Monte Carlo Integration

Motivation

Suppose we want to evaluate a definite integral $\int_D g(x)dx$, where D is some domain and g is a function. For most functions, there is no closed-form expression for such definite integrals.

A technique for evaluating a definite integral based on random sampling was developed in the 1940's by scientists working on the Manhattan project. Apparently, design of nuclear weapons requires the evaluation of complicated integrals.

Recall that the expected value of function g with respect to a distribution with pdf f is $E_f[g(x)] = \int g(x)f(x)dx$ (x is a random variable with pdf f)

If $X_1, \ldots, X_n$ are independent random variables with common density f, then $\frac{1}{n}\sum_{i=1}^n g(X_i) = E_f[g(X)] = \int g(x)f(x)dx$ as $n \to \infty$.

Roughly speaking, if $X_1, \ldots, X_n$ are independent random variables with common density f, then $\frac{1}{n}\sum_{i=1}^n g(X_i) \approx E_f[g(X)] = \int g(x)f(x)dx$ for sufficiently large n.

Example 1

We wish to evaluate $\theta = \int_0^1 e^{-x}dx = 1 - 1/e = 0.63212$. We can see how close we can come to this with samples of $n = 100$ and $n = 100000$ using Monte Carlo integration with respect to a uniform density.

Let $X_1, \ldots, X_n$ be i.i.d. from uniform distribution (0,1). Then the simple Monte Carlo estimate of θ is $\hat{\theta} = \frac{1}{n}\sum_{i=1}^n e^{-X_i}$.

```
n=100
X=runif(n,0,1)
gx=exp(-X)
thetahat=mean(gx)
thetahat
```

```
## [1] 0.6304194
```

```
n=100000
X=runif(n,0,1)
gx=exp(-X)
thetahat=mean(gx)
thetahat
```

```
## [1] 0.6316856
```

Example 2

There is nothing particularly special about the interval $(0, 1)$, we can apply the same idea to an integral on any bounded interval. Now let

$$\theta = \int_a^b g(x)dx = (b-a)\int_a^b g(x)\frac{1}{b-a}dx.$$

So now θ is just $(b-a)E[g(X)]$ when X has a uniform distribution on (a, b).

Consider the density function of a standard normal random variable $g(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$. Let θ be the probability that a standard normal variable falls between -1.96 and 1.96.

$$\theta = \int_{-1.96}^{1.96} \frac{1}{\sqrt{2\pi}}e^{-x^2/2}dx.$$

Let's try to approximate this familiar number using Monte Carlo integration.

```
n=100000
X=runif(n,-1.96,1.96)
gx=exp(-X^2/2)/sqrt(2*pi)
thetahat=(1.96-(-1.96))*mean(gx)
thetahat
```

```
## [1] 0.9502604
```

To summarize, we have considered the problem of approximating an integral $\theta = \int_a^b g(x)dx$ by expressing it as a multiple of an expected value of $g(X)$ when X has a uniform distribution on (a, b), and using what we know about sample means of a random variable.

Specifically, the algorithm is as follows.

1. Generate $X_1, X_2, \ldots, X_n$ iid Uniform (a, b)
2. Compute $\hat{\theta} = (b - a)\frac{1}{n}\sum_{i=1}^{n} g(X_i)$

Next, consider integration of a function over an unbounded interval such as $(-\infty, \infty)$, (a, ∞), or $(-\infty, b)$.

Example 3 (Method 1)

Let

$$\theta = \int_1^\infty \frac{1}{x^2[2 + \sin(x)]} dx.$$

Let's break this into parts,

$$\theta = \int_1^M \frac{1}{x^2[2 + \sin(x)]} dx + \int_M^\infty \frac{1}{x^2[2 + \sin(x)]} dx \approx \int_1^M \frac{1}{x^2[2 + \sin(x)]} dx.$$

We now compute $\hat{\theta}$ by using the 2-step algorithm above, integrating over $(1, 1000)$.

```
n=1000000
X=runif(n,1,1000)
gx=1/(X^2*(2+sin(X)))
thetahat=(1000-1)*mean(gx)
thetahat
```

```
## [1] 0.4395292
```

Example 4 (Method 2)

Another approach is to express the integral as an integral over a bounded interval. For example, rather than express θ as the integral of x that we have, express it as an integral in $y = 1/x$.

Theorem (Definite integration by substitution):

Suppose that the function h has a continuous derivative on $[a, b]$ and that g is continuous on $[h(a), h(b)]$. Let $x = h(y)$, then

$$\int_{h(a)}^{h(b)} g(x)dx = \int_a^b g(h(y))h'(y)dy.$$

In this application, we take $x = h(y) = 1/y$. Then

$$\theta = \int_1^\infty \frac{1}{x^2[2 + \sin(x)]} dx = \int_1^0 \frac{y^2}{2 + \sin(1/y)} (-1/y^2)dy = \int_0^1 \frac{1}{2 + \sin(1/y)} dy.$$

Now the integral is written as an integral over $(0, 1)$.

```r
n=1000000
Y=runif(n,0,1)
ghy=1/(2+sin(1/Y))
thetahat=mean(ghy)
thetahat
```

```
## [1] 0.4458512
```

Let's consider the standard error of $\hat{\theta}$ and how to construct a confidence interval for θ.

Standard Error of $\hat{\theta}$

Let's consider the variation in the estimator

$\hat{\theta} = \frac{b-a}{n}\sum_{i=1}^{n} g(X_i)$. Because $g(X_1), g(X_2), \ldots, g(X_n)$ are iid from the distribution of $g(X)$, we see that

$$\mathrm{E}[\hat{\theta}] = \frac{b-a}{n}\sum_{i=1}^{n} \mathrm{E}[g(X_i)] = \theta.$$

So $\hat{\theta}$ is unbiased.

$$\mathrm{Var}(\hat{\theta}) = \frac{(b-a)^2}{n^2}\sum_{i=1}^{n} \mathrm{Var}[g(X_i)] = (b-a)^2\mathrm{Var}[g(X)]/n.$$

Let $\sigma^2 = \mathrm{Var}[g(X)]$. Then σ^2 can be estimated by

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left[g(X_i) - \frac{\sum_{i=1}^{n} g(X_i)}{n}\right]^2,$$

so then $\hat{\mathrm{Var}}(\hat{\theta}) = (b-a)^2\hat{\sigma^2}/n$ and $se(\hat{\theta}) = (b-a)\hat{\sigma}/\sqrt{n}$.

A straightforward application of the Central Limit Theorem tells us that

$$\frac{\hat{\theta} - \theta}{\sqrt{\mathrm{Var}(\hat{\theta})}} \approx \frac{\hat{\theta} - \theta}{se(\hat{\theta})} \sim N(0,1).$$

From this result, we can easily derive a confidence interval for θ of level $1-\alpha$ by

$$(\hat{\theta} - z_{\alpha/2}(b-a)\hat{\sigma}/\sqrt{n}, \hat{\theta} + z_{\alpha/2}(b-a)\hat{\sigma}/\sqrt{n})$$

where $z_{\alpha/2}$ is defined by $P[Z > z_{\alpha/2}] = \alpha/2$ when $Z \sim N(0,1)$.

```r
n=1000
Y=runif(n,0,1)
ghy=1/(2+sin(1/Y))
thetahat=mean(ghy)
se=sd(ghy)/sqrt(n)
zval=qnorm(1-0.025,0,1)
L=thetahat-se*zval
U=thetahat+se*zval
L
```

```
## [1] 0.4396697
```

```r
U
```

```
## [1] 0.4632582
```

Next, we'll consider some techniques to improve on the efficiency of our definite integral estimator. Suppose we have two estimators of a parameter θ, $\hat{\theta}_1$, and $\hat{\theta}_2$. A common way to compare the relative efficiency of the two estimators is to look at their ratio of mean squared error

$$\frac{\mathrm{MSE}(\hat{\theta}_1)}{\mathrm{MSE}(\hat{\theta}_2)},$$

where MSE is defined for an estimator $\hat{\theta}$ by

$$\mathrm{MSE}(\hat{\theta}) = E[(\hat{\theta} - \theta)^2] = \mathrm{bias}^2(\hat{\theta}) + \mathrm{Var}(\hat{\theta}).$$

For two unbiased estimators, this just amounts to comparing the ratio of their variances $\frac{\mathrm{Var}(\hat{\theta}_1)}{\mathrm{Var}(\hat{\theta}_2)}$.