Monte Carlo Integration

Christoph Heindl
https://github.com/cheind/monte-carlo-integration

June 24, 2022

1 Introduction

The purpose of this document is to present Monte Carlo integration, a method to approximate the value of the definite integral

$$\int_a^b g(x)\,dx. \tag{1}$$

In Monte Carlo we frame the above integral as an expectation over a random variable for which we know that unbiased estimators exist. Assume x is random $x \sim X$ with PDF $f_X(x)$. Then multiply and divide the right hand side by $f_X(x)$ to get

$$\int_a^b \frac{g(x)}{f_X(x)} f_X(x)\,dx.$$

Further assume $f_X(x)$ is zero everywhere but in $[a, b]$, so we can change the bounds of integration to

$$\int_{-\infty}^{\infty} \frac{g(x)}{f_X(x)} f_X(x)\,dx.$$

Using LOTUS[1] we parse the above as the expected value of

$$\mathbb{E}\left[\frac{g(X)}{f_X(X)}\right]. \tag{2}$$

Invoking results from asymptotic theory we know that an unbiased estimator for the expectation is

$$\frac{1}{N}\sum_{i=1}^{N} \frac{g(\hat{x}_i)}{f_X(\hat{x}_i)}, \tag{3}$$

where $\hat{x}_i$ is the i-th random sample out of N iid samples. As N increases our approximation approaches the true value of the integral. To summarize

$$\int_a^b g(x)\,dx = \mathbb{E}\left[\frac{g(X)}{f_X(X)}\right] \approx \frac{1}{N}\sum_{i=1}^{N} \frac{g(\hat{x}_i)}{f_X(\hat{x}_i)}. \tag{4}$$

If we assume X is distributed uniformly on $[a, b]$ we have $f_X(x) = \frac{1}{b-a}$ and get

$$\int_a^b g(x)\,dx \approx \frac{1}{N}\sum_{i=1}^{N} g(\hat{x}_i)(b-a). \tag{5}$$

Geometrically, this equals taking the average area of rectangles having length $b - a$ and height $g(\hat{x}_i)$.

[1] https://en.wikipedia.org/wiki/Law_of_the_unconscious_statistician

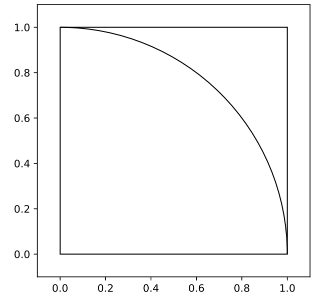

Figure 1: Quadrant of circle with bounding square.

2 Multivariate cases and areas of circles

The above result is useful in itself as it allows us to numerically integrate any function by random sampling. In this section we extend the approach to multiple dimensions and show how it relates to indicator functions.

For the area of circle, written as the integral of the circumference, we have

$$A_{\text{circle}} = \int_0^1 2\pi r \, dr.$$

To estimate the integral we proceed as follows

1. Sample N iid numbers from a uniform $[0, 1]$.
2. Compute $g(\hat{x}_i) = 2\pi\hat{x}_i$ for every sample.
3. Use Equation 5 to estimate the integral.

2.1 Multivariate Approach

A more insightful way for our purposes is to consider a different integral. Consider a quadrant of a unit circle as shown in Figure 1. The area of this quadrant can be written as the double integral

$$A_q = \int_0^1 \int_0^{\sqrt{1-x^2}} dydx.$$

Conceptually we integrate small squares of area $dydx$. While x runs from zero to one, we condition the upper limit of y on the x-position by rearranging the equation of a unit circle $x^2 + y^2 = 1$. Using the indicator function

$$\mathbb{I}_{x^2+y^2\leq 1} := \begin{cases} 1 & x^2 + y^2 \leq 1 \\ 0 & \text{else} \end{cases},$$

we rewrite the integral as

$$A_q = \int_0^1 \int_0^1 \mathbb{I}_{x^2+y^2\leq 1} \, dydx,$$

Let X and Y be random variables whose joint PDF is $f_{X,Y}$. For the same arguments as in the uni-variate case, we identify the above integral as the expectation of

$$\mathbb{E}_{X,Y}\left[\frac{g(X,Y)}{f_{X,Y}(X,Y)}\right],$$

with $g(X,Y) := \mathbb{I}_{X^2+Y^2\leq 1}$. An unbiased estimator for the A_q is therefore given by

$$A_q \approx \frac{1}{N}\sum_{i=1}^{N}\frac{g(\hat{x}_i,\hat{y}_i)}{f_{X,Y}(\hat{x}_i,\hat{y}_i)}.$$

If X and Y are independent, i.e. $f_{X,Y} = f_X(x)f_Y(y)$ and uniform on $[0,1]$ we have

$$A_q \approx \frac{(1-0)(1-0)}{N}\sum_{i=1}^{N} g(\hat{x}_i,\hat{y}_i) = \frac{1}{N}\sum_{i=1}^{N}\mathbb{I}_{x^2+y^2\leq 1},$$

and $A_{\text{circle}} = 4A_q$.

2.2 Generalization

Given a shape S, an associated indicator function $\mathbb{I}$ that determines whether a point (x,y) is inside a shape S and a bounding rectangular region B, we can approximate the area of S by

$$\begin{aligned} A_{\text{S}} &= \int_a^b \int_c^d \mathbb{I}(x,y)\,dydx \\ &\approx \frac{(b-a)(d-c)}{N}\sum_{i=1}^{N}\mathbb{I}(x,y). \end{aligned} \tag{6}$$

In other words: the area of S, A_S, is approximated by

$$A_S \approx \frac{N_{\text{inside}}}{N} A_B,$$

where N_{inside} is the number of inside points and A_B is the area of the bounding region. This concept generalizes to more than two dimensions.

3 Remarks

Up to this point we made extensively use of the uniform distribution for $f_X(x)$, because its easy to visualize and leads to simple geometric concepts. Keep in mind, though, that other distributions are suited equally well as long the PDF is zero outside the integration bounds $[a,b]$. In fact the uniform might be a bad choice (in terms of approximation error and convergence) if most of the 'mass' of $g(x)$ is concentrated in few sub-regions of $[a,b]$.

Python code is available at https://github.com/cheind/monte-carlo-integration

4 Todo

Just a few reminders for myself of what's missing:

- Prove that the Monte Carlo estimator is an unbiased estimator.
- Add importance sampling to speed up convergence.

References

[TAB17] MARCO. TABOGA. *Lectures on probability theory and mathematical statistics*. Createspace, 2017.