Monte Carlo and statistics

7th FLUKA Course

NEA Paris Sept.29-Oct.3, 2008

Overview

<u>General concepts:</u>

- Phase space
- The Boltzmann equation
- Analog vs. biased Monte Carlo calculation
- Figure of merit
- Simulation vs. integration

Sampling techniques

- discrete
- by inversion
- by rejection

Statistical errors on results: batch statistics

Phase space

- Phase Space: a concept of classical Statistical Mechanics
- Each Phase Space dimension corresponds to a particle degree of freedom
- 3 dimensions correspond to Position in (real) space: x, y, z
- 3 dimensions correspond to Momentum: p_x, p_y, p_z
  (or Energy and direction: E, θ, ϕ)
- More dimensions may be envisaged, corresponding to other possible degrees of freedom, such as quantum numbers: spin, etc.
- Each particle is represented by a point in phase space
- Time can also be considered as a coordinate, or it can be considered as an independent variable: the variation of the other phase space co-ordinates as a function of time (the trajectory of a phase space point) constitutes a particle "history"

The Boltzmann equation

- All particle transport calculations are (explicit or implicit) attempts to solve the Boltzmann Equation
- It is a balance equation in phase space: at any phase-space-point, the increment of particle phase-space-density is equal to the sum of all "production terms" minus a sum of all "destruction terms"
- Production: Sources, "Inscattering", Particle Production, Decay
- Destruction: Absorption, "Outscattering", Decay
- We can look for solutions of different type: at a number of (real or phase) space points, averages over (real or phase) space regions, projected on selected phase space hyperplanes, stationary or time-dependent

Mean of a distribution - 1

- In one dimension:

Given a variable x, distributed according to a function $f(x)$, the mean or average of another function of the same variable $A(x)$ over an interval $[a, b]$ is given by:

$$\bar{A} = \frac{\int_a^b A(x)\, f(x)\, dx}{\int_a^b f(x)\, dx}$$

Or, introducing the normalized distribution f':

$$f'(x) = \frac{f(x)}{\int_a^b f(x)\, dx}$$

$$\bar{A} = \int_a^b A(x)\, f'(x)\, dx$$

A particular case is that of $A(x) = x$: $\bar{x} = \int_a^b x\, f'(x)\, dx$

Mean of a distribution - 2

- In several dimensions: Given n variables $x, y, z, \ldots$, distributed according to the (normalized) functions $f'(x)$, $g'(y)$, $h'(z) \ldots$, the mean or average of a function of those variables $A(x, y, z \ldots)$ over an n-dimensional domain D is given by:

$$\bar{A} = \int_x \int_y \int_z \cdots \int A(x, y, z, \ldots)\, f'(x)\, g'(y)\, h'(z) \ldots dx\, dy\, dz \ldots$$

Often impossible to calculate with traditional methods, but we can sample N values of A with probability $f' \cdot g' \cdot h' \cdots$ and divide the sum of the sampled values by N:

$$S_N = \frac{\sum_1^N A(x, y, z, \ldots) f'(x)\, g'(y)\, h'(z) \ldots}{N}$$

Each term of the sum is distributed like A (Analog Monte Carlo). Integration, but also simulation!

Central limit theorem

Central Limit Theorem:

$$\lim_{N\to\infty} P(S_N) = \frac{1}{\sqrt{\frac{2\pi}{N}}\sigma_A} e^{-(S_N-\bar{A})^2/2\frac{\sigma_A^2}{N}}$$

For large values of N, the normalized sum of N independent and identically distributed random variables tends to a normal distribution with mean $\bar{A}$ and variance σ_A^2/N

$$\lim_{N\to\infty} S_N = \lim_{N\to\infty} \frac{\sum_1^N A(x,y,z,\ldots)f'(x)\,g'(y)\,h'(z)\ldots}{N} = \bar{A}$$

MC mathematical foundation

The central limit theorem is the mathematical foundation of the Monte Carlo method:

In words:

Given any observable A, that can be expressed as the result of a convolution of random processes, the average value of A can be obtained by sampling $many$ values of A according to the probability distributions of the random processes.

MC is indeed an INTEGRATION method that allows to solve multi-dimensional integrals by sampling

The accuracy of a MC estimator depends on the number of samples

Analog Monte Carlo

- In an analog Monte Carlo calculation ("honest" simulation), not only the mean of the contributions converges to the mean of the real distribution, but also the variance and all moments of higher order:

$$\boxed{\lim_{N\to\infty}\left[\frac{\sum_1^N (x-\bar{x})^n}{N}\right]^{\frac{1}{n}} = \sigma_n}$$

and fluctuations and correlations are faithfully reproduced.

Integration efficiency

- Traditional numerical integration methods (Simpson, etc.) converge to the true value as $N^{-\frac{1}{n}}$, where N = number of "points" (intervals) and n = number of dimensions
- Monte Carlo converges instead as $\frac{1}{\sqrt{N}}$

Number of dimensions	Traditional methods	Monte Carlo	Remark
$n = 1$	$\frac{1}{N}$	$\frac{1}{\sqrt{N}}$	MC not convenient
$n = 2$	$\frac{1}{\sqrt{N}}$	$\frac{1}{\sqrt{N}}$	about equivalent
$n>2$	$\frac{1}{\sqrt[n]{N}}$	$\frac{1}{\sqrt{N}}$	MC converges faster

A typical particle transport Monte Carlo problem is a 7-D problem!
x, y, z, p_x, p_y, p_z and t !!

Random numbers

- Basis for all Monte Carlo integrations are random numbers, i.e. values of a variable distributed according to a pdf (probability distribution function).
- In real world: the random outcome of a physical process
- In computer world: pseudo-random numbers
- The basic pdf is the uniform distribution:

$$f(\xi) = 1 \qquad 0 < \xi < 1$$

- Pseudo-random numbers are sequences that reproduce the uniform distribution, constructed from mathematical algorithms.
- All computers provide a pseudo-random number generator (or even several of them). In most computer languages (e.g., Fortran 90, C) a PRNG is even available as an intrinsic routine

Sampling from a distribution

Sampling from a discrete distribution

- Suppose to have a *discrete* random variable x , that can assume values $x_1, x_2, ...x_n$ with probability $p_1....p_n$.
- Assume $\sum\limits_i p_i = 1$, or normalize it.
- Divide the interval (0,1) in n subintervals, with limits

$$y_0 = 0,\ \ y_1 = p_1,\ \ y_2 = p_1 + p_2 \dots$$

- Generate a uniform random ξ
- Find in which of the i y-intervals it is
- Select the corresponding x_i as sampled value

Since ξ is uniformly random, we have

$$P(x_i) = P(y_{i-1} < \xi < y_i) = y_i - y_{i-1} = p_i$$

Sampling from a distribution

Sampling from a generic distribution

- Using one random number
- Integrate the distribution function, analytically or numerically, and normalize to 1 to obtain the normalized cumulative distribution:

$$\boxed{F(\xi) = \frac{\int_{x_{min}}^{\xi} f(x)\,\mathrm{d}x}{\int_{x_{min}}^{x_{max}} f(x)\,\mathrm{d}x}}$$

- Sample a uniform pseudo-random number ξ
- Get the desired result by finding the inverse value $\boxed{X = F^{-1}(\xi)}$, analytically or most often by interpolation (table look-up)

Since ξ is uniformly random, we have

$$P(a < x < b) = P(F(a) < \xi < F(b)) = F(b) - F(a) = \int_a^b f(x)dx$$

Example

take $f(x) = e^{-\frac{x}{\lambda}}$, $x \in [0, \infty]$

Cumulative distribution:

$$F(t) = \int_0^t e^{-\frac{x}{\lambda}}\,dx = \lambda \times (1 - e^{-\frac{t}{\lambda}})$$

Normalized:

$$F'(t) = \int_0^t \frac{e^{-\frac{x}{\lambda}}}{\lambda}dx = 1 - e^{-\frac{t}{\lambda}}$$

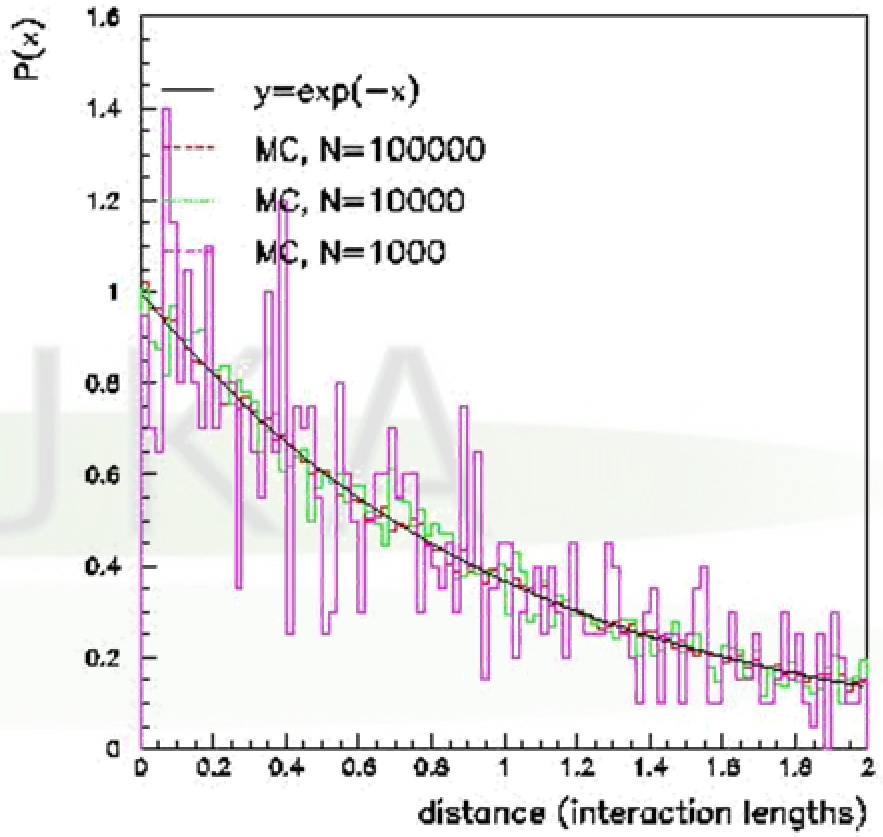
Sample a uniform $\xi \in [0, 1]$

$$1 - e^{-\frac{t}{\lambda}} = \xi$$

sample t by inverting

$$t = -\ln(\xi - 1)$$

repeat N times



Practical rule: a distribution can be sampled directly if and only if its pdf can be integrated and the integral inverted

Sampling from a distribution

- Using two random numbers (rejection technique)
- Choose a constant value C $> f(x)$ for any x
- Sample two random numbers ξ_1 and ξ_2
- If $\xi_2 < f(\xi_1)/C$, $\boxed{X = \xi_1}$,

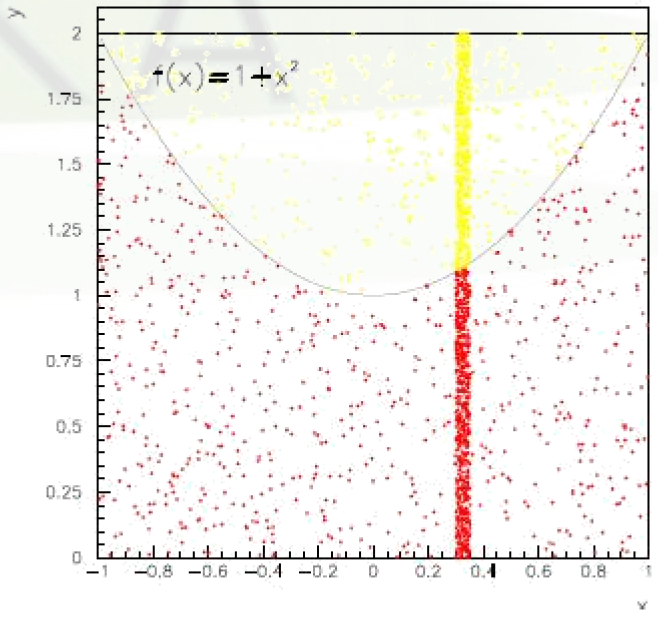
  otherwise re-sample ξ_1, ξ_2

The probability that a ξ_1 value is accepted is $f(\xi_1)/C$ for any ξ_1:

$$P(x)\,dx = P(\xi_1 = x)\,dx \cdot f(\xi_1 = x)/C$$

Since $P(\xi_1)\,dx = const$,

$$P(x)\,dx = const \cdot f(x)$$



Particle transport Monte Carlo

Application of Monte Carlo to particle transport and interaction:

- Each particle is followed on its path through matter.
- At each step the occurrence and outcome of interactions are decided by random selection from the appropriate probability distributions.
- All the secondaries issued from the same primary are transported before a new history is started.
- The accuracy and reliability of a Monte Carlo depends on the models or data on which the pdfs are based
- Statistical accuracy of results depends on the number of “histories"
- Statistical convergence can be accelerated by “biasing" techniques.

Monte Carlo Flavors –I

Microscopic Analog Monte Carlo

- uses theoretical models to describe physical processes whenever possible
- samples from actual physical phase space distributions
- predicts average quantities and all statistical moments of any order
- preserves correlations (provided the physics is correct, of course!)
- reproduces fluctuations (provided. . . see above)
- is (almost) safe and (sometimes) can be used as a “black box" (idem)

But:

- can be inefficient and converge slowly
- can fail to predict contributions due to rare events

Monte Carlo Flavors –II

Biased Monte Carlo

- samples from artificial distributions, and applies a weight to the particles to correct for the bias *(similar to an integration by a change of variable)*
- predicts average quantities, but not the higher moments (on the contrary, its goal is to *minimize* the second moment!)
- same mean with smaller variance → faster convergence
- allows sometimes to obtain acceptable statistics where an analog Monte Carlo would take years of CPU time to converge

But:

- cannot reproduce correlations and fluctuations
- ONLY privileged observables converge faster *(some regions of phase space are sampled more than others).*

Reduce variance or CPU time ?

Computer cost

A Figure of Merit:

$$\boxed{\text{Computer cost of an estimator} = \sigma^2 \times t}$$

(σ^2 = Variance, t = CPU time per primary particle)

- some biasing techniques are aiming at reducing σ^2, others at reducing t
- often reducing σ^2 increases t, and *viceversa*
- therefore, minimizing $\sigma^2 \times t$ means to reduce σ at a faster rate than t increases, or *viceversa*
- $\Longrightarrow$ the choice depends on the problem, and sometimes a combination of several techniques is most effective
- bad judgement, or excessive "forcing" on one of the two variables, can have catastrophic consequences on the other one, making computer cost "explode"

σ^2 is converging like 1/N, while t is obviously proportional to N

Monte Carlo Flavors -III

Macroscopic Monte Carlo

- Instead of simulating interactions, uses parametrizations of the reaction product distributions, obtained from fits to data and extrapolations
- Fast, especially when reactions are complex
- Can be more accurate than microscopic MC if the theory contains uncertainties/approximations

But:

- The single pdfs are reproduced, but the correlations among interaction products are not.
- Cannot be extended outside the data range.

A specific MC:…FLUKA

- FLUKA is a Microscopic MC
  - Except for the low energy neutron transport (see lecture)
- FLUKA can be used in biased mode (see lecture)

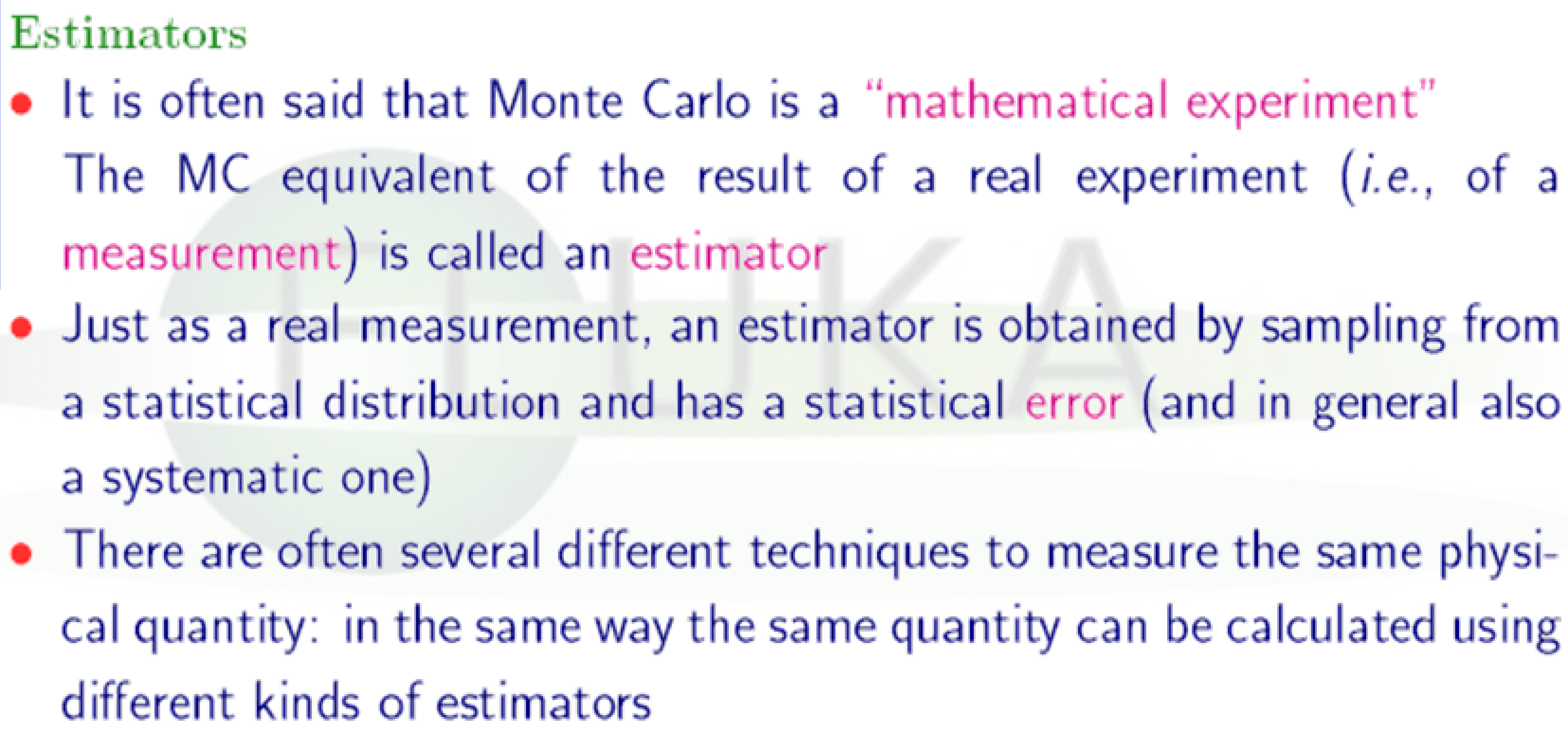
Pseudo-random numbers in FLUKA

- FLUKA uses the most up-to-date version of the Marsaglia random number generator.
- It provides sequences of 64 bit pseudo-random numbers that pass all the "randomness" tests.
- Different, independent sequences can be obtained using different SEEDS for the sequence initialization
- The initial seed can be set in the RANDOMIZE card. A default value is provided

```
RANDOMIZE Unit   Seed
```

- The code proceeds through the random sequence, each run provides the starting seed for the next one
- "parallel" runs can be performed by changing the initialization seed
- In user routines: to get a random number use the function FLRNDM(x)

Results from a MC calculation



Estimators

- It is often said that Monte Carlo is a "mathematical experiment"
  The MC equivalent of the result of a real experiment (*i.e.*, of a measurement) is called an estimator
- Just as a real measurement, an estimator is obtained by sampling from a statistical distribution and has a statistical error (and in general also a systematic one)
- There are often several different techniques to measure the same physical quantity: in the same way the same quantity can be calculated using different kinds of estimators

Statistical Errors

- Could be calculated for single histories, or for batches of several histories each
- Distribution of scoring contributions by single histories can be very asymmetric (many histories contribute zero). In FLUKA, errors are calculated in batches.
- Scoring distribution from batches tends to Gaussian for $N \to \infty$, provided $\sigma^2 \neq \infty$
- The standard deviation of a score calculated from batches is an estimate of the standard deviation of the actual distribution ("error of the mean")
- How good is such an estimate depends on the type of estimator and on the particular problem (but it converges to the true value for $N \to \infty$)

Relative error	Quality of Tally	*(from the MCNP Manual)*
50 to 100%	Garbage	
20 to 50%	Factor of a few	
10 to 20%	Questionable	
<10%	Generally reliable except for point detector *(not available in* FLUKA*)*	

Statistical Errors (batch statistics)

The variance of the mean of an estimated quantity x (e.g., fluence), calculated by N FLUKA jobs, is:

$$\boxed{\sigma^2_{<x>} = \left[\frac{1}{n}\sum_{i=1}^{N} n_i x_i^2 - \frac{1}{n^2}\left(\sum_{i=1}^{N} n_i x_i\right)^2\right]\frac{1}{N-1}}$$

where:

- n_i is the number of histories in the i^{th} job
- $n = \Sigma n_i$ is the total number of histories in the N jobs
- x_i is the is the average of x calculated in the i^{th} job: $x_i = \Sigma_{j=1}^{n_i} x_{ij}/n_i$, where x_{ij} is the contribution to x of the j^{th} history in the i^{th} job

Practical tips:

- Use always at least 5-10 batches of comparable size (it is not at all mandatory that they be of equal size)
- Never forget that the variance itself is a stochastic variable subject to fluctuations
- Be careful about the way convergence is achieved: often (particularly with biasing) apparently good statistics with few isolated spikes could point to a lack of sampling of the most relevant phase-space part
- Plot 2D and 3D distributions! In those cases the eye is the best tool in judging the quality of the result