Introductory Remarks

What is a Monte Carlo method?

(Why the name Monte Carlo?)

Invented many times!!

1) Adjoint of probability theory because of gaming (also combinatorics)

2) Early 20th century: to "verify" new results in probability

3) Fermi, von Neumann & Ulam to solve problem resulting from the atomic bomb
Approaches to Monte Carlo

Direct Simulation (game playing)

Solution of Mathematical Problems

Many ways to present Monte Carlo

a) Mathematical
   (algorithms, proofs, convergence errors)

b) Computer Science
   (algorithms, implementations, simulation)

c) Applications
   (physics, engineering, chemistry, finance)
Examples of Monte Carlo
(not to be covered in depth)

a) Financial instrument evaluation
   (very hot!)

b) Solution of Schrödinger equation, many-body
   (only known method)

c) Verification of t-distribution, etc. (Gosset, Student)

d) Buffon needle problem
Monte Carlo is Based on Probability

Probability is measure theory

Probability \sim measure \quad P(\Omega) = 1

Event \sim measurable set

Random variable \sim measurable function

Expected value \sim integral

Call A, B, C, \ldots events

\[ P(A + B + C + \ldots) \leq P(A) + P(B) + \ldots \quad \text{sub additivity} \]

A, B, C, \ldots are exclusive \leq \rightarrow = A, B, C, \ldots are exhaustive \quad r.h.s. = 1
Statistical Review

\[ P(AB) = P(A|B) \cdot P(B) \]

\[ P(A|B) \] is conditional probability

If \[ P(A|B) = P(A) \] then \( A \) and \( B \) are independent and

\[ P(AB) = P(A) \cdot P(B) \]

\[ P(ABC) = P(ABC\mid C) \cdot P(C) \]

Cumulative Distribution Function:

\[ RV \ z \quad F(y) = P(z \leq y) \]

\[ F(-\infty) = 0 \quad F(+\infty) = 1 \]

\( F \) is increasing and right-continuous

(at worst, \( F \) is a step function)
Dic

\[ \mathbb{P} [z = j] = \frac{1}{6} \]

\( j = 1, 2, 3, 4, 5, 6 \)

\[ F(y) = \mathbb{P}(z \leq y) \]

\[ = \begin{cases} 
0 & \text{if } y < 1 \\
1/6 & 1 \leq y < 2 \\
2/6 & 2 \leq y < 3 \\
3/6 & 3 \leq y < 4 \\
4/6 & 4 \leq y < 5 \\
5/6 & 5 \leq y < 6 \\
1 & y \geq 6 
\end{cases} \]

\[ \mathcal{S}(y-x_0) = \begin{cases} +\infty & \text{if } y > x_0 \\
0 & \text{otherwise} 
\end{cases} \]

\[ \int \mathcal{S}(y-x_0) \, dx = \begin{cases} 1 & \text{if } y > x_0 \\
0 & \text{if } y \leq x_0 
\end{cases} \]
Expectation

\[ E[g(z)] = \int g(cy) dF(y) \]

\[ F(y) = P(z \leq y) \]

because of \( F \) this is a Stieltjes integral

If \( F(y) = f(cy) \) exists then

\[ E[g(z)] = \int g(cy) f(cy) dy \]

Note: if \( z \) is a discrete distribution then

\[ E[g(z)] = \sum_i g(cy_i) f_i \]

\[ F(y) = \sum_i f_i H(y-y_i) \]

Heaviside function

\( f(cy) \), \( f_i \) constitute probability d.f.
Multivariate Distributions

\[ F(y, z) = P(z \leq y, z \leq z) \]

\[ G(y) = P(z \leq y) \]

Marginal Distributions

\[ H(z) = P(y \leq z) \]

"Integrate out the other variable"

If \( F(y, z) = G(y) H(z) \), then \( z \) and \( y \)

are independent

Note: \( \sum_i E[g(z_i)] = E[\sum_i g(z_i)] \)

hold for all r.v.'s \( z_i \)

BUT

\( \prod_i E[g(z_i)] = E[\prod_i g(z_i)] \)

holds only when \( z_i \)'s are independent
Note \( E[g(z)] \neq g(E[z]) \) generally.

\[
E[z] = \int y \, dF(y) = \mu
\]

is the mean of \( z \).

\( E[(z-\mu)^r] \) is the \( r \)th central moment of \( z \).

\( E[z^r] \) is the \( r \)th moment of \( z \).

\[
E[(z-\mu)^2] = \sigma^2 = \text{Var}[z]
\]

the dimensions of \( \sigma^2 \) and \( \mu^2 \) are the same.

\[
\frac{\mu}{\sqrt{\sigma^2}} \text{ is coefficient of variation (dimensionless)}
\]

Note: \( \sigma \) called standard deviation

\[
= \sqrt{\sigma^2}
\]
Multivariate
"Variances"
\[
\text{Cov}(z, \gamma) = E[(y-\mu)(z-\nu)] \\
\mu = E[z] \quad \nu = E[\gamma] \\
\text{Cov}(z, \gamma) = \text{Var}(z) \quad \text{by above} \\
\text{Assume independence} E[yz] = \mu \nu \]
\[
\text{Cov}(z, \gamma) = E[(yz - \mu z - \gamma y + \mu \nu)] \\
= E[yz] - \mu \nu \gamma - \mu \nu + \mu \nu = 0 \\

correlation \leftrightarrow \text{cov}(\ ) = 0 \\
\rho(z, \gamma) = \frac{\text{Cov}(z, \gamma)}{\sqrt{\text{Var}(z) \text{Var}(\gamma)}} \\
-1 \leq \rho \leq +1 \\
\text{negatively correlated} \quad \leftrightarrow \quad \text{positively correlated} \\
\text{uncorrelated} \quad \leftrightarrow \quad \text{independent}
\[ \text{Var} \left( \sum_{i} z_i \right) = \sum_{i} \sum_{j} \text{Cov} \left( z_i, z_j \right) \]

(left to students, it's easy!)

\[ \text{Var}[g(z_1, \ldots, z_n)] \]

where \( \mu_i = \text{E}(z_i) \)

\[ g_i = \frac{\partial g}{\partial z_i} \bigg|_{z_i = \mu_i} \]

\[ = \sum_{i=1}^{k} \sum_{j=1}^{k} g_i g_j \text{Cov} \left( z_i, z_j \right) \]

This is like a multidimensional Taylor series expansion about the means.
Some CDFs

(PDFs left to students)

\[
F(y) = \begin{cases} 
0 & y < 0 \\
\frac{y}{1} & 0 \leq y \leq 1 \\
1 & y > 1
\end{cases}
\]

uniform or rectangular distribution \( U[0,1] \)

\[U[a,b] \rightarrow F(y) = \begin{cases} 
0 & y < a \\
\frac{y-a}{b-a} & a \leq y \leq b \\
1 & y > b
\end{cases}\]

Binomial

\[F(y) = \sum_{t=y}^{n} \frac{n!}{t!(n-t)!} \ p^t \ (1-p)^{n-t}\]

\( p \) is prob of single event \( n \) times \( P \) dist. of \( n \) such trials

Poisson

\[F(y) = \sum_{t=y}^{\infty} e^{-\lambda} \ \frac{\lambda^t}{t!}\]

\( \lambda > 0 \)

dist. of number "t" of events occurring at rate \( \lambda \)
More CDFs

exponential \( F(y) = H(y)(1-e^{-\lambda y}) \)

\( \lambda \) is exponential rate in fixed time, Poisson counts numbers of these

Gaussian \( F(y) = \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt \)

\[ E[Z] = \mu \quad Var[Z] = \sigma^2 \]

\[ dF(y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt \]

\( F(y) \) is called the error function \( N(\mu, \sigma) \)

Also, have

\[ F(\bar{y}) = \int_{-\infty}^{\bar{y}} 12\pi V1 \sqrt{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\bar{y}-\mu)^2 V(t-\mu)\right] dt \]

\[ V_{ij} = Cov(Z_i, Z_j) \]

multivariate normal
Central Limit Theorem

Assume $z_i$'s are independent and identically distributed r.v.'s

with $E(z_i) = \mu_i$ and $\text{Var}(z_i) = \sigma_i^2$

$z_i = \frac{z_i - \mu_i}{\sigma_i}$ has mean 0 and variance 1

$$P \left[ \frac{1}{N} \sum_{i=1}^{N} z_i \leq y \right] \rightarrow \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-\frac{t^2}{2}} \, dt$$

as $N \to \infty$

Central Limit Theorem

$N \sim 10$ good
$N \sim 25$ good

Note when moments don't exist, other "stable" distributions can pop out

$\gamma$ has

$$F(y) = \frac{1}{\pi} \int_{-\infty}^{y} \frac{1}{1+t^2} \, dt$$

Cauchy distribution
Sampling

Want to sample \( z_1, z_2, \ldots, z_n \),

sample size \( n \)

\[
\overline{z} = \frac{\sum_{i=1}^{n} z_i}{n} = \frac{1}{n} \sum_{i=1}^{n} \overline{z}_i
\]

or

\[
\overline{z} = \frac{\sum_{i=1}^{n} z_i w_i}{\sum_{i=1}^{n} w_i}
\]

weighted sum

What about these two estimators

\( t(z) \) an estimator of \( z \)

\[
E[t(z)] = \mu \quad \text{then}
\]

\[
E[t(z) - \mu] = \beta \quad \text{is called the bias}
\]

\[
\text{Var}[t(z)] = E[(t(z) - \mu - \beta)^2] = \sigma^2
\]

taxing variance of \( t \)
Sampling Continued

Want it to be $\beta = 0$ (unbiased)

and $\sigma^2$ small as possible (minimum variance)

use minimum variance linear estimators (best weighted sum)

maximum likelihood estimates:

construct a likelihood function and choose estimate to minimize it (in $L^2$):

$$\bar{z} = \frac{1}{n} \sum_{i=1}^{n} z_i$$

$$\sigma^2 = \frac{1}{n-1} \sum_{i=1}^{n} (z_i - \bar{z})^2$$

one maximum likelihood estimates
Efficiency of Monte Carlo

\[
\overline{Z} = \frac{1}{n} \sum_{i=1}^{n} Z_i
\]

\[
\frac{\overline{Z} - \mu}{\frac{\sigma}{\sqrt{n}}} \text{ is approximately } N(0,1)
\]

So

\[
\overline{Z} \pm k \frac{\sigma}{\sqrt{n}}
\]

\[
k = 1 \quad \rho = 0.68
\]

\[
k = 2 \quad \rho = 0.90
\]

\[
k = 3 \quad \rho = 0.995
\]

to get a 99\% "confidence interval"

to be \(1/\alpha\) of its previous width requires \(\frac{\sigma}{\sqrt{n}} = \frac{\delta}{10\sqrt{n}}\) \(k = 10^2\)

So we say that Monte Carlo converges like \(n^{-1/2}\). Thus it is prudent to find ways to make \(\sigma\) as small as possible (variance reduction)
Regression

Line of best fit

\[ y = mx + b \]

\[ y_i, x_i \]

\[ L(m, b) = \sum_{i=1}^{n} (mx_i + b - y_i)^2 \]

likelihood function

\[ \frac{\partial L}{\partial m} = \frac{\partial L}{\partial b} = 0 \]

\[ L_b = \sum_{i=1}^{n} 2(mx_i + b - y_i) \]

\[ L_m = \sum_{i=1}^{n} 2(mx_i + b - y_i) x_i \]

\[ \begin{align*}
\sum_{i=1}^{n} y_i &= m \sum_{i=1}^{n} x_i + b \\
\sum_{i=1}^{n} y_i x_i &= m \sum_{i=1}^{n} (x_i)^2 + b \sum_{i=1}^{n} x_i
\end{align*} \]

solve these
Regression Example

Continued

\[ \sum_{i=1}^{n} y_i = \bar{Y} \quad \sum_{i=1}^{n} x_i = \bar{X} \]

\[ \sum_{i=1}^{n} (x_i - \bar{X})^2 = XX \quad \sum_{i=1}^{n} x_i y_i = XY \]

\[ \begin{pmatrix} \bar{X} \\ XX \end{pmatrix} \begin{pmatrix} 1 \\ \bar{X} \end{pmatrix} \begin{pmatrix} m \\ b \end{pmatrix} = \begin{pmatrix} \bar{Y} \\ XY \end{pmatrix} \]

Solution (at home) give standard linear regression results.

Sampling

Fixed Sampling (n fixed)

Sequential Sampling (something other than n fixed)

Stratified Sampling break sample space into strata