Importance Sampling

What is Importance Sampling?

What?

- Monte Carlo technique for estimating $\mathbb{E}_\pi[\phi(X)]$
- Sample from proposal q instead of target π
- Reweight samples to correct bias

Why?

- Target π difficult to sample from
- Focus sampling in important regions
- Works with unnormalized distributions
- All samples are used (unlike rejection)

How? The key identity:

$$\begin{aligned}\mathbb{E}_\pi[\phi(X)] &= \int \phi(x)\pi(x)dx \\ &= \int \phi(x)\frac{\pi(x)}{q(x)}q(x)dx \\ &= \mathbb{E}_q[\phi(X)w(X)]\end{aligned}$$

Algorithm

1. Sample $X_1, \ldots, X_n \sim q$
2. Compute $w(X_i) = \pi(X_i)/q(X_i)$
3. Estimate: $\hat{I} = \frac{1}{n}\sum_{i=1}^{n} \phi(X_i)w(X_i)$

Key Properties and Unnormalized Distributions

Properties of IS Estimator:

- **Unbiased**: $\mathbb{E}_q[\hat{I}] = \mathbb{E}_\pi[\phi(X)]$
- **Consistent**: $\hat{I} \xrightarrow{n\to\infty} \mathbb{E}_\pi[\phi(X)]$ (LLN)
- **Variance**: $\text{Var}_q[\hat{I}] = \frac{1}{n}\text{Var}_q[\phi(X)w(X)]$

Requirements:

- $q(x) > 0$ whenever $\pi(x)\phi(x) \neq 0$
- $\mathbb{E}_q[|\phi(X)w(X)|] < \infty$

Unnormalized Distributions:

When $\pi(x) = \tilde{\pi}(x)/Z$ with unknown Z:

Self-Normalized IS

- Weights: $\tilde{w}(x) = \tilde{\pi}(x)/q(x)$
- Estimator:

$$\hat{I}_{SN} = \frac{\sum_{i=1}^n \phi(X_i)\tilde{w}(X_i)}{\sum_{i=1}^n \tilde{w}(X_i)}$$

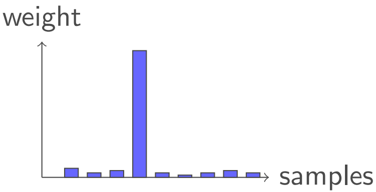
- Biased but consistent
- Bias: $O(1/n)$

Why Weight Distribution Matters

Weight Distribution Impact:

- High weight variance $\Rightarrow$ poor estimates
- Few samples dominate the sum
- Ideal case: all weights equal ($q = \pi$)
- $\mathrm{Var}_q[w(X)]$ determines convergence

Example Weight Degeneracy:



One weight dominates

Effective Sample Size (ESS)

Definition

$$\text{ESS} = \frac{\left(\sum_{i=1}^{n} w_i\right)^2}{\sum_{i=1}^{n} w_i^2}$$

Interpretation:

- Number of "equivalent" samples from π
- Range: $1 \leq \text{ESS} \leq n$
- $\text{ESS} = n$ when all weights equal, $\text{ESS} = 1$ when one weight dominates

Why ESS Matters:

- $\text{ESS} \ll n$ indicates weight degeneracy
- Low $\text{ESS} \Rightarrow$ high variance
- Monitor ESS to diagnose problems
- Rule of thumb: $\text{ESS} > n/2$ is good

Choosing Good Proposals & Dimensional Scaling

Good Proposal Properties:

1. Heavier tails than π
2. Easy to sample from
3. Similar shape to $\pi|\phi|$
4. Covers support of π
5. Minimizes $\mathrm{Var}_q[\phi(X)w(X)]$

Common Choices:

- Student-t for Gaussian targets
- Mixture distributions
- Previous MCMC output
- Laplace approximation

Curse of Dimensionality:

Gaussian Example

For $\pi = \mathcal{N}(0, I_d)$, $q = \mathcal{N}(0, \sigma^2 I_d)$:

$$\mathrm{Var}_q[w(X)] = \left(\frac{\sigma^4}{2\sigma^2 - 1}\right)^{d/2} - 1$$

Numerical Example:

d	σ	$\mathrm{Var}_q[w(X)]$
10	1.2	≈ 5.6
50	1.2	≈ 850
100	1.2	$\approx \mathbf{1.8 \times 10^4}$

Importance Sampling vs. Rejection Sampling

Aspect	Importance Sampling	Rejection Sampling
Sample usage	All samples (weighted)	Some samples rejected
Efficiency	Depends on weight variance	Depends on acceptance rate
High dimensions	Poor (variance explodes)	Very poor (accept rate $\to 0$)
Proposal req.	$q > 0$ where $\pi\phi \neq 0$	Need $Mq \geq \pi$ everywhere
Output	Weighted samples	Exact samples from π
Normalizing const.	Not required	Required (for bound M)
Bias	Unbiased (or consistent)	Unbiased (exact)
Failure mode	High variance	No samples produced

Key Insight

Both methods suffer from curse of dimensionality, but:

- IS degrades gracefully - still provides estimates (with high variance)