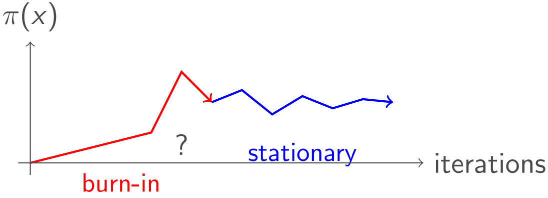
Unbiased MCMC

The Problem with Traditional MCMC

Traditional MCMC Challenges:

- Burn-in bias: Initial samples don't follow target distribution
- Convergence diagnostics: When to stop burn-in?
- Multiple chains: same issues as above and the bias prevents the consistent estimation of by averages over the independent runs.



Consequence:
Must discard unknown number of initial samples, limiting parallel efficiency

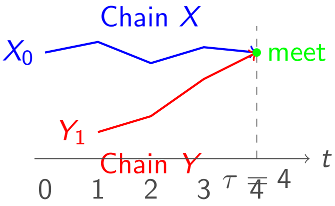
The Unbiased MCMC Solution

Key Innovation: Coupling + Debiasing

Run two coupled Markov chains that eventually meet, then use their difference to construct an unbiased estimator

The Algorithm:

1. Start chains X_1 and Y_0 from initial distribution
2. Run coupled chains until meeting time $\tau = \inf\{t \geq 1 : X_t = Y_{t-1}\}$
3. Compute debiased estimator:



Glynn–Rhee Estimator

Given a target function h, two initial values $X_0 \sim \pi_0$, $Y_0 \sim \pi_0$, and $X_1 \sim K(X_0, \cdot)$ and set the coupled kernel as

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

with marginals equal to K and meeting time τ as defined above.
Under the assumptions:

1. $E_\pi[h(X)] = \lim_{t\to\infty} E[h(X_t)]$ for $t \to \infty$
2. meeting time decrease at geometric rate, i.e. $P(\tau > t) \leq C\delta^t$
3. and that the chains stay together after they met, i.e. $X_t = Y_{t-1} \quad \forall t \geq \tau$

Then for any fixed $k \geq 0$ the **Glynn–Rhee estimator**:

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

has expectation $E_\pi[h(X)]$.

Glynn–Rhee proof sketch

- **Assumption 2**: This is the hardest to verify in practice. It requires the coupling to be efficient enough so that the meeting time has geometric tails
- **Assumption 3**: This is often easy to ensure in practice by designing the coupling appropriately

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

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

we can replace X_{t-1} by Y_{t-1} since they have the marginal distribution and swap expectation and summation due to Fubini's theorem and Assumption 2 that ensure fast Convergence of the sum. **ESE=ES**

Maximal Coupling of two Gaussians

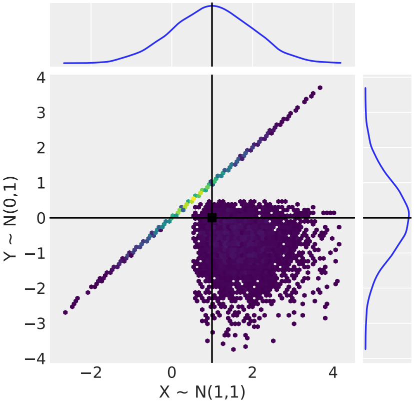


Figure: Maximal coupling of $\mathcal{N}(0,1)$ and $\mathcal{N}(1,1)$. Geometric interpretation: Maximizes mass on the diagonal.

Takeaways

- No burn-in period required
- Since the estimator is unbiased, we can generate shorter chains (in parallel) and average them to reduce variance without introducing bias.
- However, the variance can be prohibitively high. Can be reduced by introducing lag L and average I iterations.
- The cost is more complexity in both execution and implementation and the need to design effective coupling strategies.