Methods of Monte Carlo Integration

Sarah Klenha

Occidental College

March 22, 2016

Outline

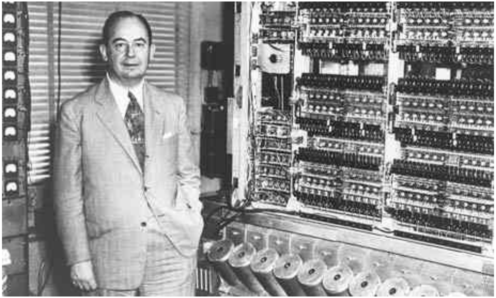
- Introduction to Monte Carlo Simulation
- Hit or Miss Method
- Sample Mean Method
- Comparison of the Two Methods
- An Example
- Conclusion

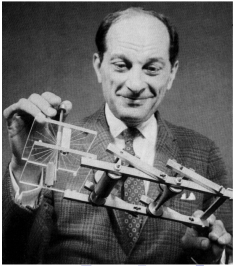
Numerical Integration

- Deterministic vs. Stochastic
- Well-known methods, such as Simpson's Rule or the trapezoidal rule are deterministic
- Monte Carlo Integration is a stochastic method

History

- First developed by John von Neumann and Stanislaw Ulam during World War II.
- Monte Carlo Simulation has applications in physics, engineering, mathematics, and finance.





Setting Up the Problem

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

- Constraints for area of integration: $0 \leq g(x) \leq c, \quad a \leq x \leq b$
- Rectangle containing area of integration: $R = \{(x, y) : a \leq x \leq b, 0 \leq y \leq c\}$
- Exact area of integration: $S = \{(x, y) : a \leq x \leq b, 0 \leq y \leq g(x)\}$

Finding θ_1

$p =$ probability any given point in R is also in S

$$p = \frac{\text{area } S}{\text{area } R} = \frac{\int_a^b g(x)dx}{c(b-a)} = \frac{I}{c(b-a)} \Rightarrow I = pc(b-a)$$

Generate a sample of N independent vectors, $(X_1, Y_1), \ldots, (X_N, Y_N)$, where $X \sim U[a,b]$ and $Y \sim U[0,c]$.

$p \approx \hat{p} = \frac{N_H}{N}$, where N_H is the number of "hits".

$$I = pc(b-a) \approx \frac{N_H}{N} c(b-a) = \theta_1$$

Parameters of θ_1

$\mathrm{E}(\theta_1) = I \Rightarrow \theta_1$ is an unbiased estimator for I.

$$\mathrm{Var}(\theta_1) = \frac{I}{N}[c(b-a) - I]$$

Setting Up the Problem

$$I = \int_a^b g(x)dx = \int_a^b \frac{g(x)}{f_X(x)} f_X(x)dx = \mathrm{E}\left[\frac{g(x)}{f_X(x)}\right]$$

where $f_X(x)$ is the pdf of random variable $X \sim U[a, b]$.

$$I = (b - a)\mathrm{E}[g(x)]$$

Finding θ_2

$\mathrm{E}[g(X)] \approx \frac{1}{N}\sum_{i=1}^{N} g(X_i) \rightarrow$ sample mean of $g(X)$

Generate a sequence of N independent samples, $X_1 \ldots, X_N$ to calculate the sample mean.

$$I \approx (b-a)\frac{1}{N}\sum_{i=1}^{N} g(X_i) = \theta_2$$

Parameters of θ_2

$\mathrm{E}(\theta_2) = I \Rightarrow \theta_2$ is an unbiased estimator for I.

$$\mathrm{Var}(\theta_2) = \frac{1}{N}\left[(b-a)\int_a^b g^2(x)dx - I^2\right]$$

Efficiency

- θ_1 and θ_2 are both unbiased estimators for I.
- We will say θ_1 is more efficient than θ_2 if,

$$\frac{t_1 \mathrm{Var}(\theta_1)}{t_2 \mathrm{Var}(\theta_2)} < 1,$$

  where $t_1 =$ time to complete Hit-or-Miss algorithm and $t_2 =$ time to complete Sample Mean algorithm.
- Both methods run in linear time so we will say $t_1 = t_2$.

Accuracy Based on Variance

- Since $t_1 = t_2$, if θ_1 is more efficient than θ_2 then,

$$\frac{\mathrm{Var}(\theta_1)}{\mathrm{Var}(\theta_2)} < 1.$$

- However, after plugging in variance equations we find,

$$\mathrm{Var}(\theta_1) > \mathrm{Var}(\theta_2).$$

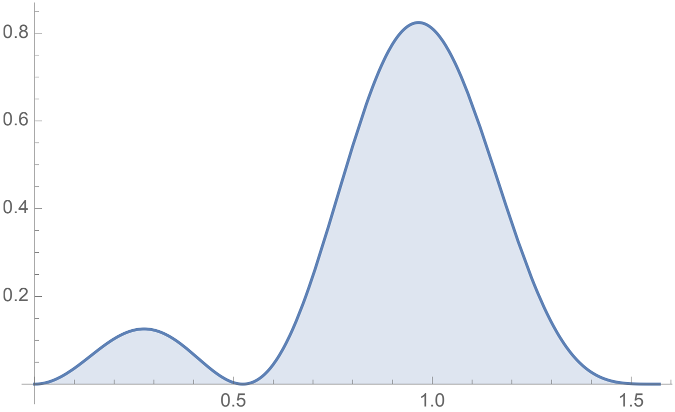
- Therefore, the Sample Mean Method is more efficient than the Hit-or-Miss Method.

The Problem

$$I = \int_0^{\pi/2} \sin^2(2x)\cos^2(3x)dx$$

- Analytical Solution: $I = \frac{\pi}{8} = 0.39827$
- Parameters: $a = 0$, $b = \frac{\pi}{2}$, $c = 1$
- Integrand will be called $g(x)$

Graph of $g(x)$



Graph of $g(x)$ with randomly generated points

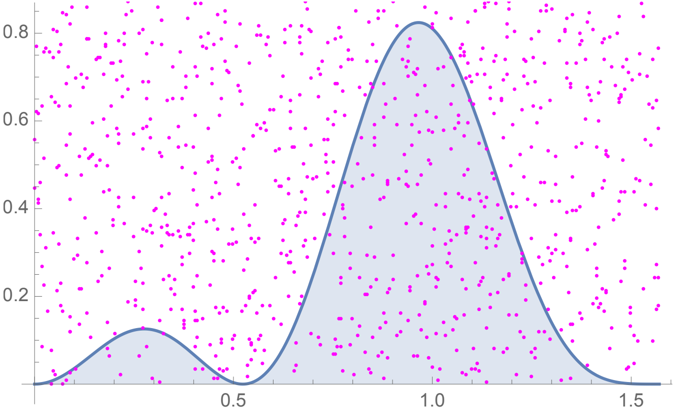
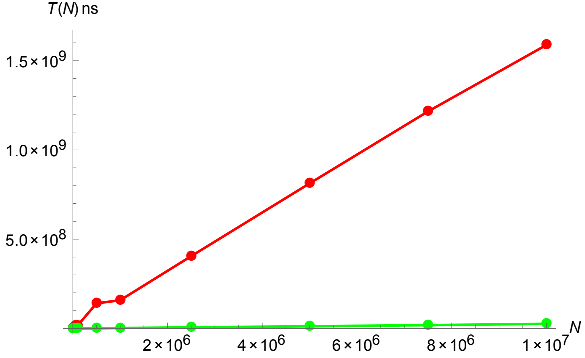


Table 1: Average values of 100 simulations

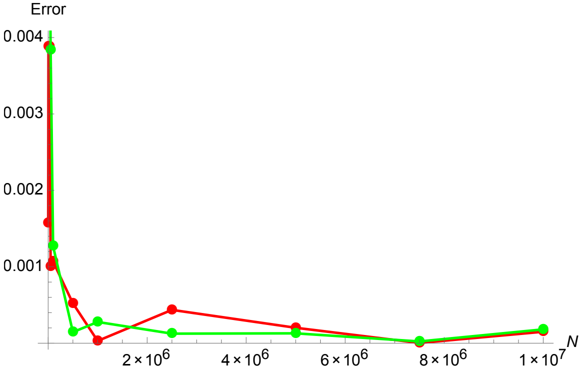
Hit-or-Miss				Sample Mean		
N	$\hat{p}$	θ_1	$Error(\theta_1)$	N	θ_2	$Error(\theta_2)$
1000	0.24800	0.39485	0.00215	1000	0.39130	0.0014
10000	0.23940	0.39190	0.0008	10000	0.39271	0.00001
50000	0.25134	0.39297	0.00027	50000	0.39275	0.0005
100000	0.24877	0.39258	0.00012	100000	0.39260	0.00001
500000	0.24964	0.39294	0.00024	500000	0.39269	0.00001
1000000	0.24996	0.39259	0.00011	1000000	0.39278	0.00008
2500000	0.24965	0.39270	0.0000	2500000	0.39269	0.00001
5000000	0.24994	0.39271	0.00001	5000000	0.39269	0.00001
7500000	0.24986	0.39271	0.00001	7500000	0.39267	0.00003
10000000	0.25006	0.39273	0.00003	10000000	0.39269	0.00001

Graph of Time Taken to Complete Algorithm



Hit-or-Miss in red, Sample Mean in green.

Graph of Error After One Iteration



Hit-or-Miss in red, Sample Mean in green

- We found the Sample Mean to be more efficient, although both methods can provide a high level of accuracy with a large simulation size.
- Pros
  - Methods do not require knowledge of the behavior of the entire function, only at points being evaluated
  - Useful for evaluating multi-dimensional integrals
  - Easy to implement
- Cons
  - Not as accurate for one-dimensional as other numerical methods.

Thanks!