Monte Carlo Integration

Jonathan Navarrete

May 20, 2017

Introduction

Beyond being able to generate random variables, we should also be able to apply these simulation methods to solve more complex problems. We'll begin with Monte Carlo Integration methods.

Topics to be covered:

1. Classic Monte Carlo Integration
2. Importance Sampling
3. Importance Sampling Resampling

Monte Carlo Integration

Given a function $g(x)$ for which we wish to integrate over [a,b], $\int_a^b g(x)dx$, we can treat this deterministic problem as a stochastic one to find the solution. Treat X as if a random variable with density $f(x)$, then the mathimatical expecation of the random variable $Y = g(X)$ is

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

If a random sample $X_1, ..., X_n$ is generated from $f(x)$, an unbiased estimator of $E[g(X)]$ is the sample mean.

$$\bar{g_n} = \frac{1}{n}\sum_{i=1}^{n} g(x_i)$$

As an example, suppose we have a function $g(x) = 3x^2$ for which we wish to integrate over the interval 0 to 2. We could apply deterministic numerical approximation methods (see R's `integrate`) or we could treat x as a random variable from a $Unif(0, 2)$ whose pdf is simply $f(x) = \frac{1}{2-0} = \frac{1}{2}$. If we now generate some $N = 1{,}000$ random values from $f(x)$ and evaluate them at $g(x)$, then take the mean, we'd be calculating the expected value of $g(x)$,

$$\begin{aligned}
\theta &= \int_0^2 g(x)dx \\
&= (\frac{2-0}{2-0}) \times \int_0^2 g(x)dx \\
&= 2 \times \int_0^2 g(x)\frac{1}{2}dx \\
&= 2 \times E[g(X)] = 2 \times \int_{-\infty}^{\infty} g(x)f(x)dx \\
&\approx 2 \times \frac{1}{n}\sum_{i=1}^{n} g(x_i) \\
&= \hat{\theta} \approx \theta
\end{aligned}$$

If we now simulate this, we will see we approximate the true solution $\int_0^2 g(x)dx = 8$

```r
N = 10000 ## sample size

g <- function(x) { 3*x^2 } ## function of interest, g(x)

X <- runif(n = N, min = 0, max = 2)  ## samples from f(x)

v = 2 * mean(g(X)) ## 2 * E[g(x)]

print(v) ## approximately 8
```

```
## [1] 8.019278
```

Now, some of you may ask, "why is it that we can use the empirical average to create an estimate?" Well, that's because the discrete expected value of $g(x)$ is $E[g(x)] = g(x_1)P(X = x_1) + ... + g(x_1)P(X = x_n)$ where $P(X = x_i) = \frac{1}{n}$ since $x_i \sim Unif(0, 2)$.

Now, to generalize the method above. Given a function $g(x)$ whose integral is well defined, for which we wish to evaluate at interval a to b. Then

$$\begin{aligned} \theta &= \int_a^b g(x)dx \\ &= (b-a)\int_a^b g(x)\frac{1}{b-a}dx \\ &= (b-a)\int_a^b g(x)f(x)dx \end{aligned}$$

where $f(x) = \frac{1}{b-a}$ is $Unif(a, b)$, and $x \sim Unif(a, b)$.

The algorithm to calculate $\hat{\theta}$ is as follows:

1. Find a density $f(x)$ from which we can sample x from.
2. Generate $x_1, ..., x_n \sim f(x)$
3. Compute $(b-a) \times \bar{g_n}$, where $\bar{g_n} = \frac{1}{n}\sum_{i=1}^n g(x_i)$

Now that we can calculate an estimate of statistic θ, $\hat{\theta}$, we should also be able to calculate the standard error in order to build confidence intervals (CIs).

The variance can be written as

$$Var(x) = \frac{1}{n}\sum_{i=1}^n (x_i - \mu)^2$$

We will use this form to calculate the variance of $\hat{\theta}$

$$\begin{aligned} &Var(g(x_i)) \\ &= \frac{1}{n}\sum_{i=1}^n (g(x_i) - \hat{\theta})^2 \\ &= \sigma^2 \end{aligned}$$

And

$$\frac{\sigma^2}{n} = \frac{1}{n^2}\sum_{i=1}^n (g(x_i) - \hat{\theta})^2$$

So, the standard error estimate is

$$\frac{\sigma}{\sqrt{n}} = \frac{1}{n}\sqrt{\sum_{i=1}^{n}(g(x_i) - \hat{\theta})^2}$$