Today: Fundamentals of Monte Carlo

What is Monte Carlo?

- Named at Los Alamos in 1940’s after the casino.

- Any method which uses \((pseudo)\)random numbers as an essential part of the algorithm.

\[ \text{Stochastic - not deterministic!} \]

- A method for doing \textit{highly dimensional integrals} by sampling the integrand.

- Often a Markov chain, called Metropolis MC.

Simple example: Buffon’s needle

Monte Carlo determination of \(\pi\)

Consider a square inscribed by circle.
Consider one quadrant.

By geometry:

\[
\frac{\text{area of } 1/4 \text{ circle}}{\text{area of square}} = \frac{\pi r^2/4}{r^2} = \frac{\pi}{4}.
\]

Simple MC like throwing darts at unit square target:

- Using RNG \(\in (0,1)\), pick a pair of coordinates \((x,y)\).
- Count \# of points (darts) in shaded section versus total.

\[
\text{Pts in shaded circle/pts in square} \sim \frac{\pi}{4}.
\]

\begin{verbatim}
hits = 0
DO n=1,NI
  x = (random #)
  y = (random #)
  distance2 = (x² + y²)
  If (distance2 \leq 1) hits = hits + 1
pi= 4 * hits/N
\end{verbatim}
MC is advantageous for high dimensional integrals
-the best general method

Consider an integral in the unit $D$-dimensional hypercube:

$$I = \int dx_1...dx_D f(x_1,...,x_D)$$

By conventional deterministic methods:
- Lay out a grid with $L$ points in each direction with $h = 1/L$
- Number of points is $N = L^D = h^{-D} \propto$ CPU time.

How does error scale with CPU time or Dimensionality?
- Error in trapezoidal rule goes as $\varepsilon = f''(x) h^2$ since
  $$f(x) = f(x_0) + f'(x_0) h + (1/2)f''(x_0)h^2 + ...$$
- “direct” CPU time $\propto \varepsilon^{-D/2}$.  ($\varepsilon \sim h^2$ and CPU $\sim h^{-D}$)
- But by sampling we find $\varepsilon^{-2}$.  ($\varepsilon \sim M^{-1/2}$ and CPU $\sim M$)
  To get another decimal place takes 100 times longer!
  But MC is advantageous for $D > 4!$

2/18/2013

**Improved Numerical Integration**

Integration methods in $D$-dimensions:
- Trapezoidal Rule: $\varepsilon \sim f''(x) h^2$
- Simpson’s Rule: $\varepsilon \sim f^4(x) h^4$
- ... generally: $\varepsilon \sim f^{(\alpha)}(x) h^{\alpha}$

And CPU time scales with sample points (grid size)
- CPU time $\sim h^D$  (e.g., 1-D like 1/L and 2-D like 1/L^2)
- Time to do integral: $T_{int} \sim \varepsilon^{-D/\alpha}$

By conventional deterministic methods:
- Lay out a grid with $L$ points in each direction with $h = 1/L$
- Number of points is $N = h^{-D} = L^D \propto$ CPU time.

Stochastic Integration (Monte Carlo)
- Monte Carlo: $\varepsilon \sim M^{-1/2}$  (sqrt of sampled points)
- CPU time: $T \sim M \propto \varepsilon^{-2}$.

> In the limit of small $\varepsilon$, MC wins if $D > 2\alpha!$
Other reasons to do Monte Carlo:
- Conceptually and practically simple.
- Comes with built in error bars.

Many methods of integration have been tried, and will be tried in this world of sin and woe. No one pretends that Monte Carlo is perfect or all-wise. Indeed, it has been said that Monte Carlo is the worst method except all those other methods that have been tried from time to time. Churchill 1947

Probability Distributions
- $P(x)dx$ = probability of observing a value in $(x,x+dx)$ is a probability distribution function (p.d.f.)
  $$\int dx \ P(x) = 1 \quad P(x) \geq 0$$

- $x$ can be either a continuous or discrete variable.
- Cumulative distribution:
  - Probability of $x<y$.
  - Useful for sampling
  $$c(y) = \int_{-\infty}^{y} dx \ P(x) \quad 0 \leq c(y) \leq 1$$

- Average or expectation
  $$\langle g(x) \rangle = \bar{g} = \int_{-\infty}^{\infty} dx \ P(x)g(x)$$

- Moments:
  - Zeroth moment $I_0=1$
  - Mean $<x>=I_1$
  - Variance $<(x-<x>)^2>=I_2-(I_1)^2$
  $$I_n = <x^n> = \int_{-\infty}^{\infty} dx \ P(x)x^n$$
Mappings of random variables

Let $p_x(x)dx$ be a probability distribution
Let $y=g(x)$ be a new variable

-- e.g., $y=g(x) = -\ln(x)$
with $0 < x \leq 1$, so $y \geq 0$

What is the pdf of $y$?
With $p_y(y)dy = p_x(x)dx$

$$p_y(y) = p_y(x) \left| \frac{dy}{dx} \right| = p_x(x) \left( \frac{dg(x)}{dx} \right)^{-1}$$

Example: $y=g(x) = -\ln(x)$

$\int p_y(y)dy = p_x(x) \left| \frac{dg(x)}{dx} \right|^{-1} dy = e^{-y} dy$

* Distributed exponentially, like in Poisson event, e.g. $y/\lambda$ has PDF of $\lambda e^{-\lambda y}$

2/18/2013

What is Mapping Doing?

Generate random deviate

$p(x)dx = \begin{cases} 
  dx & 0 < x < 1 \\
  0 & otherwise 
\end{cases}$

$\int_{-\infty}^{\infty} p(x)dx = 1$

PDF is normalized

Let $p(y)=f(y)$, then $y(x) = F^{-1}(x)$ (functional inverse), $dx/dy = f(y)$ or $x=F(y)$

- Allows a random deviate $y$ from a known probability distribution $p(y)$.
- The indefinite integral of $p(y)$ must be known and invertible.
- A uniform deviate $x$ is chosen from $(0,1)$ such that its corresponding $y$ on the definite-integral curve is desired deviate, i.e. $x = F(y)$. 

2/18/2013
Interpreting the Mapping

Let \( p(y) = f(y) \), then \( y(x) = F^{-1}(x) \), \( \frac{dx}{dy} = f(y) \) and \( x = F(y) \).

Since \( F(y) \) is area under \( p(y) = f(y) \), \( y(x) = F^{-1}(x) \) prescribes that

- Choose \( x = (0,1] \), then find value \( y \) that has that fraction \( x \) of area to left of \( y \), or \( x = F(y) \).
- Return that value of \( y \).

Example: Drawing from Poisson Distribution

\[
y = -\frac{\ln(x)}{\lambda}, \quad p(y) = \lambda e^{-\lambda y} \text{ and } x = F(y) = 1 - e^{-\lambda y}
\]

Note: \( x = 1 - e^{-\lambda y} \) \( x' = 1 - x = e^{-\lambda y} \) (which is still \( x' = (0,1] \)). So, indeed, \( y(x) = -\frac{\ln(x')}{\lambda} \).

Example: Drawing from Normal Gaussian

\[
p(y_1, y_2, \ldots) dy_1 dy_2 \ldots = p(x_1, x_2, \ldots) \left| \frac{\partial(x_1, x_2, \ldots)}{\partial(y_1, y_2, \ldots)} \right| dy_1 dy_2 \ldots
\]

**Box-Muller method:** get random deviates with normal distr. \( p(y) dy = \frac{1}{2\pi} e^{-y^2/2} dy \)

Consider: uniform deviates \((0,1)\), \( x_1 \) and \( x_2 \), and two quantities \( y_1 \) and \( y_2 \).

\[
y_1 = -\ln x_1 \cos 2\pi x_2 \quad y_2 = -\ln x_1 \sin 2\pi x_2
\]

Or, equivalently,

\[
x_1 = \exp(-[y_1^2 + y_2^2]/2) \quad x_2 = \frac{1}{2\pi} \arctan(y_2/y_1)
\]

where \( \frac{\partial(x_1, x_2, \ldots)}{\partial(y_1, y_2, \ldots)} = \left[ \begin{array}{l} \frac{1}{\sqrt{2\pi}} e^{-y_1^2/2} \\ \frac{1}{\sqrt{2\pi}} e^{-y_2^2/2} \end{array} \right] \) So, each \( y \) is independently normal distributed.

Better: Pick \( R^2 = v_1^2 + v_2^2 \) so \( x_1 = \sqrt{R^2} \) and \( \angle(v_1, v_2) = 2\pi x_2 \)

Advantage: no sine and cosine by using \( v_1/\sqrt{R^2} \) and \( v_2/\sqrt{R^2} \) and get two RNG per calculation (1 for now, 2 for next time).
Reminder: Gauss’ Central Limit Theorem

**Sample N values from** \( p(x) \, dx \), i.e. \((X_1, X_2, X_3, \ldots, X_N)\).

**Estimate mean** from \( y = (1/N) \sum x_i \).

**What is the pdf of mean?** Solve by Fourier transforms.

If you add together two random variables, you multiply together their characteristic functions:

\[
c_x(k) = \langle e^{ikx} \rangle = \int_{-\infty}^{\infty} dx \, P(x) e^{ikx} \quad \text{so} \quad c_x(k) c_y(k) = c_{x+y}(k)
\]

Then \( c_{x_1 + \ldots + x_N}(k) = c_x(k)^N \) and \( c_y(k) = c_x(k/N)^N \)

Taylor expand \( \ln[c_y(k)] = \sum_{n=1}^{\infty} \frac{(ik)^n}{n!} \kappa_n \)

Cumulants:

- **Mean**: \( \kappa_1 = \langle x \rangle = \bar{x} \)
- **Variance**: \( \kappa_2 = \langle (x-\bar{x})^2 \rangle = \sigma^2 \)
- **Skewness**: \( \kappa_3 / \sigma^3 = \langle (x-\bar{x})^3 / \sigma^3 \rangle \)
- **Kurtosis**: \( \kappa_4 / \sigma^4 = \langle (x-\bar{x})^4 / \sigma^4 \rangle > -3 \)

What happens to the reduced moments?
- The \( n=1 \) moment remains invariant.
- The rest get reduced by higher powers of \( N \).

\[
\lim_{N \to \infty} c_y(k) = e^{ik \kappa_1 - \frac{k^2}{2 \kappa_2} + \frac{k^4}{2 \kappa_4} + \ldots}
\]

\[
P(y) = (N / 2 \pi \kappa_2)^{1/2} \exp \left[ -\frac{(y - \kappa_1)^2}{2 \kappa_2} \right]
\]

Given enough averaging almost anything becomes a Gaussian distribution.

2/18/2013
Approach to normality

Gauss’ Central Limit Thm

For any population distribution, the distribution of the mean will approach Gaussian.

Figure 1. Distributions of sums of uniform random numbers, each compared with the normal distribution. (a) $R_1$, the uniform distribution. (b) $R_2$, the sum of two uniformly distributed numbers. (c) $R_3$, the sum of three uniformly distributed numbers. (d) $R_{12}$, the sum of twelve uniformly distributed numbers.

2/18/2013

Conditions on Central Limit Theorem

\[ I_n = \langle x^n \rangle = \int_{-\infty}^{\infty} dx \ P(x) x^n \]

- We need the first three moments to exist.
  - If $I_0$ is not defined $\Rightarrow$ not a pdf
  - If $I_1$ does not exist $\Rightarrow$ not mathematically well-posed.
  - If $I_2$ does not exist $\Rightarrow$ infinite variance. Important to know if variance is finite for Monte Carlo.
- Divergence could happen because of tails of distribution

\[ I_2 = \langle x^2 \rangle = \int_{-\infty}^{\infty} dx \ P(x) x^2 \]

- We need:

\[ \lim_{x \to \pm \infty} x^2 P(x) \to 0 \]

- Divergence because of singular behavior of P at finite x:

\[ \lim_{x \to 0} x P(x) \to 0 \]

2/18/2013
Multidimensional Generalization

- Suppose $r$ is an $m$ dimensional vector from a multidimensional pdf: $p(r)dr$.
- The mean is defined as before.
- The variance becomes the covariance, a positive symmetric $m \times m$ matrix:
  $$V_{ij} = \langle (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle) \rangle$$
- For sufficiently large $N$, the estimated mean ($y$) will approach the distribution:
  $$P(y)dy = \left[ \frac{2\pi}{N^m \det(v)} \right]^{1/2} \exp\left[ -(y_i - \langle y_i \rangle)^2 \frac{Nv^{-1}}{2}(y_j - \langle y_j \rangle)^2 \right]$$

2d histogram of occurrences of means

- Off-diagonal components of $v_{ij}$ are called the co-variance.
- Data can be uncorrelated, positively or negatively correlated depending on sign of $v_{ij}$
- Like a moment of inertia tensor
- 2 principal axes with variances
- Find axes with diagonalization or singular value decomposition
- Individual error bars on $x_1$ and $x_2$ can be misleading if correlated.