STAT 3503/8109 Importance Sampling Notes

Edoardo Airoldi

Scribe: Srikar Katta[*]

Fall 2020: November 30, 2020

1 Introduction

Many of the Monte Carlo methods require us to generate random samples from a distribution to approximate integrals, which are often difficult to compute. In many cases, we sample from the probability distribution that "comes with" the integral. Suppose we have a random variable $\theta \sim \mathbb{P}(\theta)$ and a function $h : \theta \to \mathbb{R}$. Then, the expectation of $h(\theta)$ would be

$$\mathbb{E}[h(\theta)] = \int_\theta h(\theta)\mathbb{P}(\theta)d\theta.$$

For this approximation, we sample from $\mathbb{P}(\theta)$. However, $\mathbb{P}(\theta)$ can also be difficult to generate sometimes, which is where importance sampling comes into play. We may also use importance sampling when we want to generate samples from a density only up to a multiplicative constant.

We will extend the example outlined earlier. First, we can define a new probability density function, $\mathbb{G}(\theta)$ with the same support as $\mathbb{P}(\theta)$. For example, if $\theta \sim Normal(0,1)$, then we could use $Normal(1,2)$ as $\mathbb{G}(\theta)$ since they both share the same support: the whole real line.

Now, we take advantage of the fact that $\frac{\mathbb{G}(\theta)}{\mathbb{G}(\theta)} = 1$, so

$$\mathbb{E}[h(\theta)] = \int_\theta h(\theta)\mathbb{P}(\theta)d\theta$$

[*]Please share any comments or suggestions with Srikar Katta at srikar@temple.edu

$$
\begin{aligned}
&= \int_\theta h(\theta)\mathbb{P}(\theta) \cdot 1 \\
&= \int_\theta h(\theta)\mathbb{P}(\theta) \cdot \frac{\mathbb{G}(\theta)}{\mathbb{G}(\theta)} \\
&= \int_\theta h(\theta)\frac{\mathbb{P}(\theta)}{\mathbb{G}(\theta)}\mathbb{G}(\theta) \\
&= \mathbb{E}_{\mathbb{G}(\theta)}\left[h(\theta)\frac{\mathbb{P}(\theta)}{\mathbb{G}(\theta)}\right],
\end{aligned}
$$

the expectation of $h(\theta)\frac{\mathbb{P}(\theta)}{\mathbb{G}(\theta)}$ with respect to the probability distribution function $\mathbb{G}(\theta)$. Because we can choose $\mathbb{G}(\theta)$ to be whatever we want – as long as it has the same support as $\mathbb{P}(\theta)$ – we can draw samples from a simpler function instead and estimate the integral more easily (i.e., simulate $\theta \sim \mathbb{G}(\theta)$). Because $\mathbb{P}, \mathbb{G}$ both integrate to 1, they are probability distributions. However, h will not necessarily integrate to 1, so it is just a function of θ.