Monte Carlo Simulation Techniques

Ji Qiang

Accelerator Modeling Program
Accelerator Technology & Applied Physics Division
Lawrence Berkeley National Laboratory

CERN Accelerator School, Thessaloniki, Greece
Nov. 13, 2018

Introduction: What is the Monte Carlo Method?

- What is the Monte Carlo method?

- Monte Carlo method is a (computational) method that relies on the use of random sampling and probability statistics to obtain numerical results for solving deterministic or probabilistic problems

"…a method of solving various problems in computational mathematics by constructing for each problem a random process with parameters equal to the required quantities of that problem. The unknowns are determined approximately by carrying out observations on the random process and by computing its statistical characteristics which are approximately equal to the required parameters."

J. H. Halton, "A retrospective and prospective survey of the Monte Carlo Method," SIAM Review, Vol. 12, No. 1 (1970).

Introduction: What can the Monte Carlo Method Do?

- Give an approximate solution to a problem that is too big, too hard, too irregular for deterministic mathematical approach
- Two types of applications:

  a) The problems that are stochastic (probabilistic) by nature:
  - particle transport,
  - telephone and other communication systems,
  - population studies based on the statistics of survival and reproduction.

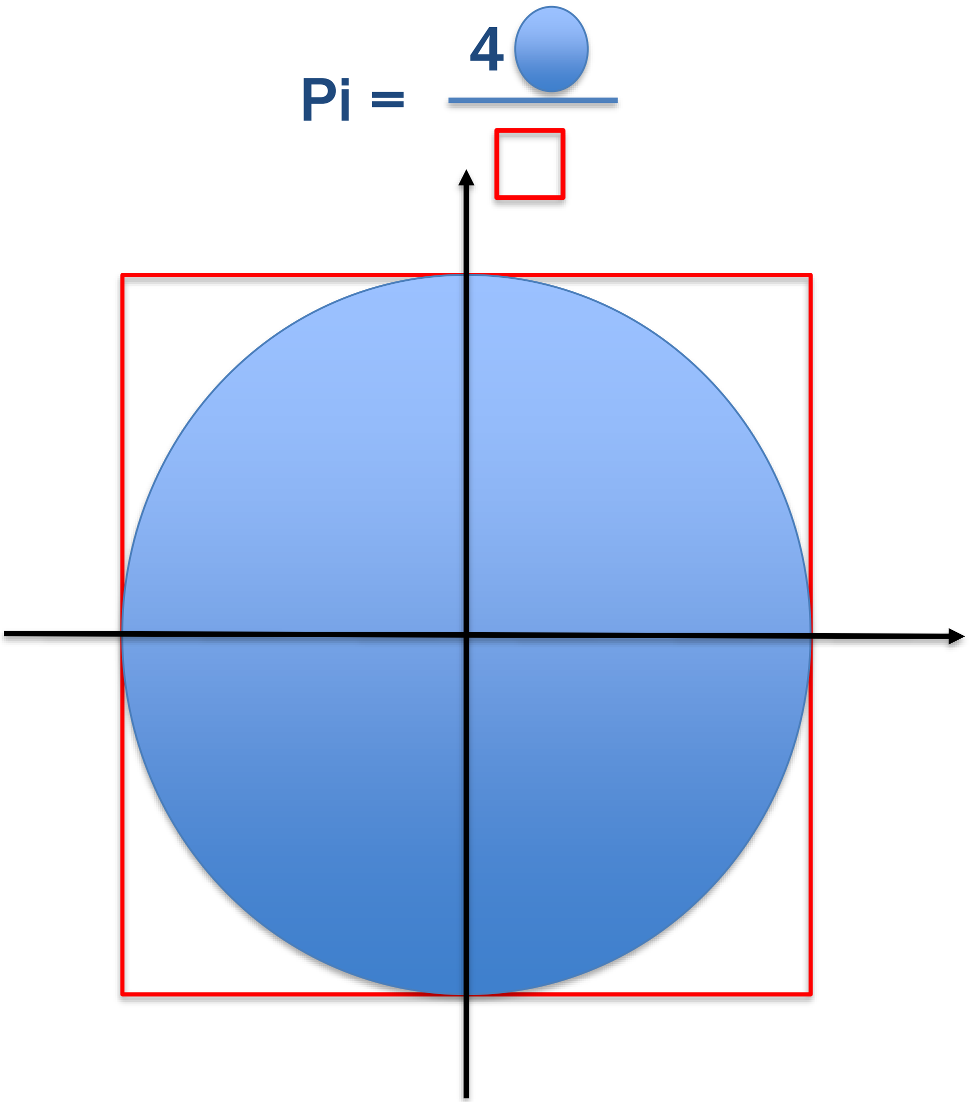
  b) The problems that are deterministic by nature:
  - the evaluation of integrals,
  - solving partial differential equations

- It has been used in areas as diverse as physics, chemistry, material science, economics, flow of traffic and many others.

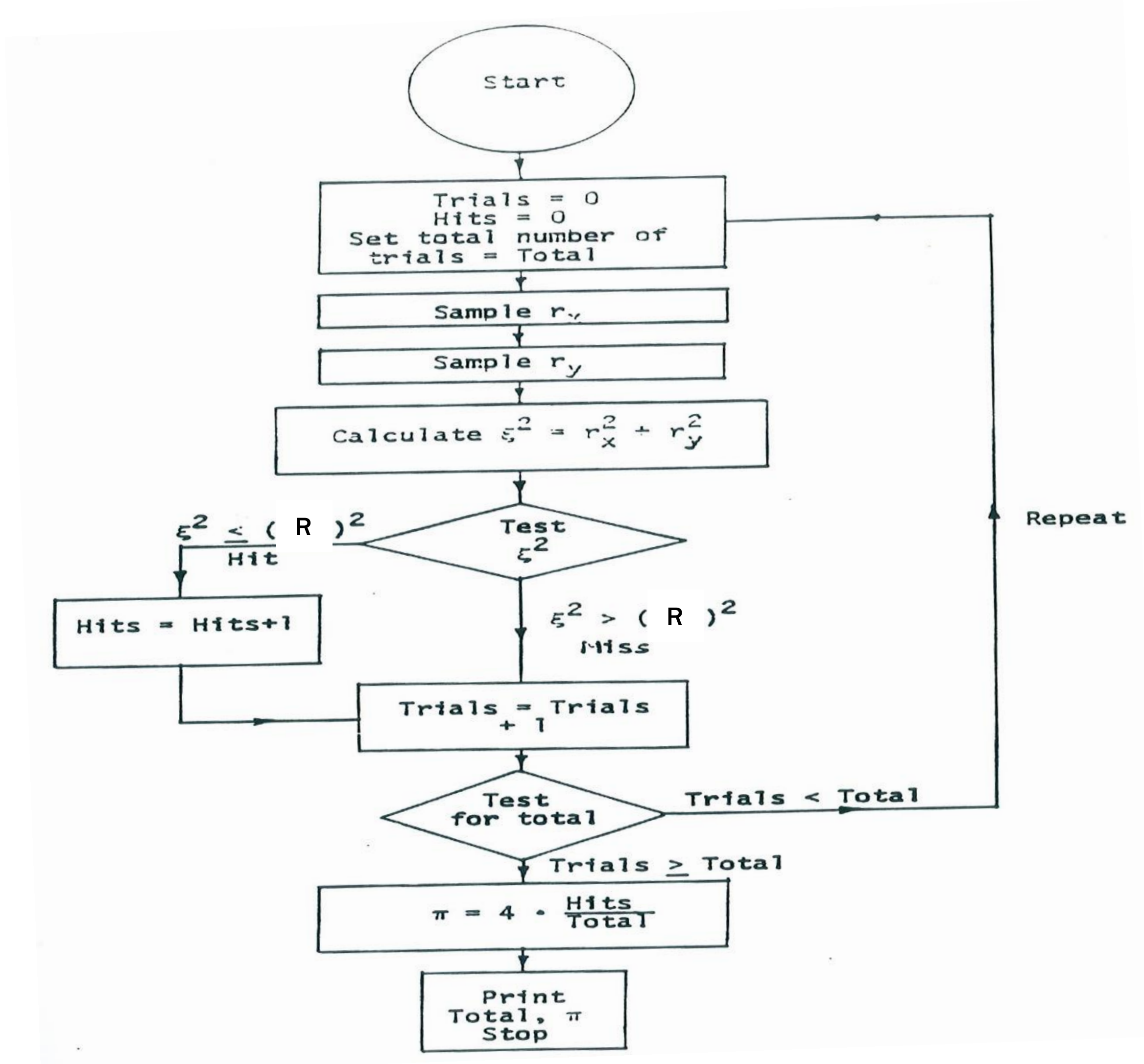
Brief History of the Monte Carlo Method

- 1772 Comte de Buffon - earliest documented use of random sampling to solve a mathematical problem (the probability of needle crossing parallel lines).
- 1786 Laplace suggested that pi could be evaluated by random sampling.
- Lord Kelvin used random sampling to aid in evaluating time integrals associated with the kinetic theory of gases.
- Enrico Fermi was among the first to apply random sampling methods to study neutron moderation in Rome.
- 1947 Fermi, John von Neuman, Stan Frankel, Nicholas Metropolis, Stan Ulam and others developed computer-oriented Monte Carlo methods at Los Alamos to trace neutrons through fissionable materials during the Manhattan project.

An Example of Monte Carlo Method: Calculation of Pi



Flow Diagram of Monte Carlo Calculation of Pi



Introduction: Basic Steps of a Monte Carlo Method

Monte-Carlo methods generally follow the following steps:

1. Define a domain of possible inputs and determine the statistical properties of these inputs

2. Generate many sets of possible inputs that follows the above properties via random sampling from a probability distribution over the domain

3. Perform deterministic calculations with these input sets

4. Aggregate and analyze statistically the results

The error on the results typically decreases as 1=1/sqrt(N)

Introduction: Major Components of a Monte Carlo Algorithm

- Probability distribution functions (pdf's) - the physical (or mathematical) system must be described by a set of pdf's.
- Random number generator - a source of random numbers uniformly distributed on the unit interval must be available.
- Sampling rule - a prescription for sampling from the specified pdf, assuming the availability of random numbers on the unit interval.
- Scoring (or tallying) - the outcomes must be accumulated into overall tallies or scores for the quantities of interest.
- Error estimation - an estimate of the statistical error (variance) as a function of the number of trials and other quantities must be determined.
- Variance reduction techniques - methods for reducing the variance in the estimated solution to reduce the computational time for Monte Carlo simulation.
- Efficient implementation on computer architectures - parallelization and vectorization

Statistics Background

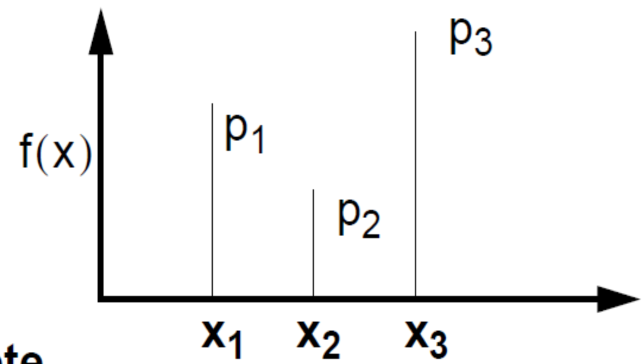
- Random variable is a real number associated with a random event whose occurring chance is determined by an underlying probability distribution.

- Discrete random variable – discrete probability distribution
  - face of a dice
  - type of reactions
  - etc
- Continuous random variable – continuous probability distribution
  - spatial position
  - time of occurrance
  - etc

Statistics Background: Discrete Random Variable

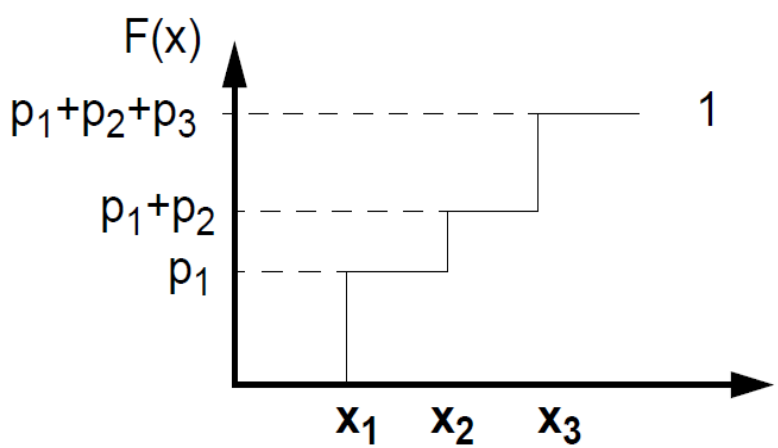
Probability Density Function (PDF) - discrete

- $f(x_i), \quad f(x_i) = p_i\delta(x - x_i)$
- $0 \le f(x_i)$
- $\sum_i f(x_i)(\Delta x_i) = 1$ or $\sum_i p_i = 1$



Cumulative Distribution Function (CDF) - discrete

- $F(x) = \sum_{x_i < x} p_i = \sum_{x_i < x} f(x_i)\Delta x_i$
- $0 \le F(x) \le 1$
-



Statistics Background: Discrete Random Variable

- If *X* is a random variable, then *g*(*X*) is also a random variable. The expectation of *g*(*X*) is defined as

$$E(g(X)) = \langle g(X) \rangle = \sum_i p_i g(x_i).$$

- From the definition of the expected value of a function, we have the property that

<constant> = constant

and that for any constants λ1, λ2 and two functions *g*1, *g*2,

$$\langle \lambda_1 g_1(X) + \lambda_2 g_2(X) \rangle = \lambda_1 \langle g_1(X) \rangle + \lambda_2 \langle g_2(X) \rangle.$$

Statistics Background: Discrete Random Variable

- An important application of expected values is to the powers of *X*.
- The *n*th moment of *X* is defined as the expectation of the *n*th power of *X*,

$$\langle X^n \rangle \equiv \sum_i p_i x_i^n;$$

- The central moments of *X* are given by

$$\langle g_n \langle X \rangle \rangle \equiv \langle (X - \mu)^n \rangle = \sum_i p_i (x_i - \langle X \rangle)^n.$$

- The second central moment has particular significance,

$$\langle (X - \mu\rangle^2 \rangle = \langle (X - \langle X \rangle)^2 \rangle = \sum_i p_i (x_i - \mu)^2$$

$$= \sum_i p_i x_i^2 - \langle X \rangle^2 = \langle X^2 \rangle - \langle X \rangle^2,$$

Statistics Background: Discrete Random Variable

- The second moments is also called the *variance of X* or var{*x*}.
- The square root of the variance is a measure of the dispersion of the random variable.
- It is referred to as the *standard deviation* and sometimes the *standard error*.
- The variance of a function of the random variable, *g*(*X*), can be determined as

$$
\begin{aligned}
\mathrm{var}\{g(X)\} &= \langle (g(X) - \langle g(X)\rangle)^2 \rangle \\
&= \sum_i p_i g^2(x_i) - \left(\sum_i p_i g(x_i)\right)^2 \\
&= \langle g(X)^2 \rangle - \langle g(X)\rangle^2 .
\end{aligned}
$$

Statistics Background: Discrete Random Variable

- Consider two real-valued functions, $g_1(X)$ and $g_2(X)$.
- They are both random variables, but they are not in general independent.
- Two random variables are said to be independent if they derive from independent events.

$$\mathrm{var}\{\lambda_1 g_1(X) + \lambda_2 g_2(X)\} = \lambda_1^2 \mathrm{var}\{g_1(X)\} + \lambda_2^2 \mathrm{var}\{g_2(X)\}$$
$$+2[\lambda_1\lambda_2\langle g_1(X) g_2(X)\rangle - \lambda_1\lambda_2\langle g_1(X)\rangle\langle g_2(X)\rangle].$$

- Let X and Y be random variables; the expectation of the product is

$$\langle XY\rangle = \sum_{i,j} p_{ij} x_i y_j.$$

- If X and Y are independent, $p_{ij} = p_{1i}.p_{2j}$ and

$$\langle XY\rangle = \sum_i p_{1i} x_i \sum_j p_{2j} y_j = \langle X\rangle\langle Y\rangle.$$

$$\mathrm{var}\{\lambda_1 X + \lambda_2 Y\} = \lambda_1^2 \mathrm{var}\{X\} + \lambda_2^2 \mathrm{var}\{Y\}.$$

Statistics Background: Discrete Random Variable

- When X and Y are not necessarily independent, we introduce a new quantity: the covariance, which is a measure of the degree of independence of the two random variables X and Y:

$$\mathrm{cov}\{X, Y\} = \langle XY \rangle - \langle X \rangle \langle Y \rangle.$$

- The covariance equals 0 when X and Y are independent and

$$\mathrm{cov}\{X, X\} = \mathrm{var}\{X\}.$$

- Note that zero covariance does not by itself imply independence of the random variables

  - Let X be a random variable that may be −1, 0, or 1 with equal probabilities, and define $Y = X^2$. Obviously,

$$\mathrm{cov}\{XY\} = \langle XY \rangle - \langle X \rangle \langle Y \rangle = 0.$$

Statistics Background: Discrete Random Variable

- The covariance can have either a positive or negative value.
- Another quantity derived from the covariance is the correlation coefficient,

$$\rho(X, Y) = \frac{\mathrm{cov}\{X, Y\}}{[\mathrm{var}\{X\}\mathrm{var}\{Y\}]^{1/2}},$$

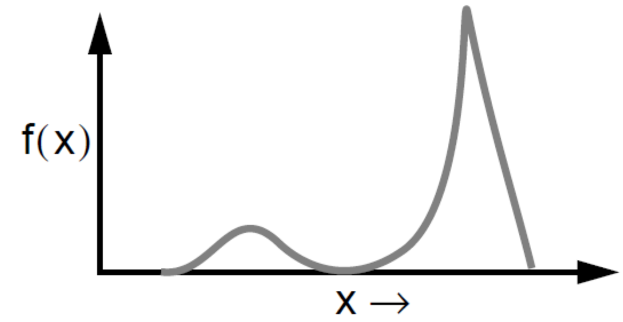
so that

$$-1 \leq \rho(X, Y) \leq 1.$$

Statistics Background: Continuous Random Variable

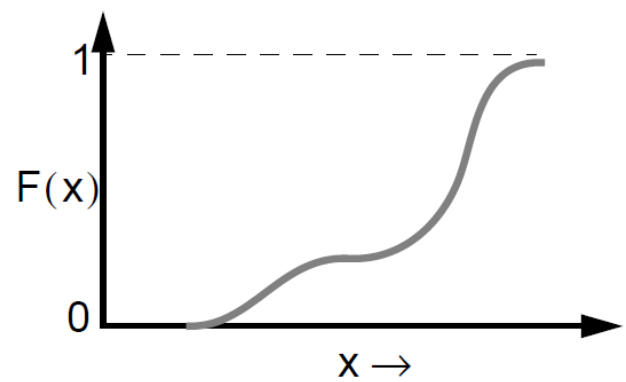
Probability Density Function (PDF) - continuous

- $f(x)$, $f(x)dx = P\{x \le x' \le x + dx\}$
- $0 \le f(x)$, $\int_{-\infty}^{\infty} f(x)dx = 1$
- Probability$\{ a \le x \le b \} = \int_a^b f(x)dx$



Cumulative Distribution Function (CDF) - continuous

- $F(x) = \int_{-\infty}^{x} f(x')dx' = P\{x' \le x\}$
- $0 \le F(x) \le 1$
- $0 \le \frac{d}{dx}F(x) = f(x)$
- $\int_a^b f(x')dx' = P\{a \le x \le b\} = F(b) - F(a)$



Statistics Background: Continuous Random Variable

- The expected value of any function of the random variable is defined as

$$E(g(X)) \equiv \int_{-\infty}^{\infty} g(x) f(x)\, \mathrm{d}x,$$

and, in particular, $E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\, \mathrm{d}x.$

- The variance of any function of the random variable is defined as

$$\mathrm{var}\{g(X)\} = E(g^2(X)) - [E(g(X))]^2.$$

1. For a random variable C, which is constant
   $\mathrm{var}\{C\} = 0$.
2. For a constant C and random variable X,
   $\mathrm{var}\{CX\} = C^2\mathrm{var}\{X\}$.
3. For independent random variables X and Y,
   $\mathrm{var}\{X + Y\} = \mathrm{var}\{X\} + \mathrm{var}\{Y\}$.

Statistics Background: Continuous Random Variable

Given the function *G* as:

$$G \equiv \sum_{i=1}^{N} \lambda_i g_i(X_i),$$

The expected value of *G* is

$$E(G) = \langle G \rangle = E\left(\sum_{i=1}^{N} \lambda_i g_i(X_i)\right) = \sum_{i=1}^{N} \lambda_i \langle g_i(X) \rangle,$$

The variance of *G* is

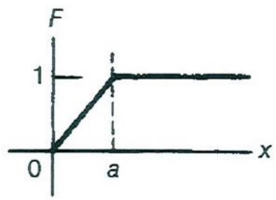
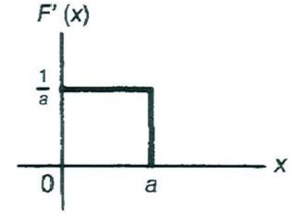
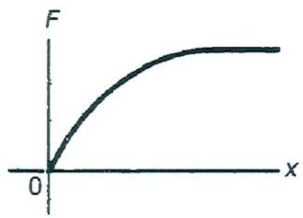
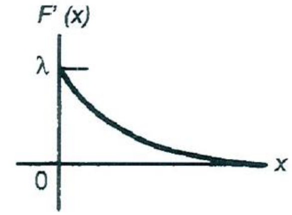
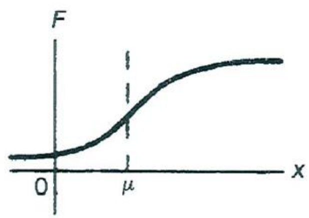
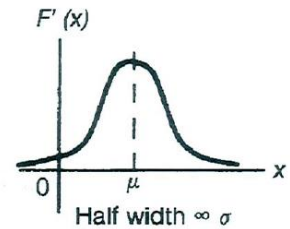
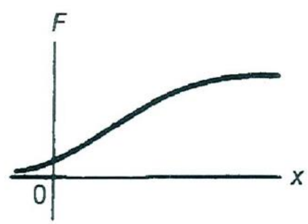
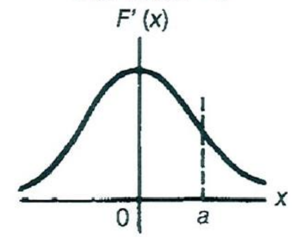
$$\mathrm{var}\{G\} = \sum_{i=1}^{N} \lambda_i^2 \mathrm{var}\{g_i(X)\}.$$

Let $\lambda_i = \frac{1}{N}$ and all the $g_i(x)$ be identical and equal to $g(x)$;

$$E(G) = E\left(\frac{1}{N}\sum_{i=1}^{N} g(X_i)\right) = \frac{1}{N}\sum_{i=1}^{N} \langle g(X) \rangle = \langle g(X) \rangle.$$

$$\mathrm{var}\{G\} = \mathrm{var}\left\{\frac{1}{N}\sum_{i=1}^{N} g(X_i)\right\} = \sum \frac{1}{N^2}\mathrm{var}\{g(X) = \frac{1}{N}\mathrm{var}\{g(X)\}.$$

Statistics Background: Some Common PDFs

Distribution function	F(x)		F′(x)		⟨x⟩	var(x)
Uniform	$0,\ x<0$ $x,\ 0\le x\le a$ $1,\ x>a$		$0,\ x<0,\ x>a$ $\frac{1}{a},\ 0<x<a$		$\frac{1}{2}a$	$\frac{1}{12}a$
Exponential	$0,\quad x<0$ $1-\exp(-\lambda x),\ x\ge 0$		$0,\quad x<0$ $1-\exp(-\lambda x),\ x\ge 0$		$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Normal $\phi(x\|\mu\sigma)$ μ = mean σ^2 = variance	$\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{x}$ $\times \exp\left\|\frac{-(t-\mu)^2}{2\sigma^2}\right\| dt$		$\frac{1}{\sigma\sqrt{2\pi}} \exp\left\|-\frac{(x-\mu)^2}{2\sigma^2}\right\| dt$		μ	σ^2
Cauchy (Lorentz)	$\frac{1}{2}+\frac{1}{\pi}\tan^{-1}\left(\frac{x}{a}\right)$		$\frac{a}{a^2+x^2}$		?	∞

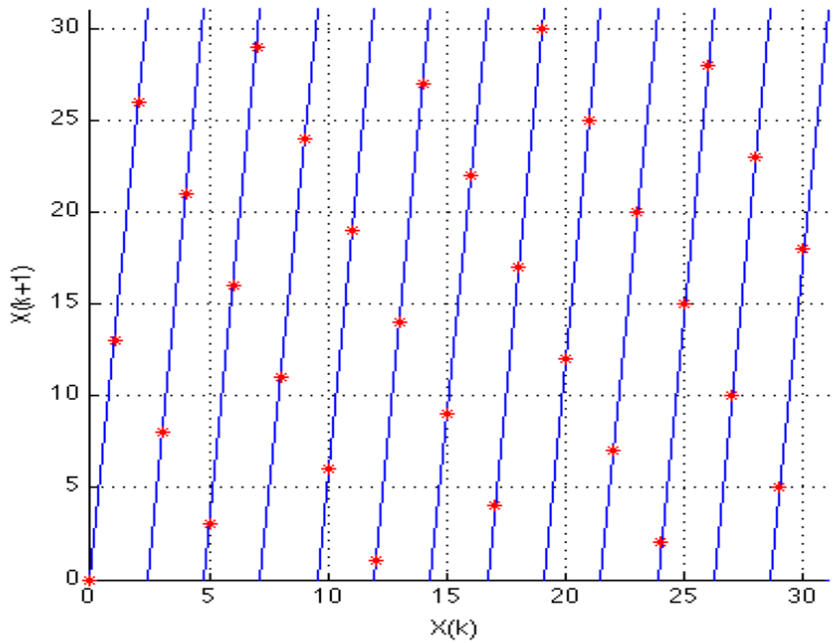
Sampling of Distribution: Pseudo-Random Number

The following formula is known as a Linear Congruential Generator or LCG.

$$x_{k+1} = (a*x_k + c) \bmod m$$

What's happening is we're drawing the line y = a * x + c "forever", but using the mod function (like wrap-around) to bring the line back into the square [0,m] x [0,m] (m is power of 2 -1). By doing so, we've induced a map on the integers 0 through m which, if we've chosen a, c and m carefully, will do an almost perfect shuffle.

Example: X(k+1) = mod(13*X(k)+0,31)



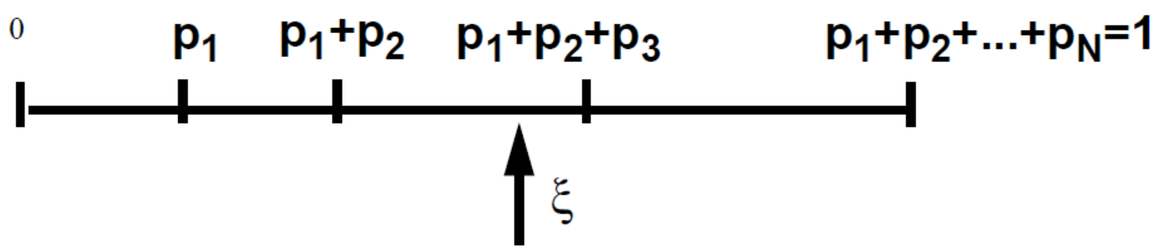
Sampling of Distribution: Pseudo-Random Number

- A more ambitious LCG has the form:
  SEED =(16807 SEED + 0) mod 2147483647
- A uniformly distributed random number between 0 and 1:
  R = SEED/2147483647
- This is the random number generator that was used in MATLAB until version 5 and ran0 in Numerical Recipe (NR). It shuffles the integers from 1 to 2,147,483,646, and then repeats itself.
- Serial correlations present in ran0.
- Ran1 in NR, uses the ran0 for its random value, but it shuffles the output to remove low-order serial correlations. A random deviate derived from the *j*th value in the sequence, I_j , is output not on the *j*th call, but rather on a randomized later call, j +32 on average.
- Ran2 in NR combines two different sequences with different periods so as to obtain a new sequence whose period is the least common multiple of the two periods. The period of ran2 is ~10^{18}.

Sampling of Distribution: Discrete Distribution

Sampling from a given discrete distribution

Given $f(x_i) = p_i$ and $\sum_i p_i = 1$, $i = 1, 2, \ldots, N$



and $0 \le \xi \le 1$, then $P(x = x_k) = p_k = P(\xi \in d_k)$ or

$$\sum_{i=1}^{k-1} p_i \le \xi < \sum_{i=1}^{k} p_i$$

Sampling of Distribution: Discrete Distribution

Example — Collision Type Sampling

Assume (for photon interactions):

$$\mu_{tot} = \mu_{cs} + \mu_{fe} + \mu_{pp}$$

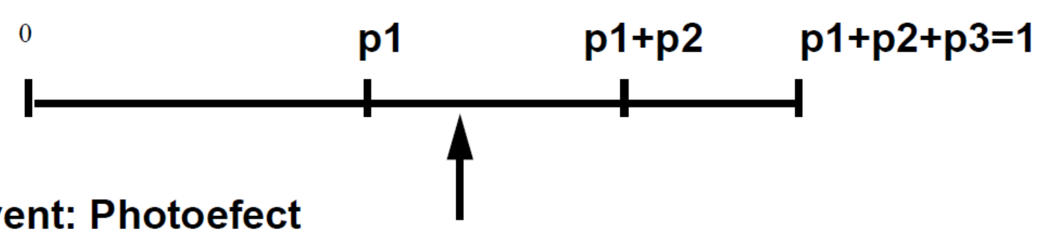
Define

$$p_1 = \frac{\mu_{cs}}{\mu_{tot}},\ p_2 = \frac{\mu_{fe}}{\mu_{tot}},\ \textbf{and}\ p_3 = \frac{\mu_{pp}}{\mu_{tot}}$$

with

$$\sum_{i=1}^{3} p_i = 1.$$

Then



Collision event: Photoefect

$$p_1 < \xi < p_1 + p_2$$

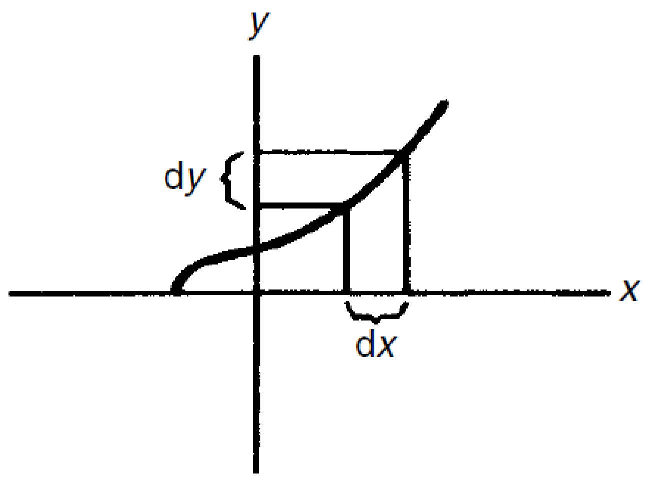
Sampling of Distribution: Transformation of Random Variables

Given that X is a random variable with pdf $f_X(x)$ and $Y = y(X)$, then

$$f_Y(y) = f_X(x)\left|\frac{\mathrm{d}x}{\mathrm{d}y}\right| = f_X(x)\left|\frac{\mathrm{d}y}{\mathrm{d}x}\right|^{-1}.$$

reflecting the fact that all the values of X in dx map into values of Y in dy

.



Sampling of Distribution: Transformation of Random Variables

Consider the linear transformation $Y = a + bX$

$$f_Y(y) = |b^{-1}| f_X\left(\frac{y-a}{b}\right).$$

Suppose X is distributed normally with mean 0 and variance 1:

$$f_X(x) = \phi'(x|0,1) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{x^2}{2}\right], \quad -\infty < x < \infty,$$

and Y is a linear transformation of X, $Y = \sigma X + \mu$. Then

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right].$$

The random variable Y is also normally distributed, but its distribution function is centered on μ and has variance σ^2.

Sampling of Distribution: Continuous Distribution

Sampling from a given continuous distribution

- If $f(y)$ and $F(y)$ represent PDF and CDF of a random variable *y*,
- if is a random number *x* distributed uniformly on [0,1] with PDF $f_x(x)=1$,
- if *y* is such that

$$F(y) = x$$

then for each x there is a corresponding y, and the variable *y* is distribute according to the probability density function $f(y)$.

Sampling of Distribution: Example 1

Sample the probability density function:

$$f_Y(y) = \frac{2}{\pi}\frac{1}{1+y^2}, \quad 0 < y < \infty;$$

The cumulative distribution function is

$$F_Y(y) = \int_0^y \frac{2}{\pi}\frac{1}{1+u^2}\,\mathrm{d}u = \frac{2}{\pi}\tan^{-1} y = \xi.$$

Solving this equation for *Y* yields

$$Y = \tan\frac{\pi}{2}\xi.$$

Sampling of Distribution: Example 2

Sample the probability density function:

$$f_R(r) = r\exp\left[-\frac{1}{2}r^2\right], \quad 0 < r < \infty,$$

The cumulative distribution function is

$$F_R(r) = \int_0^r u\exp\left[-\frac{1}{2}u^2\right]\,\mathrm{d}u = 1 - \exp\left[-\frac{1}{2}r^2\right] = \xi.$$

Solving this equation for *Y* yields

$$R = [-2\log(1-\xi)]^{\frac{1}{2}}.$$

Sampling of Distribution: Example 3

- Sample the Gaussian probability density function:

$$\phi'(y|0,1) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}y^2\right], \quad -\infty < y < \infty,$$

- Form a 2D Gaussian probability density function:

$$f(y_1, y_2) = \phi'(y_1|0,1)\,\phi'(y_2|0,1) = \frac{1}{2\pi} \exp\left[-\frac{1}{2}\left(y_1^2 + y_2^2\right)\right].$$

- Change the coordinate:

$$Y_1 = R\cos\Phi,$$

$$Y_2 = R\sin\Phi,$$

$$\phi'(y_1)\phi'(y_2)\,\mathrm{d}y_1\,\mathrm{d}y_2 = \left(\exp\left[-\frac{1}{2}r^2\right] r\,\mathrm{d}r\right)\left(\frac{1}{2\pi}\,\mathrm{d}\phi\right).$$

$$\Phi = 2\pi\xi_2.$$

$$Y_1 = [-2\log\xi_1]^{\frac{1}{2}} \cos 2\pi\xi_2,$$

$$Y_2 = [-2\log\xi_1]^{\frac{1}{2}} \sin 2\pi\xi_2;$$

Sampling of the Sum of Several Distributions

Sample the probability density function:

$$f_X(x) = \sum_{i=1}^{n} \alpha_i g_i(x); \quad \alpha_i \geq 0, \quad g_i(x) \geq 0, \quad \int_{\Omega_0} g_i(x) \neq 1, \quad \int_{\Omega_0} f_X(x)\,\mathrm{d}x = 1.$$

Form a new pdf so that the β_i are effectively probabilities for the choice of an event *i*. Let us select event *m* with probability β_m. Then sample *X* from $h_m(x)$ for that *m*.

$$h(x) = \sum_{m=1}^{n} \beta_m h_m(x). \qquad \beta_i \geq 0, \quad \sum_{i=1}^{n} \beta_i = 1.$$

$$P\{X \leq x\} = \sum_{m=1}^{n} \beta_m \int_0^x h_m(t)\,\mathrm{d}t. \qquad h_i(x) \geq 0, \quad \int_{\Omega_0} h_i(x)\,\mathrm{d}x = 1,$$

$$f_X(x) = \sum_{i=1}^{n} \alpha_i \left[\int_{\Omega_0} g_i(u)\,\mathrm{d}u\right]\left[\frac{g_i(x)}{\int g_i(w)\,\mathrm{d}w}\right].$$

$$h_i(x) = \frac{g_i(x)}{\int g_i(w)\,\mathrm{d}w} \qquad \beta_i = \alpha_i \int_{\Omega_0} g_i(u)\,\mathrm{d}u$$

Sampling of the Sum of Several Distributions: Example

Sample the probability density function:

$$f_X(x) = \frac{3}{5}\left(1 + x + \frac{1}{2}x^2\right), \quad 0 < x < 1;$$

Rewrite the probability density function as:

$$f_X(x) = \frac{3}{5} \times 1 + \frac{3}{5} \times \frac{1}{2} \times 2x + \frac{3}{5} \times \frac{1}{2} \times \frac{1}{3} \times 3x^2$$

The resulting β *and h(x) are* :

$$\beta_1 = \tfrac{3}{5}, \quad h_1 = 1,$$
$$\beta_2 = \tfrac{3}{10}, \quad h_2 = 2x,$$
$$\beta_3 = \tfrac{1}{10}, \quad h_3 = 3x^2,$$

$$\sum_{i=1}^{n=3} \beta_i = 1.$$

1. sample ξ_0;
2. if $\xi_0 \le \frac{6}{10}$, $i = 1$;
3. else if $\xi_0 \le \frac{9}{10}$, $i = 2$;
4. else $i = 3$.

Sampling of Some Common PDFs

Probability Density Function		*Direct Sampling Method*
Linear: (L1, L2)	$f(x) = 2x, \quad 0 < x < 1$	$x \leftarrow \sqrt{\xi}$
Exponential: (E)	$f(x) = e^{-x}, \quad 0 < x$	$x \leftarrow -\log \xi$
2D Isotropic: (C)	$f(\vec{\rho}) = \frac{1}{2\pi}, \quad \vec{\rho} = (u, v)$	$u \leftarrow \cos 2\pi\xi_1$ $v \leftarrow \sin 2\pi\xi_1$
3D Isotropic: (I1, I2)	$f(\vec{\Omega}) = \frac{1}{4\pi}, \quad \vec{\Omega} = (u, v, w)$	$u \leftarrow 2\xi_1 - 1$ $v \leftarrow \sqrt{1-u^2}\cos 2\pi\xi_2$ $w \leftarrow \sqrt{1-u^2}\sin 2\pi\xi_2$
Maxwellian: (M1, M2, M3)	$f(x) = \frac{2}{T\sqrt{\pi}}\sqrt{\frac{x}{T}}\, e^{-x/T}, \quad 0 < x$	$x \leftarrow T(-\log\xi_1 - \log\xi_2 \cos^2 \frac{\pi}{2}\xi_3)$
Watt Spectrum: (W1, W2, W3)	$f(x) = \frac{2e^{-ab/4}}{\sqrt{\pi a^3 b}}\, e^{-x/a} \sinh\sqrt{bx}, \quad 0 < x$	$w \leftarrow a(-\log\xi_1 - \log\xi_2 \cos^2 \frac{\pi}{2}\xi_3)$ $x \leftarrow w + \frac{a^2 b}{4} + (2\xi_4 - 1)\sqrt{a^2 b w}$
Normal: (N1, N2)	$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$	$x \leftarrow \mu + \sigma\sqrt{-2\log\xi_1}\cos 2\pi\xi_2$

Sampling of Complex Distribution

For a pdf whose cdf is not analytically available, one can numerically calculate:

$$F(y_n) = \int_0^{y_n} f_Y(y)\,\mathrm{d}y = \frac{n}{N}, \quad n = 0, 1, 2, \ldots, N,$$

For a uniformly sampled random number ξ from u(0,1), find *n* such that

$$F(y) = \xi$$

$$\frac{n}{N} < \xi < \frac{n+1}{N}.$$

The value for *Y* may be calculated by linear interpolation

$$Y = y_n + (y_{n+1} - y_n)u,$$

$$u = N\xi - n, \quad 0 < u < 1.$$

This method corresponds to approximating a pdf by a piecewise-constant function with the area of each piece a fixed fraction.

Sampling of Multi-Dimensional Distribution

- If the random variable in each dimensional is independent of each other, the sampling of multi-dimensional pdf can be done in each dimension.
- if the marginal and conditional functions can be determined, sampling the multivariate distribution will then involve sampling the sequence of univariate distributions.

$$m(x) \equiv \int_{-\infty}^{\infty} f(x, y)\, \mathrm{d}y, \qquad f(y|x) = \frac{f(x, y)}{\int_{-\infty}^{\infty} f(x, y)\, \mathrm{d}y} = \frac{f(x, y)}{m(x)}.$$

$$P\{X \le x\} = \int_{-\infty}^{x} \int_{-\infty}^{\infty} f(t, y)\, \mathrm{d}y\, \mathrm{d}t = \int_{-\infty}^{x} m(t)\, \mathrm{d}t,$$

Example:

$$f(\mathbf{x}) = \frac{1}{8\pi} \mathrm{e}^{-r}, \qquad \text{where } r = \sqrt{x^2 + y^2 + z^2}. \qquad f(\mathbf{x})\, \mathrm{d}V = \frac{1}{2} r^2 \mathrm{e}^{-r}\, \mathrm{d}r \frac{\sin\theta\, \mathrm{d}\theta}{2} \frac{\mathrm{d}\phi}{2\pi},$$

$$R = -\sum_{i=1}^{3} \log \xi_i = -\log(\xi_1 \times \xi_2 \times \xi_3).$$

$$\Phi = 2\pi\xi_5. \qquad \cos\Theta = 2\xi_4 - 1.$$

$$Z = R\cos\Theta, \quad X = R\sin\Theta\cos\Phi, \quad Y = R\sin\Theta\sin\Phi.$$

Sampling of Distribution: Rejection Method

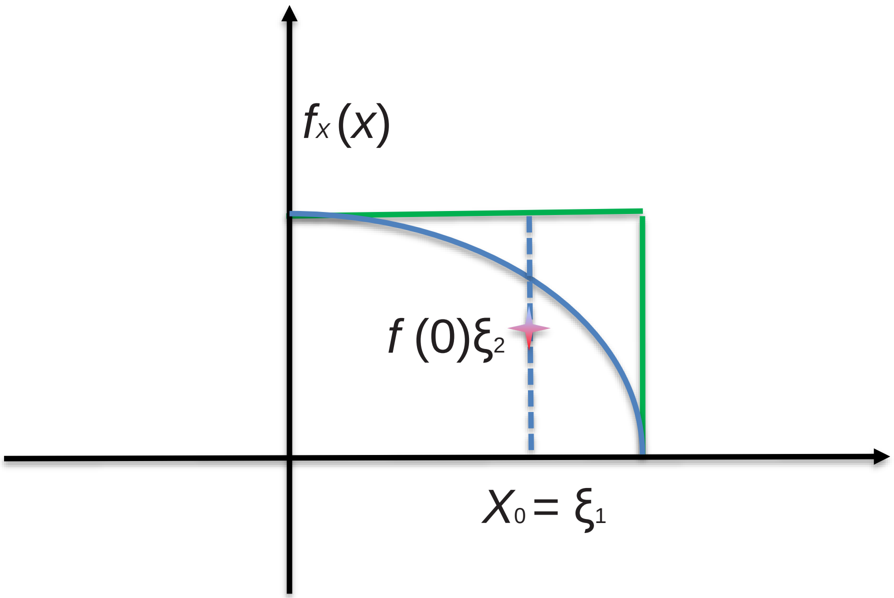
- *Rejection method*
  - Generate a uniform random number x_0 between xmin and xmax.
  - Generate another uniform random number ξ_2 between 0 and 1.

$$\xi_2 \le \frac{f(x_0)}{f_{\max}}; \quad accept\ x_0$$

$$otherwise; \quad reject\ x_0$$

- Apply to complex distribution function

Sampling of Distribution: Geometric View of the Rejection Method



Stated in geometric way, points are chosen uniformly in the smallest rectangle that encloses the curve $f_X\,(x)$. The ordinate of such a point is $X_0 = \xi_1$; the abscissa is $f\,(0)\xi_2$. Points lying above the curve are rejected. Points below are accepted; their ordinates $X = X_0$ have distribution $f_X\,(x)$.

Sampling of Distribution: Rejection Method

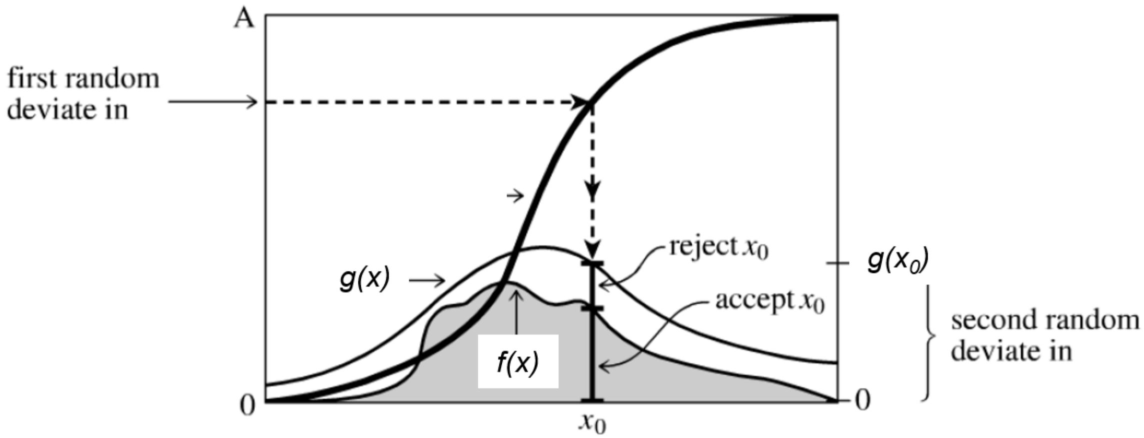
- The goal is to sample an *X* from a pdf, *f* (*x*).
- We can more easily sample a random variable *Z* from pdf *g*(*z*).
- This *Z* is accepted, *X* = *Z*, with probability *h*(*z*); else sample another *Z*.
- Then *X* has a pdf proportional to *h*(*z*)*g*(*z*).

$$P\{Z < x|\text{success}\} = \text{distribution of } Z'\text{s coming from a rejection algorithm}$$

$$= \frac{\int_{-\infty}^{x} h(z)g(z)\ \mathrm{d}z}{\int_{-\infty}^{\infty} h(z)g(z)\ \mathrm{d}z};$$

$$h(z) = \frac{f(z)/g(z)}{B_h}, \qquad \frac{h(z)g(z)}{\int_{-\infty}^{\infty} h(t)g(t)\ \mathrm{d}t} = \frac{f(x)}{B_h \int_{-\infty}^{\infty} h(t)g(t)\ \mathrm{d}t} = f(x).$$

Sampling of Distribution: Rejection Method



1. Find a comparison function $g(x)$ that can be sampled, so that $g(x) \geq f(x), \quad \forall x$
2. Draw a random deviate x_0 from $f(x)$
3. Draw a uniform random deviate y_0 from $\mathcal{U}(0, g(x_0))$
4. If $y_0 < f(x_0)$, accept x_0, otherwise discard it
5. Repeat 2.–4. until you have enough values

- The efficiency of rejection method depends on $f(x)/g(x)$

Sampling of Distribution: Rejection Method Example

1) Sample the following pdf:

$$f_X(x) = \frac{4}{\pi} \frac{1}{1 + x^2}, \quad 0 < x < 1,$$

1. $X_0 = \xi_1$;
2. if $\xi_2 > \frac{1}{(1+X_0^2)}$, then repeat from 1; else $X = X_0$.

2) Sample a uniform distribution inside a unit circle:

1. $X = \xi_1$ and $Y = \xi_2$;
2. if $(X^2 + Y^2 > 1)$, reject and repeat from 1;
3. otherwise set $\Phi = \tan^{-1} Y/X$.

$$\cos\Phi = \frac{X}{(X^2 + Y^2)^{\frac{1}{2}}} = \frac{\xi_1}{(\xi_1^2 + \xi_2^2)^{\frac{1}{2}}},$$

$$\sin\phi = \frac{Y}{(X^2 + Y^2)^{\frac{1}{2}}} = \frac{\xi_2}{(\xi_1^2 + \xi_2^2)^{\frac{1}{2}}}.$$

Sampling of Distribution: Markov Chain Monte Carlo (MCMC)

- The Markov Chain Monte Carlo (MCMC) is a very simple and powerful method.
- It can be used to sample essentially any distribution function regardless of analytic complexity in any number of dimensions.
- Complementary disadvantages are that sampling is correct only asymptotically and that successive variables produced are correlated, often very strongly.
- Some initial samplings will be thrown away (also called burn-in phase).
- A set of random variables xi ; i = 1; : : : is a Markov chain if:

$$P(x_{i+1} = x|x_1; \dots ; x_i) = P(x_{i+1} = x|x_i)$$

in other words, the distribution of X(i+1) depends only on the previous draw, and is independent of X(0);X(1); : : : ;X(i-1)

Sampling of Distribution: MCMC

- Ergodicity: A Markov chain is ergodic if it satisfies the following conditions:

  - Irreducibile: Any set A can be reached from any other set B with nonzero probability
  - Positive recurrent: For any set A, the expected number of steps required for the chain to return to A is nite
  - Aperiodic: For any set A, the number of steps required to return to A must not always be a multiple of some value k

  It means that all possible states will be reached at some time.

- Reversibility/Detailed balance: A Markov chain is reversible if there exists a distribution f(x) such that: $f(x_{i+1})P(x_{i+1}|x_i) = f(x_i)P(x_i|x_{i+1})$; for all i.

Sampling of Distribution: MCMC

If a Markov chain is reversible:

$$\int_\alpha f(\alpha)P(x_{i+1}=\beta|x_i=\alpha) = \int_\alpha f(\beta)P(x_{i+1}=\alpha|x_i=\beta) =$$

$$= f(\beta)\int_\alpha P(x_{i+1}=\alpha|x_i=\beta) = f(\beta)$$

This property is also called detailed balance. $f(x)$ is then the equilibrium distribution of the Markov chain.

Provided that a Markov chain is ergodic it will converge to a unique stationary distribution, also known as an equilibrium distribution.

This stationary distribution is determined entirely by the transition probabilities of the chain; the initial value of the chain is irrelevant in the long run.

Sampling of Distribution: Metropolis MCMC

1. Choose a proposal distribution $p(x)$ and an initial value x_1
2. Select a candidate step using $p(x)$, so that: $\hat{x} = x_i + \varepsilon,\ \varepsilon \sim p(x)$
3. If $f(\hat{x}) > f(x_i)$ accept $x_{i+1} = \hat{x}$, otherwise accept $x_{i+1} = \hat{x}$ with a probability $f(\hat{x})/f(x_i)$, else reject and start again at 2.
4. Continue at 2. until you have enough values
5. Discard the early values (burning phase) which are influenced by the choice of x_1

Gaussian update: $p(x) = \mathcal{N}(0, \sigma)$

Uniform update: $p(x) = \mathcal{U}(-v, v)$

Sampling of Distribution: Metropolis-Hastings MCMC

- A symmetric proposal distribution might not be optimal
- Boundary effects: less time is spent close to boundaries, which might not be well sampled
- A correction factor, the Hastings ratio, is applied to correct for the bias
- The Hastings ratio usually speeds up convergence
- The choice of the proposal distribution becomes however more important

Sampling of Distribution: Metropolis-Hastings MCMC

- The proposed update is $\hat{x} \sim Q(x; x_i)$
- The probability to accept the next point, A is modified so that instead of:

$$A = \min\left(1, \frac{f(\hat{x})}{f(x_i)}\right)$$

the probability is corrected by the Hastings ratio:

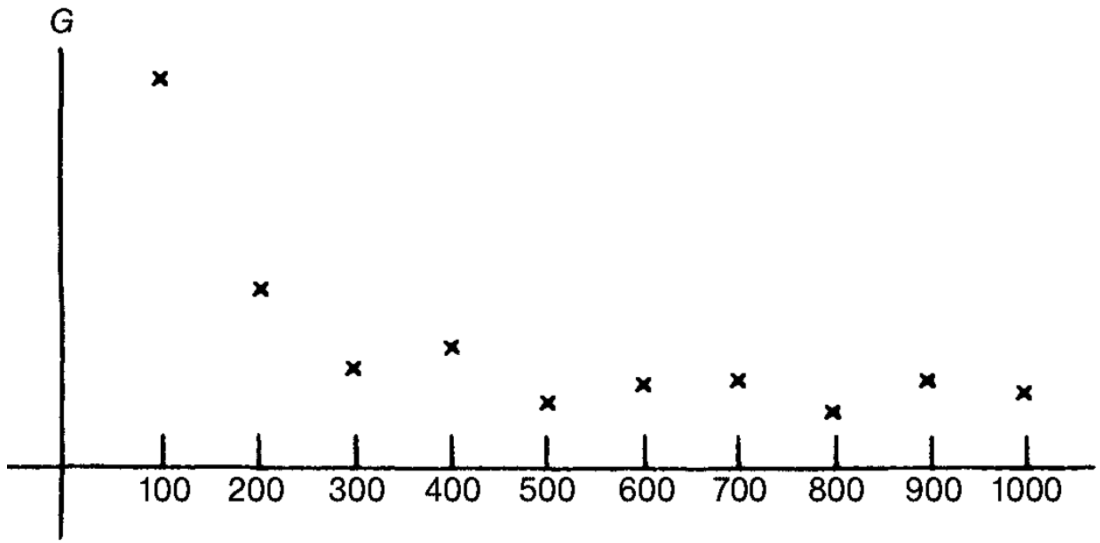
$$A = \min\left(1, \frac{f(\hat{x})}{f(x_i)} \frac{Q(\hat{x}; x_i)}{Q(x_i; \hat{x})}\right)$$

- If $Q(\alpha; \beta) = Q(\beta; \alpha)$, this is the Metropolis algorithm

Sampling of Distribution: Some Practice Considerations in MCMC

- Check the acceptance ratio: Values between 30 and 70% are conventionally accepted
- Discard the burn-in phase: The autocorrelation function is a standard way to check if the initial value has become irrelevant or not
- The width of the proposal distribution (e.g. for a Gaussian update or for a uniform update) should be tuned during the burn-in phase to set the rejection fraction in the right range.
- Reflection can be used when an edge of f (x) is reached.
- Thinning the chain: In order to break the dependence between draws in the Markov chain, one might keep only every dth draw of the chain.

Sampling of Distribution: Convergence of Markov Chain



- Monitor behavior of <*G*> with length of the Metropolis random walk.

- When the variance of multiple chains is much less than the variance within the chains.

Numerical Integration: Application of the Monte Carlo Method

Given the following integral:

$$G = \int_{\Omega_0} g(x) f(x)\, dx, \qquad f(x) \geq 0, \quad \int_{\Omega_0} f(x)\, dx = 1,$$

We draw a set of variables $X_1, X_2, \ldots, X_N$ from $f(x)$ (i.e. we "sample" the probability distribution function $f(x)$ and form the arithmetic mean:

$$G_N = \frac{1}{N} \sum_i g(X_i).$$

The integration result will be:

$$G = G_N + \text{error}.$$

with

measure the spread of g(x)

$$|\text{error}| = \epsilon \cong \frac{\sigma_1}{N^{\frac{1}{2}}}, \qquad \sigma_1^2 = \int g^2(x) f(x)\, dx - G^2.$$

The error will decrease as 1/sqrt(N) independent of the dimensionality of the integral. This is the key advantage of the MC over numerical quadrature.

Numerical Integration: Importance Sampling for Variance Reduction

Rewrite the integral as:

$$G = \int \left[\frac{g(x)f(x)}{\tilde{f}(x)} \right] \tilde{f}(x)\, \mathrm{d}x, \qquad \tilde{G}_N = \frac{1}{N} \sum_{i=1}^{N} \frac{g(X_i)f(X_i)}{\tilde{f}(X_i)}$$

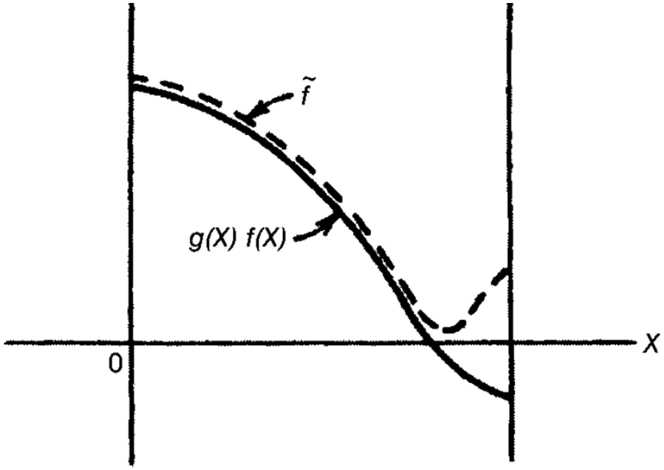
The new variance will be:

$$\mathrm{var}\{G\}_{\tilde{f}} = \int \left[\frac{g^2(x)f^2(x)}{\tilde{f}^2(x)} \right] \tilde{f}(x)\, \mathrm{d}x - G^2.$$

The optimal $\tilde{f}(x)$:

$$\tilde{f}(x) = \frac{g(x)f(x)}{G}.$$

In practice, a ‘‘similar’’ functions will reduce the variance.



Numerical Integration: Importance Sampling Example

Given the following integral:

$$G = \int_0^1 \cos\left(\frac{\pi x}{2}\right) \mathrm{d}x.$$

A straightforward Monte Carlo algorithm would be to sample X_i uniformly on (0, 1) with $f_1(x) = 1$, and to sum the quantity:

$$g_1(x) = \cos\left(\frac{\pi x}{2}\right). \qquad \mathrm{var}\{g_1\} = 0.0947\ldots.$$

If we approximate the original function as:

$$\cos\left(\frac{\pi x}{2}\right) = 1 - \frac{\pi^2 x^2}{8} + \frac{\pi^4 x^4}{2^4 4!} - \cdots$$

$$\tilde{g} = \frac{g_1}{\tilde{f}} = \frac{2}{3}\frac{\cos\left(\frac{\pi x}{2}\right)}{1 - x^2} \qquad \tilde{f}(x) = \frac{3}{2}\left(1 - x^2\right).$$

The variance of the new function is:

$$\mathrm{var}\{\tilde{g}\} = 0.000990.$$

Two orders of magnitude reduction of the variance!

Numerical Integration: Correlation Methods for Variance Reduction

Consider the following integral:

$$G = \int g(x) f(x)\, dx$$

Rewrite the integral as:

$$G = \int (g(x) - h(x)) f(x)\, dx + \int h(x) f(x)\, dx,$$

has analytical solution

If:

$$\text{var}\{(g(x) - h(x))\}_f \ll \text{var}\{g(x)\}_f,$$

$$G \cong \int h(x) f(x)\, dx + \frac{1}{N} \sum_{i=1}^{N} [g(x_i) - h(x_i)],$$

- In particular, if $|g(x) - h(x)|$ is approximately constant for different values of $h(x)$, then correlated sampling will be more efficient than importance sampling.
- Conversely, if $|g(x) - h(x)|$ is approximately proportional to $|h(x)|$, then importance sampling would be the appropriate method to use.

Numerical Integration: Correlation Methods for Variance Reduction Example

Consider the following integral:

$$G = \int_0^1 e^x \, dx,$$

The variance of g is

var{g} = 0.242

Rewrite the integral as:

$$\int_0^1 e^x \, dx = \int_0^1 (e^x - (1 + x)) \, dx + \frac{3}{2},$$

The new variance is

$$\text{var}\{e^x - (1 + x)\} = 0.0437,$$

More than order of magnitude reduction of the variance!

Numerical Integration: Antithetic Variates

Consider the following integral:

$$G = \int_0^1 g(x)\, dx,$$

Exploits the decrease in variance that occurs when random variables are negatively correlated and rewrite the integral as:

$$G = \int_0^1 \frac{1}{2}[g(x) + g(1-x)]\, dx.$$

$$G_N = \frac{1}{N}\sum_{i=1}^{N} \frac{1}{2}[g(x_i) + g(1-x_i)],$$

Give exactly *G* with zero variance for linear *g*. For nearly linear functions, this method will substantially reduce the variance.

For example: $G = \int_0^1 e^{-x}\, dx;$ Var(G) = 0.242.

The variance of the rewritten integral: Var(G_N) = 0.0039

More than order of magnitude reduction of the variance!

Sampling of Distribution: Non-Random Sampling

- *Quasi-Monte Carlo Sampling*
  - sampling a distribution can be generated from the transformation of sampling a uniform distribution
- A non-random sequence that has low discrepancy (a measure of deviation from uniformity) can be used to simulate the uniform distribution.
- Hammersley/Halton sequence in *p+1* dimension is defined as follows:

$$X = \{(j-1/2)/N, \Phi_2(j), \Phi_3(j), .., \Phi_r(j), ..., \Phi_p(j)\},\ j = 1, ..., N.$$

$$j = a_0 + a_1 r^1 + \cdots$$

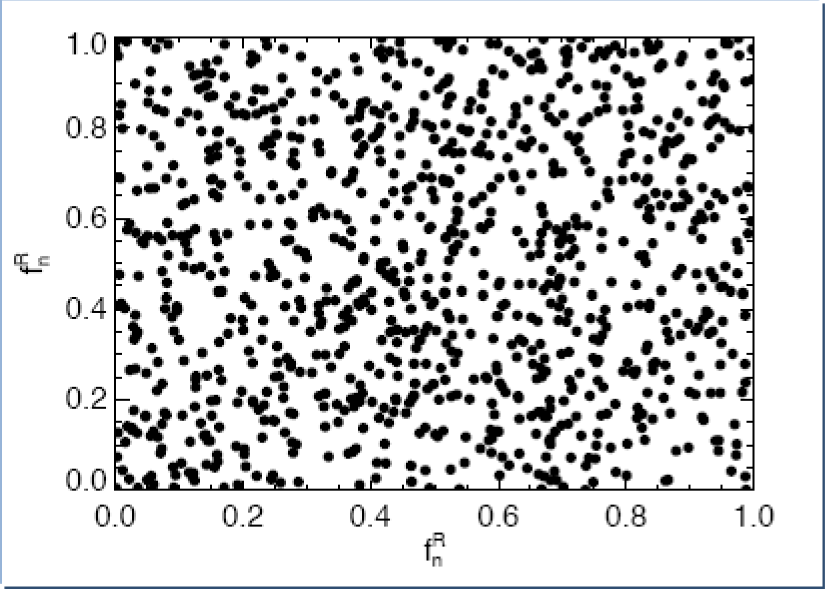
$$\Phi_r(j) = a_0 r^{-1} + a_1 r^{-2} + \cdots$$

ϕ(j) is the radical inversion function in the base of a prime number *r*.
Example: using base 3, and j = 1, 2, 3, 4 one obtains the sequence:

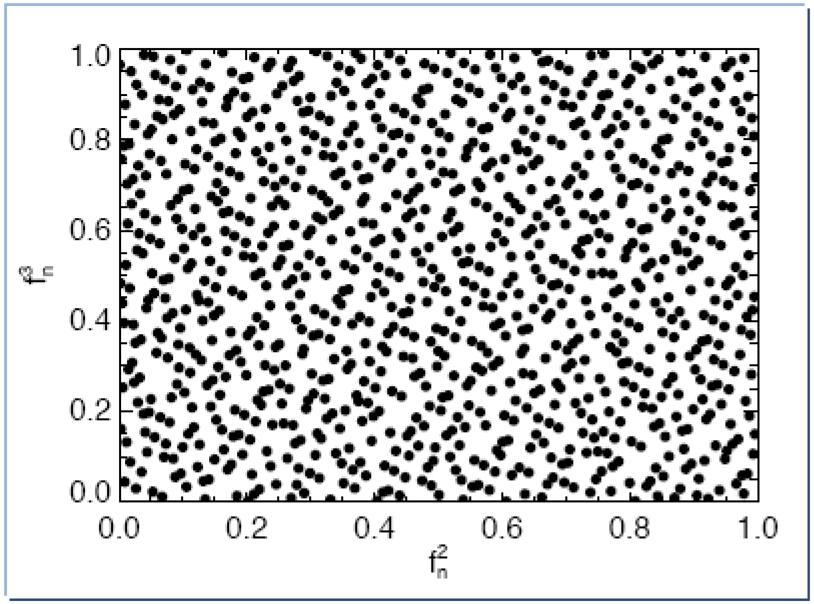
$$\Phi_3(1) = 1/3, \Phi_3(2) = 2/3, \Phi_3(3) = 1/9, \Phi_3(4) = 4/9$$

- Fluctuation of this type of sequence scales as 1/N whereas a random Monte Carlo sampling scales as 1/sqrt(N).

Sampling of Distribution: Non-Random Sampling

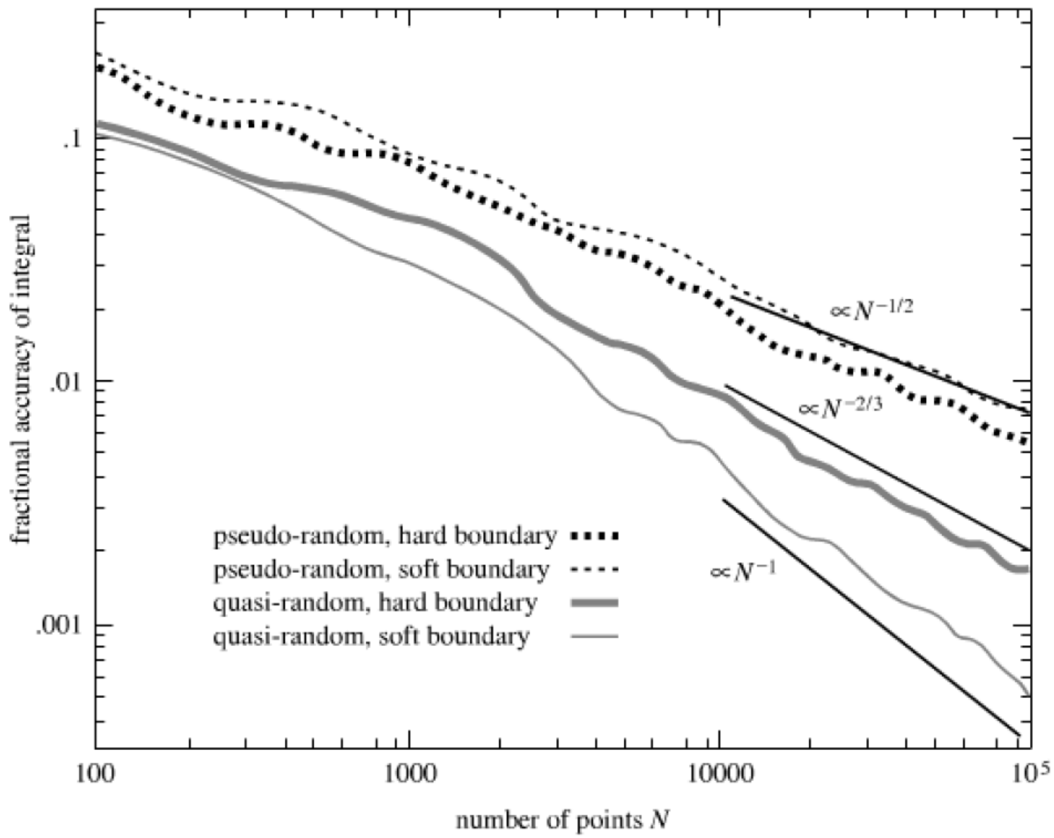


Random Monte Carlo Sampling



Hammersley Sequence with base 2 and 3

Numerical Integration Using Quasi-Random Sampling



Convergence in some cases of numerical integration can reach 1/N

Summary

- Brief introduction to the Monte-Carlo method
- Brief review some statistic backgrounds
- Several methods of sampling of distribution
  - direct inversion
  - rejection method
  - Markov chain Monte Carlo
- Numerical integration using the Monte-Carlo method and variance reduction
  - Importance sampling
  - Correlation method
  - Antithetic variate method
  - Non-random sampling method