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.

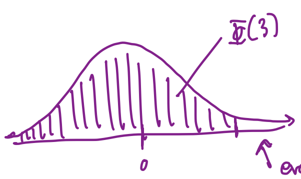
↳ more efficient

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

2.1 The Problem

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).



↑ events out here are rare, so may not get many values out here to occur by chance ⇒ not many values to contribute to our 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.

For very rare events, extremely 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 $g(X)$,

= {x : f(x) > 0}

X ~ f

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

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 includes support of f.

$$\theta = E[g(X)] = \int_{\mathcal{X}} g(x) \frac{f(x)}{\phi(x)} \phi(x)\, dx$$

$$= E\left[g(Y) \frac{f(Y)}{\phi(Y)} \mathbb{I}(Y \in \mathcal{X})\right]$$

$Y \sim \phi$

ϕ is called the importance sampling function

ϕ MUST be a density (i.e. must integrate to 1 and always ≥ 0).

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

1. Sample $X_1, \dots, X_m \sim \phi$
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 not easy to sample from f).

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 a die m times

(2) Estimate of $P(X=1)$ could be proportion of ones in sample.

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

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

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

$EY = \sum_{i=1}^{6} \mathbb{I}(Y=1) \cdot \frac{1}{6} = \frac{1}{6}$

$\text{Var } Y = p(1-p) = \frac{1}{6} \cdot \frac{5}{6} = \boxed{\frac{5}{36}}$

Expected # of 1's in m rolls:

$E\left[\sum Y_i\right] = \sum_{i=1}^{m} EY_i = \frac{m}{6}$.

Properties of resulting estimator → Proportion of 1's in sample:

$E\left[\frac{\sum Y_i}{m}\right] = \frac{1}{m} E\left(\sum Y_i\right) = \frac{m}{6} \cdot \frac{1}{m} = \frac{1}{6}$.

$\text{Var}\left[\frac{\sum Y_i}{m}\right] = \frac{1}{m^2} \text{Var} \sum Y_i = \frac{1}{m^2} \sum \text{Var } Y_i = \frac{1}{m^2} \cdot m \cdot \boxed{\frac{5}{36}} = \frac{5}{36m}$

We can consider the "coefficient of variation" $\quad CV[X] = \frac{\sqrt{\text{Var}[X]}}{E[X]}$. ← relative measure of variation used in chemistry and physics, etc.

So $CV\left[\frac{\sum Y_i}{m}\right] = \frac{\sqrt{\text{Var}(\sum Y_i / m)}}{E(\sum Y_i / m)} = \frac{\sqrt{5/36m}}{1/6}$ (← estimator)

If we want coefficient of variation = 5%, then

$$\frac{\sqrt{5/36m}}{1/6} = 0.05$$

$$\frac{5}{36m} = \left[\frac{1}{6} \cdot (0.05)\right]^2$$

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

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

This increases prob. of rolling a 1 to 0.5 but now we aren't drawing from the target distribution (a fair die).

Now:
$P(X=1) = \frac{1}{2}$
$P(X=2) = P(X=3) = 0$
$P(X=4) = P(X=5) = P(X=6) = \frac{1}{6}$

We can correct this by

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

Then the expectation of the sample mean of Y

$$E\left[\frac{1}{m}\sum_{i=1}^{m} Y_i\right] = \frac{1}{m}\sum_{i=1}^{m} EY_i = EY_1 = \frac{1}{3}\cdot\frac{1}{2} + 0\left[0+0+\frac{1}{6}+\frac{1}{6}+\frac{1}{6}\right] = \frac{1}{6} \checkmark$$

But the variance is

$$E(Y^2) = \frac{1}{3^2}\cdot\frac{1}{2} = \frac{1}{18}$$

$$\text{Var}\left[\frac{1}{m}\sum_{i=1}^{m} Y_i\right] = \frac{1}{m^2}\sum \text{Var}\, Y_i = \frac{1}{m}\text{Var}\, Y_1 = \frac{1}{m}\left[\frac{1}{18} - \left(\frac{1}{6}\right)^2\right] = \frac{1}{36m}$$

So to achieve CV of 5% we would only need

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

$\vdots$

$$m = \frac{1}{36\left(\frac{1}{6}\cdot .05\right)^2} = 400 \text{ rolls.}$$

The die rolling example is successful because the importance sampling function (rolling the die w/ 3 ones) is used to over-sample a proportion of the state space that recieves lower probability under the target dsn. and importance weighting corrects the bias.

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.