Outline

1. Basic Scheme of Monte Carlo Simulation
2. Bias and Variance Reduction
3. Examples
Basics of Monte Carlo (MC) Simulation

- Let $X$ denote a random variable (or a random vector or stochastic process). The goal is to calculate $\mu = E(h(X))$ for some function $h$.
- i.i.d. = independent and identically distributed

Plain Monte Carlo

1. Generate a sample of size $n$: $(X_1, X_2, \ldots, X_n)$, i.i.d. with the same distribution as that of $X$.
2. Estimate $\mu$: Use the sample average

$$\hat{\mu} = \frac{1}{n}[h(X_1) + h(X_2) + \cdots + h(X_n)].$$

3. Justification: Law of Large Numbers ensures that

$$\frac{1}{n}[h(X_1) + h(X_2) + \cdots + h(X_n)] \to E(h(X)) = \mu.$$
Remark

- $\hat{\mu}$ is an unbiased estimator of $\mu$; namely,
  \[ E(\hat{\mu}) = \mu. \]

- Let $A \subseteq \mathbb{R}$ be any subset of real numbers.
  \[ P(X \in A) = E(h(X)) \]
  where
  \[ h(x) = \begin{cases} 
    1 & \text{if } x \in A \\
    0 & \text{if } x \notin A.
  \end{cases} \]

The plain MC can be used to estimate the probability $P(X \in A)$.

- Take $A = (-\infty, x]$, the plain MC can be used to estimate the cumulative probability $P(X \leq x)$. 
Let $H = h(X)$ have a finite variance $\text{var}(H) = \sigma_H^2 < \infty$.

- $\text{var}(\hat{\mu}) = \frac{\sigma_H^2}{n} \to 0$, as $n \to \infty$.
- Central Limit Theorem ensures that
  $$P\left(\frac{\hat{\mu} - \mu}{\sigma_H/\sqrt{n}} \leq x\right) \to \Phi(x), \text{ as } n \to \infty.$$  

- That is, if $Z \sim N(0, 1)$, then

$$\mu \approx \underbrace{\hat{\mu}}_{1\text{st order estimate}} + \underbrace{Z \frac{\sigma_H}{\sqrt{n}}}_{\text{error estimate}}.$$
Confidence Interval

More precisely, a $100(1 - \alpha)\%$ confidence interval for $\mu$ is given by

\[
\left( \hat{\mu} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \hat{\mu} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right)
\]

where $z_{\alpha/2}$ is the $100\alpha/2$ percentage point.
The standard deviation $\sigma_H$ is rarely known in practice.

- Let $H_i = h(X_i)$, $i = 1, \ldots, n$, Use

$$s^2_H = \frac{1}{n-1} \sum_{i=1}^{n} (H_i - \hat{\mu})^2$$

to estimate $\sigma^2_H$.

- Replace $\sigma_H$ by $s_H$, a $100(1 - \alpha)$% confidence interval for $\mu$ becomes

$$\left( \hat{\mu} - t_{\alpha/2} \frac{s_H}{\sqrt{n}}, \hat{\mu} + t_{\alpha/2} \frac{s_H}{\sqrt{n}} \right)$$

where $t_{\alpha/2}$ is the $100\alpha/2$ percentage point of the t distribution with $n - 1$ degrees of freedom.

- $s_H/\sqrt{n}$ is known as the standard error.
Variance Reduction

- When comparing two unbiased estimates, the one with smaller variance is more efficient (smaller standard error).
- There are many variance reduction methods.
- Importance Sampling: Let $X \sim F$ with density $f(x)$, and $Y \sim G$ with density $g(x)$. To estimate $\mu = E_F[h(X)]$, consider

$$\mu = \int_{0}^{\infty} h(x)f(x)dx = \int_{0}^{\infty} h(x)\frac{f(x)}{g(x)}g(x)dx = E_G\left[h(Y)\frac{f(Y)}{g(Y)}\right].$$

Use an estimator

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} h(Y_i)\frac{f(Y_i)}{g(Y_i)}$$

based on the sample $(Y_1, \ldots, Y_n)$ sampling from $G$ to achieve variance reduction.
Biased-ness

- Systematic bias is a more serious problem.
- For example, bias occurs when we approximate continuous-time process (such as Brownian motion) by using discrete-time processes.
Example: Bias due to discretization

- Consider the geometric Brownian motion:
  \[ S_t = S_0 e^{(r-\sigma^2/2)t+\sigma B_t}, \ t \geq 0. \]

- We want to estimate \( \mu = E[\max_{0 \leq t \leq T} S_t] \); that is, the average largest value of \( S_t \) over \([0, T]\).

- Discretize \([0, T]\) by using \( 0 = t_0 < t_1 < \cdots < t_m = T \). Compute
  \[ S_{t_0} = S_0, \]
  \[ S_{t_{i+1}} = S_{t_i} e^{(r-\sigma^2/2)(t_{i+1}-t_i)+\sigma(B_{t_{i+1}}-B_{t_i})}, \ i = 0, 1, \ldots, m, \]
  where \( B_{t_{i+1}} - B_{t_i} \sim N(0, t_{i+1} - t_i) \), and \( B_{t_{i+1}} - B_{t_i}, i = 0, 1, \ldots, m, \) are independent.

- Estimate \( \max_{0 \leq t \leq T} S_t \) using \( \max_{0 \leq i \leq m} S_{t_i} \).

- Observe that \( \max_{0 \leq i \leq m} S_{t_i} \leq \max_{0 \leq t \leq T} S_t \), and thus such an estimate always underestimate \( \mu = E[\max_{0 \leq t \leq T} S_t] \).
Example: Bivariate Normal Distribution

- Let $Z_1$ and $Z_2$ be independent and have the standard normal distribution: $Z_1 \sim \mathcal{N}(0, 1), \quad Z_2 \sim \mathcal{N}(0, 1)$.
- Let $C$ denote a $2 \times 2$ matrix:

$$C = \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix}$$

- The bivariate normal vector $(X_1, X_2)$ is defined as follows

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix} \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix}$$

$$= \begin{pmatrix} \mu_1 + C_{11}Z_1 + C_{12}Z_2 \\ \mu_2 + C_{21}Z_1 + C_{22}Z_2 \end{pmatrix}.$$  

- The mean value vector $(E(X_1), E(X_2))' = (\mu_1, \mu_2)'$.
- The covariance matrix $\Sigma = CC'$. 
Simulate a sample vector \((x_1, x_2)\) of the bivariate normal vector

\[
\begin{pmatrix}
X_1 \\
X_2
\end{pmatrix}
= \begin{pmatrix}
C_{11} & C_{12} \\
C_{21} & C_{22}
\end{pmatrix}
\begin{pmatrix}
Z_1 \\
Z_2
\end{pmatrix}.
\]

- Let \(\Phi(x)\) be the cdf of the standard normal distribution \(N(0, 1)\).
- Observe that
  \[
  \Phi(Z_1) = U_1, \quad \Phi(Z_2) = U_2
  \]
  have the uniform distributions over \([0, 1]\).
- Select two numbers \(u_1\) and \(u_2\) independently at random from \([0, 1]\). Then calculate \(z_1 = \Phi^{-1}(u_1)\) and \(z_2 = \Phi^{-1}(u_2)\).
- Calculate
  \[
  \begin{pmatrix}
x_1 \\
x_2
\end{pmatrix}
= \begin{pmatrix}
C_{11} & C_{12} \\
C_{21} & C_{22}
\end{pmatrix}
\begin{pmatrix}
z_1 \\
z_2
\end{pmatrix}.
  \]
Assume that the stock price follows the geometric Brownian motion:

\[ S_t = S_0 e^{(r-\sigma^2/2)t+\sigma B_t}, \quad t \geq 0. \]

We want to estimate the price of the Asian option with strike price \( K \):

\[ v = E \left[ e^{-rT} \left( \sum_{i=0}^{m} S_{t_i} - K \right)^+ \right] \]

where \( 0 = t_0 < t_1 < \cdots < t_m = T \) are a fixed set of dates.
Example: Asian Option (cont’d)

Algorithm:

1. Discretize $[0, T]$ by using fixed dates $0 = t_0 < t_1 < \cdots < t_m = T$. Compute

   
   \[
   S_{t_0} = S_0,
   \]

   \[
   S_{t_{i+1}} = S_{t_i} e^{(r - \sigma^2/2)(t_{i+1} - t_i) + \sigma (B_{t_{i+1}} - B_{t_i})}, \quad i = 0, 1, \ldots, m,
   \]

   where $B_{t_{i+1}} - B_{t_i} \sim N(0, t_{i+1} - t_i)$, and $B_{t_{i+1}} - B_{t_i}, i = 0, 1, \ldots, m,$ are independent.

2. Calculate the discounted payoff

   \[
   \nu = e^{-rT} \left( \sum_{i=0}^{m} S_{t_i} - K \right)^+. 
   \]

3. Repeat (1) and (2) $n$ times to obtain $n$ i.i.d. copies $\nu_1, \ldots, \nu_n$. Then estimate $\nu$ using

   \[
   \hat{\nu} = \frac{1}{n} (\nu_1 + \cdots + \nu_n). 
   \]
Example: Spread Call Option

- Assume that the two stock prices follow the correlated geometric Brownian motions:

  \[ X_t = X_0 e^{(r - \sigma_1^2/2)t + \sigma_1 W_t}, \quad t \geq 0. \]

  \[ Y_t = Y_0 e^{(r - \sigma_2^2/2)t + \sigma_2 B_t}, \quad t \geq 0, \]

  where \((W_t, B_t)\) has a bivariate normal distribution with means zero and convariance matrix

  \[ \Sigma = t \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}, \quad t > 0. \]

- We want to estimate the price of the spread call option with strike price \(K\):

  \[ v = E \left[ e^{-rT} (X_T - Y_T - K)^+ \right]. \]

  \((X_T, Y_T)\) spread
Example: Spread Call Option (cont’d)

At time $T > 0$,

$$X_T = X_0 e^{(r - \sigma_1^2/2)T + \sigma_1 W_T},$$

$$Y_T = Y_0 e^{(r - \sigma_2^2/2)T + \sigma_2 B_T},$$

where $(W_T, B_T)$ has a bivariate normal distribution with means zero and covariance matrix

$$\Sigma = T \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}.$$

Let

$$C = \sqrt{T} \begin{pmatrix} 1 & 0 \\ \rho & \sqrt{1 - \rho^2} \end{pmatrix}, \quad t > 0.$$

Verify that $CC' = \Sigma$ (Cholesky decomposition)
Example: Spread Call Option (cont’d)

Algorithm:

1. Simulate a normal sample \((w_1, w_2)\) with zero means and covariance matrix \(\Sigma\).

2. Calculate

\[
x = X_0 e^{(r - \sigma_1^2/2)T + \sigma_1 w_1},
\]

\[
y = Y_0 e^{(r - \sigma_2^2/2)T + \sigma_2 w_2}.
\]

3. Calculate the discounted payoff

\[
v = e^{-rT}(x - y - K)^+.
\]

4. Repeat (1), (2) and (3) \(n\) times to obtain \(n\) i.i.d. copies \(v_1, \ldots, v_n\). Then estimate \(v\) using

\[
\hat{v} = \frac{1}{n}(v_1 + \cdots + v_n).
\]
Example: Simulate Value-at-Risk

Let $X_1, X_2, \ldots$ be a random sequence with following correlation structure:

Given $X_i = x$, $X_{i+1} \sim N(0, x^2)$.

Observe that

$$X_{i+1} \overset{d}{=} X_i Z, \text{ where } Z \sim N(0, 1).$$

This is a simple form of scale mixture models, and widely used in time series analysis (ARCH models).

This is also a simple example of Markov chains.
Algorithm 1:
1. Simulate $x_1$ from $N(0, 1)$.
2. For $i = 2, 3, \ldots, m$, simulate $z$ from $N(0, 1)$. Calculate
   \[ x_i = x_{i-1} z \]
3. Calculate $s = \sum_{i=1}^{m} x_i$.

Repeat Algorithm 1 $n$ times to generate $n$ i.i.d. copies $s_1, s_2, \ldots, s_n$. 
Example: Simulate Value-at-Risk (cont’d)

- Again let $S = \sum_{i=1}^{m} X_i$.
- We want to estimate $\text{VaR}(S)$; that is,

$$P(S \leq \text{VaR}(S)) = \alpha.$$ 

Algorithm 2

1. Arrange $s_1, s_2, \ldots, s_n$ in the increasing order

$$s_{(1)} \leq s_{(2)} \leq \cdots \leq s_{(n)} \text{ (called order statistics)}.$$

2. Set $\widehat{\text{VaR}}(S) = s_{([n\alpha])}$, where $[np]$ denotes the integer part of $np$. 

Haijun Li
Math 416/516: Stochastic Simulation  
Weeks 4-5  
20 / 21
Enrico Fermi (1930s) first experimented with the simulation method while studying neutron diffusion, but did not publish anything on it.

While working in Los Alamos National Laboratory on nuclear physics, Stanislaw Ulam in 1946 initiated the idea of using random simulation experiments. Later in 1946, Ulam described the idea to John von Neumann to plan large-scale actual calculations.

Von Neumann chose the code name “Monte Carlo” for the project. The name refers to the Monte Carlo Casino in Monaco where Ulam’s uncle would borrow money to gamble.

Monte Carlo methods were central to the simulations required for the Manhattan Project.

Today, Monte Carlo simulations are used everywhere in science and engineering.