pi(x) \propto \exp(-\|x-\mu\|^\alpha)$, $\alpha \geq 1$.). Otherwise, the chain converges at a subgeometric (e.g. polynomial) rate.*

Reversibility and positivity

The definition of a WPI involved $\mathcal{E}(P^*P, f)$. In general P^*P is a complex object.

Reversibility and positivity

The definition of a WPI involved $\mathcal{E}(P^*P, f)$. In general P^*P is a complex object.

Recall that P is reversible if $P^* = P$, and furthermore P is positive if for all $f \in L^2(\mu)$, $\langle f, Pf \rangle \geq 0$.

Reversibility and positivity

The definition of a WPI involved $\mathcal{E}(P^*P, f)$. In general P^*P is a complex object.

Recall that P is reversible if $P^* = P$, and furthermore P is positive if for all $f \in L^2(\mu)$, $\langle f, Pf \rangle \geq 0$.

Lemma ([Baxendale (2005)], [Andrieu and Vihola (2015)])

The following MCMC chains are positive: Random Walk Metropolis with Gaussian proposals, the Independence Sampler, and pseudo-marginal MH built from any of these chains.

Reversibility and positivity

The definition of a WPI involved $\mathcal{E}(P^*P, f)$. In general P^*P is a complex object.

Recall that P is reversible if $P^* = P$, and furthermore P is positive if for all $f \in L^2(\mu)$, $\langle f, Pf \rangle \geq 0$.

Lemma ([Baxendale (2005)], [Andrieu and Vihola (2015)])

The following MCMC chains are positive: Random Walk Metropolis with Gaussian proposals, the Independence Sampler, and pseudo-marginal MH built from any of these chains.

Theorem [Andrieu, Lee, Power, Wang (2022)]

For a positive kernel P, we have for all $f \in \mathrm{L}^2(\mu)$,

$$\mathcal{E}(P, f) \leq \mathcal{E}(P^2, f).$$

Therefore a WPI for P implies a WPI for $P^*P = P^2$.

Comparison of Markov chains

WPIs

We have discussed WPIs of the form:

$$\|f\|_2^2 \leq s\mathcal{E}(P, f) + \beta(s)\Phi(f), \quad \forall s > 0, \quad \forall f \in \mathrm{L}_0^2(\mu). \tag{1}$$

Comparison of Markov chains

WPIs

We have discussed WPIs of the form:

$$\|f\|_2^2 \le s\mathcal{E}(P, f) + \beta(s)\Phi(f), \quad \forall s > 0, \quad \forall f \in \mathrm{L}_0^2(\mu). \tag{1}$$

In order to compare Markov chains, we will consider a more general form of inequalities.

General comparison inequality

For two (reversible) Markov kernels P, P_2:

$$\mathcal{E}(P_1, f) \le s\mathcal{E}(P_2, f) + \beta(s)\Phi(f), \quad \forall s > 0, \quad \forall f \in \mathrm{L}_0^2(\mu). \tag{2}$$

Comparison of Markov chains

WPIs

We have discussed WPIs of the form:

$$\|f\|_2^2 \leq s\mathcal{E}(P, f) + \beta(s)\Phi(f), \quad \forall s > 0, \quad \forall f \in \mathrm{L}_0^2(\mu). \tag{1}$$

In order to compare Markov chains, we will consider a more general form of inequalities.

General comparison inequality

For two (reversible) Markov kernels P, P_2:

$$\mathcal{E}(P_1, f) \leq s\mathcal{E}(P_2, f) + \beta(s)\Phi(f), \quad \forall s > 0, \quad \forall f \in \mathrm{L}_0^2(\mu). \tag{2}$$

Indeed, (1) is a special case of (2) where $P_1(x, \mathrm{d}y) = \mu(\mathrm{d}y)$ corresponds to perfect sampling.

Comparison of Markov chains

WPIs

We have discussed WPIs of the form:

$$\|f\|_2^2 \le s\mathcal{E}(P, f) + \beta(s)\Phi(f), \quad \forall s > 0, \quad \forall f \in \mathrm{L}_0^2(\mu). \tag{1}$$

In order to compare Markov chains, we will consider a more general form of inequalities.

General comparison inequality

For two (reversible) Markov kernels P, P_2:

$$\mathcal{E}(P_1, f) \le s\mathcal{E}(P_2, f) + \beta(s)\Phi(f), \quad \forall s > 0, \quad \forall f \in \mathrm{L}_0^2(\mu). \tag{2}$$

Indeed, (1) is a special case of (2) where $P_1(x, \mathrm{d}y) = \mu(\mathrm{d}y)$ corresponds to perfect sampling.

Intuition: (2) gives a bound on the convergence of P_2 relative to the convergence rate of P_1.

General comparison result

Theorem ([Andrieu, Lee, Power, Wang (2022)])

Let P_1, P_2 be two μ-invariant Markov kernels on $\mathsf{E} \times \mathcal{F}$. *Assume that for any* $(x, B) \subset \mathsf{E} \times \mathcal{F}$,

$$P_2(x, B \backslash \{x\}) \geq \int_{B \backslash \{x\}} \epsilon(x, y) P_1(x, \mathrm{d}y),$$

for some $\epsilon : \mathsf{E}^2 \to (0, \infty)$.

Then for any $s > 0$, $f \in \mathrm{L}_0^\infty(\mu) \subset \mathrm{L}_0^2(\mu)$,

$$\mathcal{E}(P_1, f) \leq s\, \mathcal{E}(P_2, f) + \frac{1}{2} \mu \otimes P_1(A(s)^\complement \cap \{X \neq Y\}) \Phi(f),$$

where $A(s) := \{(x, y) \in \mathsf{E}^2 : s\, \epsilon(x, y) > 1\}$, *and*

$$\Phi(f) := \|f\|_{\mathrm{osc}}^2.$$

General comparison result - remarks

What we need to show is

$$P_2(x, B\backslash\{x\}) \geq \int_{B\backslash\{x\}} \epsilon(x,y) P_1(x, \mathrm{d}y),$$

for some $\epsilon : \mathsf{E}^2 \to (0, \infty)$.

General comparison result - remarks

What we need to show is

$$P_2(x, B\backslash\{x\}) \geq \int_{B\backslash\{x\}} \epsilon(x, y) P_1(x, \mathrm{d}y),$$

for some $\epsilon : \mathsf{E}^2 \to (0, \infty)$.
The theorem then immediately gives us a (relative) WPI off the back of this.

General comparison result - remarks

What we need to show is

$$P_2(x, B\backslash\{x\}) \geq \int_{B\backslash\{x\}} \epsilon(x, y) P_1(x, \mathrm{d}y),$$

for some $\epsilon : \mathsf{E}^2 \to (0, \infty)$.
The theorem then immediately gives us a (relative) WPI off the back of this.

If $\inf_{x,y} \epsilon(x, y) > 0$, then we will get a SPI / geometric convergence.

General comparison result - remarks

What we need to show is

$$P_2(x, B\backslash\{x\}) \geq \int_{B\backslash\{x\}} \epsilon(x,y) P_1(x, \mathrm{d}y),$$

for some $\epsilon : \mathsf{E}^2 \to (0, \infty)$.
The theorem then immediately gives us a (relative) WPI off the back of this.

If $\inf_{x,y} \epsilon(x,y) > 0$, then we will get a SPI / geometric convergence.

If we aren't interested in comparisons, but want want an (absolute) WPI, note that if $P(x, \mathrm{d}y) = \pi(\mathrm{d}y)$ represents perfect sampling, then $\mathcal{E}(P, f) = \|f\|_2^2$, and there is the useful representation (exercise!)

$$\|f\|_2^2 = \frac{1}{2} \int \pi(x)\pi(y)[f(y) - f(x)]^2 \, \mathrm{d}x \, \mathrm{d}y.$$

Independence Sampler: geometric case

The Independence Sampler (IS) is one the simplest MCMC methods: given target π, at each step sample proposal $Y_i \sim q$, and accept with probability

$$\alpha(X_{i-1}, Y_i) = 1 \wedge \frac{q(X_{i-1})\pi(Y_i)}{\pi(X_{i-1})q(Y_i)} = 1 \wedge \frac{w(Y_i)}{w(X_{i-1})}.$$

Independence Sampler: geometric case

The Independence Sampler (IS) is one the simplest MCMC methods: given target π, at each step sample proposal $Y_i \sim q$, and accept with probability

$$\alpha(X_{i-1}, Y_i) = 1 \wedge \frac{q(X_{i-1})\pi(Y_i)}{\pi(X_{i-1})q(Y_i)} = 1 \wedge \frac{w(Y_i)}{w(X_{i-1})}.$$

Well-known that there is a spectral gap if and only if the weights are bounded:

$$w(x) := \frac{\pi(x)}{q(x)} \leq M, \quad \forall x \in \mathcal{X}.$$

(E.g. if so, then you can just do rejection sampling.)

Independence Sampler: geometric case

The Independence Sampler (IS) is one the simplest MCMC methods: given target π, at each step sample proposal $Y_i \sim q$, and accept with probability

$$\alpha(X_{i-1}, Y_i) = 1 \wedge \frac{q(X_{i-1})\pi(Y_i)}{\pi(X_{i-1})q(Y_i)} = 1 \wedge \frac{w(Y_i)}{w(X_{i-1})}.$$

Well-known that there is a spectral gap if and only if the weights are bounded:

$$w(x) := \frac{\pi(x)}{q(x)} \leq M, \quad \forall x \in \mathcal{X}.$$

(E.g. if so, then you can just do rejection sampling.)The kernel is

$$P(x, \mathrm{d}y) = q(y) \cdot 1 \wedge \frac{w(y)}{w(x)} \,\mathrm{d}y + (1 - \alpha(x))\delta_x(\mathrm{d}y).$$

Independence Sampler exercises

$$P(x, \mathrm{d}y) = q(y) \cdot 1 \wedge \frac{w(y)}{w(x)} \, \mathrm{d}y + (1 - \alpha(x))\delta_x(\mathrm{d}y), \quad w(x) := \frac{\pi(x)}{q(x)}.$$

Independence Sampler exercises

$$P(x, \mathrm{d}y) = q(y) \cdot 1 \wedge \frac{w(y)}{w(x)} \, \mathrm{d}y + (1 - \alpha(x))\delta_x(\mathrm{d}y), \quad w(x) := \frac{\pi(x)}{q(x)}.$$

1. Use the general comparison theorem to establish a WPI for the Independence Sampler (i.e. show that $P(x, y) \geq \epsilon(x, y)\pi(y)$ for an appropriate function $\epsilon(x, y)$).
2. Directly deduce a WPI for the Independence Sampler using the useful representations below.

Independence Sampler exercises

$$P(x, \mathrm{d}y) = q(y) \cdot 1 \wedge \frac{w(y)}{w(x)} \,\mathrm{d}y + (1 - \alpha(x))\delta_x(\mathrm{d}y), \quad w(x) := \frac{\pi(x)}{q(x)}.$$

1. Use the general comparison theorem to establish a WPI for the Independence Sampler (i.e. show that $P(x, y) \geq \epsilon(x, y)\pi(y)$ for an appropriate function $\epsilon(x, y)$).
2. Directly deduce a WPI for the Independence Sampler using the useful representations below.

Useful representations (exercise!):

$$\|f\|_2^2 = \frac{1}{2} \int \pi(x)\pi(y)[f(y) - f(x)]^2 \,\mathrm{d}x \,\mathrm{d}y,$$

$$\mathcal{E}(P, f) = \frac{1}{2} \int \pi(x)\pi(y) \left(w^{-1}(x) \wedge w^{-1}(y)\right) [f(x) - f(y)]^2 \,\mathrm{d}x \,\mathrm{d}y.$$

Independence Sampler: subgeometric case

Suppose the weights are unbounded, namely

$$\sup_{x \in \mathcal{X}} w(x) = \sup_{x \in \mathcal{X}} \frac{\pi(x)}{q(x)} = \infty.$$

Independence Sampler: subgeometric case

Suppose the weights are unbounded, namely

$$\sup_{x \in \mathcal{X}} w(x) = \sup_{x \in \mathcal{X}} \frac{\pi(x)}{q(x)} = \infty.$$

The problematic region is where $w(x)$ is large, or equivalently, where $w^{-1}(x)$ is close to 0.

Independence Sampler: subgeometric case

Suppose the weights are unbounded, namely

$$\sup_{x\in\mathcal{X}} w(x) = \sup_{x\in\mathcal{X}} \frac{\pi(x)}{q(x)} = \infty.$$

The problematic region is where $w(x)$ is large, or equivalently, where $w^{-1}(x)$ is close to 0.

The previous representation of $\mathcal{E}(P, f)$ immediately gives us a WPI for the subgeometric IS: take

$$\epsilon(x, y) = \left(w^{-1}(x) \wedge w^{-1}(y)\right),$$

then

$$A(s) = \left\{(x, y) \in \mathsf{E} \times \mathsf{E} : \left(w^{-1}(x) \wedge w^{-1}(y)\right) \geq 1/2\right\}.$$

Pseudo-marginal derivation

The WPI for pseudo-marginal is derived in almost the same way as for the Independence Sampler!

We work on the augmented state space $\mathcal{X} \times [0, \infty)$.
The pseudo-marginal kernel $\tilde{P}$ is given by: $\mathfrak{r}(x, y)$ is the standard MH acceptance ratio,

$$\tilde{P}(x, w; \mathrm{d}y, \mathrm{d}u) = \left[1 \wedge \left\{\mathfrak{r}(x, y)\frac{u}{w}\right\}\right] q(x, \mathrm{d}y) Q_y(\mathrm{d}u) + \delta_{x,w}(\mathrm{d}y, \mathrm{d}u)\tilde{\rho}(x, w),$$

Pseudo-marginal derivation

The WPI for pseudo-marginal is derived in almost the same way as for the Independence Sampler!

We work on the augmented state space $\mathcal{X} \times [0, \infty)$.
The pseudo-marginal kernel $\tilde{P}$ is given by: $\mathfrak{r}(x, y)$ is the standard MH acceptance ratio,

$$\tilde{P}(x, w; \mathrm{d}y, \mathrm{d}u) = \left[1 \wedge \left\{\mathfrak{r}(x, y)\frac{u}{w}\right\}\right] q(x, \mathrm{d}y) Q_y(\mathrm{d}u) + \delta_{x,w}(\mathrm{d}y, \mathrm{d}u)\tilde{\rho}(x, w),$$

which we compare to the standard MH, on the extended state space:

$$P(x, w; \mathrm{d}y, \mathrm{d}u) \left[1 \wedge \mathfrak{r}(x, y)\right] q(x, \mathrm{d}y)\, u\, Q_y(\mathrm{d}u) + \delta_{x,w}(\mathrm{d}y, \mathrm{d}u)\tilde{\rho}(x, w)$$

Pseudo-marginal derivation

The WPI for pseudo-marginal is derived in almost the same way as for the Independence Sampler!

We work on the augmented state space $\mathcal{X} \times [0, \infty)$.
The pseudo-marginal kernel $\tilde{P}$ is given by: $\mathfrak{r}(x, y)$ is the standard MH acceptance ratio,

$$\tilde{P}(x, w; \mathrm{d}y, \mathrm{d}u) = \left[1 \wedge \left\{\mathfrak{r}(x, y)\frac{u}{w}\right\}\right] q(x, \mathrm{d}y) Q_y(\mathrm{d}u) + \delta_{x,w}(\mathrm{d}y, \mathrm{d}u)\tilde{\rho}(x, w),$$

which we compare to the standard MH, on the extended state space:

$$P(x, w; \mathrm{d}y, \mathrm{d}u) \left[1 \wedge \mathfrak{r}(x, y)\right] q(x, \mathrm{d}y)\, u\, Q_y(\mathrm{d}u) + \delta_{x,w}(\mathrm{d}y, \mathrm{d}u)\tilde{\rho}(x, w)$$

The problematic region is when w gets large. We take

$$\epsilon(x, w; y, u) = w^{-1} \wedge u^{-1}.$$