Importance Sampling

What, Why, and How

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

How? The key identity:

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

Algorithm

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

Requirements, Properties and Unnormalized case

Requirements:

- $q(x) > 0$ whenever $\pi(x) > 0$

Properties of IS Estimator:

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

Unnormalized Distributions:

When $\pi(x) = \tilde{\pi}(x)/Z_\pi$ and $q(x) = \tilde{q}(x)/Z_q$:

Self-Normalized IS

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

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

- Biased but consistent

$$\mathbb{E}_q[\hat{I}_n] = \mathbb{E}_q\left[\frac{1}{N}\sum_{i=1}^{N}\phi(x_i)w(x_i)\right] = \frac{1}{N}N\mathbb{E}_q[\phi(X)w(X)] = \mathbb{E}_q[\phi(X)w(X)] = \int \phi(x)w(x)q(x)\,dx = \int \phi(x)\frac{p(x)}{q(x)}q(x)\,dx = \int \phi(x)\,p(x)\,dx = \mathbb{E}_p[\phi(X)]$$

Importance sampling diagnostics

In extreme settings, one of the w_i may be vastly larger than all the others and then we have effectively only got one observation.

- High Variance from Large Weights. If $\pi(x) \gg q(x)$ in some regions, you get huge weights $w(x) = \frac{\pi(x)}{q(x)}$. Even rare samples from these regions cause massive variance.

On the other end of the spectrum all the weights could be very small if q places too much mass in regions where π is negligible, i.e. $q(x) \gg \pi(x)$.

- Most weights ≈ 0 (negligible contribution)

$$\mathrm{Var}_q[\hat{I}_n] = \frac{1}{N}\mathrm{Var}_q[\phi(X)w(X)] = \frac{1}{N}\left(\mathbb{E}_q[\phi^2(X)w^2(X)] - (\mathbb{E}_p[\phi(X)w(X)])^2\right)$$

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

Note: If all weights are close to 0, ESS will still be close to n even though weights are not informative. One way to check for this $\sum_{i=1}^{N} w_i \approx N$ (since $\mathbb{E}_q[w(X)] = 1$).

What Makes a Good Proposal Distribution?

Ideal: $q(x) \propto |\phi(x)|\pi(x)$ (as it minimizes variance $V_q(\phi(X)w(X))$)

Practical guidelines:

- Heavy tails: q should have heavier tails than π (importance region coverage)
- Support: $q(x) > 0$ wherever $\phi(x)\pi(x) \neq 0$
- Similar shape: q should roughly match the shape of π, especially where $|\phi(x)|$ is large

Dimensional Scaling

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:

If we set $d = 100$, $\sigma = 1.2$, then $\mathrm{Var}_q[w(X)]$ is approximately 1.8×10^4.