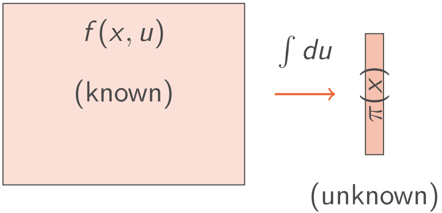
The Challenge: Intractable Marginals

The Problem:

- Target: $\pi(x) = \int f(x, u) du$
- $f(x, u)$ is known (complete data)
- Integral is intractable
- Standard MCMC requires exact $\pi(x)$

Examples:

- Hidden Markov Models
- Mixed Effects Models
- Phylogenetics
- Topic Models



Key Insight: We can estimate $\pi(x)$ unbiasedly!

The Pseudo-marginal Solution

Key Prerequisites

For pseudo-marginal MCMC to be applicable, we need:

1. Ability to **evaluate** $f(x, u)$ pointwise for any (x, u)
2. Ability to **sample** from an importance distribution $q_x(\cdot)$ over the u-space
3. The importance distribution must have appropriate support: $q_x(u) > 0$ whenever $f(x, u) > 0$

Importance Sampling Estimator:

$$\hat{\pi}(x) = \frac{1}{N} \sum_{i=1}^{N} \frac{f(x, U_i)}{q_x(U_i)}, \quad U_i \sim q_x(\cdot)$$

Key Property: $\mathbb{E}[\hat{\pi}(x)] = \pi(x)$ (unbiased!)

The Magic: Replace π with $\hat{\pi}$ in MH ratio!

$$\alpha = \min\left\{1, \frac{\hat{\pi}(y)q(y, x)}{\hat{\pi}(x)q(x, y)}\right\}$$

Result: Still targets correct $\pi(x)$!

Why It Works: Extended Target

One can think of estimator (the "pseudo-marginal") as the product of the true target and a random variable:

$$\hat{\pi}(x) = \pi(x) Z_x$$

where Z_x satifies:

1. is non-negative: $Z_x \geq 0$,
2. has density $g_x()$: $\int_0^\infty g_x(z)dz = 1$
3. has expectation 1: $\mathbb{E}[Z_x] = \int_0^\infty z g_x(z)dz = 1.$

Extended Target Construction:

$$\bar{\pi}(x, z) = \pi(x) \cdot z \cdot g_x(z)$$

where $g_x(z)$ is the density of Z_x

Key Property:

$$\int \bar{\pi}(x, z)dz = \pi(x)$$

Intuition:

- Run exact MCMC on (x, z) space
- Marginal in x gives correct target
- z represents the "noise" in estimates

Pseudo-marginal MCMC Algorithm

Given $(X^{(t-1)}, \hat{\pi}^{(t-1)})$:

1. **Propose:** $Y \sim q(X^{(t-1)}, \cdot)$
2. **Estimate:**
   - Sample $U_i \sim q_Y(\cdot)$
   - $\hat{\pi}(Y) = \frac{1}{N} \sum_i \frac{f(Y, U_i)}{q_Y(U_i)}$
3. **Accept with probability:**

$$\alpha = \min \left\{ 1, \frac{\hat{\pi}(Y) q(Y, X^{(t-1)})}{\hat{\pi}^{(t-1)} q(X^{(t-1)}, Y)} \right\}$$

4. **Update:**
   - If accept: $(X^{(t)}, \hat{\pi}^{(t)}) = (Y, \hat{\pi}(Y))$
   - Else: $(X^{(t)}, \hat{\pi}^{(t)}) = (X^{(t-1)}, \hat{\pi}^{(t-1)})$

Critical Points:

- Store estimates with states! In the next iteration, use the stored $\hat{\pi}(X^{(t-1)})$.
- Fresh randomness for each proposal. Every time you propose a new state Y, you must generate a completely new, independent estimate $\hat{\pi}(Y)$ using fresh random samples.
- Works with *any* MH proposal q

Equivalence to MH on Extended Space

Theorem (Equivalence)

Metropolis-Hastings on the extended target $\bar{\pi}$ with proposal $\bar{q}$ is equivalent to the pseudo-marginal algorithm using estimates $\hat{\pi}$.

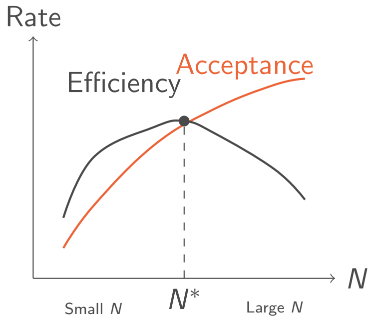
Proof Sketch: The MH acceptance ratio on the extended space is:

$$\begin{aligned}
\alpha_{ext} &= \min\left\{1, \frac{\bar{\pi}(y, w)\bar{q}((y, w), (x, z))}{\bar{\pi}(x, z)\bar{q}((x, z), (y, w))}\right\} \\
&= \min\left\{1, \frac{\pi(y) \cdot w \cdot g_y(w) \cdot q(y, x) \cdot g_x(z)}{\pi(x) \cdot z \cdot g_x(z) \cdot q(x, y) \cdot g_y(w)}\right\} \\
&= \min\left\{1, \frac{\pi(y) \cdot w \cdot q(y, x)}{\pi(x) \cdot z \cdot q(x, y)}\right\} = \min\left\{1, \frac{\hat{\pi}(y)q(y, x)}{\hat{\pi}(x)q(x, y)}\right\} = \alpha_{pm}
\end{aligned}$$

In the last step, we used $\hat{\pi}(x) = \pi(x)z$ and $\hat{\pi}(y) = \pi(y)w$, which is exactly the pseudo-marginal acceptance probability.

The Variance-Efficiency Trade-off

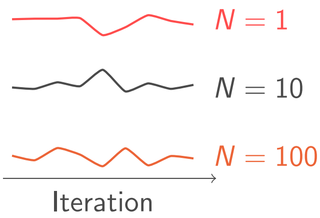
Choice of N (sample size):



Small N: High variance, poor mixing
Large N: Expensive per iteration

Optimal for Random Walk:

- $\text{Var}(Z_x) \approx 3.283$
- Acceptance rate $\approx 7\%$
- Much lower than standard MCMC (23%)!



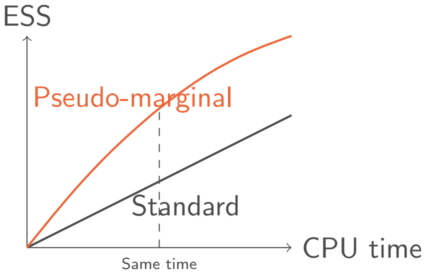
Comparison: Standard vs Pseudo-marginal MCMC

Standard MCMC on (x, u):

- × High dimensional
- × Slow mixing in u
- ✓ No tuning of N

Pseudo-marginal on x:

- ✓ Lower dimensional
- ✓ Integrates out u
- × Need to choose N
- × More complex implementation



Key Message: Pseudo-marginal can be more efficient despite noise!

Practical Guidelines

When to use:

- Marginal likelihood intractable
- Can evaluate $f(x, u)$ pointwise
- Good importance distribution available
- Dimension of x moderate

Implementation checklist:

- ☐ Store estimates with states
- ☐ Use fresh randomness
- ☐ Monitor acceptance rate
- ☐ Tune N for $\approx 7\%$ acceptance
- ☐ Use log-scale for stability

Common pitfalls:

- Recomputing old estimates
- Using N too large
- Poor importance distribution
- Numerical overflow/underflow

Rule of Thumb

Choose N such that:

$$\mathrm{CV}[\hat{\pi}(x)/\pi(x)] \approx 1.7$$

This gives $\mathrm{Var}(Z_x) \approx 3.3$

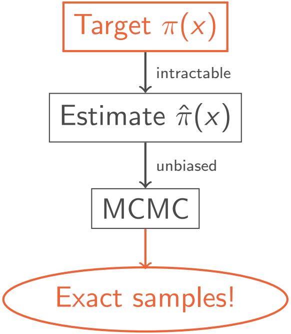
Summary

Pseudo-marginal MCMC:

- Enables exact inference with unbiased estimates
- Theoretically elegant (extended target)
- Practically powerful

Key papers:

- Beaumont (2003) - Introduction
- Andrieu & Roberts (2009) - Theory
- Sherlock et al. (2015) - Optimal scaling



Take-home message:
Noise + Unbiasedness = Exactness