蒙特卡罗方法
(Monte Carlo simulation)

引言 (Introduction)

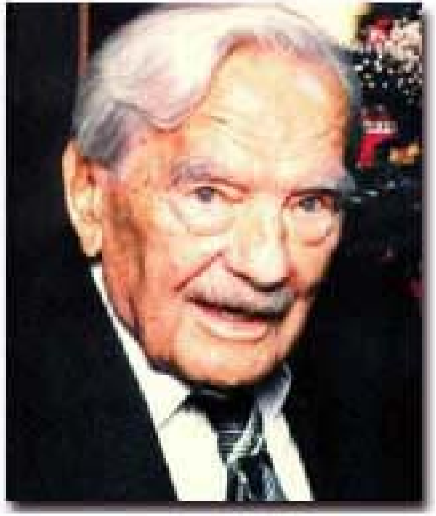
Monte Carlo方法：

亦称统计模拟方法，**statistical simulation method**

→利用随机数进行数值模拟的方法

Monte Carlo名字的由来：

- 是由**Metropolis**在二次世界大战期间提出的：**Manhattan**计划，研究与原子弹有关的中子输运过程；
- **Monte Carlo**是摩纳哥（**monaco)**的首都，该城以赌博闻名



Nicholas Metropolis (1915-1999)

Monte-Carlo, Monaco

Monte Carlo模拟的应用：

自然现象的模拟：

宇宙射线在地球大气中的传输过程；

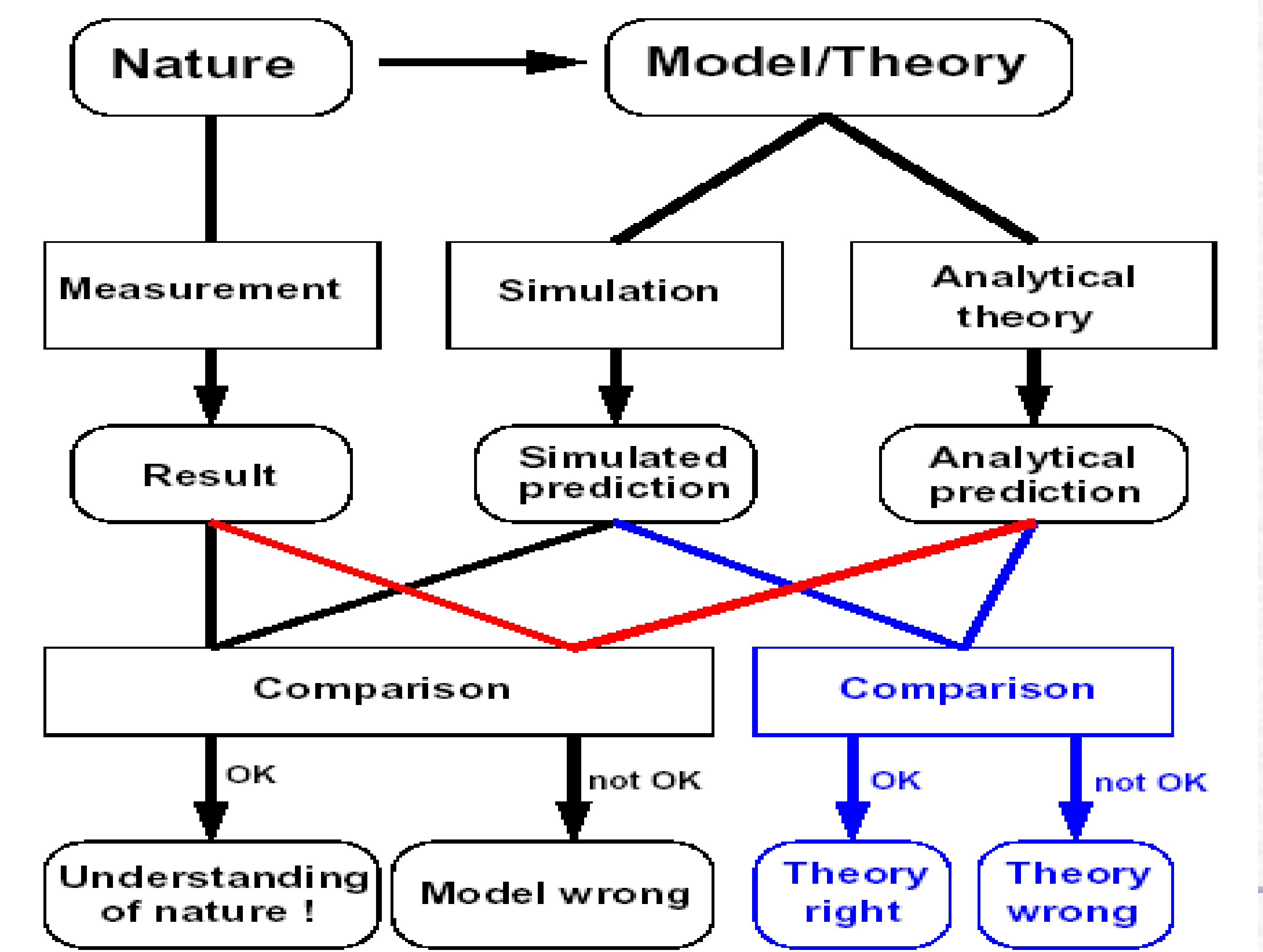
高能物理实验中的核相互作用过程；

实验探测器的模拟

数值分析：

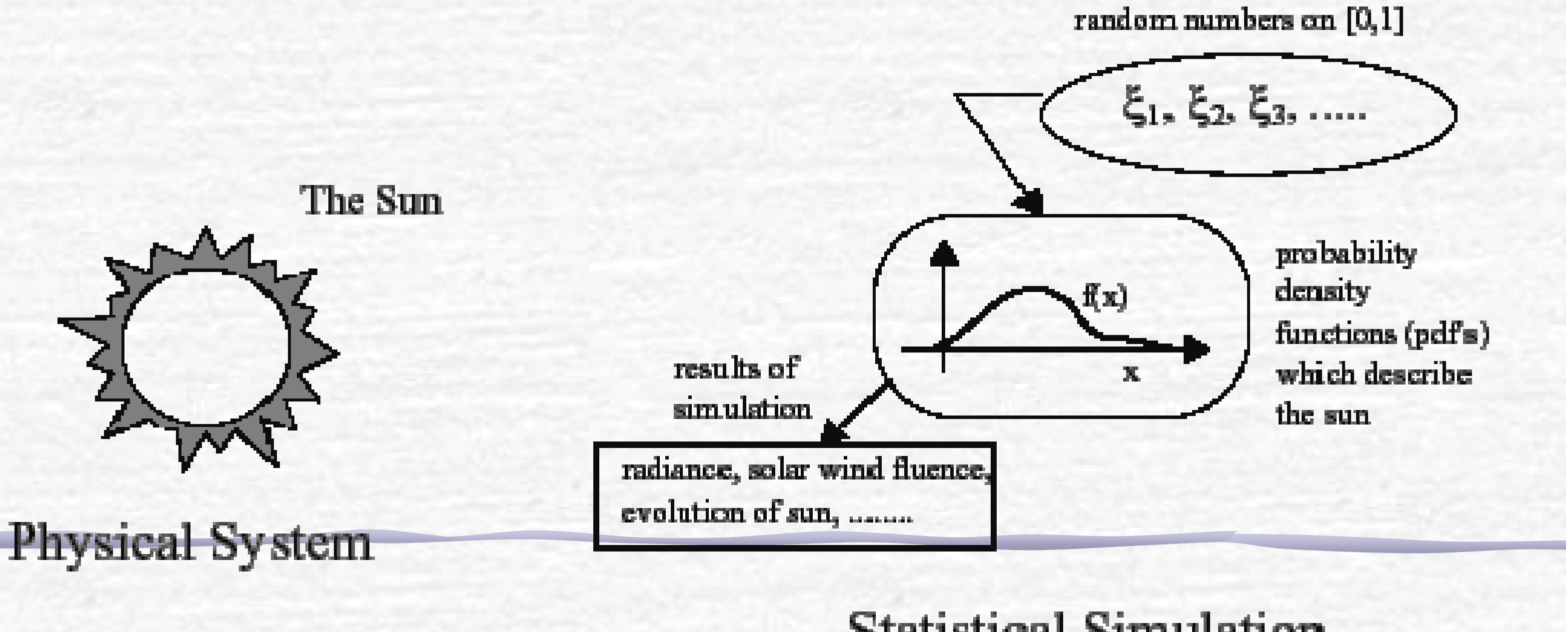
利用Monte Carlo方法求积分

Monte Carlo模拟在物理研究中的作用



Monte Carlo模拟的步骤：

1. 根据欲研究的物理系统的性质，建立能够描述该系统特性的理论模型，导出该模型的某些特征量的概率密度函数；
2. 从概率密度函数出发进行随机抽样，得到特征量的一些模拟结果；
3. 对模拟结果进行分析总结，预言物理系统的某些特性。



注意以下两点：

- **Monte Carlo**方法与数值解法的不同**:**
  - ✓ **Monte Carlo**方法利用随机抽样的方法来求解物理问题;
  - ✓数值解法:从一个物理系统的数学模型出发,通过求解一系列的微分方程来的导出系统的未知状态;

- **Monte Carlo**方法并非只能用来解决包含随机的过程的问题:
  - ✓许多利用**Monte Carlo**方法进行求解的问题中并不包含随机过程

    例如:用**Monte Carlo**方法计算定积分**.**

    对这样的问题可将其转换成相关的随机过程, 然后用**Monte Carlo**方法进行求解

Monte Carlo算法的主要组成部分

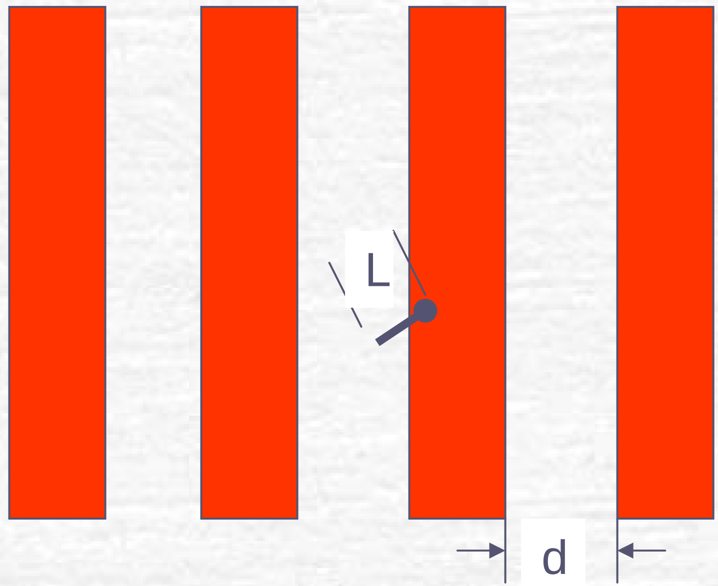
- ✓概率密度函数(pdf)— 必须给出描述一个物理系统的一组概率密度函数;
- ✓随机数产生器—能够产生在区间[0,1]上均匀分布的随机数
- ✓抽样规则—如何从在区间[0,1]上均匀分布的随机数出发,随机抽取服从给定的pdf的随机变量;
- ✓模拟结果记录—记录一些感兴趣的量的模拟结果
- ✓误差估计—必须确定统计误差（或方差）随模拟次数以及其它一些量的变化；
- ✓减少方差的技术—利用该技术可减少模拟过程中计算的次数；
- ✓并行和矢量化—可以在先进的并行计算机上运行的有效算法

Monte Carlo方法简史

简单地介绍一下Monte Carlo方法的发展历史

1、Buffon投针实验：

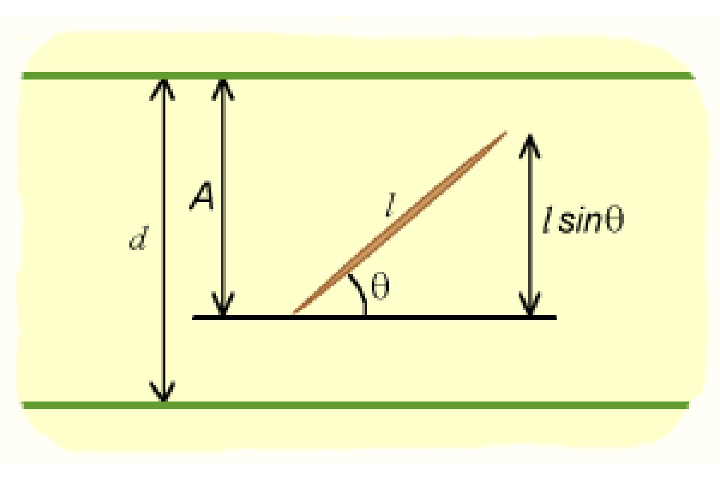
1768年，法国数学家Comte de Buffon利用投针实验估计π的值



$$p = \frac{2L}{\pi d}$$

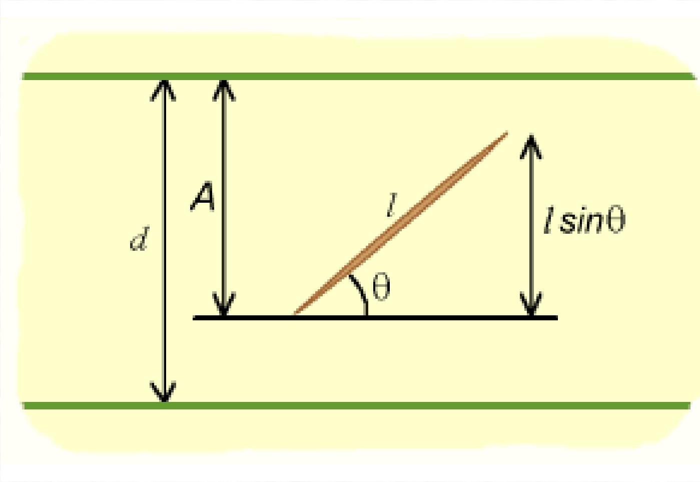
Problem of Buffon's needle:

If a needle of length *l* is dropped at random on the middle of a horizontal surface ruled with parallel lines a distance *d>l* apart, what is the probability that the needle will cross one of the lines?



Solution:

The positioning of the needle relative to nearby lines can be described with a random vector which has components:



$$A \in [0, d)$$
$$\theta \in [0, \pi)$$

The random vector is uniformly distributed on the region $[0,d)\times[0,\pi)$. Accordingly, it has probability density function $1/d\pi$

The probability that the needle will cross one of the lines is given by the integral

$$p = \int_0^{\pi} \int_0^{l \sin\theta} \tfrac{1}{d\pi}\, dA\, d\theta = \frac{2l}{d\pi}$$

2、1930年，Enrico Fermi利用Monte Carlo方法研究中子的扩散，并设计了一个Monte Carlo机械装置，Fermiac,用于计算核反应堆的临界状态

3、Von Neumann是Monte Carlo方法的正式奠基者,他与Stanislaw Ulam合作建立了概率密度函数、反累积分布函数的数学基础，以及伪随机数产生器。在这些工作中，Stanislaw Ulam意识到了数字计算机的重要性

→合作起源于Manhattan工程：利用ENIAC(Electronic Numerical Integrator and Computer)计算产额

随机数

什么是随机数？

- 单个的数字不是随机数
- 是指一个数列，其中的每一个体称为随机数，其值与数列中的其它数无关；
- 在一个均匀分布的随机数中，每一个体出现的概率是均等的；
  - 例如：在[0,1]区间上均匀分布的随机数序列中，0.00001与0.5出现的机会均等

随机数应具有的基本特性

- 考虑一个对高能粒子反应过程的模拟：需用随机数确定：
  - 出射粒子的属性：能量、方向、…
  - 粒子与介质的相互作用
- 对这一过程的模拟应满足以下要求（相空间产生过程）：
  - 出射粒子的属性应是互不相关的，即每一粒子的属性的确定独立于其它的粒子的属性的确定；
  - 粒子的属性的分布应满足物理所要求的理论分布；
- 所模拟的物理过程要求随机数应具有下列特性：
  - **随机数序列应是独立的、互不相关的(uncorrelated)**：

    即序列中的任一子序列应与其它的子序列无关；

- **长的周期(long period)：**

  实际应用中，随机数都是用数学方法计算出来的，这些算法具有周期性，即当序列达到一定长度后会重复；

- **均匀分布的随机数应满足均匀性(Uniformity)：**

  随机数序列应是均匀的、无偏的，即：如果两个子区间的“面积”相等，则落于这两个子区间内的随机数的个数应相等。

  例如：对[0,1)区间均匀分布的随机数，如果产生了足够多的随机数，而有一半的随机数落于区间[0,0.1]➔不满足均匀性

  如果均匀性不满足，则会出现序列中的多组随机数相关的情况➔均匀性与互不相关的特性是有联系的

- 有效性（**Efficiency):**

  模拟结果可靠

  - ➔模拟产生的样本容量大

    - ➔所需的随机数的数量大

      - ➔随机数的产生必须快速、有效，最好能够进行并行计算。

线性乘同余方法（Linear Congruential Method)

948年由Lehmer提出的一种产生伪随机数的方法，是最常用的方法

、递推公式：

$$I_{n+1} = (aI_n + c) \bmod m$$

其中：

I_0： 初始值（种子seed)

a： 乘法器 （multiplier)

c： 增值（additive constant)

m： 模数（modulus)

mod：取模运算：(aI_n+c)除以m后的余数

$$a, c \geq 0$$
$$m > I_0, a, c$$

a, c和m皆为整数

➔产生整型的随机数序列,随机性来源于取模运算

如果c=0 ➔ 乘同余法：速度更快，也可产生长的随机数序列

、实型随机数序列：

$$r_n = \frac{I_n}{float(m)} \to [0,1)$$

$$r_n = \frac{I_n}{float(m-1)} \to [0,1]$$

$$I_n < m$$

$$I_n \le m-1$$

、特点：

1）最大容量为m：$0 \le I_n \le m$

2）独立性和均匀性取决于参数a和c的选择

例：$a=c=I_0=7$, m=10 ➔ 7,6,9,0,7,6,9,0,…

、模数*m*的选择：

- *m* 应尽可能地大，因为序列的周期不可能大于*m*;
- 通常将*m*取为计算机所能表示的最大的整型量，在32位计算机上，$m=2^{31}=2\text{x}10^9$

、乘数因子*a*的选择：

1961年，M. Greenberger证明：用线性乘同余方法产生的随机数序列具有周期m的条件是：

1. c和m为互质数；
2. a-1是质数p的倍数，其中p是a-1和m的共约数；
3. 如果m是4的倍数，a-1也是4的倍数。

例：a=5,c=1,m=16,I0=1 →周期=m=16

1,6,15,12,13,2,11,8,9,14,7,4,5,10,3,0,1,6,15, 12,13,2,..

Monte Carlo积分

Monte Carlo法的重要应用领域之一：计算积分和多重积分

适用于求解：

1. 被积函数、积分边界复杂，难以用解析方法或一般的数值方法求解；
2. 被积函数的具体形式未知，只知道由模拟返回的函数值。

本章内容：

用Monte Carlo法求定积分的几种方法：
均匀投点法、期望值估计法、重要抽样法、半解析法、…

Goal: Evaluate an integral:

$$I = \int_a^b g(x)\, dx$$

Why use random methods?

Computation by "deterministic quadrature" can become expensive and inaccurate.

- grid points add up quickly in high dimensions
- bad choices of grid may misrepresent g(x)

- **Monte Carlo method can be used to compute integral of any dimension *d* (*d*-fold integrals)**

- **Error comparison of *d*-fold integrals**

  - **Simpson's rule,…**

    approximating the integral of a function *f* using quadratic polynomials

    $$E \propto N^{-1/d}$$

    $$\int_{x_0}^{x} f(x)dx = \int_{x_0}^{x_0+h} f(x)dx \approx \frac{1}{3}h[f(x_0)+4f(x_1)+f(x_2)]$$

    $$x_1 - x_0 = x_2 - x_1 = h$$

  - **Monte Carlo method**

    $$E \propto N^{-\frac{1}{2}}$$

    purely statistical,
    not rely on the dimension !

  - **Monte Carlo method WINS, when *d* >> 3**

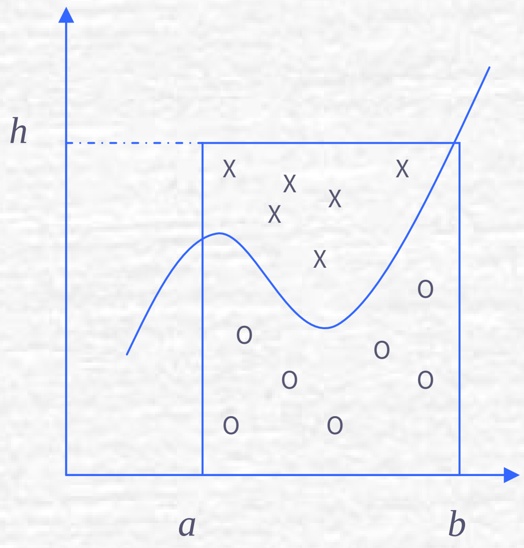
- **Hit-or-Miss Method**
- **Sample Mean Method**
- **Variance Reduction Technique**
- **Variance Reduction using Rejection Technique**
- **Importance Sampling Method**

Hit-or-Miss Method

- Evaluation of a definite integral

$$I = \int_a^b \rho(x)dx$$

$$h \geq \rho(x) \ \ \text{for any } x$$



- Probability that a random point reside inside the area

$$p = \frac{I}{(b-a)h} \approx \frac{M}{N}$$

$$I \approx (b-a)h\frac{M}{N}$$

- N : Total number of points
- M : points that reside inside the region

Hit-or-Miss Method

Sample uniformly from the rectangular region

$$[a,b]x[0,h]$$

The probability that we are below the curve is

$$p := \frac{I}{h(b-a)}$$

So, if we can estimate p, we can estimate I:

$$\hat{I} = \hat{p}h(b-a)$$

where $\hat{p}$ is our estimate of p

Hit-or-Miss Method

We can easily estimate p:

- throw N “uniform darts” at the rectangle
- let M be the number of times you end up under the curve $y=g(x)$
- let $\hat{p} = \dfrac{M}{N}$

Hit-or-Miss Method

Start

Set N : large integer

$M = 0$

Loop N times:

Choose a point x in $[a,b]$

$$x = (b-a)u_1 + a$$

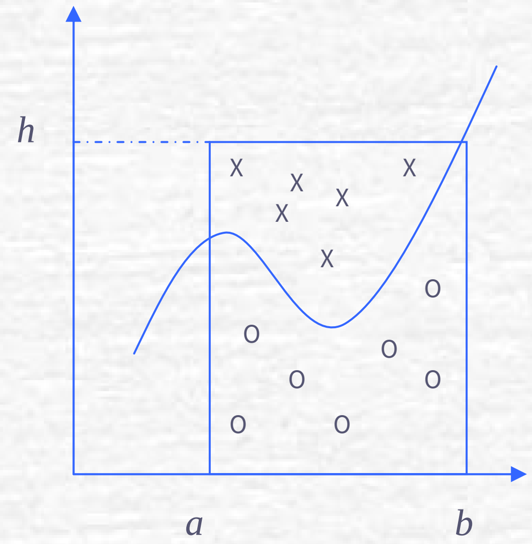
Choose a point y in $[0,h]$

$$y = hu_2$$

if [x,y] reside inside then M = M+1

$I = (b-a)\,h\,(M/N)$

End



Hit-or-Miss Method

- **Error Analysis of the Hit-or-Miss Method**
  - **It is important to know how accurate the result of simulations are**
  - **note that M is binomial(M,p)**

$$E(M) = Np \quad \sigma^2(M) = Np(1-P)$$

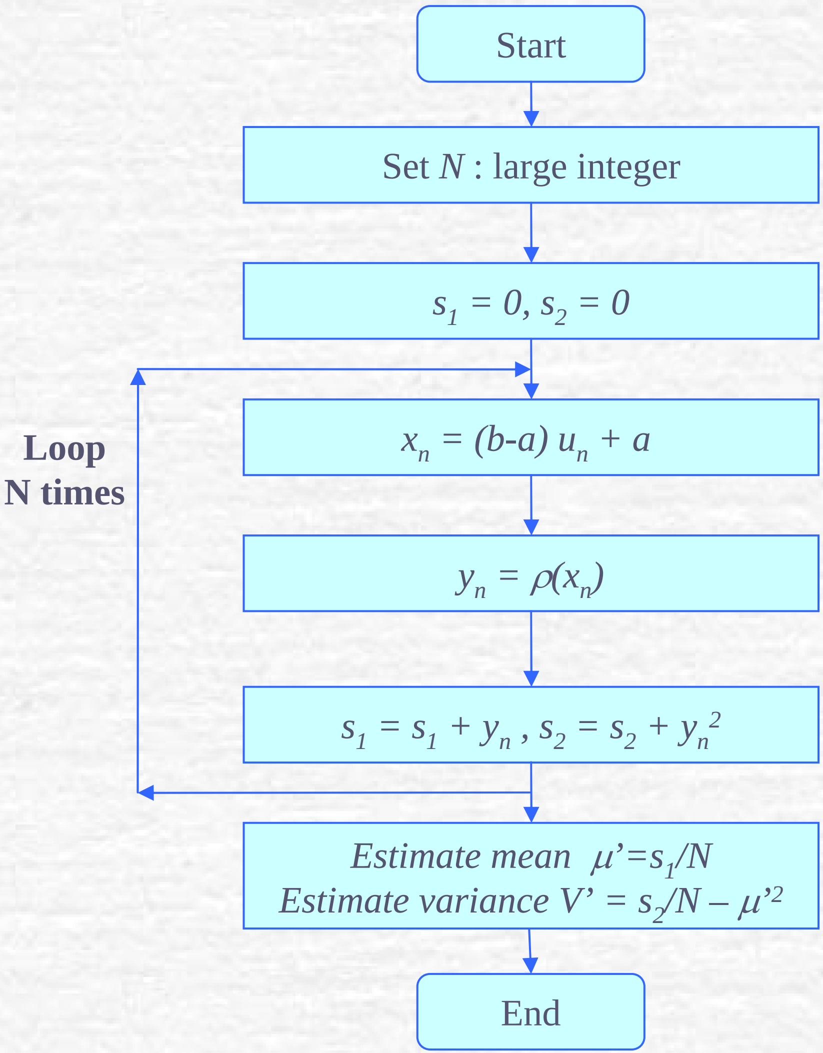
$$E(\hat{I}) = E\left[\hat{p}h(b-a)\right] = E\left[\frac{M}{N}h(b-a)\right] = \frac{h(b-a)}{N}E(M) = h(b-a)p = I$$

$$\sigma^2(\hat{I}) = \sigma^2\left[\hat{p}h(b-a)\right] = \sigma^2\left[\frac{M}{N}h(b-a)\right] = \frac{h^2(b-a)^2}{N^2}\sigma^2(M) = \frac{h^2(b-a)^2 p(1-}{N}$$

$$\sigma(\hat{I}) = h(b-a)\sqrt{\frac{p(1-p)}{N}} \propto N^{-\frac{1}{2}}$$

$$p \approx \hat{p} = \frac{M}{N} \qquad \sigma(\hat{I}) = \frac{h(b-a)}{N}\sqrt{M(1-\frac{M}{N})}$$

Sample Mean Method



$$I_{error}^{SM} = (b-a) \times 0.6745 \sqrt{\frac{V'}{N}}$$

Sample Mean Method

$$I = \int_a^b g(x)\,dx$$

Write this as:

$$I = \int_a^b g(x)\,dx = (b-a)\int_a^b g(x)\,\frac{1}{b-a}\,dx$$

$$= (b-a)\,E\left[g(X)\right]$$ where X~unif(a,b)

Sample Mean Method

$$I = (b - a)\,E\left[g(X)\right] \qquad \text{where } X \sim unif(a,b)$$

So, we will estimate I by estimating E[g(X)] with

$$\hat{E}[g(X)] = \frac{1}{n}\sum_{i=1}^{n} g(X_i)$$

where $X_1, X_2, \ldots, X_n$ is a random sample from the uniform(a,b) distribution.

Sample Mean Method

Example:

$$I = \int_0^3 e^x \, dx$$

(we know that the answer is $e^3 - 1$ 19.08554)

- write this as

$$I = 3\int_0^3 e^x \frac{1}{3} dx = 3E[e^X]$$

where X~unif(0,3)

Sample Mean Method

- write this as

$$I = 3\int_0^3 e^x \frac{1}{3} dx = 3E[e^X]$$

where X~unif(0,3)

estimate this with

$$\hat{I} = 3\hat{E}[e^X] = 3\frac{1}{n}\sum_{i=1}^{n} e^{X_i}$$

where X_1, X_2, ..., X_n are n independent unif(0,3)'s.

Sample Mean Method

Simulation Results: true = 19.08554, n=100,000

Simulation	$\hat{I}$
1	19.10724
2	19.08260
3	18.97227
4	19.06814
5	19.13261

Sample Mean Method

Don't ever give an estimate without a confidence interval!

This estimator is "unbiased":

$$E[\hat{I}] = E\left[(b-a)\frac{1}{n}\sum_{i=1}^{n} g(X_i)\right]$$

$$= (b-a)\frac{1}{n}\sum_{i=1}^{n} E[g(X_i)]$$

$$= (b-a)\frac{1}{n}\int_a^b g(x)\frac{1}{b-a}\,dx = \int_a^b g(x)\,dx = I$$

Sample Mean Method

$$\sigma^2_{\hat{I}} := \mathrm{Var}[\hat{I}] = \mathrm{Var}\left[(b-a)\frac{1}{n}\sum_{i=1}^{n} g(X_i)\right]$$

$$= \frac{(b-a)^2}{n^2}\mathrm{Var}\left[\sum_{i=1}^{n} g(X_i)\right]$$

$$= \frac{(b-a)^2}{n^2}\sum_{i=1}^{n}\mathrm{Var}[g(X_i)] \qquad \text{(indep)}$$

$$= \frac{(b-a)^2}{n}\mathrm{Var}[g(X_1)] \qquad \text{(indent)}$$

$$= \frac{(b-a)^2}{n}\int_a^b \left(g(x) - E[g(X)]\right)^2 \frac{1}{b-a}\,dx$$

$$= \frac{(b-a)^2}{n}\int_a^b \left(g(x) - \frac{I}{b-a}\right)^2 \frac{1}{b-a}\,dx$$

Sample Mean Method

- an approximation

$$s_{\hat{I}}^{2} = \frac{(b-a)^{2}}{n} \frac{\sum_{i=1}^{n}\left(g(x_{i}) - \frac{\hat{I}}{b-a} \right)}{n-1}$$

Sample Mean Method

$$\hat{I} = (b-a)\frac{1}{n}\sum_{i=1}^{n} g(X_i)$$

- $X_1, X_2, \ldots, X_n$ iid -> $g(X_1), g(X_2), \ldots, g(X_n)$ iid
- Let $Y_i = g(X_i)$ for $i = 1,2,\ldots,n$

Then

$$\hat{I} = (b-a)\,\overline{Y}$$

and we can once again invoke the CLT.

Sample Mean Method

For n "large enough" (n>30),

$$\hat{I} \approx N(I, \sigma_{\hat{I}}^2)$$

So, a confidence interval for I is roughly given by

$$\hat{I} \pm z_{\alpha/2}\, \sigma_{\hat{I}}$$

but since we don't know $\sigma_{\hat{I}}$, we'll have to be content with the further approximation:

$$\hat{I} \pm z_{\alpha/2}\, s_{\hat{I}}$$

Sample Mean Method

By the way...

No one ever said that you have to use the uniform distribution

Example: $$I := \int_a^b x^{1/2}\, e^{-2x}\, dx$$

$$= \frac{1}{2}\int_0^\infty x^{1/2}\, 2e^{-2x}\, I_{[a,b]}(x)\, dx$$

$$= \frac{1}{2} E\left[X^{1/2}\, I_{[a,b]}(X)\right]$$

where X~exp(rate=2).

Sample Mean Method

Comparison of Hit-and-Miss and Sample Mean Monte Carlo

- Let $\hat{I}_{HM}$ be the hit-and-miss estimator of I
- Let $\hat{I}_{SM}$ be the sample mean estimator of I

Then

$$\mathrm{Var}(\hat{I}_{HM}) \geq \mathrm{Var}(\hat{I}_{SM})$$

Sample Mean Method

Comparison of Hit-and-Miss and Sample Mean Monte Carlo

Sample mean Monte Carlo is generally preferred over Hit-and-Miss Monte Carlo because:

- ❖ the estimator from SMMC has lower variance
- ❖ SMMC does not require a non-negative integrand (or adjustments)
- ❖ H&M MC requires that you be able to put g(x) in a "box", so you need to figure out the max value of g(x) over [a,b] and you need to be integrating over a finite integral.

- **Hit-or-Miss Method**
- **Sample Mean Method**
- **Variance Reduction Technique**
- **Variance Reduction using Rejection Technique**
- **Importance Sampling Method**

Variance Reduction Technique

Introduction

- Monte Carlo Method and Sampling Distribution
  - Monte Carlo Method : Take values from random sample
  - From central limit theorem,

$$\overline{\mu} = \mu \qquad \overline{\sigma}^2 = \sigma^2 / N$$

  - 3σ rule

$$P(\mu - 3\overline{\sigma} \le \overline{X} \le \mu - 3\overline{\sigma}) \approx 0.9973$$

  - Most probable error

$$Error \approx \pm \frac{0.6745\sigma}{\sqrt{N}}$$

- Important characteristics

$$Error \propto 1/\sqrt{N} \qquad Error \propto \sigma$$

Variance Reduction Technique

Introduction

- Reducing error
  - *100 samples reduces the error order of 10
  - Reducing variance → Variance Reduction Technique
- The value of variance is closely related to how samples are taken
  - Unbiased sampling
  - Biased sampling
    - More points are taken in important parts of the population

Variance Reduction Technique Motivation

- If we are using sample-mean Monte Carlo Method
  - Variance depends very much on the behavior of $\rho(x)$
    - $\rho(x)$ varies little → variance is small
    - $\rho(x)$ = const → variance=0
- Evaluation of a integral

$$I' = (b-a)\mu_{\overline{Y}} = \frac{b-a}{N}\sum_{n=1}^{N}\rho(x_n)$$

  - Near minimum points → contribute less to the summation
  - Near maximum points → contribute more to the summation
  - More points are sampled near the peak →"importance sampling strategy"

Variance Reduction Technique

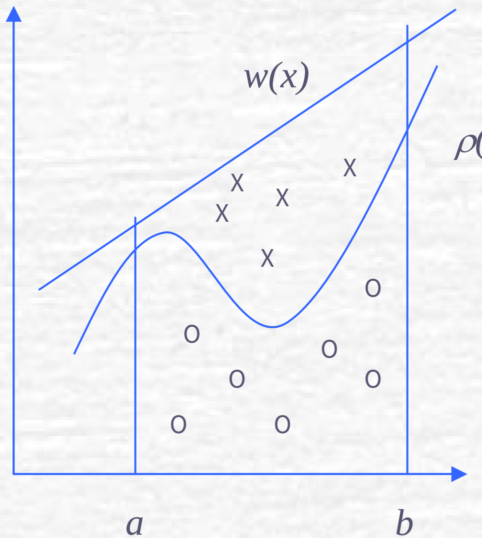
- Variance Reduction for Hit-or-Miss method
  - In the domain [a,b] choose a comparison function

$$w(x) \geq \rho(x)$$

$$W(x) = \int_{-\infty}^{x} w(x)dx$$

$$A = \int_{a}^{b} w(x)dx$$

$$Au = w(x) \longrightarrow x = W^{-1}(Au)$$



- Points are generated on the area under *w(x)* function
  - Random variable that follows distribution *w(x)*

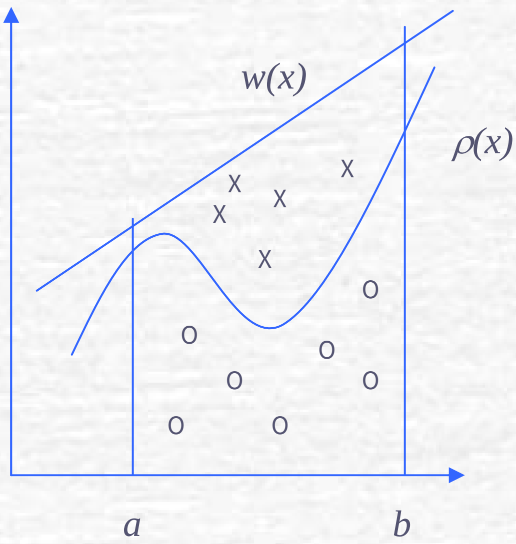
Variance Reduction Technique

Points lying above *ρ(x)* is rejected

$$I \approx A\frac{N'}{N}$$

$$y_n = w(x_n)u_n$$

$$q_n = \begin{cases} 1 \text{ if } y_n \le \rho(x_n) \\ 0 \text{ if } y_n > \rho(x_n) \end{cases}$$



q	1	0
P(q)	r	1-r

$$r = I / A$$

Variance Reduction Technique

- Error Analysis

$$E(Q) = r, \qquad V(Q) = r(1-r)$$

$$I = Ar = AE(Q)$$

$$I_{error}^{RJ} \approx 0.67\sqrt{\frac{I(A-I)}{N}}$$

Hit or Miss method

$$A = (b-a)h$$

Error reduction

$$A \to I \quad \text{then} \quad \text{Error} \to 0$$

Variance Reduction Technique

Start

Set N : large integer

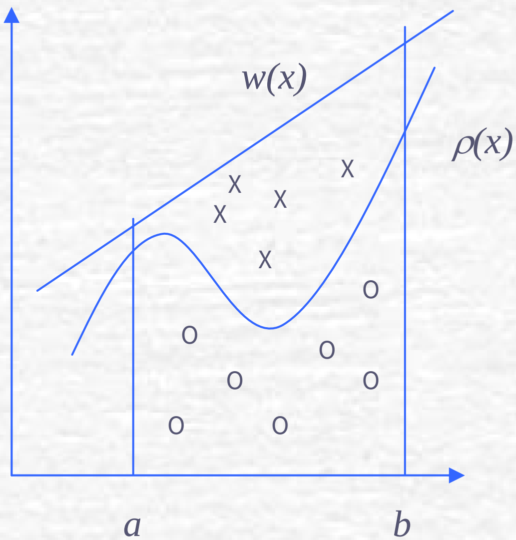
$N' = 0$

Loop N times

- Generate u_1, $x = W^{-1}(Au_1)$
- Generate u_2, $y = u_2\, w(x)$
- If $y <= f(x)$ accept value $N' = N'+1$
  Else : reject value

$I = (b-a)\, h\, (N'/N)$

End



$$I^{RJ}_{error} \approx 0.67\sqrt{\frac{I'(A-I')}{N}}$$

Importance Sampling Method

- Basic idea
  - Put more points near maximum
  - Put less points near minimum

- F(x) : transformation function (or weight function_

$$F(x) = \int_{-\infty}^{x} f(x)dx$$

$$y = F(x)$$

$$x = F^{-1}(y)$$

Importance Sampling Method

$$dy/dx = f(x) \rightarrow dx = dy/f(x)$$

$$I = \int_a^b \frac{\rho(x)}{f(x)} dy = \int_a^b \left[\frac{\rho(x)}{f(x)} \right] f(x) dx$$

$$\gamma(x) = \frac{\rho(x)}{f(x)}$$

$$<\eta>_f = \int_a^b \eta(x) f(x) dx$$

if we choose $f(x) = c\rho(x)$,then variance will be small
The magnitude of error depends on the choice of $f(x)$

$$I = \int_a^b \gamma(x) f(x) dx = <\gamma>_f$$

Importance Sampling Method

- Estimate of error

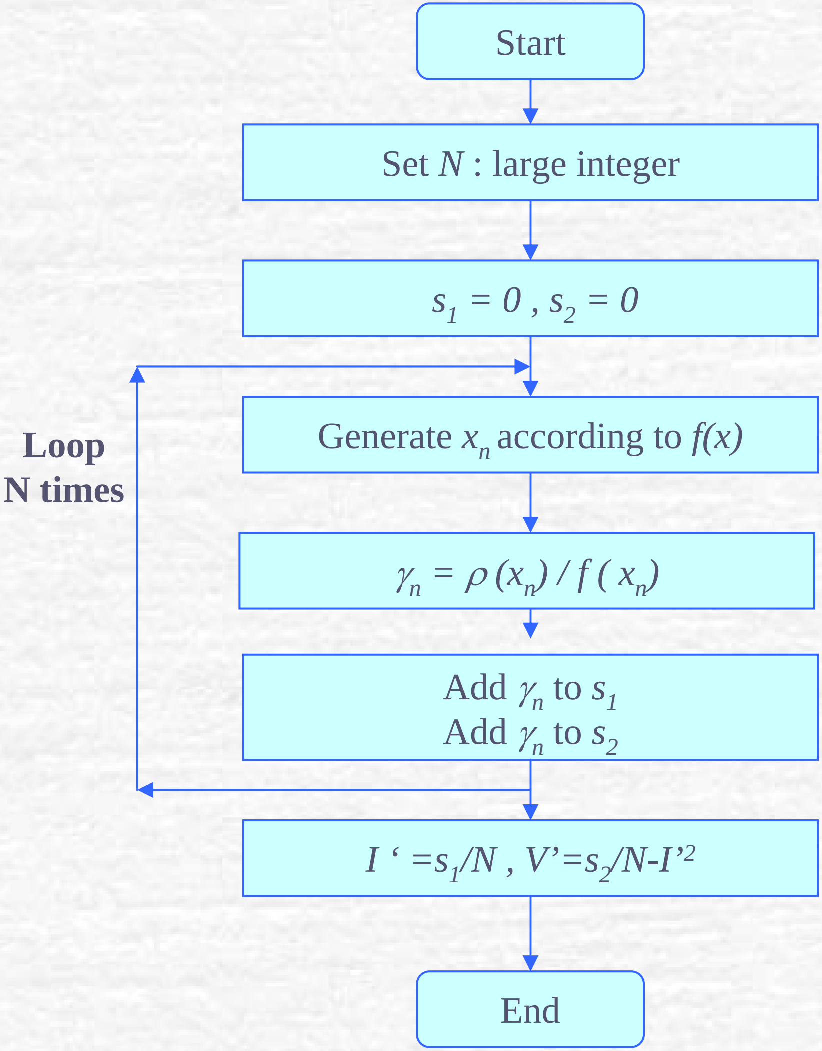
$$I = <\gamma>_f \approx \frac{1}{N}\sum_{n=1}^{N}\gamma(x_n)$$

$$I_{error} = 0.67\sqrt{\frac{V_f(\gamma)}{N}}$$

$$V_f(\gamma) = <\gamma^2>_f - (<\gamma>_f)^2$$

$$I_{error}^{IS} = 0.67\sqrt{\frac{<\gamma^2>_f - I^2}{N}}$$

Importance Sampling Method



$$I^{IS}_{error} = 0.67\sqrt{\frac{V'}{N}}$$