Monte Carlo Integration with R

General idea:

We wish to integrate,

```
            I(f)=Int_{a}^{b} f(x) dx
```

1. Choose a pdf g(x) on [a,b].
2. Generate data X_1,X_2,...,X_n from g(x).
3. Estimate I(f) by:

```
             1                  f(X_i)
            --- Sum_{i=1}^{n} --------
             n                  g(X_i)
```

```
1. I(f)=Int_{0}^{1}(phi(x)), phi(x) standard normal pdf.
---------------------------------------------------------
#Exact Answer:
> pnorm(1)-pnorm(0)
[1] 0.3413447
```

Monte Carlo: Since [a,b]=[0,1], we can use g(x) ~ Unif[0,1]

```
Integral <- function(n){
X <- runif(n)                           #g(x) ~ Unif[0,1]
Y <- exp(-X^2/2)/sqrt(2*pi)
Int <- sum(Y)/n
Error <- Int-(pnorm(1)-pnorm(0))
list(Int,Error)}

> Integral(1000)  #n=1000
[[1]]             -------
```

```
[1] 0.3406526

[[2]]
[1] -0.0006921899



> Integral(100000) #n=100000
[[1]]              ----------
[1] 0.3413795

[[2]]
[1] 3.474216e-05



2. I(f)=Int_{a}^{b}(phi(x)), phi(x) standard normal pdf.
--------------------------------------------------------

Integral <- function(n,a,b){
X <- runif(n,a,b)
Y <- (exp(-X^2/2)/sqrt(2*pi))/(1/(b-a))   #g(x) ~ Unif[a,b]
Int <- sum(Y)/n
Error <- Int-(pnorm(b)-pnorm(a))
list("Int"=Int,"Error"=Error)}


> Integral(1000,-1,1)  #n=1000
$Int                  ---------
[1] 0.6859197

$Error
[1] 0.003230224



> Integral(10000,-2,2) #n=10,000
$Int                  -------------
```

```
[1] 0.9598045

$Error
[1] 0.005304792



> Integral(1000000,-2,2) #n=1,000,000
$Int                      ------------
[1] 0.9547782

$Error
[1] 0.0002784936


$Int
[1] 0.9546407

$Error
[1] 0.0001409268
```

```
3. I(f)=Int_{a}^{b}(phi(x)), phi(x) N(0,1) pdf. Now also supply 95% CI
   for I(f).
------------------------------------------------------------------------

CLT: Asymptotically,


Ihat(f) ~ N(I(f), Var)

      1                f^2(x)
Var=  ---[ Int_{a}^{b} ------- dx - I^2(f)]
      n                 g(x)
```

```
Integral <- function(n,a,b){
X <- runif(n,a,b)
Y <- (exp(-X^2/2)/sqrt(2*pi))/(1/(b-a))      #g(x) ~ Unif[a,b]
Int <- sum(Y)/n
Error <- Int-(pnorm(b)-pnorm(a))
X <- runif(n,a,b)
YY <- ((exp(-X^2/2)/sqrt(2*pi))/(1/(b-a)))^2
SE <- sqrt((sum(YY)/n-Int^2)/n)
CI <- c(Int-1.96*SE,Int+1.96*SE)
list("Int"=Int,"Error"=Error, "SE"=SE, "CI"=CI)}


> Integral(1000,-1.96,1.96)  #n=1000
$Int:
[1] 0.9439135

$Error:
[1] -0.006090713

$SE:
[1] 0.01499378

$CI:
[1] 0.9145257 0.9733013  #OK. True Prob.=pnorm(b)-pnorm(a)=0.9500042


> Integral(10000,-1.96,1.96)  #n=10000
$Int:
[1] 0.9516164

$Error:
[1] 0.001612226

$SE:
[1] 0.004461732
```

```
$CI:
[1] 0.9428714 0.9603614   #OK. True Prob.=pnorm(b)-pnorm(a)=0.9500042
```

```
4. I(f)=Int_{a}^{b}(f(x)), arbitrary f. Also 95% CI for I(f).
----------------------------------------------------------------

Note: (a,b) is an arbitrary interval, finite or infinite
      Assume finite. Then can use g ~ Unif[a,b].
```

```
f <- function(x){x}              #Function to be integrated over [a,b]
                                 #is f(x)=x.

Integral <- function(n,a,b,h=f){ #Large n gives better approximation.
X <- runif(n,a,b)                #Use Unif[a,b] as reference pdf.
Y <- (h(X))/(1/(b-a))
Int <- sum(Y)/n
X <- runif(n,a,b)
YY <- (h(X)/(1/(b-a)))^2
SE <- sqrt((sum(YY)/n-Int^2)/n)
CI <- c(Int-1.96*SE,Int+1.96*SE)
list("Int"=Int,"SE"=SE, "CI"=CI)}



> Integral(1000,0,1,f)
$Int:
[1] 0.5080465

$SE:
[1] 0.009077023

$CI:
[1] 0.4902555 0.5258374         #OK. Contains 0.5.
```

```
> f1 <- function(x){exp(-x)}
>
> Integral(1000,0,10,f1)
$Int
[1] 1.094521

$SE
[1] 0.06225339

$CI
[1] 0.9725047 1.2165380   #OK. Contains approx 1.


----------

         @@@@@@#Simpler: Use f directly (before used h=f)#@@@@@


f <- function(x){x}            #Function to be integrated over [a,b].

Integral <- function(n,a,b,f){ #Large n gives better approximation.
X <- runif(n,a,b)              #Use Unif[a,b] as reference pdf.
Y <- (f(X))/(1/(b-a))
Int <- sum(Y)/n                #Int. Approx.
X <- runif(n,a,b)              #SE of Int and 95% CI
YY <- (f(X)/(1/(b-a)))^2
SE <- sqrt((sum(YY)/n-Int^2)/n)
CI <- c(Int-1.96*SE,Int+1.96*SE)
list("Int"=Int,"SE"=SE, "95% CI"=CI)}


> Integral(1000,0,1,f)
$Int:
[1] 0.4957296

$SE:
[1] 0.009197327
```

```
$"95% CI":
[1] 0.4777029 0.5137564   #OK. Contains 0.5

----------------------------

> f <- function(x){x^2}
> Integral(1000,0,1,f)
$Int:
[1] 0.3307101

$SE:
[1] 0.009225372

$"95% CI":
[1] 0.3126284 0.3487919       #OK. Contains 0.3333333


--------------------
f <- function(x){4*sqrt(1-x^2)}
> Integral(1000,0,1,f)
$Int:
[1] 3.152334

$SE:
[1] 0.0278814

$"95% CI":
[1] 3.097686 3.206981
-----------------------------


5. Comparison With Numerical Integration
--------------------------------------------

> f <- function(x){ exp(-x^2/2)/sqrt(2*pi)}  #MC=Monte Carlo
> Integral(1000,-1.96,1.96,f)
$Int:
[1] 0.9489701
```

```
$SE:
[1] 0.01447981

$"95% CI":
[1] 0.9205897 0.9773506



> f <- function(x){ exp(-x^2/2)/sqrt(2*pi)}   #NI=Numerical Integration
> integrate(f,-1.96,1.96)
0.9500042 with absolute error < 1.0e-11

----------------

> f <- function(x){ exp(-x^2/2)/sqrt(2*pi)}    #MC
> Integral(1000,-3,3,f)                        #n=1000
$Int
[1] 1.031649

$SE
[1] 0.02547308

$‘95% CI‘
[1] 0.9817218 1.0815763   ##True int is pnorm(3)-pnorm(-3)=0.9973002




> Integral(100000,-3,3,f)                      #MC
$Int: 1.001041
                                #n=100000
$SE
[1] 0.00263771

$‘95% CI‘
[1] 0.995871 1.006211   ##True int is pnorm(3)-pnorm(-3)=0.9973002
```

```
> f <- function(x){ exp(-x^2/2)/sqrt(2*pi)}    #NI
> integrate(f,-3,3)
0.9973002 with absolute error < 9.3e-07



NOTE:
> pnorm(3)-pnorm(-3)   #<--------"True Integral" is from NI!!!
[1] 0.9973002

---------------------------------

f <- function(x) sin(x)           #MC
> Integral(1000,0,1,sin)
$Int:
[1] 0.4567833

$SE:
[1] 0.008425967

$"95% CI":
[1] 0.4402684 0.4732982



> f <- function(x) sin(x)           #NI
> integrate(f,0,1)
0.4596977 with absolute error < 5.1e-15


------------------------------   Some More MC Examples -----
Simpler:
> Integral(1000,0,pi,sin)
$Int
[1] 1.997878
```

```
$SE
[1] 0.03293127

$‘95% CI‘
[1] 1.933333 2.062423



> Integral(1000,0,pi,sin)
$Int:
[1] 2.055601

$SE:
[1] 0.02115842

$"95% CI":
[1] 2.014130 2.097071
------------------------------
> Integral(1000,0,pi,cos)
$Int:
[1] 0.07335976

$SE:
[1] 0.07020089

$"95% CI":
[1] -0.06423399  0.21095351
--------------------------------
> f <- function(x){exp(-x)}
> Integral(1000,0,2,f)
$Int:
[1] 0.8762992

$SE:
[1] 0.01385853

$"95% CI":
[1] 0.8491365 0.9034620
```

```
---------------------------

6. Discontinuous Functions
---------------------------------------

f <- function(x){ifelse( (x < 1 ),1,2)} #Discontinuous function.
x <- seq(0,2,0.01)
plot(x,f(x), type="l")

> Integral(1000,0,2,f)     #MC
$Int:
[1] 3.012

$SE:
[1] 0.03486913

$"95% CI":
[1] 2.943657 3.080343  ##True I(f)=3


Integral(1000000,0,2,f)  #MC
$Int
[1] 3.000164

$SE
[1] 0.0009988894

$‘95% CI‘
[1] 2.998206 3.002122


> integrate(f,0,2)               #NI
3 with absolute error < 3.3e-14



                                  OR
f <- function(x){ifelse( (x < 1 ) |  (x > 2),1,2)} #Discontinuous fun.
x <- seq(0,3,0.01)
```

```
plot(x,f(x), type="l")



> Integral(100000,0,3,f)
$Int:
[1] 3.99888                  #MC n=100000

$SE:
[1] 0.004452425

$"95% CI":
[1] 3.990153 4.007607      ##True I(f)=4


> Integral(500000,0,3,f)    #MC n=500,000
$Int:
[1] 4.001754

$SE:
[1] 0.001990105

$"95% CI":
[1] 3.997853 4.005655




> integrate(f,0,3)             #NI
4 with absolute error < 4.4e-15
```

```
7. Special Case:  I(f)=Int_{a}^{b}(f(x)), f(x)=k(x)*g(x)
                                     g(x) pdf. on (a,b).
--------------------------------------------------------
```

Consider:
f(x)=x^2*exp(-x), k(x)=x^2, g(x)=exp(-x)
I(f)=Int_{0}^{infinity}(f(x))=Gamma(3)=2

```
k <- function(x){x^2}    #Function to be integrated w.r.t. g(x)=exp(-x)

Integral <- function(n,k){
U <- runif(n)
X <- -log(1-U)             #g ~ Exponential(1)
Int <- sum(k(X))/n         #Int. Approx.
Int}

> Integral(1000,k)
[1] 2.105587
[1] 2.068958
[1] 2.045131
> Integral(10000,k)
[1] 2.00496
-------------------------------------------
```

8. Importance Sampling

As in 7, suppose we want to integrate (i.e. get E[k(X)], X~g)

I(f)= E(k(X))=Int_{a}^{b} k(x)*g(x),

where g(x) is an inconvenient pdf. Use another reference pdf h(x):

Int k(x)*g(x) = Int [k(x)*g(x)/h(x)]*h(x)

and sample from h(x).

##NOTE: For the method to work, the tail of h(x) must be heavier than that of g(x).

Take: h(x)=3*exp(-3*x), which is exponential(3).

```
      k <- function(x){x^2}, g(x)=exp(-x), and E(X^2)=2

Integral <- function(n,k){
X <- rexp(n,3)   #X ~ exponential(3)
Int <- sum(k(X)*exp(-X)/(3*exp(-3*X)))/n
Int}

> Integral(1000,k)
[1] 0.7478736        #True 2
> Integral(1000,k)
[1] 2.429151
> Integral(1000,k)
[1] 7.381861         #Evidently, we pay a price for change of measure if
> Integral(1000,k)    there is a tails problem!!! Tail of h(x) is too thin
[1] 1.640584          relative to g(x).
> Integral(1000,k)
[1] 1.210625
> Integral(1000,k)
[1] 1.98542

## To show h(x)=3*exp(-3*x) has thinner tailes than exp(-x):
x <- seq(0,3,0.01)
plot(x,exp(-x),type="l")
lines(x,3*exp(-3*x),type="p")
```

Now change: Take g(x)=3*exp(-3*x), and h(x)=exp(-x)
Then h(x) has a heavier tail than g(x).

Wish to integrate k(x)*g(x) on (0,ininity).

```
k <- function(x){x^2}

Integral <- function(n,k){
X <- rexp(n)   #X ~ exponential(1)
Int <- sum(k(X)*3*exp(-3*X)/(exp(-X)))/n
Int}

> 2/9 = TRUE E(X^2) = 2/3^2
[1] 0.2222222  #<--- Correct answer.

> Integral(1000,k)
[1] 0.2174634
[1] 0.2209648         #OK here!
[1] 0.228226
[1] 0.2236857
```

9. Comparisons Between Reference Distributions on [0,Infinity) in Monte Carlo Integration as in part 1.

```
------------------------------------------------------------------
```

```
f <- function(x){exp(-x)}      #To be integrated over [0,Infinity).
                                Integral=1.
```

Reference pdf is Gamma(shape,scale).

Must be careful. Get different approximations for different shapes and scales. Some OK some not.

```
Integral <- function(n,f,shape,scale){
s <- shape; lam <- scale
X <- rgamma(n,s)/lam
Y <- (lam^s)*(X^(s-1))*exp(-lam*X)/gamma(s)
Int <- sum(f(X)/Y)/n
X <- rgamma(n,s)/lam
Y <- (lam^s)*(X^(s-1))*exp(-lam*X)/gamma(s)
SE <- sqrt((sum((f(X)/Y)^2)/n-Int^2)/n)
CI <- c(Int-1.96*SE,Int+1.96*SE)
list("Int"=Int,"SE"=SE, "95% CI"=CI)}

-----------------
> Integral(1000,f,0.8,1)     ###Gamma(0.8,1)
$Int:
[1] 0.9938465     #OK
$"95% CI":
[1] 0.9797706 1.0079225
---------------
> Integral(1000,f,2,1)       ###Gamma(2,1)
$Int:
[1] 0.9891774     #OK
$"95% CI":
[1] 0.9110158 1.0673389
---------------------
> Integral(1000,f,8,1)       ###Gamma(8,1)
$Int:
[1] 0.3100254     #NOT OK
$"95% CI":
[1] 0.03716331 0.58288752
------------------------
> Integral(10000,f,8,1)      ###Gamma(8,1)
$Int:
[1] 1.458507       #NOT OK
$"95% CI":
[1] 1.236363 1.680651
--------------------------
> Integral(1000,f,1.8,1)     ###Gamma(1.8,1)
$Int:
[1] 0.9988329     #OK
$"95% CI":
[1] 0.8825257 1.1151401
------------------------
> Integral(1000,f,0.5,0.5)   ###ChiSq(1)
$Int:
```

```
[1] 0.9934391     #OK
$"95% CI":
[1] 0.9602447 1.0266335
--------------------------
> Integral(1000,f,0.5,5)     ###Gamma(0.5,5)
$Int:
[1] 0.5796288     #NOT OK
$"95% CI":
[1] 0.3436768 0.8155807
------------------------------
> Integral(1000,f,4,0.5)     ###ChiSq(8)
$Int:
[1] 0.6486151      #NOT OK
$"95% CI":
[1] 0.09757695 1.19965319
--------------------------------
> Integral(1000,f,2,0.5)     ###ChiSq(4)
$Int:
[1] 1.011917      #OK
$"95% CI":
[1] 0.879809 1.144025
----------------------------
> Integral(1000,f,1,0.5)     ###Exponential(0.5)
$Int:
[1] 0.994716     #OK
$"95% CI":
[1] 0.9595706 1.0298615
----------------------------
> Integral(1000,f,1,10)      ###Exponential(10)
$Int:
[1] 0.6177593     #NOT OK
$"95% CI":
[1] -0.5367878  1.7723064

----------------------------
> Integral(100000,f,0.8,1)   ###Gamma(0.8,1)
$Int:
[1] 0.9999284    $##Good
```

```
$"95% CI":
[1] 0.9982812 1.0015756
----------------------------
> Integral(100000,f,0.5,0.5) ###ChiSq(1)
$Int:
[1] 1.000163     $##Good
$"95% CI":
[1] 0.9973289 1.0029961
----------------------------
> Integral(1000,f,10,10)     ###Gamma(10,10)
$Int:
[1] 0.7003642     #NO
$"95% CI":
[1] 0.6454564 0.7552720
---------------------------
```