Some motivation for unbiased estimators, and progress on constructing them with couplings

Pierre E. Jacob

ESSEC BUSINESS SCHOOL

Journée AppliBUGS

19 June 2026

Outline

1 Monte Carlo estimators and bias

2 Couplings of MCMC algorithms
- Coupling of two random variables
- Coupling of two Markov chains

3 Unbiased estimators and convergence diagnostics
- Unbiased estimation
- Convergence diagnostics

4 Bias removal for importance sampling

Outline

1 Monte Carlo estimators and bias

2 Couplings of MCMC algorithms
- Coupling of two random variables
- Coupling of two Markov chains

3 Unbiased estimators and convergence diagnostics
- Unbiased estimation
- Convergence diagnostics

4 Bias removal for importance sampling

Monte Carlo with perfect samples

Target probability distribution π.

Test function h, with expectation with respect to π:

$$\pi(h) = \int h(x)\pi(dx).$$

If we can sample $X_1, \ldots, X_N \sim \pi$ independently, then the Monte Carlo estimator

$$\hat{\pi}_N(h) = \frac{1}{N}\sum_{n=1}^{N} h(X_n)$$

is unbiased: $\mathbb{E}[\hat{\pi}_N(h)] = \pi(h)$.
If $\pi(h^2) < \infty$, $\hat{\pi}_N(h) \to \pi(h)$ at the Monte Carlo rate, CLT, etc.

Monte Carlo with perfect samples

Target probability distribution π.

Test function h, with expectation with respect to π:

$$\pi(h) = \int h(x)\pi(dx).$$

If we can sample $X_1, \ldots, X_N \sim \pi$ independently, then the Monte Carlo estimator

$$\hat{\pi}_N(h) = \frac{1}{N}\sum_{n=1}^{N} h(X_n)$$

is unbiased: $\mathbb{E}[\hat{\pi}_N(h)] = \pi(h)$.

If $\pi(h^2) < \infty$, $\hat{\pi}_N(h) \to \pi(h)$ at the Monte Carlo rate, CLT, etc.

Monte Carlo with perfect samples

Target probability distribution π.

Test function h, with expectation with respect to π:

$$\pi(h) = \int h(x)\pi(dx).$$

If we can sample $X_1, \ldots, X_N \sim \pi$ independently, then the Monte Carlo estimator

$$\hat{\pi}_N(h) = \frac{1}{N}\sum_{n=1}^{N} h(X_n)$$

is unbiased: $\mathbb{E}[\hat{\pi}_N(h)] = \pi(h)$.
If $\pi(h^2) < \infty$, $\hat{\pi}_N(h) \to \pi(h)$ at the Monte Carlo rate, CLT, etc.

Markov chain Monte Carlo

MCMC: $X_0 \sim \pi_0$, then $X_t | X_{t-1} \sim P(X_{t-1}, \cdot)$ for $t \geq 1$.

Standard MCMC estimator of $\pi(h)$: $t^{-1} \sum_{s=0}^{t-1} h(X_s)$.

- $P^t(x, \cdot) = \int P^{t-1}(x, dx') P(x', \cdot)$: distribution of $X_t | X_0 = x$.
- $\pi_t = \pi_0 P^t = \int \pi_0(dx) P^t(x, \cdot)$: marginal distribution of X_t.

Since $\pi_t \neq \pi$, for any fixed t, MCMC is biased:

$$\mathbb{E}[t^{-1} \sum_{s=0}^{t-1} h(X_s)] \neq \pi(h).$$

Still, under conditions, convergence at the Monte Carlo rate.
The bias is asymptotically negligible in the MSE.
So, who cares about the bias?

Markov chain Monte Carlo

MCMC: $X_0 \sim \pi_0$, then $X_t|X_{t-1} \sim P(X_{t-1}, \cdot)$ for $t \geq 1$.

Standard MCMC estimator of $\pi(h)$: $t^{-1} \sum_{s=0}^{t-1} h(X_s)$.

- $P^t(x, \cdot) = \int P^{t-1}(x, dx')P(x', \cdot)$: distribution of $X_t|X_0 = x$.
- $\pi_t = \pi_0 P^t = \int \pi_0(dx) P^t(x, \cdot)$: marginal distribution of X_t.

Since $\pi_t \neq \pi$, for any fixed t, MCMC is biased:

$$\mathbb{E}[t^{-1} \sum_{s=0}^{t-1} h(X_s)] \neq \pi(h).$$

Still, under conditions, convergence at the Monte Carlo rate.
The bias is asymptotically negligible in the MSE.
So, who cares about the bias?

Markov chain Monte Carlo

MCMC: $X_0 \sim \pi_0$, then $X_t|X_{t-1} \sim P(X_{t-1}, \cdot)$ for $t \geq 1$.

Standard MCMC estimator of $\pi(h)$: $t^{-1}\sum_{s=0}^{t-1} h(X_s)$.

- $P^t(x,\cdot) = \int P^{t-1}(x, dx')P(x',\cdot)$: distribution of $X_t|X_0 = x$.
- $\pi_t = \pi_0 P^t = \int \pi_0(dx)P^t(x,\cdot)$: marginal distribution of X_t.

Since $\pi_t \neq \pi$, for any fixed t, MCMC is biased:

$$\mathbb{E}[t^{-1}\sum_{s=0}^{t-1} h(X_s)] \neq \pi(h).$$

Still, under conditions, convergence at the Monte Carlo rate.
The bias is asymptotically negligible in the MSE.
So, who cares about the bias?

Markov chain Monte Carlo

MCMC: $X_0 \sim \pi_0$, then $X_t | X_{t-1} \sim P(X_{t-1}, \cdot)$ for $t \geq 1$.

Standard MCMC estimator of $\pi(h)$: $t^{-1} \sum_{s=0}^{t-1} h(X_s)$.

- $P^t(x, \cdot) = \int P^{t-1}(x, dx') P(x', \cdot)$: distribution of $X_t | X_0 = x$.
- $\pi_t = \pi_0 P^t = \int \pi_0(dx) P^t(x, \cdot)$: marginal distribution of X_t.

Since $\pi_t \neq \pi$, for any fixed t, MCMC is biased:

$$\mathbb{E}[t^{-1} \sum_{s=0}^{t-1} h(X_s)] \neq \pi(h).$$

Still, under conditions, convergence at the Monte Carlo rate.
The bias is asymptotically negligible in the MSE.
So, who cares about the bias?

Markov chain Monte Carlo

MCMC: $X_0 \sim \pi_0$, then $X_t|X_{t-1} \sim P(X_{t-1}, \cdot)$ for $t \geq 1$.

Standard MCMC estimator of $\pi(h)$: $t^{-1} \sum_{s=0}^{t-1} h(X_s)$.

- $P^t(x, \cdot) = \int P^{t-1}(x, dx')P(x', \cdot)$: distribution of $X_t|X_0 = x$.
- $\pi_t = \pi_0 P^t = \int \pi_0(dx)P^t(x, \cdot)$: marginal distribution of X_t.

Since $\pi_t \neq \pi$, for any fixed t, MCMC is biased:

$$\mathbb{E}[t^{-1} \sum_{s=0}^{t-1} h(X_s)] \neq \pi(h).$$

Still, under conditions, convergence at the Monte Carlo rate.
The bias is asymptotically negligible in the MSE.
So, who cares about the bias?

Markov chain Monte Carlo

MCMC: $X_0 \sim \pi_0$, then $X_t | X_{t-1} \sim P(X_{t-1}, \cdot)$ for $t \geq 1$.

Standard MCMC estimator of $\pi(h)$: $t^{-1} \sum_{s=0}^{t-1} h(X_s)$.

- $P^t(x, \cdot) = \int P^{t-1}(x, dx') P(x', \cdot)$: distribution of $X_t | X_0 = x$.
- $\pi_t = \pi_0 P^t = \int \pi_0(dx) P^t(x, \cdot)$: marginal distribution of X_t.

Since $\pi_t \neq \pi$, for any fixed t, MCMC is biased:

$$\mathbb{E}[t^{-1} \sum_{s=0}^{t-1} h(X_s)] \neq \pi(h).$$

Still, under conditions, convergence at the Monte Carlo rate.
The bias is asymptotically negligible in the MSE.
So, who cares about the bias?

If it was not for the bias...

Rosenthal, 2000.

Suppose that the chain is exactly stationary after k steps.

You could run R chains for $\ell > k$ steps, average h over the last $\ell - k$ states of each chain, then average over the R chains.

- Parallel computation over the R chains.
- Average estimator converges at Monte Carlo rate.
- Confidence intervals are easily obtained.
- Loss of efficiency relative to long-run MCMC controlled by ℓ.

If it was not for the bias...

Rosenthal, 2000.

Suppose that the chain is exactly stationary after k steps.

You could run R chains for $\ell > k$ steps, average h over the last $\ell - k$ states of each chain, then average over the R chains.

- Parallel computation over the R chains.
- Average estimator converges at Monte Carlo rate.
- Confidence intervals are easily obtained.
- Loss of efficiency relative to long-run MCMC controlled by ℓ.

If it was not for the bias...

Rosenthal, 2000.

Suppose that the chain is exactly stationary after k steps.

You could run R chains for $\ell > k$ steps, average h over the last $\ell - k$ states of each chain, then average over the R chains.

- Parallel computation over the R chains.
- Average estimator converges at Monte Carlo rate.
- Confidence intervals are easily obtained.
- Loss of efficiency relative to long-run MCMC controlled by ℓ.

Self-normalised importance sampling

Quantity of interest: $\pi(h) = \int h(x)\pi(dx)$ for some test function h. Global approximation $q \approx \pi$, such that $\pi \ll q$.

1. Sample $X_1, \ldots, X_N \overset{\text{ind}}{\sim} q$.
2. Compute $\omega(X_n) = \pi(X_n)/q(X_n)$ for all n.
3. Compute SNIS estimator

$$\hat{H} = \frac{\sum_{n=1}^{N} \omega(X_n) h(X_n)}{\sum_{n=1}^{N} \omega(X_n)}.$$

Easy to parallelize, confidence intervals from CLT + Δ-method.

Asymptotic bias of SNIS

Deligiannidis, Jacob, Khribch, Wang (2026),
and Daudel & Roueff (2026):

Assuming $q(|(h - \pi(h)) \cdot \omega^2|^{1+\epsilon}) < \infty$, and $q(\omega^{-\eta}) < \infty$ for some $\eta > 0$, we have

$$\lim_{N\to\infty} N \times \mathbb{E}\left[\hat{H} - \pi(h)\right] = -\int (h(x) - \pi(h))\, \omega^2(x) q(dx).$$

Bias of SNIS $\sim N^{-1}$. But who cares about the bias?

Asymptotic bias of SNIS

Deligiannidis, Jacob, Khribch, Wang (2026),
and Daudel & Roueff (2026):

Assuming $q(|(h-\pi(h))\cdot\omega^2|^{1+\epsilon})<\infty$, and $q(\omega^{-\eta})<\infty$ for some $\eta>0$, we have

$$\lim_{N\to\infty} N \times \mathbb{E}\left[\hat{H}-\pi(h)\right] = -\int \left(h(x)-\pi(h)\right)\omega^2(x)q(dx).$$

Bias of SNIS $\sim N^{-1}$. But who cares about the bias?

Nested expectation

Problem formalized in Andradóttir & Glynn (2002/2016), *Computing Bayesian Means Using Simulation*:

$$\text{QoI} = \mathbb{E}_1[H(X_1)], \quad \text{where } H(x_1) = \mathbb{E}_{2|1}[h(X_1, X_2)|X_1 = x_1],$$

for a test function h. How do we numerically approximate QoI?

Cutting feedback in Bayesian modular inference naturally leads to the nested expectation problem.

Example: X_1: missing data, X_2: parameters.
Or: X_1: weather forecast, X_2: biological parameters.

Nested expectation

Problem formalized in Andradóttir & Glynn (2002/2016), *Computing Bayesian Means Using Simulation*:

$$\text{QoI} = \mathbb{E}_1[H(X_1)], \quad \text{where } H(x_1) = \mathbb{E}_{2|1}[h(X_1, X_2)|X_1 = x_1],$$

for a test function h. How do we numerically approximate QoI?

Cutting feedback in Bayesian modular inference naturally leads to the nested expectation problem.

Example: X_1: missing data, X_2: parameters.
Or: X_1: weather forecast, X_2: biological parameters.

Nested approximations

Assume:

- we can generate X_1, M times for a cost M,
- we can generate estimator $\hat{H}(x_1)$ for any x_1, with
  - cost: N for each x_1,
  - bias: $\mathbb{E}[\hat{H}(x_1)] = H(x_1) + \beta(x_1)/N^\gamma$,
  - variance: $\mathbb{V}[\hat{H}(x_1)] = \sigma^2(x_1)/N$.

Estimator of QoI: $M^{-1}\sum_{m=1}^{M} \hat{H}(X_1^m)$, with cost $\approx M \times N =: C$.

What is the rate of convergence as $C = M \times N \to \infty$?

Optimal allocation leads to MSE of order $C^{-2\gamma/(2\gamma+1)}$.
Unbiased estimator for $\hat{H}$ would result in MSE $\sim C^{-1}$.

Nested approximations

Assume:

- we can generate X_1, M times for a cost M,
- we can generate estimator $\hat{H}(x_1)$ for any x_1, with
  - cost: N for each x_1,
  - bias: $\mathbb{E}[\hat{H}(x_1)] = H(x_1) + \beta(x_1)/N^\gamma$,
  - variance: $\mathbb{V}[\hat{H}(x_1)] = \sigma^2(x_1)/N$.

Estimator of QoI: $M^{-1}\sum_{m=1}^{M} \hat{H}(X_1^m)$, with cost $\approx M \times N =: C$.

What is the rate of convergence as $C = M \times N \to \infty$?

Optimal allocation leads to MSE of order $C^{-2\gamma/(2\gamma+1)}$.
Unbiased estimator for $\hat{H}$ would result in MSE $\sim C^{-1}$.

Nested approximations

Assume:

- we can generate X_1, M times for a cost M,
- we can generate estimator $\hat{H}(x_1)$ for any x_1, with
  - cost: N for each x_1,
  - bias: $\mathbb{E}[\hat{H}(x_1)] = H(x_1) + \beta(x_1)/N^\gamma$,
  - variance: $\mathbb{V}[\hat{H}(x_1)] = \sigma^2(x_1)/N$.

Estimator of QoI: $M^{-1}\sum_{m=1}^{M} \hat{H}(X_1^m)$, with cost $\approx M \times N =: C$.

What is the rate of convergence as $C = M \times N \to \infty$?

Optimal allocation leads to MSE of order $C^{-2\gamma/(2\gamma+1)}$.
Unbiased estimator for $\hat{H}$ would result in MSE $\sim C^{-1}$.

Nested approximations

Assume:

- we can generate X_1, M times for a cost M,
- we can generate estimator $\hat{H}(x_1)$ for any x_1, with
  - cost: N for each x_1,
  - bias: $\mathbb{E}[\hat{H}(x_1)] = H(x_1) + \beta(x_1)/N^\gamma$,
  - variance: $\mathbb{V}[\hat{H}(x_1)] = \sigma^2(x_1)/N$.

Estimator of QoI: $M^{-1}\sum_{m=1}^{M} \hat{H}(X_1^m)$, with cost $\approx M \times N =: C$.

What is the rate of convergence as $C = M \times N \to \infty$?

Optimal allocation leads to MSE of order $C^{-2\gamma/(2\gamma+1)}$.
Unbiased estimator for $\hat{H}$ would result in MSE $\sim C^{-1}$.

Nested approximations

Assume:

- we can generate X_1, M times for a cost M,
- we can generate estimator $\hat{H}(x_1)$ for any x_1, with
  - cost: N for each x_1,
  - bias: $\mathbb{E}[\hat{H}(x_1)] = H(x_1) + \beta(x_1)/N^\gamma$,
  - variance: $\mathbb{V}[\hat{H}(x_1)] = \sigma^2(x_1)/N$.

Estimator of QoI: $M^{-1}\sum_{m=1}^{M} \hat{H}(X_1^m)$, with cost $\approx M \times N =: C$.

What is the rate of convergence as $C = M \times N \to \infty$?

Optimal allocation leads to MSE of order $C^{-2\gamma/(2\gamma+1)}$.
Unbiased estimator for $\hat{H}$ would result in MSE $\sim C^{-1}$.

Outline

1. Monte Carlo estimators and bias
2. Couplings of MCMC algorithms
   - Coupling of two random variables
   - Coupling of two Markov chains
3. Unbiased estimators and convergence diagnostics
   - Unbiased estimation
   - Convergence diagnostics
4. Bias removal for importance sampling

Example: Metropolis–Rosenbluth–Teller–Hastings

With Markov chain at state X_t,

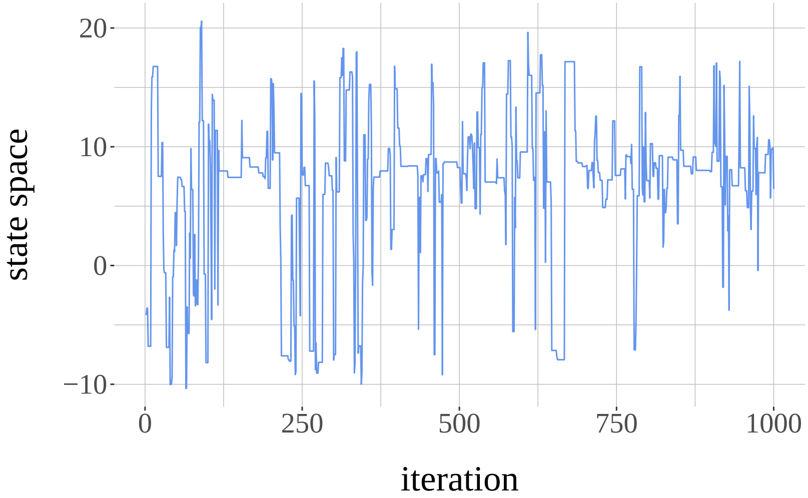
1. propose $X^\star \sim q(X_t, \cdot)$,

2. sample $U \sim \text{Uniform}(0,1)$,

3. if
$$U \leq \frac{\pi(X^\star)q(X^\star, X_t)}{\pi(X_t)q(X_t, X^\star)},$$
set $X_{t+1} = X^\star$, otherwise set $X_{t+1} = X_t$.

Hastings, 1970.

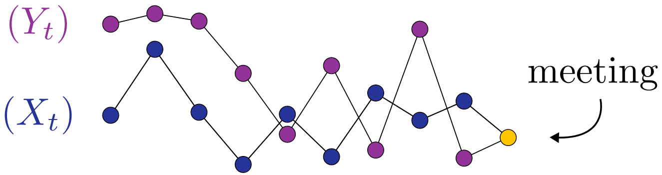
Example: traceplot



Goal: successful couplings

Consider the problem of generating two chains (X_t) and (Y_t), such that

- marginally, each chain starts from π_0 and evolves according to P,
- jointly, there exists $\tau < \infty$ such that $X_t = Y_t$ for all $t \geq \tau$.

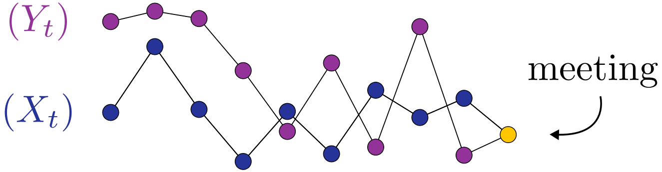


How are we going to achieve this?

Goal: successful couplings

Consider the problem of generating two chains (X_t) and (Y_t), such that

- marginally, each chain starts from π_0 and evolves according to P,
- jointly, there exists $\tau < \infty$ such that $X_t = Y_t$ for all $t \geq \tau$.



How are we going to achieve this?

Outline

1 Monte Carlo estimators and bias

2 Couplings of MCMC algorithms
- Coupling of two random variables
- Coupling of two Markov chains

3 Unbiased estimators and convergence diagnostics
- Unbiased estimation
- Convergence diagnostics

4 Bias removal for importance sampling

Couplings of random variables

A coupling of two random variables $X \sim p$ and $Y \sim q$ refers to a joint distribution for (X, Y) with marginals p and q.

For continuous variables, the independent coupling $p \otimes q$ is such that $\mathbb{P}(X = Y) = 0$.

Other couplings exist with a non-zero $\mathbb{P}(X = Y)$.

Couplings of random variables

Rejection sampling from q to p. Let $M \geq \sup_x p(x)/q(x)$.

1. Draw $X^\star \sim q$.
2. Draw $U \sim \text{Uniform}(0, 1)$.
3. If $U \leq p(X^\star)/(Mq(X^\star))$, then set $Y = X^\star$. Otherwise repeat.

Let X be the first proposal, and Y the outcome.

Then $X \sim q$, $Y \sim p$, and $\mathbb{P}(X = Y) = 1/M > 0$.

Couplings of random variables

Rejection sampling from q to p. Let $M \geq \sup_x p(x)/q(x)$.

1. Draw $X^\star \sim q$.
2. Draw $U \sim \text{Uniform}(0, 1)$.
3. If $U \leq p(X^\star)/(Mq(X^\star))$, then set $Y = X^\star$. Otherwise repeat.

Let X be the first proposal, and Y the outcome.

Then $X \sim q$, $Y \sim p$, and $\mathbb{P}(X = Y) = 1/M > 0$.

Couplings of random variables

For two distributions p and q,

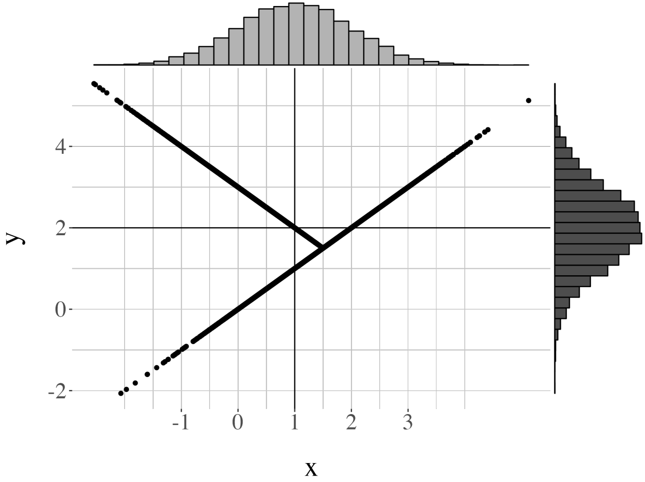
$$|p - q|_{\mathrm{TV}} = \sup_{h:\, |h|\leq 1} |p(h) - q(h)|$$
$$= 1 - \sup_{(X,Y)\in \text{couplings}(p,q)} \mathbb{P}(X = Y).$$

TV distance is zero *iff* $p = q$, and one *iff* supports are disjoint.

The supremum is attained by maximal couplings (not unique).

Reflection-maximal coupling

For two Normals with different means, same variance.



Bubley, Dyer, Jerrum 1998, Eberle, Guillin & Zimmer, 2019.

Reflection-maximal coupling

A coupling of Normal(μ_1, Σ) and Normal(μ_2, Σ).

1. Let $v = \Sigma^{-1/2}(\mu_1 - \mu_2)$ and $e = v/|v|$.
2. Sample $Z_1 \sim$ Normal$(0_d, I_d)$, and $W \sim$ Uniform$(0, 1)$.
3. If $W \leq \varphi(Z_1 + v)/\varphi(Z_1)$, set $Z_2 = Z_1 + v$.
   Otherwise set $Z_2 = (I_d - 2ee^T)Z_1$.
4. Set $X_1 = \Sigma^{1/2}Z_1 + \mu_1, X_2 = \Sigma^{1/2}Z_2 + \mu_2$, return (X_1, X_2).

Then: $X_1 \sim$ Normal(μ_1, Σ), $X_2 \sim$ Normal(μ_2, Σ),
and $\mathbb{P}(X_1 = X_2)$ is maximal.

Reflection-maximal coupling

A coupling of Normal(μ_1, Σ) and Normal(μ_2, Σ).

1. Let $v = \Sigma^{-1/2}(\mu_1 - \mu_2)$ and $e = v/|v|$.
2. Sample $Z_1 \sim \text{Normal}(0_d, I_d)$, and $W \sim \text{Uniform}(0,1)$.
3. If $W \leq \varphi(Z_1 + v)/\varphi(Z_1)$, set $Z_2 = Z_1 + v$.
   Otherwise set $Z_2 = (I_d - 2ee^T)Z_1$.
4. Set $X_1 = \Sigma^{1/2} Z_1 + \mu_1, X_2 = \Sigma^{1/2} Z_2 + \mu_2$, return (X_1, X_2).

Then: $X_1 \sim \text{Normal}(\mu_1, \Sigma)$, $X_2 \sim \text{Normal}(\mu_2, \Sigma)$,
and $\mathbb{P}(X_1 = X_2)$ is maximal.

Outline

1 Monte Carlo estimators and bias

2 Couplings of MCMC algorithms
- Coupling of two random variables
- Coupling of two Markov chains

3 Unbiased estimators and convergence diagnostics
- Unbiased estimation
- Convergence diagnostics

4 Bias removal for importance sampling

Metropolis–Rosenbluth–Teller–Hastings

At each iteration t, Markov chain at state X_t,

1. propose $X^\star \sim q(X_t, \cdot)$,
2. sample $U \sim \text{Uniform}(0, 1)$,
3. if
$$U \leq \frac{\pi(X^\star)q(X^\star, X_t)}{\pi(X_t)q(X_t, X^\star)},$$
set $X_{t+1} = X^\star$, otherwise set $X_{t+1} = X_t$.

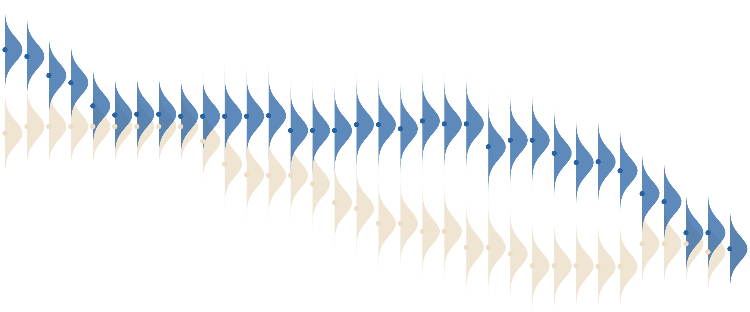
How to propagate two chains from states X_t and Y_t such that the event $\{X_{t+1} = Y_{t+1}\}$ can happen?

Metropolis–Rosenbluth–Teller–Hastings

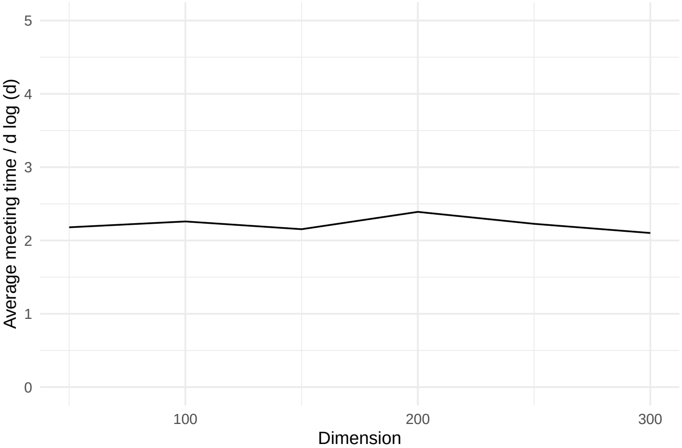
At each iteration t, two Markov chains at states X_t, Y_t,

1. propose $(X^\star, Y^\star)$ from max coupling of $q(X_t, \cdot)$, $q(Y_t, \cdot)$,
2. sample $U \sim \text{Uniform}(0, 1)$,
3. if
$$U \leq \frac{\pi(X^\star) q(X^\star, X_t)}{\pi(X_t) q(X_t, X^\star)},$$
set $X_{t+1} = X^\star$, otherwise set $X_{t+1} = X_t$,
if
$$U \leq \frac{\pi(Y^\star) q(Y^\star, Y_t)}{\pi(Y_t) q(Y_t, Y^\star)},$$
set $Y_{t+1} = Y^\star$, otherwise set $Y_{t+1} = Y_t$.

Proposal distributions of two MRTH chains



Meeting times of random walk MRTH



In line with Andrieu, Lee, Power & Wang (2024) on convergence of random walk MRTH. See also Papp & Sherlock, 2025.

Outline

1 Monte Carlo estimators and bias

2 Couplings of MCMC algorithms
- Coupling of two random variables
- Coupling of two Markov chains

3 Unbiased estimators and convergence diagnostics
- Unbiased estimation
- Convergence diagnostics

4 Bias removal for importance sampling

Outline

1 Monte Carlo estimators and bias

2 Couplings of MCMC algorithms
- Coupling of two random variables
- Coupling of two Markov chains

3 Unbiased estimators and convergence diagnostics
- Unbiased estimation
- Convergence diagnostics

4 Bias removal for importance sampling

Unbiased estimators

Breakthrough: Glynn & Rhee, 2014.

Idea: ergodic averages converge slowly (e.g. $t^{-1/2}$),
but marginal distributions tend to converge quickly (often ρ^t).

Application to MCMC:
Agapiou, Roberts & Vollmer, 2018.
Jacob, O'Leary & Atchadé, 2020,
Douc, Jacob, Lee & Vats (2026).

Coupled chains with a lag

Sample (X_0, Y_0) from $\bar{\pi}_0$,

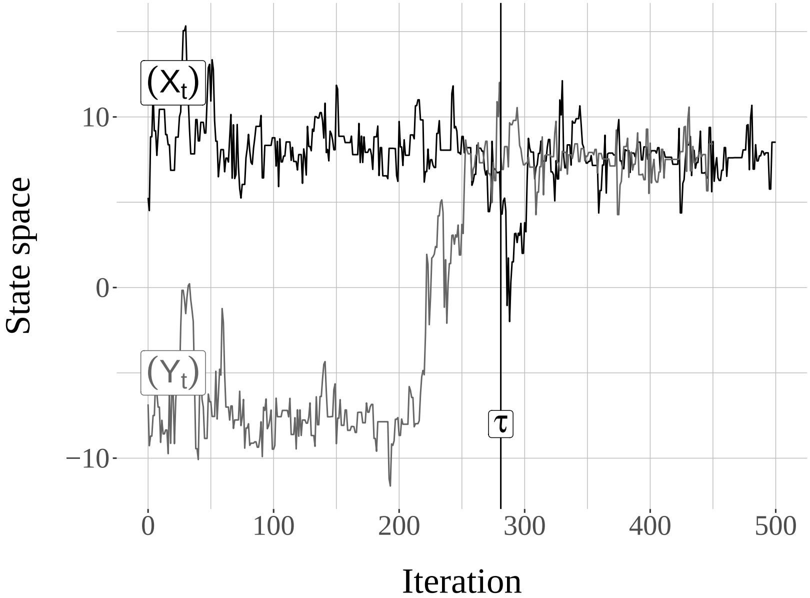
sample $X_t | X_{t-1}$ with transition P, for $t = 1, \ldots, L$,

for $t > L$, sample $(X_t, Y_{t-L}) | (X_{t-1}, Y_{t-L-1})$ with transition $\bar{P}$.

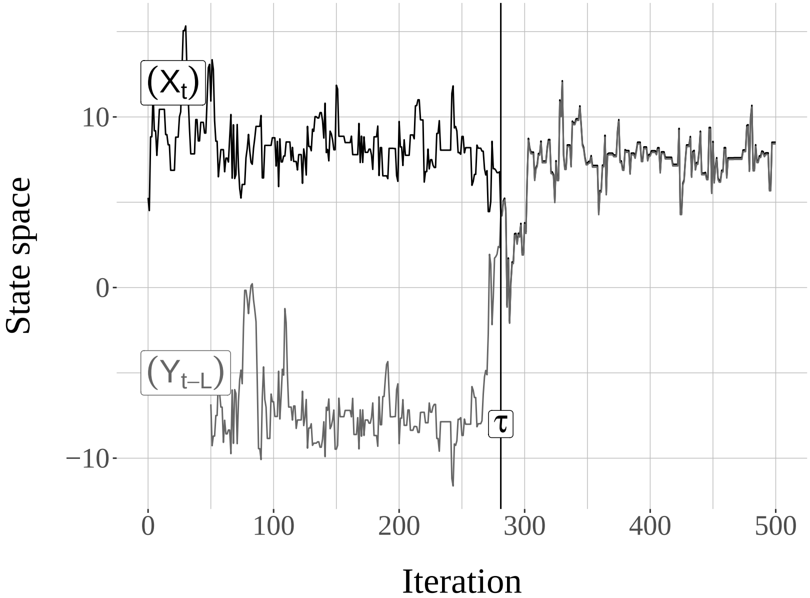
$\bar{\pi}_0$ and $\bar{P}$ must be such that

- X_t and Y_t have the same marginal distribution for all $t \geq 0$,
- $\exists$ finite meeting time τ such that $X_t = Y_{t-L}$ for $t \geq \tau$.

Coupled chains with a lag



Coupled chains with a lag



Unbiased estimators via telescopic sums

Limit as a telescopic sum, for all $k \geq 0$,

$$\begin{aligned}\pi(h) &= \lim_{t\to\infty} \pi_t(h)\\ &= \pi_k(h) + \sum_{j=1}^{\infty} \pi_{k+jL}(h) - \pi_{k+(j-1)L}(h).\end{aligned}$$

Since for all $t \geq 0$, X_t and Y_t have the same distribution π_t,

$$= \mathbb{E}[h(X_k)] + \sum_{j=1}^{\infty} \mathbb{E}[h(X_{k+jL}) - h(Y_{k+(j-1)L})].$$

If we can swap expectation and limit,

$$= \mathbb{E}[h(X_k) + \sum_{j=1}^{\infty}(h(X_{k+jL}) - h(Y_{k+(j-1)L}))].$$

Unbiased estimators via telescopic sums

Limit as a telescopic sum, for all $k \geq 0$,

$$\begin{aligned}\pi(h) &= \lim_{t\to\infty} \pi_t(h) \\ &= \pi_k(h) + \sum_{j=1}^{\infty} \pi_{k+jL}(h) - \pi_{k+(j-1)L}(h).\end{aligned}$$

Since for all $t \geq 0$, X_t and Y_t have the same distribution π_t,

$$= \mathbb{E}[h(X_k)] + \sum_{j=1}^{\infty} \mathbb{E}[h(X_{k+jL}) - h(Y_{k+(j-1)L})].$$

If we can swap expectation and limit,

$$= \mathbb{E}[h(X_k) + \sum_{j=1}^{\infty} (h(X_{k+jL}) - h(Y_{k+(j-1)L}))].$$

Unbiased estimators via telescopic sums

Limit as a telescopic sum, for all $k \geq 0$,

$$\begin{aligned}\pi(h) &= \lim_{t\to\infty} \pi_t(h) \\ &= \pi_k(h) + \sum_{j=1}^{\infty} \pi_{k+jL}(h) - \pi_{k+(j-1)L}(h).\end{aligned}$$

Since for all $t \geq 0$, X_t and Y_t have the same distribution π_t,

$$= \mathbb{E}[h(X_k)] + \sum_{j=1}^{\infty} \mathbb{E}[h(X_{k+jL}) - h(Y_{k+(j-1)L})].$$

If we can swap expectation and limit,

$$= \mathbb{E}[h(X_k) + \sum_{j=1}^{\infty}(h(X_{k+jL}) - h(Y_{k+(j-1)L}))].$$

A first unbiased signed measure

A signed measure constructed as

$$\hat{\pi} = \delta_{X_k} + \sum_{j=1}^{\infty} (\delta_{X_{k+jL}} - \delta_{Y_{k+(j-1)L}})$$

is unbiased, in the sense that $\mathbb{E}[\hat{\pi}(h)] = \pi(h)$.

You can generate $\hat{\pi}$ R times independently, and average.

- Parallelizable.
- Monte Carlo rate.
- Confidence intervals are easily obtained.
- But not efficient relative to long-run MCMC.

Jacob, O'Leary & Atchadé, 2020.

A first unbiased signed measure

A signed measure constructed as

$$\hat{\pi} = \delta_{X_k} + \sum_{j=1}^{\infty} (\delta_{X_{k+jL}} - \delta_{Y_{k+(j-1)L}})$$

is unbiased, in the sense that $\mathbb{E}[\hat{\pi}(h)] = \pi(h)$.

You can generate $\hat{\pi}$ R times independently, and average.

- Parallelizable.
- Monte Carlo rate.
- Confidence intervals are easily obtained.
- But not efficient relative to long-run MCMC.

Jacob, O'Leary & Atchadé, 2020.

Unbiased MCMC: a more efficient approximation

Jacob, O'Leary & Atchadé, 2020. Vanetti & Doucet, 2020.
Douc, Jacob, Lee & Vats, 2026.
Run the coupled chains with lag L, for $\max(\tau, \ell)$ iterations, where $\ell > k$. Unbiased signed measure:

$$\hat{\pi} = \frac{1}{\ell - k + 1} \sum_{t=k}^{\ell} \delta_{X_t} + \sum_{t=k+L}^{\tau-1} \frac{v_t}{\ell - k + 1} \left(\delta_{X_t} - \delta_{Y_{t-L}} \right)$$

with

$$v_t = \lfloor (t-k)/L \rfloor - \lceil \max(L, t-\ell)/L \rceil + 1.$$

Asymptotic variance of unbiased MCMC

Assumptions:

- For some $\kappa > 1$, $\mathbb{E}_{\pi\otimes\pi}[\tau^\kappa] < \infty$,
- $h \in L^m(\pi)$ for some $m > 2\kappa/(\kappa - 1)$.

Then for any $k, L \in \mathbb{N}$,

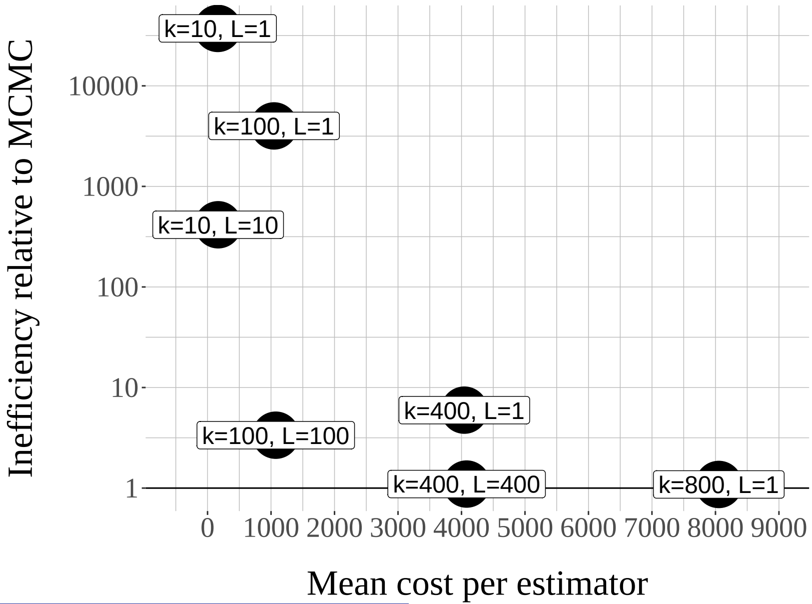
$$\sqrt{\ell - k + 1}\,(\hat{\pi}(h) - \pi(h)) \xrightarrow[\ell\to\infty]{d} \text{Normal}(0, v(P, h)),$$

where $v(P, h)$ is the same asymptotic variance as in

$$\sqrt{t}(t^{-1}\sum_{s=0}^{t-1} h(X_s) - \pi(h)) \xrightarrow[t\to\infty]{d} \text{Normal}(0, v(P, h)).$$

Douc, Jacob, Lee & Vats, 2026.

Inefficiency of unbiased MCMC



Outline

1. Monte Carlo estimators and bias
2. Couplings of MCMC algorithms
   - Coupling of two random variables
   - Coupling of two Markov chains
3. Unbiased estimators and convergence diagnostics
   - Unbiased estimation
   - Convergence diagnostics
4. Bias removal for importance sampling

Bounds on distance to stationarity

With the same coupled lagged chains as before, we can write:

$$|\pi_t - \pi|_{\mathrm{TV}} \leq \sum_{j=1}^{\infty} |\pi_{t+jL} - \pi_{t+(j-1)L}|_{\mathrm{TV}}$$

$$\leq \sum_{j=1}^{\infty} \mathbb{P}(X_{t+jL} \neq Y_{t+(j-1)L})$$

$$= \sum_{j=1}^{\infty} \mathbb{E}[\mathbb{1}(\tau > t + jL)].$$

Swapping infinite sum and expectation,

$$|\pi_t - \pi|_{\mathrm{TV}} \leq \mathbb{E}\left[\sum_{j=1}^{\infty} \mathbb{1}(\tau > t + jL)\right] = \mathbb{E}[\max(0, \lceil(\tau - L - t)/L\rceil)].$$

Biswas, Jacob & Vanetti, 2019.

Bounds on distance to stationarity

With the same coupled lagged chains as before, we can write:

$$
\begin{aligned}
|\pi_t - \pi|_{\mathrm{TV}} &\leq \sum_{j=1}^{\infty} |\pi_{t+jL} - \pi_{t+(j-1)L}|_{\mathrm{TV}} \\
&\leq \sum_{j=1}^{\infty} \mathbb{P}(X_{t+jL} \neq Y_{t+(j-1)L}) \\
&= \sum_{j=1}^{\infty} \mathbb{E}[\mathbb{1}(\tau > t + jL)].
\end{aligned}
$$

Swapping infinite sum and expectation,

$$
|\pi_t - \pi|_{\mathrm{TV}} \leq \mathbb{E}\left[\sum_{j=1}^{\infty} \mathbb{1}(\tau > t + jL)\right] = \mathbb{E}[\max(0, \lceil(\tau - L - t)/L\rceil)].
$$

Biswas, Jacob & Vanetti, 2019.

Bounds on distance to stationarity

With the same coupled lagged chains as before, we can write:

$$
\begin{aligned}
|\pi_t - \pi|_{\mathrm{TV}} &\leq \sum_{j=1}^{\infty} |\pi_{t+jL} - \pi_{t+(j-1)L}|_{\mathrm{TV}} \\
&\leq \sum_{j=1}^{\infty} \mathbb{P}(X_{t+jL} \neq Y_{t+(j-1)L}) \\
&= \sum_{j=1}^{\infty} \mathbb{E}[\mathbb{1}(\tau > t + jL)].
\end{aligned}
$$

Swapping infinite sum and expectation,

$$
|\pi_t - \pi|_{\mathrm{TV}} \leq \mathbb{E}\left[\sum_{j=1}^{\infty} \mathbb{1}(\tau > t + jL)\right] = \mathbb{E}[\max(0, \lceil(\tau - L - t)/L\rceil)].
$$

Biswas, Jacob & Vanetti, 2019.

Bounds on distance to stationarity

With the same coupled lagged chains as before, we can write:

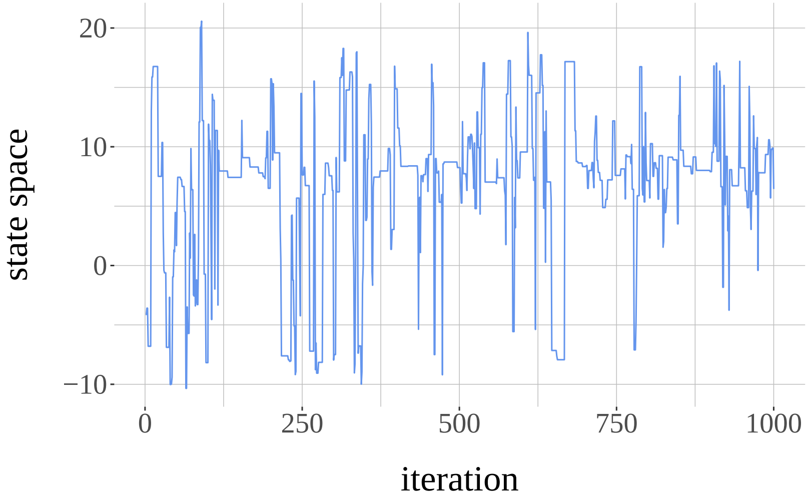
$$\begin{aligned} |\pi_t - \pi|_{\mathrm{TV}} &\leq \sum_{j=1}^{\infty} |\pi_{t+jL} - \pi_{t+(j-1)L}|_{\mathrm{TV}} \\ &\leq \sum_{j=1}^{\infty} \mathbb{P}(X_{t+jL} \neq Y_{t+(j-1)L}) \\ &= \sum_{j=1}^{\infty} \mathbb{E}[\mathbb{1}(\tau > t + jL)]. \end{aligned}$$

Swapping infinite sum and expectation,

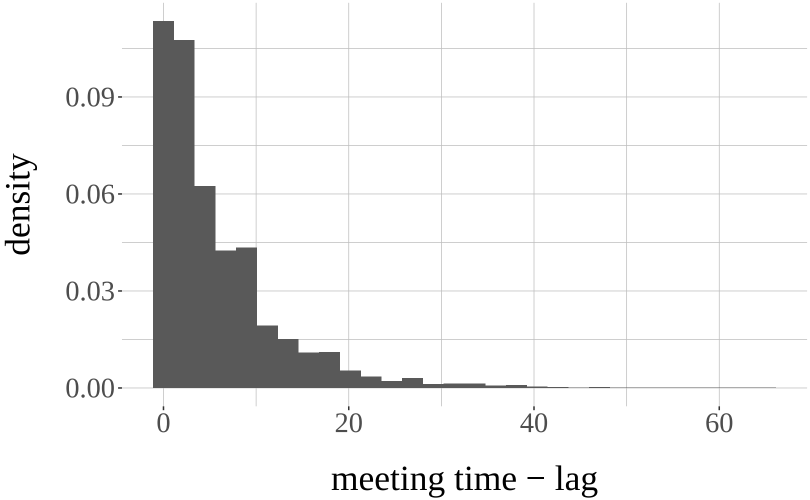
$$|\pi_t - \pi|_{\mathrm{TV}} \leq \mathbb{E}\left[\sum_{j=1}^{\infty} \mathbb{1}(\tau > t + jL)\right] = \mathbb{E}[\max(0, \lceil (\tau - L - t)/L \rceil)].$$

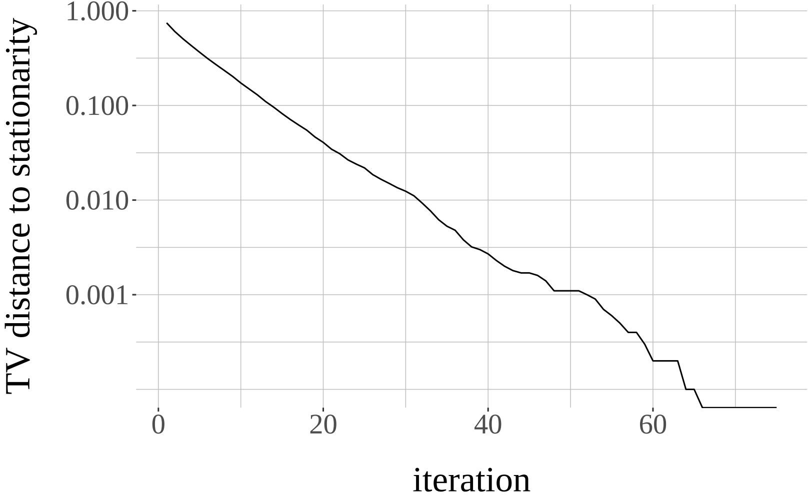
Biswas, Jacob & Vanetti, 2019.

Example



Example





Application to phylogenetics in Ryder, Kelly & Clarté, 2023.

Outline

1 Monte Carlo estimators and bias

2 Couplings of MCMC algorithms
- Coupling of two random variables
- Coupling of two Markov chains

3 Unbiased estimators and convergence diagnostics
- Unbiased estimation
- Convergence diagnostics

4 Bias removal for importance sampling

Self-normalised importance sampling

1. Sample $X_1, \ldots, X_N \overset{\text{ind}}{\sim} q$.
2. Compute $\omega(X_n) = \pi(X_n)/q(X_n)$ for all n.
3. Compute SNIS estimator

$$\hat{H} = \frac{\sum_{n=1}^{N} \omega(X_n) h(X_n)}{\sum_{n=1}^{N} \omega(X_n)}.$$

Bias of order N^{-1}.

Particle independent MRTH

Andrieu, Doucet & Holenstein 2010: define

$$\bar{q}(x_1, \ldots, x_N) = \prod_{n=1}^{N} q(x_n),$$

$$\bar{\pi}(x_1, \ldots, x_N) = \left(\frac{1}{N}\sum_{k=1}^{N}\omega(x_k)\right)\prod_{n=1}^{N} q(x_n) = \hat{Z}_N(\mathbf{X})\bar{q}(\mathbf{X}),$$

and implement MRTH with proposal $\bar{q}$ and target $\bar{\pi}$.

1. Draw $\mathbf{X}^\star \sim \bar{q}$.
2. Compute the acceptance probability:

$$\alpha(\mathbf{X}, \mathbf{X}^\star) = \min\left\{1, \frac{\hat{Z}_N(\mathbf{X}^\star)}{\hat{Z}_N(\mathbf{X})}\right\}.$$

3. Accept $\mathbf{X}^\star$ or not with probability $\alpha(\mathbf{X}, \mathbf{X}^\star)$.

Particle independent MRTH

Andrieu, Doucet & Holenstein 2010: define

$$\bar{q}(x_1, \ldots, x_N) = \prod_{n=1}^{N} q(x_n),$$

$$\bar{\pi}(x_1, \ldots, x_N) = \left(\frac{1}{N}\sum_{k=1}^{N}\omega(x_k)\right)\prod_{n=1}^{N} q(x_n) = \hat{Z}_N(\mathbf{X})\bar{q}(\mathbf{X}),$$

and implement MRTH with proposal $\bar{q}$ and target $\bar{\pi}$.

1. Draw $\mathbf{X}^\star \sim \bar{q}$.
2. Compute the acceptance probability:

$$\alpha(\mathbf{X}, \mathbf{X}^\star) = \min\left\{1, \frac{\hat{Z}_N(\mathbf{X}^\star)}{\hat{Z}_N(\mathbf{X})}\right\}.$$

3. Accept $\mathbf{X}^\star$ or not with probability $\alpha(\mathbf{X}, \mathbf{X}^\star)$.

Common draws coupling of PIMRTH

Liu, 1996. Roberts & Rosenthal, 2011.

Two chains in states $\mathbf{X}$ and $\mathbf{Y}$.

1. Draw U from a Uniform$(0, 1)$ distribution.
2. If $U < \hat{Z}_N(\mathbf{X}^\star)/\hat{Z}_N(\mathbf{X})$, set $\mathbf{X}' = \mathbf{X}^\star$, otherwise set $\mathbf{X}' = \mathbf{X}$.
3. If $U < \hat{Z}_N(\mathbf{X}^\star)/\hat{Z}_N(\mathbf{Y})$, set $\mathbf{Y}' = \mathbf{X}^\star$, otherwise set $\mathbf{Y}' = \mathbf{Y}$.
4. Return $(\mathbf{X}', \mathbf{Y}')$

We show that this coupling is maximal, i.e. leads to the smallest meeting times possible.

Unbiased MCMC with that coupling + a few tricks
$\Rightarrow$ Coupling-based Unbiased Importance Sampling.

Deligiannidis, Jacob, Khribch & Wang, 2026.

Common draws coupling of PIMRTH

Liu, 1996. Roberts & Rosenthal, 2011.

Two chains in states $\mathbf{X}$ and $\mathbf{Y}$.

1. Draw U from a Uniform$(0, 1)$ distribution.
2. If $U < \hat{Z}_N(\mathbf{X}^\star)/\hat{Z}_N(\mathbf{X})$, set $\mathbf{X}' = \mathbf{X}^\star$, otherwise set $\mathbf{X}' = \mathbf{X}$.
3. If $U < \hat{Z}_N(\mathbf{X}^\star)/\hat{Z}_N(\mathbf{Y})$, set $\mathbf{Y}' = \mathbf{X}^\star$, otherwise set $\mathbf{Y}' = \mathbf{Y}$.
4. Return $(\mathbf{X}', \mathbf{Y}')$

We show that this coupling is maximal, i.e. leads to the smallest meeting times possible.

Unbiased MCMC with that coupling + a few tricks
$\Rightarrow$ Coupling-based Unbiased Importance Sampling.

Deligiannidis, Jacob, Khribch & Wang, 2026.

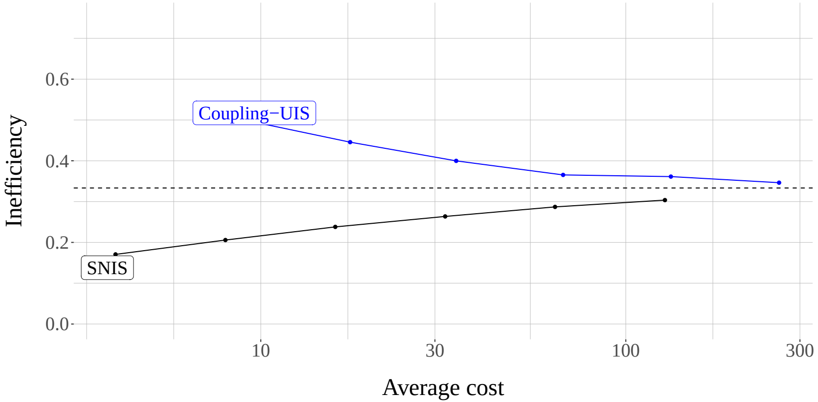
Coupling-UIS

Algorithm 6 Coupling-based unbiased importance sampling (Coupling-UIS).

1. Sample $\mathbf{X}^N = (X_1, \ldots, X_N) \sim \bar{q}$ and $\mathbf{Y}^N = (Y_1, \ldots, Y_N) \sim \bar{q}$ independently.
2. If $\hat{Z}_N(\mathbf{X}^N) < \hat{Z}_N(\mathbf{Y}^N)$, swap $\mathbf{X}^N$ and $\mathbf{Y}^N$, so that $\hat{Z}_N(\mathbf{X}^N) \geq \hat{Z}_N(\mathbf{Y}^N)$ henceforth.
3. Set $\tau \leftarrow 1$. Initialize
$$\tilde{F}_N^u \leftarrow \frac{1}{2}\hat{F}_N(\mathbf{X}^N) + \frac{1}{2}\hat{F}_N(\mathbf{Y}^N).$$
4. Draw $U \sim \text{Uniform}(0,1)$. Update estimate, with $\hat{F}_N^{\text{RB}}$ as in (44):
$$\tilde{F}_N^u \leftarrow \tilde{F}_N^u + \frac{1}{2}\left(\hat{F}_N^{\text{RB}}(\mathbf{X}^N, \mathbf{Y}^N) - \hat{F}_N(\mathbf{Y}^N)\right).$$
If $U \leq \hat{Z}_N(\mathbf{Y}^N)/\hat{Z}_N(\mathbf{X}^N)$, go to step 5.
5. Iterate until both conditions in (e) and (f) are satisfied:
   a) $\tau \leftarrow \tau + 1$.
   b) Sample $\mathbf{X}^{N,\star} = (X_1^\star, \ldots, X_N^\star) \sim \bar{q}$.
   c) $\tilde{F}_N^u \leftarrow \tilde{F}_N^u + \frac{1}{2}(\hat{F}_N^{\text{RB}}(\mathbf{X}^N, \mathbf{X}^{N,\star}) - \hat{F}_N^{\text{RB}}(\mathbf{Y}^N, \mathbf{X}^{N,\star}))$.
   d) Draw $U \sim \text{Uniform}(0,1)$.
   e) If $U \leq \hat{Z}_N(\mathbf{X}^{N,\star})/\hat{Z}_N(\mathbf{X}^N)$, set $\mathbf{X}^N \leftarrow \mathbf{X}^{N,\star}$.
   f) If $U \leq \hat{Z}_N(\mathbf{X}^{N,\star})/\hat{Z}_N(\mathbf{Y}^N)$, set $\mathbf{Y}^N \leftarrow \mathbf{X}^{N,\star}$.
6. Set cost $\leftarrow N(\tau + 1)$.
7. Return $\tilde{F}_N^u$.

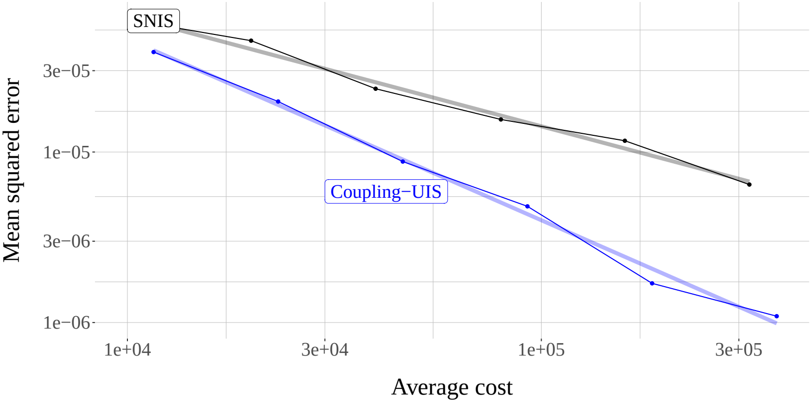
SNIS versus Coupling-UIS

Coupling-UIS: proposed bias-removed SNIS.
Inefficiency: Mean squared error $\times$ expected cost.



$\Rightarrow$ Coupling-UIS is less efficient for fixed N, but not as $N \to \infty$.

SNIS versus Coupling-UIS within nested expectation



$\Rightarrow$ Unbiased estimators lead to faster convergence rates.

References

Atchadé & Jacob,
Unbiased Markov Chain Monte Carlo: what, why, and how,
in the 2nd edition of the *Handbook of Markov Chain Monte Carlo*
https://arxiv.org/abs/2406.06851

Deligiannidis, Jacob, Khribch & Wang, *On importance sampling and independent Metropolis-Hastings with an unbounded weight function*
https://arxiv.org/abs/2411.09514

Andradóttir & Glynn, *Computing Bayesian Means using Simulation*, 2002/2016.

Thank you for your attention!