Hit and Miss Method

$$I = \int_a^b g(x)\,dx$$

Area of region S under $g(x)$ curve.

$$f_{\hat{\mathbf{x}}\hat{\mathbf{y}}}(x,y) = \begin{cases} \frac{1}{c(b-a)} & \text{if } (x,y) \in \Omega \\ 0 & \text{if } (x,y) \notin \Omega \end{cases}$$

Probability p that (x,y) lies in S is:

$$p = \int_\Omega f_{\hat{\mathbf{x}}\hat{\mathbf{y}}}(x,y)\,dx\,dy = \frac{1}{c(b-a)} \int_S dx\,dy = \frac{I}{c(b-a)}$$

Assuming $0 \leq g(x) \leq c$

Generate randomly N point $(x_1, y_1), (x_2, y_2), ..., (x_N, y_N)$

$\hat{N}_A$ number of points in S

$\hat{N}_A$ follows the binomial distribution

$$\hat{\theta}_1 = \frac{c(b-a)}{N} \hat{N}_A$$

the mean value equals the integral:

$$\langle \hat{\theta}_1 \rangle = \frac{c(b-a)}{N} \langle \hat{N}_A \rangle = c(b-a)p = I$$

$\hat{\theta}_1$ is an *unbiased estimator* of I.

$$\sigma[\hat{\theta}_1] = \frac{c(b-a)}{N} \sqrt{Np(1-p)} = c(b-a)\sqrt{\frac{p(1-p)}{N}}$$

Replace p by its sample value: $\hat{p} = \hat{N}_A / N$.

$$I = \hat{\theta}_1 \pm \sigma[\hat{\theta}_1] = c(b-a)\hat{p} \pm c(b-a)\sqrt{\frac{\hat{p}(1-\hat{p})}{N}}$$

Relative error:

$$\frac{\sigma[\hat{\theta}_1]}{< \hat{\theta}_1 >} = \sqrt{\frac{1-p}{pN}}$$

decreases for large p.

Take $c = max(g(x))$.

```
      subroutine mc1(g,a,b,c,n,r,s)
      external g
      na=0
      do 1 i=1,n
      u=ran_u()
      v=ran_u()
      if (g(a+(b-a)*u).gt.c*v) na=na+1
1     continue
      p=real(na)/n
      r=(b-a)*c*p
      s=sqrt(p*(1.-p)/n)*c*(b-a)
      return
      end
```

Sampling methods: Uniform Sampling

$$I = \int_a^b g(x)\, dx$$

$$I = \int_a^b (b-a) g(x) \frac{1}{b-a}\, dx \equiv (b-a) \int g(x) f_{\hat{\mathbf{x}}}(x)\, dx$$

$$I = (b-a) E_{\mathbf{x}}[g(x)]$$

$\hat{\mathbf{x}}$ es $\hat{U}(a,b)$:

$$f_{\hat{\mathbf{x}}}(x) = \begin{cases} \frac{1}{b-a} & \text{if } a \le x \le b \\ 0 & \text{otherwise} \end{cases}$$

Sample mean:

$$\hat{\mu}_N[g(x)] = \frac{1}{N} \sum_{i=1}^{N} g(x_i)$$

r.v. $\hat{\theta}_2 = (b-a)\hat{\mu}_N$ from sampling:

$$\hat{\theta}_2 = (b-a) \frac{1}{N} \sum_{i=1}^{N} g(x_i)$$

$x_i,\ i = 1, 2, ..., N$ uniformly distributed in the interval $[a, b]$.

$$\langle \theta_2 \rangle = I$$

$$I = \hat{\theta}_2 \pm \sigma[\hat{\theta}_2] = \hat{\theta}_2 \pm \frac{(b-a)\sigma[g(x)]}{\sqrt{N}}$$

$$\hat{\sigma}_N^2 = \left[\frac{1}{N} \sum_{i=1}^{N} g(x_i)^2 - \left(\frac{1}{N} \sum_{i=1}^{N} g(x_i) \right)^2 \right]$$

```
      subroutine mc2(g,a,b,n,r,s)
      external g
      r1=0.
      s1=0.
      do 1 i=1,n
      u=ran_u()
      g0=g(a+(b-a)*u)
      r1=r1+g0
      s1=s1+g0*g0
1     continue
      r1=r1/n
      s1=sqrt((s1/n-r1*r1)/n)
      r=(b-a)*r1
      s=(b-a)*s1
      return
      end
```

Method of Importance Sampling

$$I = \int_a^b g(x) f_{\hat{\mathbf{x}}}(x)\, dx$$

$$I = E_{\mathbf{x}}[g(x)]$$

$\hat{\mathbf{x}}$ distributed according $f_{\hat{\mathbf{x}}}(x)$. Sample mean:

$$\hat{\mu}_N[g(x)] = \frac{1}{N} \sum_{i=1}^{N} g(x_i)$$

$x_i,\ i = 1, 2, ..., N$ are values of the r.v. $\hat{\mathbf{x}}$ distributed according to $f_{\hat{\mathbf{x}}}(x)$

Unbiased estimator $\langle \hat{\mu}_N[g(x)] \rangle = I$.

$$I = \hat{\mu}_N[g(x)] \pm \sigma[\hat{\mu}_N[g(x)]]$$

$$I = \hat{\mu}_N[g(x)] \pm \frac{\sigma[g(x)]}{\sqrt{N}}$$

$$\sigma^2[g(x)] = \int f_{\hat{\mathbf{x}}} g(x)^2\, dx - \left(\int f_{\hat{\mathbf{x}}} g(x)\, dx\right)^2$$

Use sample variance $\hat{\sigma}_N^2$:

$$\hat{\sigma}_N^2 = \left[\frac{1}{N} \sum_{i=1}^{N} g(x_i)^2 - \left(\frac{1}{N} \sum_{i=1}^{N} g(x_i)\right)^2\right]$$

How do we generate $\hat{\mathbf{x}}_i$?

If u_i is a $\hat{U}(0,1)$ variable, then:

$$x_i = F_{\hat{\mathbf{x}}}^{-1}(u_i)$$

```
      subroutine mc3(g,n,r,s)
      external g
      r=0.
      s=0.
      do 1 i=1,n
      x=ran_f()
      g0=g(x)
      r=r+g0
      s=s+g0*g0
1     continue
      r=r/n
      s=sqrt((s/n-r*r)/n)
      return
      end
```

Efficiency of an integration method

Computer time needed to compute a given integral with a given error ϵ.

So far:

$$\epsilon = \frac{\sigma}{\sqrt{N}}$$

t is the computer time needed to add a contribution to the estimator

total computer time $Nt \propto t\sigma^2$

Relative efficiency of method 1 and 2:

$$e_{12} = \frac{t_1\sigma_1^2}{t_2\sigma_2^2}$$

Uniform sampling is more efficient that hit and miss method.

Hit and miss:

$$\sigma_1^2 = c(b-a)I - I^2$$

Uniform sampling:

$$\sigma_2^2 = (b-a)\int g(x)^2\,dx - I^2$$

$$\sigma_1^2 - \sigma_2^2 = (b-a)\left[cI - \int g(x)^2\,dx\right]$$

The condition $0 \leq g(x) \leq c$ implies:

$$\int g(x)^2 \, dx \leq c \int g(x) \, dx = cI$$

and $\sigma_1 \geq \sigma_2$. Furthermore, in general, $t_1 > t_2$, and it follows $e_{12} > 1$.

Advantages and disadvantages of the Monte-Carlo integration

Very singular function

$n-$dimensional integral, with high n

$$\int_\Omega g(x_1, x_2, \dots, x_n) f_{\hat{\mathbf{x}}_1 \dots \hat{\mathbf{x}}_n}(x_1, x_2, \dots, x_n) \, dx_1 \, dx_2 \dots \, dx_n \approx \frac{1}{N} \sum_{i=1}^{N} g(x_1, x_2, \dots, x_n)$$

$(x_1, x_2, \dots, x_n)$ is a random vector distributed according the probability density function $f_{\hat{\mathbf{x}}_1 \dots \hat{\mathbf{x}}_n}(x_1, x_2, \dots, x_n)$.

Everything is valid for sums:

$$\sum_i g_i = \int g(x) f_{\hat{\mathbf{x}}}(x) \, dx$$

by considering a discrete random variable

$$f_{\hat{\mathbf{x}}}(x) = \sum_i p_i \delta(x - x_i)$$

and

$$g_i = g(x_i)$$