2 Importance Sampling

Can we do better than the simple Monte Carlo estimator of

$$\theta = E[g(X)] = \int g(x)f(x)dx \approx \frac{1}{m}\sum_{i=1}^{m} g(X_i)$$

where the variables $X_1, \ldots, X_m$ are randomly sampled from f?

Yes!!

Goal: estimate integrals with lower variance than the simplest Monte Carlo approach.

↳ more efficient estimator

To accomplish this, we will use *importance sampling*.

2.1 The Problem

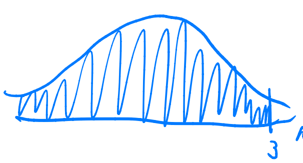
rare event

If we are sampling an event that doesn't occur frequently, then the naive Monte Carlo estimator will have high variance.

Example 2.1 Monte Carlo integration for the standard Normal cdf. Consider estimating $\Phi(-3)$ or $\Phi(3)$. (HW 6)

$\parallel$

$P(X \le 3)$



← events out here are rare ⇒ we may not get a lot of samples in the MC estimator.

We want to improve accuracy by causing rare events to occur more frequently than they would under the naive Monte Carlo sampling framework, thereby enabling more precise estimation.

place more "importance" on the rare events than we normally would by upweighting their chance of occurence and the correcting out estimator.

For very rare events, large reductions in the variance of the MC estimator are possible.

2.2 Algorithm

Consider a density function $f(x)$ with support $\mathcal{X}$. Consider the expectation of $g(X)$,

$$\theta = E[g(X)] = \int_{\mathcal{X}} g(x) f(x) dx.$$ where $X \sim f$.

Importance part.

Let $\phi(x)$ be a density where $\phi(x) > 0$ for all $x \in \mathcal{X}$. Then the above statement can be rewritten as

↑ support of ϕ covers the support of f.

downweighting back to what we need w/ f.

$$\theta = E[g(X)] = \int_{\mathcal{X}} g(x) \frac{f(x)}{\phi(x)} \underbrace{\phi(x)}_{\text{upweighting rare events}} dx = E\left[g(X) \boxed{\frac{f(X)}{\phi(X)}}\right]$$ where $X \sim \phi$

importance weight

ϕ is called the importance sampling function

ϕ must be a density (integrate to 1, and be ≥ 0 always).

An estimator of ϕ is given by the *importance sampling algorithm*:

1. Sample $X_1, \ldots, X_m$ from ϕ

2. Compute

$$\hat{\theta} = \frac{1}{m} \sum_{i=1}^{m} g(X_i) \frac{f(X_i)}{\phi(X_i)}.$$

For this strategy to be convenient, it must be

1. easy to sample from ϕ
2. easy to evaluate f (even if its not easy to sample from f).

Let X = result of rolling 1 fair six-sided die.

Example 2.2 Suppose you have a fair six-sided die. We want to estimate the probability that a single die roll will yield a 1. Want to estimate $P(X=1)$.

We could

① Roll the die m times

② Use a point estimator of $P(X=1)$ as proportion of ones in the sample.

This is a MC approach.
- You are sampling $X_1, \dots, X_m$ from f.
- estimate $P(X=1)$
$\frac{1}{m}\sum_{i=1}^{m} I(X_i=1)$.

The variance of the estimator is $\frac{5}{36m}$ if the die is fair.

Why?

$$\mathcal{X} = \{1, \dots, 6\} \quad f(x) = \begin{cases} 1/6 & x = 1, \dots, 6 \\ 0 & \text{o.w.} \end{cases}$$

$$\text{Define } Y = \begin{cases} 1 & \text{if } X = 1 \\ 0 & \text{o.w.} \end{cases} \Rightarrow Y \sim \text{Bernoulli}\left(\tfrac{1}{6}\right).$$

$$EY = p = \frac{1}{6}$$

$$\text{Var}\,Y = p(1-p) = \frac{1}{6}\left(\frac{5}{6}\right) = \frac{5}{36}$$

Expected # of 1s in m rolls:

$$E\left[\sum_{i=1}^{m} Y\right] = \sum_{i=1}^{m} EY = \frac{m}{6}$$

$$\hat{\theta} = \frac{1}{m}\sum_{i=1}^{m} Y_i \leftarrow \text{proportion of 1s in our sample.}$$

$$E(\hat{\theta}) = E\left(\frac{1}{m}\sum_{i=1}^{m} Y_i\right) = \frac{1}{m} E\left(\sum_{i=1}^{m} Y_i\right) = \frac{1}{m}\cdot\frac{m}{6} = \frac{1}{6}.$$

$$\text{Var}(\hat{\theta}) = \text{Var}\left(\frac{1}{m}\sum_{i=1}^{m} Y_i\right) = \frac{1}{m^2}\sum_{i=1}^{m}\text{Var}\,Y_i = \frac{1}{m^2}\sum_{i=1}^{m}\frac{5}{36} = \frac{5}{36m}$$

relative measure of variability in a r.v.

used a lot in chemistry and physics.

We can consider the "coefficient of variation" $CV[X] = \frac{\sqrt{\text{Var}[X]}}{E[X]}$

If we want a CV of 5%, then what value of m do we need?

$$CV\left(\frac{\sum Y_i}{m}\right) = \frac{\sqrt{\text{Var}\left(\frac{\sum Y_i}{m}\right)}}{E\left(\frac{\sum Y_i}{m}\right)} = \frac{\sqrt{5/36m}}{1/6} \overset{\text{want}}{=} .05$$

$$\frac{5}{36m} = \left[\frac{1}{6}(.05)\right]^2$$

$$m = \frac{5}{36\left[\frac{1}{6}(.05)\right]^2} = 2000 \text{ rolls.}$$

want to lower this.

To reduce the # of rolls, we could consider biasing the die by replacing the faces bearing 2 and 3 with additional 1s.

This increases the probability of rolling a 1 to 0.5, but now we aren't sampling from the target dsn (a fair die roll).

Now $P(X=1) = \frac{1}{2}$

$P(X=2) = P(X=3) = 0$

$P(X=4) = P(X=5) = P(X=6) = \frac{1}{6}$

Can correct this by

- weighting each roll of a 1 by $\frac{1}{3}$
- Let $Y_i = \begin{cases} \frac{1}{3} & \text{if } X=1 \\ 0 & \text{otherwise.} \end{cases}$

Then the expectation of the sample mean of $\frac{\sum Y_i}{m}$ is

$$E\left[\frac{\sum Y_i}{m}\right] = \frac{1}{m}\sum_{i=1}^{m} EY_i = EY = \frac{1}{3}\cdot\frac{1}{2} + \underbrace{0\left[0+0+\frac{1}{6}+\frac{1}{6}+\frac{1}{6}\right]}_{=0} = \frac{1}{6}$$

But the variance is

$$Var\left[\frac{\sum Y_i}{m}\right] = \frac{1}{m^2}\sum_{i=1}^{m} Var Y_i = \frac{1}{m} Var Y = \frac{1}{m}\left[\frac{1}{18} - \left(\frac{1}{6}\right)^2\right] = \frac{1}{36m}$$

$$EY^2 = \left(\frac{1}{3}\right)^2 \cdot \frac{1}{2} = \frac{1}{18}$$

So to acheive a CV of 5%, we would need only

$$\frac{\sqrt{\frac{1}{36m}}}{\frac{1}{6}} = .05$$

solve for m $\quad m = \frac{1}{36\left(\frac{1}{6}\times .05\right)^2} = 400$ rolls.

2.3 Choosing ϕ

In order for the estimators to avoid excessive variability, it is important that $f(x)/\phi(x)$ is bounded and that ϕ has heavier tails than f.

Example 2.3

Example 2.4

A rare draw from ϕ with much higher density under f than under ϕ will receive a huge weight and inflate the variance of the estimate.

Strategy –

Example 2.5

The importance sampling estimator can be shown to converge to θ under the SLLN so long as the support of ϕ includes all of the support of f.

2.4 Compare to Previous Monte Carlo Approach

Common goal –

Step 1 Do some derivations.

a. Find an appropriate f and g to rewrite your integral as an expected value.

b. For **importance sampling** only,

   Find an appropriate ϕ to rewrite θ as an expectation with respect to ϕ.

Step 2 Write pseudo-code (a plan) to define estimator and set-up the algorithm.

- For **Monte Carlo integration**

    1.

    2.

- For **importance sampling**

    1.

    2.

Step 3 Program it.

2.5 Extended Example

In this example, we will estimate $\theta = \int_0^1 \frac{e^{-x}}{1+x^2} dx$ using MC integration and importance sampling with two different importance sampling distributions, ϕ.