Monte Carlo Integration 1

Mihir Arjunwadkar

Centre for Modeling and Simulation
Savitribai Phule Pune University

Many ways to ⟨evaluate⟩ an integral

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

- Analytically

  Gradshteyn and Ryzhik!

- Symbolically

  Wolfram Alpha, Open-source computer algebra systems, ...

- Numerically

  Trapezoidal, Simpson, Gaussian quadrature, ...

- **Stochastically**

  Monte Carlo integration: A collection of methods to *estimate* integrals and expectation values

Consider this computation:

```
g <- function( x ) { x^2 } # integrand
n <- 10                    # sample size

mean( g( runif( n ) ) )
```

Has this computation anything to do with the integral below?

$$I = \int_0^1 x^2 dx = \frac{x^3}{3}\Big|_0^1 = \frac{1}{3}$$

Try the above code yourselves ...

An example

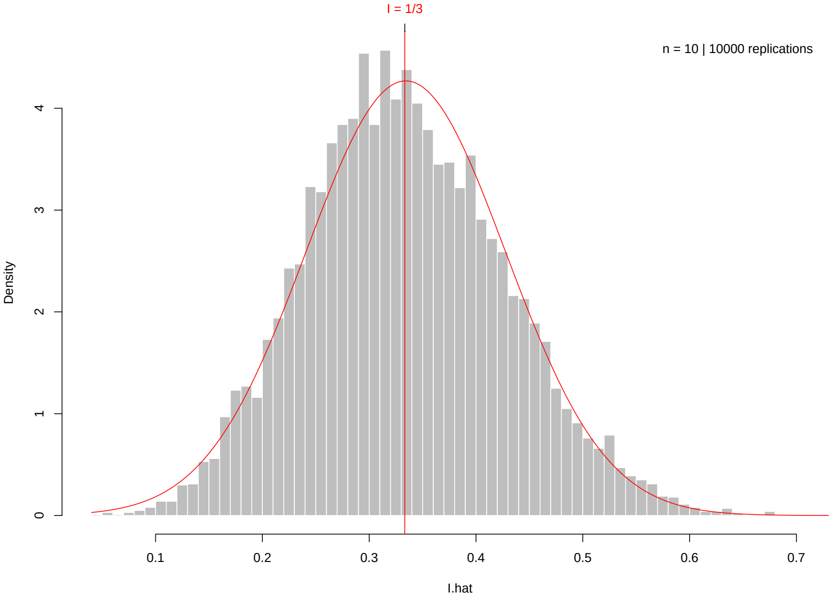
... run it multiple times ...

```
g <- function( x ) { x^2 } # integrand
n <- 10                    # sample size
m <- 10000                 # replications

I.hat <- replicate( m, mean( g( runif( n ) ) ) )
```

... plot a histogram ...

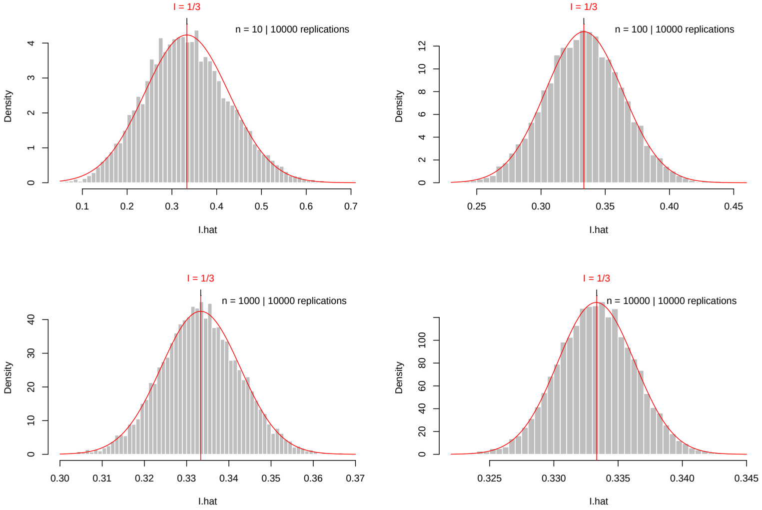
```
h <- hist( I.hat, 'FD', plot = FALSE )
pdf( 'basic-mc-2_2.pdf', height = 8.3, width = 11.7, paper = 'a4r' )
 op <- par( mar = c( 4.1, 4.1, 2.1, 0.1 ) )
  plot( h, col = 'gray', border = 'white', main = '', freq = FALSE )
  curve( dnorm( x, mean( I.hat ), sd( I.hat ) ),
         n = 501, add = TRUE, col = 'red' )

  text( max( I.hat ), max( h$density ),
        paste( 'n =', n, '|', m, 'replications' ), pos = 2 )

  abline( v = 1/3, col = 'red' ) # mark the true value of the integral
  axis( 3, at = 1/3, label = 'I = 1/3', col.axis = 'red' )
 par( op )
dev.off()
```

An example



I = 1/3
n = 10 | 10000 replications
Density
0
1
2
3
4
0.1
0.2
0.3
0.4
0.5
0.6
0.7
I.hat

... explore behaviour for larger values of `n` ...



Estimates appear to be normally distributed around $I = 1/3$!

Write

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

with

$$h(x) = (b-a)g(x) \quad \text{and} \quad f(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}.$$

Written this way, we see that I is the expected value of h with respect to the Uniform(a, b) PDF f; i.e.,

$$I = \mathrm{E}_f(h).$$

Why we saw what we saw?

An *estimator* for $\mathrm{E}_f(h)$ is

$$\widehat{I}_n = \frac{1}{n}\sum_{i=1}^{n} h(X_i),$$

with

$$X_1, X_1, \ldots, X_n \sim f \text{ IID}.$$

This, in fact, is the plug-in estimator for $\mathrm{E}_f(h)$.

The computation we did earlier was
repeated *estimation* of the integral $I = \mathrm{E}_f(h)$
using the above estimator.

- Indeed, if the integrand $h(x)f(x)$ is $\langle$sufficiently well-behaved$\rangle$ and satisfies $\langle$appropriate conditions$\rangle$, then we expect $\widehat{I}_n$ to follow Law of Large Numbers & Central Limit Theorem. (more details here)

- An estimator for the variance of $\widehat{I}_n$ is

$$\widehat{\text{Var}}\left(\widehat{I}_n\right) = \frac{1}{n} \times \frac{1}{n-1} \sum_{i=1}^{n} \left(h(X_i) - \widehat{I}_n\right)^2 .$$

- Using approximate or asymptotic normality of $\widehat{I}_n$, we have the usual prescription for a $1-\alpha$ confidence interval on I:

$$\widehat{I}_n \pm z_{\alpha/2} \sqrt{\widehat{\text{Var}}\left(\widehat{I}_n\right)},$$

where $z_{\alpha/2} = \Phi^{-1}(1-\alpha/2)$ and Φ^{-1} is the standard normal inverse CDF (i.e., quantile function).

Basic Monte Carlo integration procedure

Consider integrals that can be written in the form

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

where f is *any* valid* PDF defined over $\langle a, b \rangle$.

f need not necessarily be Uniform(a, b).

Limits a, b need not necessarily be finite.

This link has a nice interpretation of MC integration with uniform sampling

*Conditions apply

Basic Monte Carlo integration procedure

- Generate random numbers $X_1, X_2, \ldots, X_n \sim f$ IID
- Estimate the integral I as $\widehat{I}_n = \frac{1}{n} \sum_{i=1}^{n} h(X_i)$
- Estimate variance of $\widehat{I}_n$ as $\widehat{\text{Var}}\left(\widehat{I}_n\right) = \frac{1}{n} \times \frac{1}{n-1} \sum_{i=1}^{n} \left(h(X_i) - \widehat{I}_n\right)^2$
- Compute a $1 - \alpha$ confidence interval for I. For example, if $\widehat{I}$ is approximately or asymptotially normal, then $\widehat{I}_n \pm z_{\alpha/2} \sqrt{\widehat{\text{Var}}\left(\widehat{I}_n\right)}$

Conditions and reasons for why it works remain the same as before.

Estimating the Gamma function

$$\Gamma(z) = \int_0^\infty x^{z-1} e^{-x} dx.$$

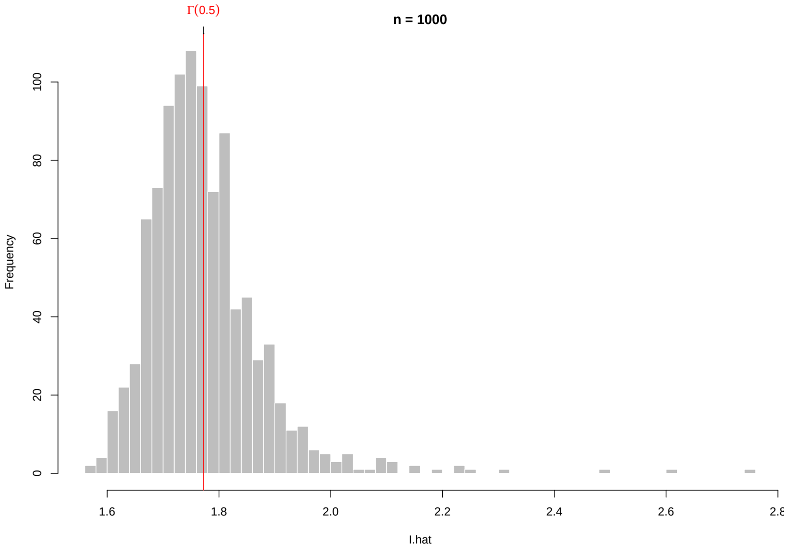
We may identify

$h(x) = x^{z-1}$, together with
$f(x) = e^{-x}$ as the Exp(1) PDF which is defined over $[0, \infty)$.

Example 1.1: Estimating $\Gamma(1/2)$

```
h <- function( x, z ) { x^( z - 1 ) } # part of the integrand
z <- 0.5                              # argument for h()

n <- 1000 # sample size
m <- 1000 # of replications

I.hat <- replicate( m, mean( h( rexp( n ), z ) ) )

pdf( 'gamma.pdf', height = 8.3, width = 11.7, paper = 'a4r' )
 op <- par( mar = c( 4.1, 4.1, 2.1, 0.1 ) )
  hist( I.hat, 'FD', col = 'gray', border = 'white', main = paste( 'n =', n ) )

  # mark the true value of the integral
  abline( v = gamma( z ), col = 'red' )
  axis( 3, at = gamma( z ),
        label = parse( text = paste( 'Gamma(', z, ')' ) ), col.axis = 'red' )
 par( op )
dev.off()
```

Example 1.1: Estimating Γ(1/2)



Example 1.2: Estimating $\Gamma(1/2)$

In the preceding example, what will happen if you identify

$h(x) = 1$

and

$f(x) = \text{PDF}$ for the $\Gamma(k, \theta)$ distribution with $k = z, \theta = 1$?

Try it yourself!

Example 2: "Deriving" EDF from MCI

Consider estimating the CDF of $X \sim f_X$; i.e.,

$$P(x) = \text{Prob}(X \leq x) = \int_{-\infty}^{x} f_X(t)dt = \int h(t) f_X(t) dt$$

with the natural identification

$$h(t) = \begin{cases} 1 & t \leq x \\ 0 & \text{otherwise} \end{cases}.$$

To estimate the integral $P(x)$, we _*conceptually*_ generate random observations

$$X_1, X_2, \ldots, X_n \sim f_X,$$

and estimate $P(x)$ as

$$\widehat{P}_n(x) = \frac{1}{n} \sum_{i=1}^{n} h(X_i) = \frac{1}{n} \sum_{i=1}^{n} \left\{ \begin{array}{ll} 1 & X_i \leq x \\ 0 & \text{otherwise} \end{array} \right\} = \frac{\#\text{ of observations} \leq x}{n}.$$

This is nothing but the Empirical Distribution Function, an estimator for the CDF!

Consider again

$$I = \int_a^b g(x)dx = \int h(x)f(x)dx.$$

What if

- f is difficult or impossible to sample from?
- regions which contribute to the integral most is a low-density region under f?

Reformulate I in an alternate form:

$$I = \int \left(h(x) \frac{f(x)}{q(x)} \right) q(x) dx = \int w(x) q(x) dx = \mathrm{E}_q(w),$$

where q is a PDF over the same support/domain, but is easier to sample compared to f — i.e., a relatively more efficient random number generator is available for q.

Importance sampling: Apply the basic Monte Carlo procedure to this reformulation of I. Justification for why it should work under appropriate conditions remains the same.

$\widehat{I_n}$ may have infinite variance! Why?

$$\mathrm{E}_q(w^2) = \int w^2(x)q(x)dx = \int \left(h(x)\frac{f(x)}{q(x)}\right)^2 q(x)dx = \int \frac{h^2(x)f^2(x)}{q(x)}dx.$$

If q goes to zero at a faster rate than $h^2 f^2$, we expect $\mathrm{E}_q(w^2) \to \infty$!

Hence we expect the variance of $\widehat{I_n} \to \infty$ as well.

Importance sampling: Avoiding the catch

- In general, choose $q(x)$ to have **thicker** tails than f:

  If $q(x) \ll f(x)$ over some set of x values, then the ratio f/q will be large, leading to large variance.

- q should have shape as close to f as possible:

  It can be shown that the variance of $\widehat{I_n}$ is minimized when $q = f$

  (AoS2004 Theorem 24.5)

Importance sampling vs. rejection sampling

	importance sampling	**rejection sampling**
purpose	estimating integrals	generating random numbers
PDF	integrating PDF f	target PDF f
sampling	sampling PDF q	envelope PDF e, k s.t. $ke \geq f$
performance measure	$\widehat{\text{Var}}\left(\widehat{I}_n\right)$	acceptance probability
best performance	$q = f$	$ke = f$

Estimating normal tail probability: Basic MCI

$$I = \text{Prob}(Z > 3) = \frac{1}{\sqrt{2\pi}} \int_3^\infty \exp\left(-\frac{x^2}{2}\right) \approx 0.0013$$

Write

$$I = \int h(x) f(x) dx,$$

with

$$h(x) = \begin{cases} 1 & x > 3 \\ 0 & \text{otherwise} \end{cases},$$

and

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right).$$

Sample $X_1, \ldots, X_n \sim N(0,1)$
Estimate I as $\widehat{I}_n = \frac{1}{n} \sum_{i=1}^n h(X_i)$.

Estimating normal tail probability: Basic MCI

n	$\widehat{I}_n$	$\sqrt{\widehat{\text{Var}}\left(\widehat{I}_n\right)}$	absolute error
5	0.000000000000	0.0000000	0.0013499
10	0.000000000000	0.0000000	0.0013499
20	0.000000000000	0.0000000	0.0013499
50	0.000000000000	0.0000000	0.0013499
100	0.010000000000	0.0100000	0.0086501
500	0.002000000000	0.0020000	0.0006501
1000	0.002000000000	0.0014135	0.0006501
5000	0.001800000000	0.0005995	0.0004501
10000	0.001300000000	0.0003603	0.0000499
100000	0.001260000000	0.0001122	0.0000899
1000000	0.001346000000	0.0000367	0.0000039
10000000	0.001313600000	0.0000115	0.0000363

Most sample points are wasted, because the tail probability is low. To realize an event with probability ≈ 0.001, we need n of the order of 1000 or more.

Estimating normal tail probability: Importance sampling

$$I = \text{Prob}(Z > 3) = \frac{1}{\sqrt{2\pi}} \int_3^\infty \exp\left(-\frac{x^2}{2}\right) \approx 0.0013$$

Write

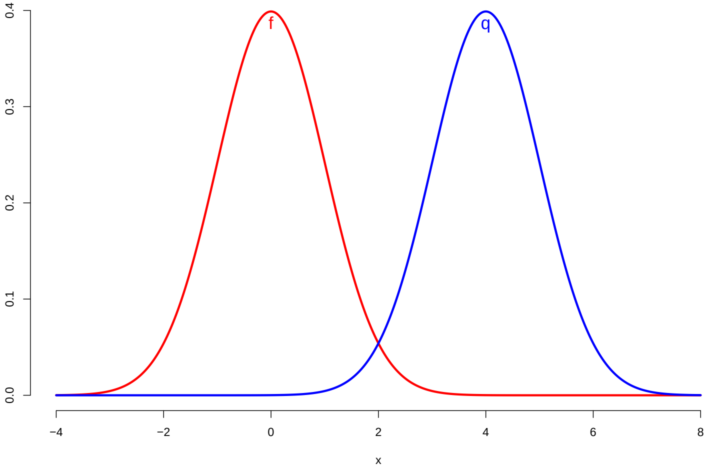
$$I = \int \left(h(x) \frac{f(x)}{q(x)} \right) q(x) dx,$$

with

$$q = N(4,1) \text{ PDF} = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-4)^2}{2}\right).$$

Sample $X_1, \ldots, X_n \sim N(4,1)$
Estimate I as $\widehat{I_n} = \frac{1}{n} \sum_{i=1}^n h(X_i) f(X_i)/q(X_i)$.

Estimating normal tail probability: Importance sampling



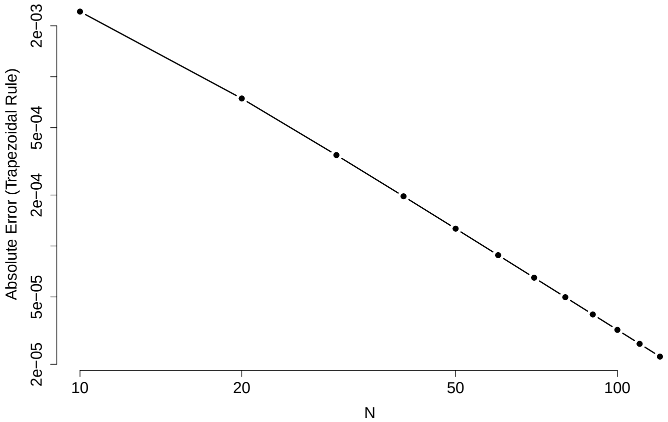
Estimating normal tail probability: Importance sampling

n	$\widehat{I}_n$	$\sqrt{\widehat{\text{Var}}\left(\widehat{I}_n\right)}$	absolute error
5	0.002917014316	0.0014843	0.0015671
10	0.003017586990	0.0018823	0.0016677
20	0.001461115145	0.0006851	0.0001112
50	0.001628391994	0.0005013	0.0002785
100	0.001205269400	0.0003142	0.0001446
500	0.001434046970	0.0001397	0.0000841
1000	0.001619157501	0.0001120	0.0002693
5000	0.001361899535	0.0000445	0.0000120
10000	0.001390816326	0.0000318	0.0000409
100000	0.001361292352	0.0000099	0.0000114
1000000	0.001341089919	0.0000031	0.0000088
10000000	0.001349609965	0.0000010	0.0000003

Yields better estimates of I compared to the basic MCI procedure.

Why use Monte Carlo integration at all?

Apply Trapezoidal Rule to $I = \frac{1}{\sqrt{2\pi}} \int_3^\infty e^{-x^2/2} dx$:



- Trapezoidal Rule: Absolute error of $\sim 1 \times 10^{-5}$ with only about 100 points.
- Gaussian Quadrature: far better approximation in far fewer # of points!

Then why use Monte Carlo integration at all?