Chapter 5

Multiple Random Variables

Monte Carlo Methods

Outline

Monte Carlo Integration

Monte Carlo methods is a collection of computational algorithms that use stochastic simulations to approximate solutions to questions that are very difficult to solve analytically.

This approach has seen widespread use in fields as diverse as statistical physics, astronomy, population genetics, protein chemistry, and finance.

Monte Carlo Integration

Let $X_1, X_2, \ldots$ be independent random variables uniformly distributed on the interval $[a, b]$ and write f_X for their common density.

Then, by the law of large numbers, for n large we have that

$$\overline{g(X)}_n = \frac{1}{n}\sum_{i=1}^{n} g(X_i) \approx Eg(X_1) = \int_a^b g(x)f_X(x)\ dx = \frac{1}{b-a}\int_a^b g(x)\ dx.$$

$$\text{Thus,}\quad Ig = \int_a^b g(x)\ dx \approx (b-a)\overline{g(X)}_n = \widehat{Ig}_n.$$

Recall that in calculus, we defined the average of g to be

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

We can now interpret this integral as an expected value.

Monte Carlo Integration

Thus, Monte Carlo integration leads to a procedure for estimating integrals.

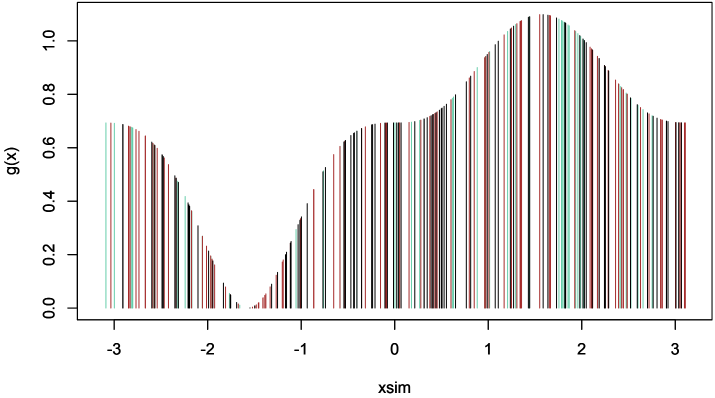
- Simulate uniform random variables $X_1, X_2, \ldots, X_n$ on the interval $[a, b]$,
- Evaluate $g(X_1), g(X_2), \ldots, g(X_n)$.
- Average this values and multiply by $b - a$ to estimate the integral.

Let $g(x) = \ln(2 + \sin^3(x))$ (`g<-function(x) log(2+sin(x)^3)`) for $x \in [-\pi, \pi]$, to find $\int_{-\pi}^{\pi} g(x)\, dx$.

The three steps above become the following R code.

```
> sim<-runif(250,-pi,pi)
> gsim<-g(sim)
> 2*pi*mean(gsim)
[1] 4.155533
```

Monte Carlo Integration



Monte Carlo integration of $g(x) = \ln(2 + \sin^3(x))$ Simulations of 50, 150, and 250 values are shown.

Monte Carlo Integration

The variance $\text{Var}(\widehat{Ig_n}) = \text{Var}((b-a)\overline{g(X)}_n) = \frac{(b-a)^2}{n}\text{Var}(g(X_1))$ where

$$\sigma^2 = \text{Var}(g(X_1)) = E(g(X_1) - \mu_{g(X_1)})^2 = \frac{1}{b-a}\int_a^b (g(x) - \mu_{g(X_1)})^2 \, dx.$$

Typically this integral is more difficult to estimate than $\int_a^b g(x)\, dx$. However we can use the simulation to estimate the standard deviation of $\widehat{Ig}$.

```
> sd(gsim)*2*pi/sqrt(250)
[1] 0.126093
```

R does integration numerically using the `integrate` command.

```
 >(Ig<-integrate(g,-pi,pi)); (mean(gsim)*2*pi-Ig$value)/(sd(gsim)*2*pi/sqrt(250))
 4.083949 with absolute error < 1.7e-06
[1] 0.5677076
```

Thus, the estimate is off by approximately 0.567 standard deviations.

Monte Carlo Integration

With only a small change in the algorithm, we can also use this to evaluate multivariate integrals. For example, in three dimensions, the integral

$$\int_{a_1}^{b_1}\int_{a_2}^{b_2}\int_{a_3}^{b_3} g(x,y,z)\,dz\,dy\,dx \approx (b_1-a_1)(b_2-a_2)(b_3-a_3)\frac{1}{n}\sum_{i=1}^{n} g(X_i,Y_i,Z_i).$$

where $X_i \sim U(a_1,b_1)$, $Y_i \sim U(a_2,b_2)$, and $Z_i \sim U(a_3,b_3)$

Example. To estimate

$$\int_{-2}^{2}\int_{1/2}^{1}\int_{0}^{1} \frac{e^{-x^2/2y}}{x^2z+1}\,dz\,dy\,dx$$

```
> x<-runif(250,-2,2);y<-runif(250,1/2,1);z<-runif(250)
> g<-exp(-x^2/(2*y))/(x^2*z+2)
> 4*0.5*1*mean(g)
[1] 0.4550264
```

Monte Carlo Integration

Monte Carlo integration uses the averages of a simulated random sample and consequently, its value is itself random. To obtain a sense of the distribution of the approximations to the integral

$$\int_0^8 \frac{1+e^{-x/2}}{\sqrt[3]{x}}\,dx,$$

we perform 1000 simulations using 250 uniform random variables.

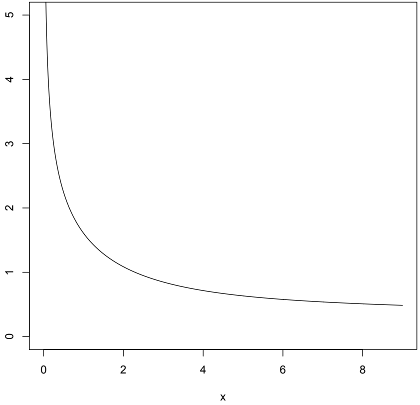
```
> Ighat<-numeric(1000)
> for (i in 1:1000){x<-runif(250,0,8);Ighat[i]<-8*mean((1+exp(-x/2))/x^(1/3))}

> mean(Ighat)
[1] 8.120468
> sd(Ighat)
[1] 0.4715746
```

Monte Carlo Integration

To reduce the standard deviation, we can

- Increase the size of the simulation.
  - An increase from 250 to 1000 decreases the variance by a factor of 4 and thus the standard deviation by a factor of 2.
- Concentrate the values of x where the function g changes rapidly.
  - Such a strategy is called importance sampling.



The graph of $g(x)$

Importance Sampling

Goal. Reduce the standard deviation in the approximation of the integral

$$\int_a^b g(x)dx$$

Write $g(x) = w(x)f_X(x)$ where

- $f_X(x)$ is a density function that captures the change in $g(x)$ *and* has an easy to determine distribution function $F_X(x)$.
  - f_X is called the importance sampling function or the proposal density.
  - w is called the importance sampling weight.

Now, simulate $X_1, X_2, \ldots, X_n$ independent random variables with common density f_X. Then by the law of large numbers.

$$\frac{1}{n}\sum_{i=1}^{n} w(X_i) \approx Ew(X_1) = \int_a^b w(x)f_X(x)\ dx = \int_a^b g(x)\ dx.$$

Importance Sampling

$$\int_0^8 \frac{1+e^{-x/2}}{\sqrt[3]{x}}\,dx = \int_0^8 (1+e^{-x/2})\frac{1}{\sqrt[3]{x}}\,dx$$

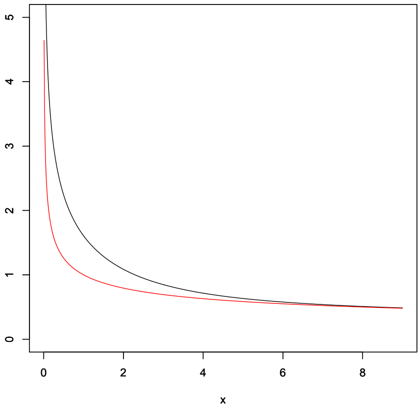
The distribution function

$$F_x(x) = c\int_0^x \frac{1}{\sqrt[3]{t}}\,dt = \frac{3c}{2}t^{2/3}\Big|_0^x = \frac{3c}{2}x^{2/3}$$

Now, $1 = F_X(8) = \frac{3c}{2}8^{2/3} = \frac{3c}{2}4 = 6c.$

So, $c = 1/6$ and $f_X(x) = \frac{1}{6\sqrt[3]{x}}$ is a density

and distribution function $F_X(x) = \frac{1}{4}x^{2/3}$ on $[0, 8]$.



The graph of $g(x)$ (black) and the proposal density $f_X(x)$ (red)

Importance Sampling

$$\int_a^b g(x)\ dx = \int_a^b w(x) f_X(x)\ dx$$
$$\int_0^8 \frac{1+e^{-x/2}}{\sqrt[3]{x}}\ dx = \int_0^8 6(1+e^{-x/2})\frac{1}{6}\frac{1}{\sqrt[3]{x}}\ dx$$

So,

the density $f_X(x) = \dfrac{1}{6\sqrt[3]{x}}$ and the weight function $w(x) = 6(1+e^{x/2})$.

To simulate the X_i we use the probability transform.

$$u = F_X(x) = \frac{1}{4}x^{2/3}. \quad \text{Thus, } x = (4u)^{3/2}.$$

Importance Sampling

For the probability transform in R we enter `u<-runif(250);x<-(4*u)^3/2`. Thus, for 1000 importance sampling approximations, we find

```
> ISg<-numeric(1000)
> for (i in 1:1000){u<-runif(250);x<-(4*u)^(3/2);ISg[i]<-mean(6*(1+exp(-x/2)))}

> mean(ISg)
[1] 8.132918
> sd(ISg)
[1] 0.1164385
```

Compare this with simple Monte Carlo.

```
> sd(Ig)
[1] 0.4715746
```

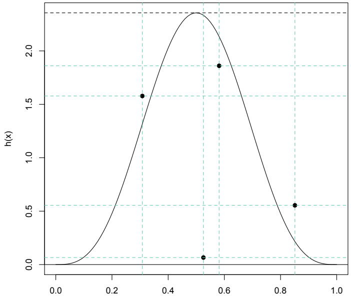
The standard deviation is reduced by a factor of ~ 4 and thus, we would need to increase the number of simulations by a factor of ~ 16 for simple Monte Carlo to meet the same standard deviation.

Rejection Sampling

We now apply Monte Carlo ideas to estimate probabilities.

Let Y be a continuous random variable with density f_Y on an interval $[0,1]$ and choose c greater than the maximum value $f_Y(y)$. Here is the rejection sampling algorithm

1. Choose a point $(U,V) \in [0,1] \times [0,c]$ uniformly on the rectangle.
2. If the value is below the graph of f_Y then choose $Y = y$, otherwise reject the value and try again.



Four choices of uniformly distributed on the rectangle. Two above the curve f_Y are rejected. The two below are simulation from the distribution of Y

Rejection Sampling

Exercise. $P\{V \le y | U < f_Y(V)\} = P\{Y \le y\}$.

$$P\{V \le y | U < f_Y(V)\} = \frac{P\{V \le y, U < f_Y(V)\}}{P\{U < f_Y(V)\}}.$$

and

$$P\{V \le y, U < f_Y(V)\} = \int_0^y \int_0^{f_Y(v)} \frac{1}{c}\, du\, dv = \frac{1}{c} \int_0^y f_Y(v) dv = \frac{1}{c} P\{Y \le y\}.$$

To find $P\{U < f_Y(V)\}$, take $y = 1$ in the previous computation.

$$P\{U < f_Y(V)\} = \frac{1}{c} P\{Y \le 1\} = \frac{1}{c}.$$

Taken together, we find that $P\{V \le y | U < f_Y(V)\} = P\{Y \le y\}$.

Note that the probability of accepting the choice is $1/c$.

Rejection Sampling

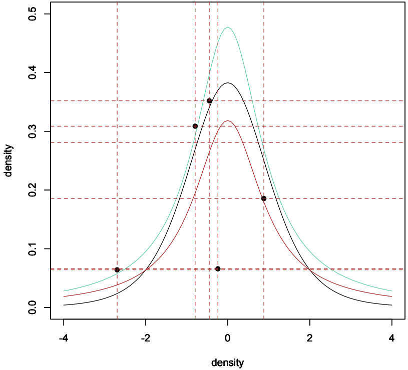
To improve on this:

- Choose a second density f_V that looks like f_Y with a distribution F_V that is easy to determine.
- Use the probability transform F_V on Y.
- Then, the transformed Y is close to uniform resulting in samples are often accepted.

The algorithm to generate a random sample is as follows:

1. Set $c \geq \sup_y \frac{f_Y(y)}{f_V(y)}$.
2. Generate two independent samples - $U \sim U[0,1]$ and V with density f_V.
3. If $U < f_Y(V)/(cf_V(v))$, set $Y = V$. Otherwise, return to step 1.

Rejection Sampling



The goal is to create a sample from Y with the density shown. Find a second random variable V whose distribution is easy to find. Multiply the density of V by a constant c so that it lies above the density of Y. Sample according the density of V and cU. Plot the point (cU, V). Accept those samples that are found below the density of Y.

Rejection Sampling

To show that these sample have the same distribution of Y, we compute

$$P\{V \le y | U < f_Y(V)/(cf_V(V))\} = \frac{P\{V \le y, U < f_Y(V)/(cf_V(V)\}}{P\{U < f_Y(V)/(cf_V(V))\}}.$$

The numerator $P\{V \le y, U < f_Y(V)/(cf_V(v))\}$

$$\begin{aligned} &= \int_{-\infty}^{y} \int_{-\infty}^{f_Y(v)/(cf_V(v))} f_V(v) du\, dv = \int_{-\infty}^{y} \frac{f_Y(v)}{cf_V(v)} f_V(v) dv \\ &= \frac{1}{c} \int_{-\infty}^{y} f_Y(v) du\, dv = \frac{1}{c} P\{Y \le y\} \end{aligned}$$

The value of the denominator, $1/c$, can be found taking $y \to \infty$ in the previous computation. Therefore,

$$P\{V \le y | U < f_Y(V)/(cf_V(V))\} = P\{Y \le y\}.$$

Rejection Sampling

Example. We will find samples for Y, the t distribution with 6 degrees of freedom. We will use the Cauchy distribution with density and distribution

$$f_V(v) = \frac{1}{\pi}\frac{1}{1+v^2}, \quad F_V(v) = \frac{1}{\pi}\arctan(v) + \frac{1}{2}, \quad F_V^{-1}(q) = \tan\left(\pi\left(q - \frac{1}{2}\right)\right)$$

The choice of $c = 3/2$ places $cf_V(v)$ above $f_Y(v)$.

```
> f<-function (v) 1/(pi*(1+v^2))
> Q<-function (q) tan(pi*(q-1/2))
> v<-Q(runif(200)) #samples for the Cauchy distribution
> u<-runif(200)
> sum(u<dt(v,6)/(c*f(v))) #number of accepted samples
[1] 141
> va<-v[u<dt(v,6)/(c*f(v))] #values of accepted samples
```

Rejection Sampling

```
> plot(ecdf(va),xlim=c(-4,4),ylim=c(0,1),col="aquamarine3")
> par(new=TRUE)
> curve(pt(x,6),-4,4,ylim=c(0,1))
```