Unbiased MCMC

- Standard MCMC estimators are biased for any fixed number of iterations.
- We do not know in practice how many iterations are needed to reduce bias to an acceptable level.
- Burn-in period is often used to reduce bias, but it is difficult to choose appropriately.
- Unbiased MCMC estimators can be constructed using coupling techniques.
- Glynn–Rhee estimator provides a way to obtain unbiased estimates of expectations with respect to the target distribution.
- Key assumptions include moment conditions, geometric tail bounds on meeting times, and the property that chains stay together after meeting.
- Since the estimator is unbiased, we can generate shorter chains (in parallel) and average them to reduce variance without introducing bias.
- The cost is more complexity in implementation and the need to design effective coupling strategies.

Unbiased MCMC

- MCMC algorithms provide estimators that are consistent as the number of iterations grows large but potentially biased for any fixed number of iterations.
- **Unbiased estimation using coupling** (Glynn and Rhee, 2014, Jacob et al., 2020 & 2024)

$$X_0 \sim \pi_0, \quad Y_0 \sim \pi_0, \quad X_1 \sim K(X_0, \cdot)$$
$$(X_{t+1}, Y_t) \sim \bar{K}((X_t, Y_{t-1}), (\cdot, \cdot)), \quad \forall t \geq 1$$

- **Assumption 1**: $E_\pi[h(X)] = \lim_{t \to \infty} E[h(X_t)]$ and there is $\eta > 0$ and $D < \infty$ that $E[|h(X_t)|^{2+\eta}] \leq D$ for all t.
- **Assumption 2**: meeting/coupling time

  $$\tau = \inf\{t \geq 1 : X_t = Y_{t-1}\} \text{ satisfies } \Pr(\tau > t) \leq C\delta^t \text{ for all}$$

  t, for some constant $C < \infty$ and $\delta < 1$.
- **Assumption 3**: the chains stay together after meeting: $X_t = Y_{t-1}$ for all $t \geq \tau$.

Glynn–Rhee estimator - assumptions

- **Assumption 1**:
- **Assumption 2**:
- **Assumption 3**:

Glynn–Rhee debiasing formula

Theorem

Under assumptions A1-A3, for any fixed $k \geq 0$

$$E_\pi[h(X)] = E[h(X_k) + \sum_{t=k+1}^{\tau-1} [h(X_t) - h(Y_{t-1})]]$$

$$E_\pi[h(X)] \underset{\text{by A1}}{=} \lim_{t\to\infty} E[h(X_t)] = E[h(X_k)] + \sum_{t=k+1}^{\infty} \{E[h(X_t)] - E[h(X_{t-1})]\}$$

$$\underset{\text{by A1 and A2}}{=} E\left[h(X_k) + \sum_{t=k+1}^{\infty} [h(X_t) - h({\color{red}Y_{t-1}})]\right] \underset{\text{by A3}}{=} E\left[h(X_k) + \sum_{t=k+1}^{{\color{red}\tau-1}} [h(X_t) - h(Y_{t-1})]\right]$$

Example of Coupling: Maximal Coupling

Algorithm 1 Sampling a coupling of p and q, with parameter $\eta \in (0, 1]$. The coupling maximizes $\mathbb{P}(X = Y)$ when $\eta = 1$, but the variance of the cost is bounded when $\eta < 1$.

```
Sample X ~ p.
Sample W ~ Uniform(0, 1).
if W ≤ min(η, q(X)/p(X)) then
    set Y = X.
else
    while true do
        sample Y* ~ q and W* ~ Uniform(0, 1) until W* > ηp(Y*)/q(Y*)
    end while
end if
Return (X, Y).
```

Maximal Coupling of two Gaussians

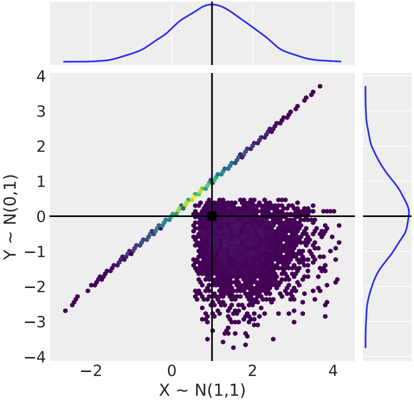

Figure: Maximal coupling of $\mathcal{N}(0,1)$ and $\mathcal{N}(1,1)$. Geometric interpretation: Maximizes mass on the diagonal.

Coupling

Couplings of MCMC algorithms can be devised using maximal couplings, reflection couplings, and common random numbers. We have focused on couplings that can be implemented without further analytical knowledge about the target distribution or about the MCMC kernels. However, these couplings might result in prohibitively large meeting times, either because the marginal chains mix slowly or because the coupling strategy is ineffective

Coupling

Algorithm 2 Successful coupling of chains with lag L and length ℓ. Coupled initial distribution: $\tilde{\pi}_0$, transition: P, coupled transition: $\tilde{P}$, meeting time $\tau = \inf\{t \geq L : X_t = Y_{t-L}\}$.

Sample (X_0, Y_0) from $\tilde{\pi}_0$.
if $L \geq 1$ **then**
    for $t = 1, \ldots, L$ **do**
        sample X_t from $P(X_{t-1}, \cdot)$
    end for
end if
for $t \geq L$ **do**
    while true **do**
        sample (X_{t+1}, Y_{t-L+1}) from $\tilde{P}((X_t, Y_{t-L}), \cdot)$ until $X_{t+1} = Y_{t-L+1}$ and $t+1 \geq \ell$
    end while
end for