# 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. Review: 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 $\int_0^2 g(x)dx = 8$

```
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] 8.015489
```

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 $\theta$, $\hat{\theta}$, 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

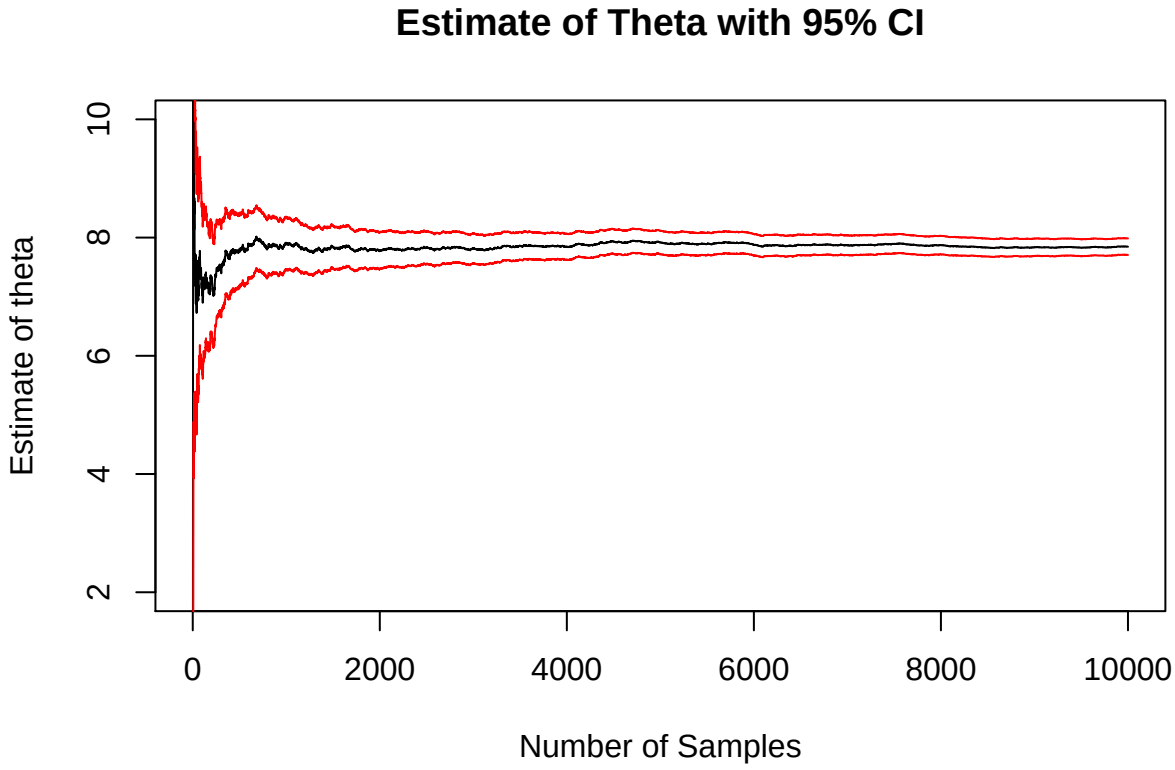
$$\frac{\sigma^2}{n} = \frac{1}{n^2}\sum_{i=1}^n (g(x_i) - \hat{\theta})^2$$

So, the standard error estimate is

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

We can now calculate the standard error for the former example.

```
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)
g_values <- 2 * g(X)

cumMean <- function(x, n){
  num = cumsum(x)
  denom = 1:n
  result = num/denom
  return(result)
}


cumSE <- function(x, n){
  m = mean(x)
  num = sqrt(cumsum((x - m)^2))
  denom = 1:n
  result = num/denom ## cummulative mean of (x_i - theta)^2
  return(result)
}

thetas = cumMean(g_values, N)
SE = cumSE(g_values, N)

plot(x = 1:N, y = thetas, type = "l", ylim = c(2, 10), xlab = "Number of Samples",
     ylab = "Estimate of theta",
     main = "Estimate of Theta with 95% CI")
lines(x = 1:N, y = thetas + 1.96*SE, col = "red") ## CI
lines(x = 1:N, y = thetas - 1.96*SE, col = "red") ## CI
```




```
## final estimate
thetaHat = mean(g_values)
se <- sd(x = g_values)/sqrt(N)
ci <- thetaHat + 1.96*c(-1,1) * se

print("Theta estimate: ")
```

```
## [1] "Theta estimate: "
```

```
print(thetaHat)
```

```
## [1] 7.844576
```

```
print("Confidence Intervals")
```

```
## [1] "Confidence Intervals"
```

```
print(ci)
```

```
## [1] 7.704906 7.984246
```

The principle of the Monte Carlo method for approximating $E_f[g(X)] = \int_{\mathcal{X}} g(x)f(x)dx$ is to generate a sample $\mathbf{X} = (\mathbf{x_1}, ..., \mathbf{x_n})$ from pdf $f(x)$ and using the empirical average of $g(x)$ as an approximation. This Monte Carlo estimate **almost surely** converges to $E_f[g(X)]$ by the Strong Law of Large Numbers.

### Strong Law of Large Numbers

The SLLN says that for a given random sample $X_1, ..., X_n$ the empirical mean, $\bar{X}$, can be used as an estimator for $E[x] = \mu$, and given a **sufficiently** large sample, this estimate approximates the true expectation $\mu$ almost certainly with a probability value of 1. Or more formally,

$$P(\lim_{n\to\infty} \bar{X} \to E[X]) = 1$$

See reference: Strong Law of Large Numbers

### Central Limit Theorem

The CLT implies that for our Monte Carlo estimator $\hat{\theta}$,

$$\frac{\hat{\theta} - E[\hat{\theta}]}{\sqrt{Var(\hat{\theta})}} \sim N(0,1)$$

as $n \to \infty$. See: Central Limit Theorem

### Example: Normal p-value estimate for unbounded region

$$\Phi(x) = P(X \leq x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} \, e^{(t-\mu)^2/2\sigma} dx$$

Now, suppose that $\mu = 1$ and $\sigma = 1$, such that we are only dealing with the standard Normal CDF. We are interested in calculating a Monte Carlo estimate for $\Phi(x)$

$$\Phi(x) = P(X \leq x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \, e^{t^2/2} dx$$

Our first problem is dealing with bounds $(-\infty, x)$. We cannot direcly apply our MC integration algorithm directly because the integral range $(-\infty, x)$, but a way around this issue is by breaking this problem into two sections exploiting the Normal distribution's symmetry around $\mu$. The two sections will be $(-\infty, 0)$ and $(0, x)$. With the given symmetry property, we know that any random variable from $N(\mu, \sigma)$ will have an equal probability of being on the positive end or negative. Thus, let's set $P(X < 0) = 1/2$. Now, all we need to concern ourselves with is the region $(0, x)$.

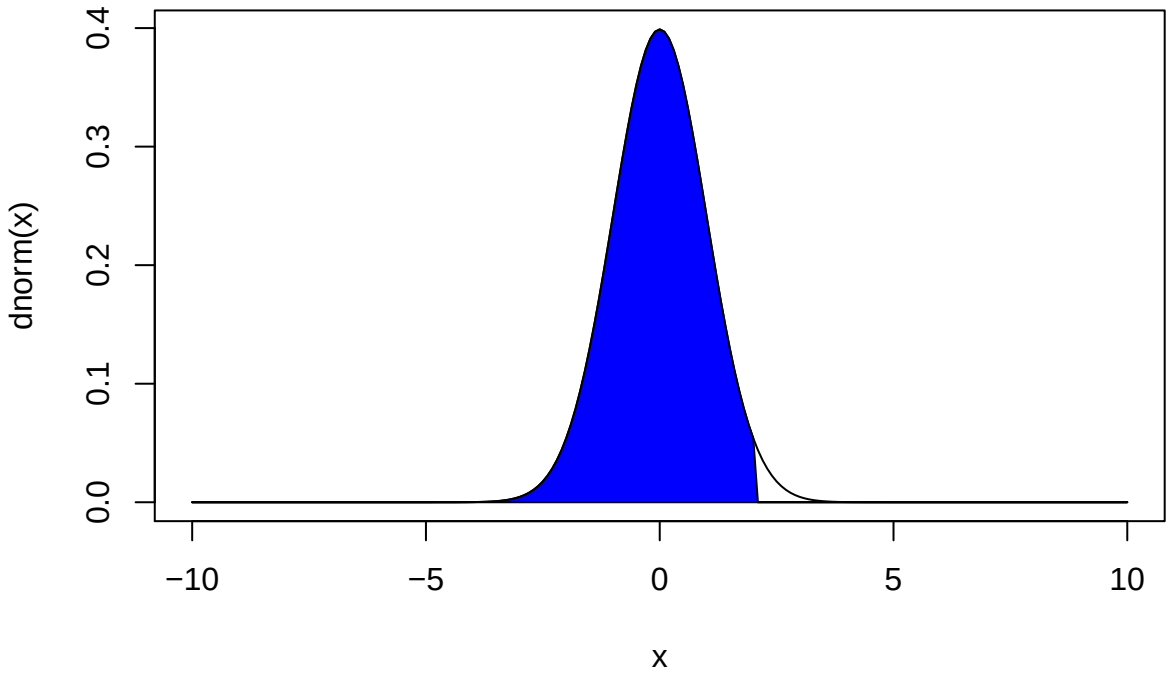
Now, assuming $x > 0$, then we use the Monte Carlo estimator as is

$$\begin{aligned}
\Phi(x) &= P(X \leq x) \\
&= P(X < 0) + \int_{0}^{x} \frac{1}{\sqrt{2\pi}} \, e^{t^2/2} dx \\
&= \frac{1}{2} + \int_{0}^{x} \frac{1}{\sqrt{2\pi}} \, e^{t^2/2} dx
\end{aligned}$$

However, what if $x < 0$? Well, then, again exploiting the symmetry property of $N(\mu, \sigma)$, we esimate $\Phi(X \leq x)$ by $1 - \Phi(-x)$. I'll illustrate how this will be done.

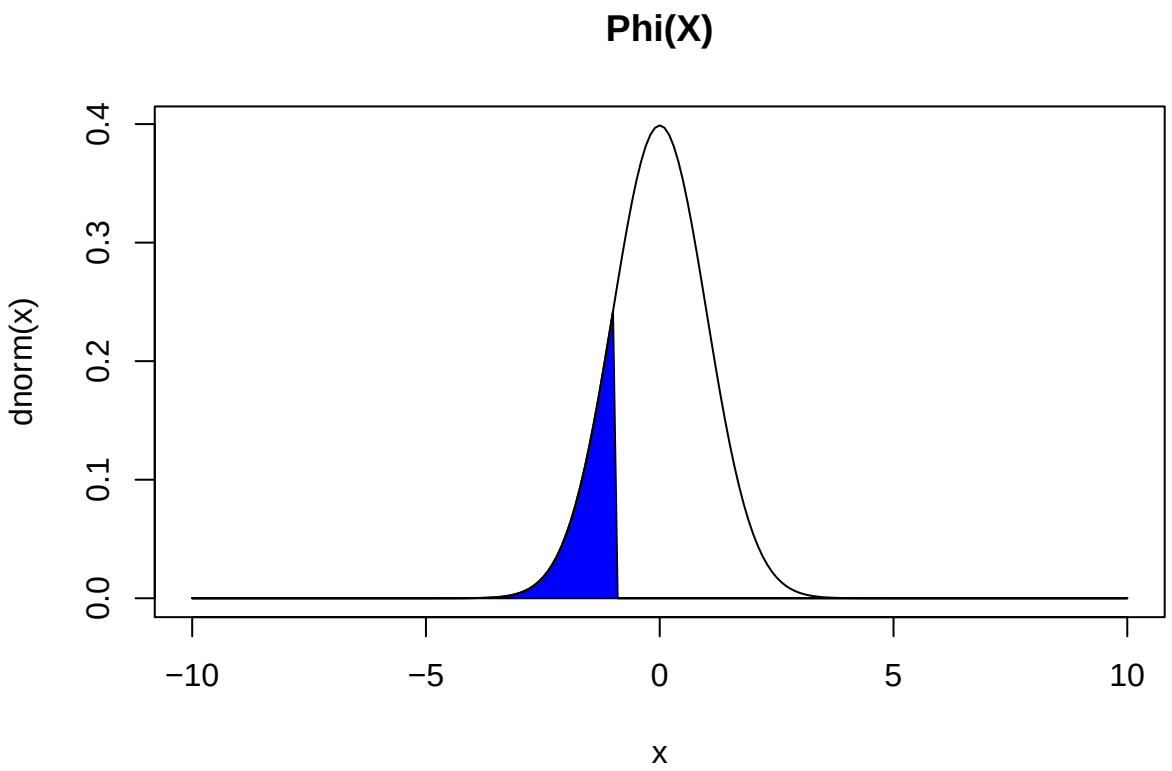
For example, if $x = 2$, then we would calculate a MC estimate for $\Phi(X \leq 2)$.

```
x <- seq(from = -10, to = 10, by = 0.1)

out = dnorm(x)
out[x > 2] = 0

plot(x, dnorm(x), type = "l")
polygon(x, out, col = "blue")
```




However, if $x = -1$, then we would use the symmetry property of the Normal distribution to calculate $\Phi(-1)$ by $1 - \Phi(-(-1)) = 1 - \Phi(1)$

```
x <- seq(from = -10, to = 10, by = 0.1)

out = dnorm(x)
out[x > -1] = 0

plot(x, dnorm(x), type = "l", main = "Phi(X)")
polygon(x, out, col = "blue")
```




```
out = dnorm(x)
out[x < 1] = 0

plot(x, dnorm(x), type = "l", main = "1 - Phi(-X)")
polygon(x, out, col = "red")
```