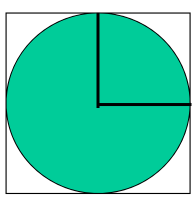
Today: Fundamentals of Monte Carlo

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.
  Stochastic - not deterministic!
- A method for doing *highly dimensional integrals* by sampling the integrand.
- Often a Markov chain, called Metropolis MC.

Simple example: Buffon's needle
Monte Carlo determination of π



Consider a square inscribed by circle.
Consider one quadrant.

By geometry: $$\frac{\text{area of 1/4 circle}}{\text{area of square}} = \frac{\pi r^2/4}{r^2} = \pi/4.$$

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.

Pts in shaded circle/pts in square $\sim \pi/4$.

```
hits = 0
DO  n=1,N{
    x = (random #)
    y = (random #)
    distance2 = (x^2 + y^2)
   If (distance2 ≤ 1)  hits = hits + 1}
pi= 4 * hits/N
```

MC is advantageous for high dimensional integrals -the best general method

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

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

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

$$f(x)=f(x_0) + f'(x_0)\, h + (1/2) f''(x_0) h^2 + \ldots$$

- "direct" CPU time $\propto \varepsilon^{-D/2}$. ($\varepsilon \sim h^2$ and CPU $\sim h^{-D}$)
- But by sampling we find ε^{-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 3

Improved Numerical Integration

Integration methods in D-dimensions:

- Trapeziodal Rule: $\varepsilon \sim f^{(2)}(x)\, h^2$
- 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:

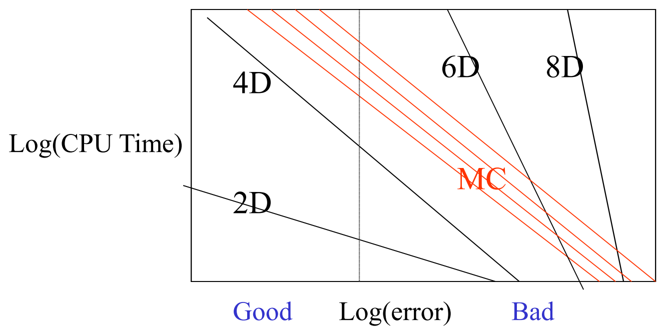
- Lay out a grid with L points in each direction with $h=1/L$
- Number of points is $N=h^{-D}=L^D \propto$ CPU time.

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 ε, MC wins if $D > 2\alpha$!

2/18/2013 4



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.)

$$\int dx\ P(x) = 1 \qquad P(x) \geq 0$$

- x can be either a continuous or discrete variable.
- Cumulative distribution:

Probability of x<y.
Useful for sampling

$$c(y) = \int_{-\infty}^{y} dx\ P(x) \quad 0 \leq c(y) \leq 1$$

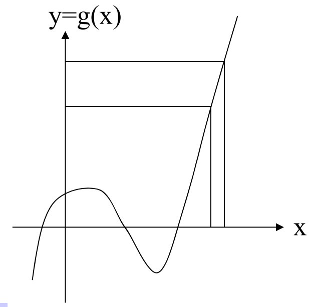
- Average or expectation

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

- 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 $p_y(y)dy=p_x(x)dx$

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

What happens when g is not monotonic?

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

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

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

What is Mapping Doing?

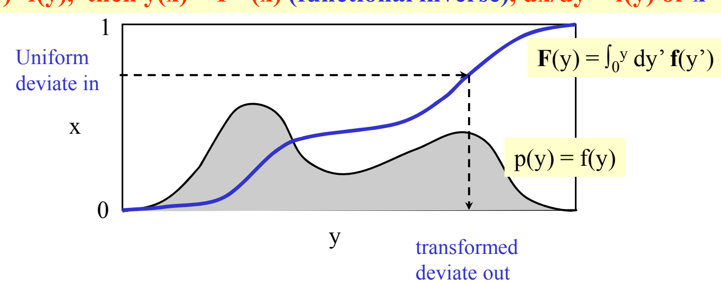
Generate random deviate

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

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

PDF is normalized

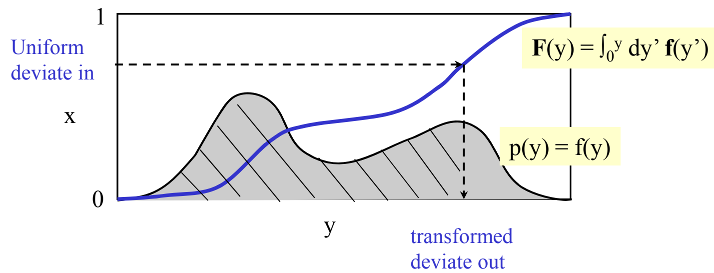
Let $p(y)=f(y)$, then $y(x) = F^{-1}(x)$ (functional inverse), $dx/dy = f(y)$ or $x=F(y)$



- 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)***.

Interpreting the Mapping

Let **p(y)=f(y), then y(x) = F^{-1}(x), 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= – ln(x)/λ , p(y)= $\lambda e^{-\lambda y}$ and x=F(y) = 1 – $e^{-\lambda y}$

Note: **x = 1 – $e^{-\lambda y}$** or x' = 1- x = $e^{-\lambda y}$ (which is still x'=(0,1]).
So, indeed, y(x) = – ln(x')/λ

2/18/2013 9

Example: Drawing from Normal Gaussian

$$p(y_1,y_2,\ldots)dy_1dy_2\ldots = p(x_1,x_2,\ldots)\left|\frac{\partial(x_1,x_2,\ldots)}{\partial(y_1,y_2,\ldots)}\right|dy_1dy_2\ldots$$

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.

$$y_1 = \sqrt{-\ln x_1}\cos 2\pi x_2 \qquad y_2 = \sqrt{-\ln x_1}\sin 2\pi x_2$$

Or, equivalently,

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

where $\frac{\partial(x_1,x_2,\ldots)}{\partial(y_1,y_2,\ldots)} = -\left[\frac{1}{\sqrt{2\pi}}e^{-y_1^2/2}\right]\left[\frac{1}{\sqrt{2\pi}}e^{-y_2^2/2}\right]$

So, each y is independently normal distributed.

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

Advantage: no sine and cosine by using $v_1/\sqrt{R^2}$ and $v_2/\sqrt{R^2}$ and
And get two RNG per calculation (1 for now, 2 for next time)

2/18/2013 10

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 $y = (1/N)\Sigma x_i$.

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

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

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

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

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

cumulants

2/18/2013 11

Cumulants: κ_n

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

What happens to the reduced moments? $\tilde{\kappa}_n = \kappa_n N^{1-n}$

- The n=1 moment remains invariant.
- The rest get reduced by higher powers of N.

$$\lim_{N\to\infty} \quad c_y(k) = e^{ik\kappa_1 - k^2\kappa_2/2N - ik^3\kappa_3/6N^2 \ldots}$$

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

Given enough averaging almost anything becomes a Gaussian distribution.

2/18/2013 12

Approach to normality

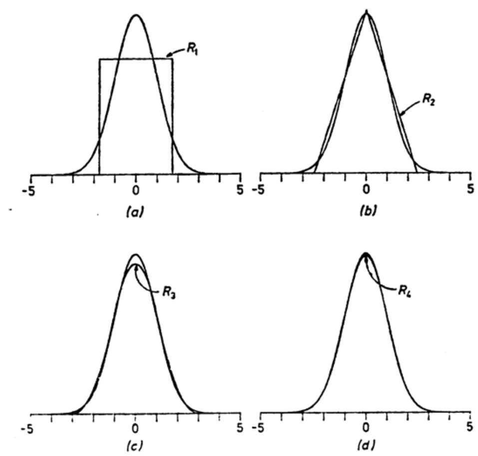


Figure 1. Distributions of sums of uniform random numbers, each compared with the normal distribution. (*a*) R_1, the uniform distribution. (*b*) R_2, the sum of two uniformly distributed numbers. (*c*) R_3, the sum of three uniformly distributed numbers. (*d*) R_{12}, the sum of twelve uniformly distributed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

Gauss'
Central Limit Thm

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

Conditions on Central Limit Theorem

$$I_n = < x^n > = \int_{-\infty}^{\infty} dx \, P(x) x^n$$

- 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

$$I_2 = < x^2 > = \int_{-\infty}^{\infty} dx \, P(x) x^2$$

- We need: $\lim_{x \to \pm\infty} \quad x^3 P(x) \to 0$
- Divergence because of singular behavior of P at finite x:

$$\lim_{x \to 0} \quad x P(x) \to 0$$

Multidimensional Generalization

- Suppose **r** is an m dimensional vector from a multidimensional pdf: p(**r**)d^m^r.
- The mean is defined as before.
- The variance becomes the covariance, a positive symmetric m x m matrix :

$$V_{i,j} = <(x_i - < x_i >)(x_j - < x_j >)>$$

- For sufficiently large N, the estimated mean (**y**) will approach the distribution:

$$P(y)dy = \left[\frac{2\pi}{N}\det(v)\right]^{-1/2} \exp[-(y_i - < y_i >)\frac{Nv^{-1}}{2}(y_j - < y_j >)]$$

2d histogram of occurrences of means

- Off-diagonal components of ν_{ij} are called the co-variance.
- Data can be uncorrelated, positively or negatively correlated depending on sign of ν_{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.