Chapter 3 - Monte Carlo Methods

Variance Reduction Methods.
Importance Sampling.

Prof. Alex Alvarez, Ali Raisolsadat

School of Mathematical and Computational Sciences
University of Prince Edward Island

Importance Sampling Method Rationale

- In the standard Monte Carlo method, some simulated random values will have more impact in the parameter that is being estimated.
- The rationale of the **Importance Sampling** method is to generate more of these "relevant" values and fewer not-so-relevant values.
- This bias (towards "relevant" values) is corrected by considering a weighted sum. There will be more "relevant" values but they are given less weight compared to the Basic Monte Carlo method.

Importance Sampling (Mathematical Justification)

Let X be a random vector with $X \sim f(x)$.

$$\theta = E_f[g(X)] = \int_{\mathbb{R}^d} g(x)f(x)dx$$

The notation E_f means the expected value under the probability density f. We want to improve the standard MC estimator.
Let $h(x)$ be another density function such that $f(x) = 0$ whenever $h(x) = 0$, then:

$$\begin{aligned}
\theta = E_f[g(X)] &= \int_{\mathbb{R}^d} g(x)f(x)dx \\
&= \int_{\mathbb{R}^d} \frac{g(x)f(x)}{h(x)} h(x)dx \\
&= E\left(\frac{g(X)f(X)}{h(X)}\right) \quad \text{when} \quad X \sim h(x) \\
&= E_h\left(\frac{g(X)f(X)}{h(X)}\right)
\end{aligned}$$

We just proved that $\theta = E_f[g(X)] = E_h\left(\frac{g(X)f(X)}{h(X)}\right)$.

Then, in order to a approximate θ we could also apply a basic Monte Carlo estimator based on the second expected value above.

Algorithm 1 Monte Carlo Estimation via Importance Sampling

Input: Number of samples n, target density $f(x)$, proposal density $h(x)$, and function $g(x)$
for $j = 1$ to n **do**
    Generate $X_j \sim h(x)$
    Compute weight $w_j = \frac{f(X_j)}{h(X_j)}$
    Compute sample value $Y_j = g(X_j) \cdot w_j$
end for
Output: $\hat{\theta}_n^{IS} = \frac{1}{n}\sum_{j=1}^{n} Y_j$

Remark: The estimator is unbiased, since

$$E_h[\hat{\theta}_n^{IS}] = \theta$$

Importance Sampling Remarks

- The variance of the importance sampling estimator depends strongly on the choice of the function h.
- If h is poorly chosen, the variance of the estimator may actually increase instead of decrease.
- Ideally, h should be chosen such that the function $\frac{g(X)f(X)}{h(X)}$ is nearly constant (therefore with a small variance).
- In other words, a good choice of h should be a function with shape similar to $g(x)f(x)$, meaning that more samples are generated in regions where $g(x)f(x)$ is larger.
- At the same time, we should be able to sample efficiently from density h.
- In the textbook you can find a more detailed analysis of the MSE for the importance sampling estimator.

Importance Sampling Example

Problem: Estimate

$$\theta = \mathbb{E}\big(e^{-U^4}\big) = \int_0^1 e^{-x^4}\,dx, \qquad U \sim \text{Uniform}(0,1)$$

Basic Monte Carlo estimator:

$$U_1, \ldots, U_n \overset{\text{i.i.d.}}{\sim} \text{Uniform}(0,1), \qquad \hat{\theta}_n^{\text{MC}} = \frac{1}{n}\sum_{j=1}^{n} e^{-U_j^4}$$

To reduce variance we choose a proposal density $h(x)$ on $[0,1]$ that resembles $g(x) = e^{-x^4}$. A convenient (unnormalized) choice is $\tilde{h}(x) = e^{-x}$ on $[0,1]$. Normalize:

$$\int_0^1 e^{-x}\,dx = 1 - e^{-1}, \qquad C = \frac{1}{1-e^{-1}} = \frac{e}{e-1}$$

so the normalized proposal is

$$h(x) = Ce^{-x}, \qquad x \in [0,1]$$

Inverse CDF for h. The CDF of h is

$$F(x) = \int_0^x Ce^{-t}\,dt = C\big(1 - e^{-x}\big), \qquad 0 \le x \le 1$$

and the inverse is:

$$F^{-1}(x) = \ln\left(\frac{C}{C - x}\right)$$

Importance Sampling Estimator: The original density is $f(x) = 1$ on $[0, 1]$, so the importance weight is

$$w(x) = \frac{f(x)}{h(x)} = \frac{1}{Ce^{-x}} = \frac{e^x}{C},$$

$$\hat{\theta}_n^{\mathrm{IS}} = \frac{1}{n}\sum_{j=1}^{n} g(X_j)w(X_j) = \frac{1}{nC}\sum_{j=1}^{n} e^{X_j - X_j^4},$$

where $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} h(x) = Ce^{-x}$.

Homework

Let $X \sim N(0,1)$ and $A = [3,4]$.

1. Use Basic Monte Carlo to approximate the probability $P(X \in A)$.
2. Use an Importance Sampling estimate for the probability $P(X \in A)$ by generating samples from the distribution $N(3.5,1)$ (meaning this is the density h).

Hint: The probability $P(X \in A)$ is equal to $E(1_A(X))$ where the random variable 1_A is defined as

$$1_A(X) = \begin{cases} 1 & \text{if } X \in A \\ 0 & \text{if } X \notin A \end{cases}$$