Importance Sampling

Motivation of IS

Let's consider another Monte Carlo technique for evaluating $\theta = \int_A g(x)dx$. Here, the region of integration is the set A, which maybe a bounded interval, an unbounded interval, or some union of intervals.

IS

Let $\phi(x)$ be a density for the random variable X which takes values only in A so that $\int_A \phi(x)dx = 1$. Then

$$\int_{x \in A} g(x)dx = \int_{x \in A} g(x)\frac{\phi(x)}{\phi(x)}dx = \int_{x \in A} \frac{g(x)}{\phi(x)}\phi(x)dx = E_\phi\left(\frac{g(X)}{\phi(X)}\right),$$

so long as $\phi(x) \neq 0$ for any $x \in A$ for which $g(x) \neq 0$, and where E_ϕ denotes the expectation with respect to the density ϕ. This gives a Monte Carlo estimator:

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} \frac{g(X_i)}{\phi(X_i)},$$

where $X_1, X_2, \ldots, X_n$ is a simple random sample from ϕ.

Variance

It can be shown that $\text{Var}(\hat{\theta})$ is minimized when $\phi(x) \; \propto \; |g(x)|$ (see Rubinstein 1981, p.123).

For $g(x) \geq 0, \; \forall x \in A$, $\alpha\phi(x) = g(x)$ where α is some constant of proportionality, it is clear that we have $g(x)/\phi(x) = \alpha$, $\forall \; \{x : \phi(x) > 0\}$ so $E(g(X)/\phi(X)) = \alpha$ and hence the Monte Carlo variance would be zero.

How to choose importance sampling function

In summary, a good importance sampling function $\phi(x)$ should have the following properties:

1. $\phi(x) > 0$ whenever $g(x) \neq 0$
2. $\phi(x)$ should be close to being proportional to $|g(x)|$
3. it should be easy to simulate values from $\phi(x)$

Extension

Note that $g(x)$ is any arbitrary function, so it certainly includes the integrand of a standard expectation. For example, with $X \sim f_X$ we might be interested in $E(g(X))$ for some function r so we could use

$$\theta = E(g(X)) = \int g(x) f_X(x) dx = \int \frac{g(x) f_X(x)}{\phi(x)} \phi(x) = E_\phi\left(\frac{g(x) f_X(x)}{\phi(x)} \right)$$

and then go searching about for a suitable $\phi(x)$ that is close to proportional to $g(x) f_X(x)$.

Example 1:

Consider the function $g(x) = 10\exp(-2|x-5|)$. Suppose that we want to calculate $E(g(X))$, where $X \sim \text{Uniform}(0, 1)$. That is, we want to calculate the integral

$$\int_0^{10} \exp(-2|x-5|)dx = \int_0^{10} 10\exp(-2|x-5|)\frac{1}{10}dx.$$

The true value for this integral is about 1.

Monte Carlo Integration

```r
n=100000
x <- runif(n,0,10)
r <- 10*exp(-2*abs(x-5))
thetahat=mean(r)
se=sd(r)/sqrt(n)
c(thetahat,se*100)
```

```
## [1] 1.000624 0.633330
```

IS

We can re-write the integral as

$$\int_0^{10} \exp(-2|x-5|)dx$$

$$= \int_0^{10} 10\exp(-2|x-5|)\frac{1/10}{\frac{1}{\sqrt{2\pi}}e^{-(x-5)^2/2}}\frac{1}{\sqrt{2\pi}}e^{-(x-5)^2/2}dx,$$

where $g(x) = 10\exp(-2|x-5|)$,$f(x) = \frac{1}{10}$, and $\phi(x) = \frac{1}{\sqrt{2\pi}}e^{-(x-5)^2/2}$.

Conti.

```
n=100000
x=rnorm(n,mean=5,sd=1)
g <-  10*exp(-2*abs(x-5))
f <-  dunif(x, 0, 10)
phi=  dnorm(x, mean=5, sd=1)
gfphi=g*f/phi
thetahat=mean(gfphi)
se=sd(gfphi)/sqrt(n)
c(thetahat,se*100)
```

```
## [1] 0.9992798 0.1910513
```

Example 2:

As another example we will consider estimating the moments of a distribution we are unable to sample from. Let

$$f(x) = \frac{1}{2}e^{-|x|}$$

which is called the double exponential density. The CDF is

$$F(x) = \frac{1}{2}e^{-x}(I(x \leq 0)) + (1 - e^{-x}/2)I(x > 0)$$

which is a piecewise function and difficult to invert. Suppose you want to estimate $E(X^2)$ for this distribution, which is support on $\mathcal{R}$. That is,

$$\int_{-\infty}^{\infty} x^2 \frac{1}{2} e^{-|x|} dx$$

Conti.

We can re-write this as

$$\int_{-\infty}^{\infty} x^2 \frac{\frac{1}{2}e^{-|x|}}{\frac{1}{\sqrt{8\pi}}e^{-x^2/8}} \sqrt{8\pi}e^{-x^2/8} dx$$

$$E\left(X^2 \frac{\frac{1}{2}e^{-|x|}}{\frac{1}{\sqrt{8\pi}}e^{-x^2/8}}\right)$$

by the sample mean of this quantity.

R

```
n=100000
x <- rnorm(n, sd=2)
g=x^2
f=.5 * exp(-abs(x))
phi=dnorm(x, sd=2)
theta.hat=mean(g*f/phi)
theta.hat
```

```
## [1] 1.982394
```

The true value for this integral is 2, so importance sampling has done the job here.

Example 3:

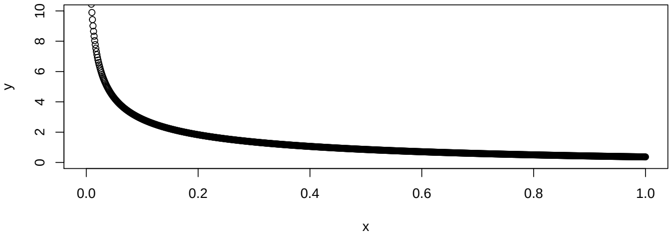
If the function $g(x)$ is unbounded then ordinary Monte Carlo may have a large variance. We may be able to use importance sampling to turn a problem with an unbounded random variable into a problem with a bounded random variable.

Evaluate

$$\theta = \int_0^1 x^{-\alpha} e^{-x} dx, \quad 0 < \alpha < 1$$

R

```r
par(mar=c(15, 10, 10, 3))
x=seq(0,1,,1000)
y=x^(-0.5)*exp(-x)
plot(x,y,ylim=c(0,10),xlim=c(0,1),lty=1)
```



R

```r
## approximation to theta
integrate(function(x) x^(-0.5)*exp(-x),0,1)
```

```
## 1.493648 with absolute error < 9.4e-06
```

```r
#if a=0.5
#MC
a=0.5
n=10000
x=runif(n,0,1)
g=x^(-a)*exp(-x)
thetahat=mean(g)
se=sd(g)/sqrt(n)
c(thetahat,se*100)
```

```
## [1] 1.528011 2.725424
```

IS

??

R

```r
#compared to IS
a=0.5
n=10000
u=runif(n,0,1)
x=u^(1/(1-a))
g=x^(-a)*exp(-x)
f=dunif(x,0,1)
phi=(1-a)*(x^(-a))
thetahat=mean(g*f/phi)
se=sd(g*f/phi)/sqrt(n)
c(thetahat,se*100)
```

```
## [1] 1.4970268 0.4025747
```

Example 4:

Let's consider estimating the parameter

$$\theta = \int_0^1 \frac{e^{-x}}{1+x^2} dx.$$

We can think of this as

$$\theta = E(g(X)) = \int_{-\infty}^{\infty} g(x) f_X(x) dx$$

where $g(x) = e^{-x}/(1+x^2)$ and $f(x)$ is the uniform density $f(x) = 1$ for $x \in (0,1)$ and 0 otherwise.

A few different possible importance sampling functions

1. $\phi_1(x) = 1, \ 0 < x < 1.$
2. $\phi_2(x) = e^{-x}, \ 0 < x < \infty.$
3. $\phi_3(x) = \frac{e^{-x}}{1-e^{-1}}, \ 0 < x < 1.$
4. $\phi_4(x) = \frac{4}{(1+x^2)\pi}, \ 0 < x < 1.$

R

```
n=10000
se=numeric(4)
theta.hat=numeric(4)
# try phi_1
x=runif(n)
g=exp(-x)/(1+x^2)
f=(x>0)*(x<1)
phi=1
gfphi=g*f/phi
theta.hat[1]=mean(gfphi)
se[1]=sd(gfphi)/sqrt(n)
# try phi_2
x=rexp(n,1)
g=exp(-x)/(1+x^2)
f=(x>0)*(x<1)
phi=exp(-x)
gfphi=g*f/phi
theta.hat[2]=mean(gfphi)
```

R

```
#try phi_3
u=runif(n)
x=-log(-u*(exp(1)-1)+exp(1))+1
g=exp(-x)/(1+x^2)
f=(x>0)*(x<1)
phi=exp(-x)/(1-exp(-1))
gfphi=g*f/phi
theta.hat[3]=mean(gfphi)
se[3]=sd(gfphi)/sqrt(n)
#try phi_4
u=runif(n)
x=tan(pi*u/4)
g=exp(-x)/(1+x^2)
f=(x>0)*(x<1)
phi=4/((1+x^2)*pi)
gfphi=g*f/phi
theta.hat[4]=mean(gfphi)
se[4]=sd(gfphi)/sqrt(n)
```

R

```r
## Look at theta.hat and 100*se
round(rbind(theta.hat,100*se),3)
```

```
##            [,1]  [,2]  [,3]  [,4]
## theta.hat 0.527 0.520 0.525 0.524
##           0.246 0.421 0.096 0.141
```

Not surprisingly, ϕ_3 and ϕ_4 resulted in the smallest standard errors.