Monte Carlo Methods in High Energy Physics

Peter Uwer

HUMBOLDT-UNIVERSITÄT ZU BERLIN

PHYSICS AT THE TERA SCALE
Helmholtz Alliance

HEPTOOLS
Tools and Precision Calculations for Physics Discoveries at Colliders

Computational Theoretical Particle Physics
SFB TR9

Contents

Introduction

Simple definition:

"A Monte Carlo technique is any technique making use of random numbers to solve a problem"

[F.James '80]

Two different classes of problems:

1. Problems which are intrinsically of probabilistic nature

   → direct simulation, application of MC method appears naturally

1. Strictly deterministic problems

   → Application of MC method more tricky, need to map non-stochastic problem to a stochastic one from which information on the original problem can be inferred

Introduction

Historically: The starting point of large scale application of MC methods were the simulations

Some of the pioneers:

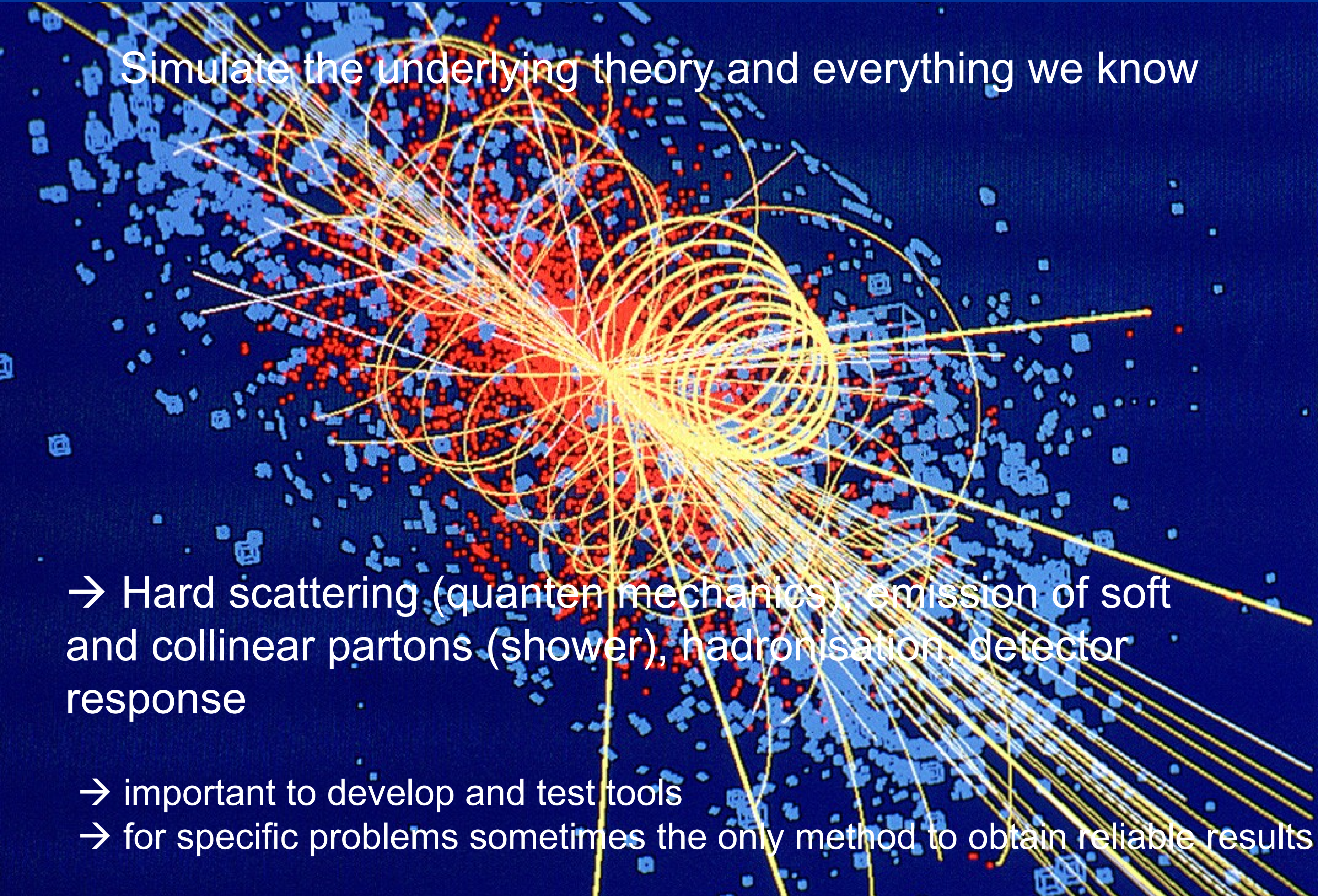
Fermi, Ulam, vNeumann, Metropolis,…

→ Manhattan project…

Examples:

- Neutron scattering and absorption in a nuclear reactor
- Simulation of LHC physics
- Ising model
- Buffon's needle

Simulation of LHC physics



Simulate the underlying theory and everything we know

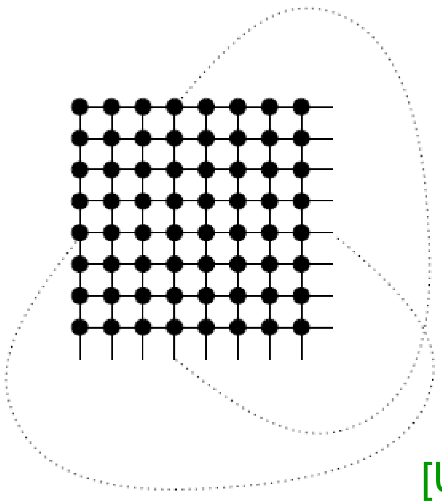
→ Hard scattering (quanten mechanics), emission of soft and collinear partons (shower), hadronisation, detector response

→ important to develop and test tools
→ for specific problems sometimes the only method to obtain reliable results

Ising model

Simple model of ferromagnetism

Consider d-dimensional cubic lattice with spin at each lattice site



[U.Wolff]

Hamiltonian:

$$H = -\epsilon \sum s(x)s(y) + B \sum s(x)$$

next neighbor interaction

coupling to external B field

$s(x)$ spin {+1,-1} at position x

From Thermodynamics/statistical mechanics:
Probability to find specific configuration of spins:

$$p(\{s\}) \sim \exp\left(-\frac{H(s)}{k_b T}\right) = \exp\left(-\beta H(s)\right)$$

Boltzmann factor

Ising model

Expectation values:

$$\langle O \rangle \sim \sum_{\{s\}} O(s) \exp(-\beta H(s))$$

Note:

Not possible to evaluate by brute force:
for a 10x10 lattice we have already 2^{100} configurations

2^{100} x 10^{-9} s = 10^{15} years

assume 1 FLOP on a GFLOP machine

Way out:

Generate random configurations distributed as

$$p(\{s\}) \sim \exp\left(-\frac{H(s)}{k_b T}\right) = \exp(-\beta H(s))$$

Ising model

Sounds easy but…

No algorithm known which generates configurations independent from each other following the Boltzmann factor

(→ general problem for some specific solutions see later)

Solution here:

Metropolis algorithm

→ many downloadable simulations available…

(→Example)

Ising model very important:

- Simple model for ferromagnetism
- Interesting testground for MC methods, in particular because 1- and 2-dim. are exactly solvable → Cluster algorithm [Wolff]
- Highly non-trivial test for (pseudo) random numbers

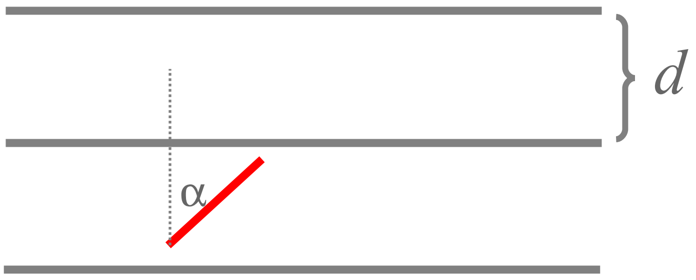
Buffon's needle

Goal:

Calculate the value of π

Buffon's (1777) idea:

Throw *randomly* a needle on a pattern
of equidistant parallel lines,
count the total number of throws together with
the number of hits (=the needle touches a line)



(needle with length d)

$$2\frac{\text{hits}+\text{miss}}{\text{hits}} \to \pi$$

Buffon's needle

From a simulation I obtained:

n	result
10	2.85714,2.5,2.85714,3.33333,3.33333,10,3.33333,2.85714,4,2.22222
100	3.1746,3.33333,3.50877,3.33333,2.85714,3.125,3.38983,2.7027, 3.63636,3.125
1000	3.19489,3.2,3.15956,3.10078,3.09119,3.1746,3.07692,3.08642, 3.11526,3.10078
10000	3.1294,3.13087,3.16556,3.16456,3.16506,3.10849,3.11915,3.10318, 3.16006,3.10849
100000	3.13529,3.12774,3.14347,3.13416,3.14916,3.14268, 3.15288,3.14001,3.13996,3.1508
1000000	3.1416, 3.14053, 3.14156, 3.14013, 3.14159, 3.14656, 3.14291, 3.13951, 3.14345, 3.14075

Comments:

- Good news: In principle it works, but
- Bad news: it does not work very well

→ general feature of hit-and-miss MC's, for improvements be patient
how do we estimate how well it will work?

Buffon's needle

- We are actually doing a MC integration, since what we calculate is essentially the area under $(1-\cos(x))$

  This is in a certain sense actually true for every MC, every MC method can be understood as MC integration

- Although the method might not work very well we have still solved a non-stochastic problem using MC methods

Exercise:

Simulate Buffon's needle

Exercise: Buffon's needle

Buffon's needle

1. Simulate Buffon's needle, determine the value of π using nmax = 1000, 10000,1000000, and 10000000 shots. A shot consist of throwing the angle and the position.

2. Run the simulation 10 time for each nmax and try to get an naiv estimate of the uncertainty.

3. Run the simulation 1000 times for each nmax and plot the result as an histogramm.

4. The number of hits follow a binomial distribution, try to understand why the methods works not very well.

Basic terminology

Random events

An elementary random event is an event which we cannot decomposed into simpler events and cannot be predicted in advance.

Examples are: flipping a coin or throwing a dice.

Probability

If we have $X_1, \ldots X_N$ possible events we can attribute a probability $p_i = p(X_i)$ to each event with the usual properties:

- $0 \leq p_i \leq 1$
- $p(E_i \text{ and/or } E_j) \leq p_i + p_j$
- for mutually exclusive events: $p(E_i \text{ and } E_j) = 0$ and $p(E_i \text{ or } E_j) = p_i + p_j$
- for an exhaustive list of exclusive events: $p(\text{some } E_i) = \sum p_i = 1$

[Kolmogorov]

Basic terminology

Conventions:

$p(X_1 + X_2 + \ldots)$ probability to find event out of list

$p(X_1 \times X_2 \times \ldots)$ probability to find X_1 and X_2 and so on

Conditional probability:

Consider two elementary random events (X, Y)

$p(X_i|Y_j)$ is the probability to find X_i when Y_j happened

We have:

$$p(X_i \times Y_j) = p(X_i|Y_j)p(Y_j) = p(Y_j|X_i)p(X_i)$$

Bayes theorem

→ basis of bayesian approach, subjective probability

Basic terminology

Random variable

For a set of events which are exhaustive and exclusive we may characterize each event by number x.
The number x is called a random variable

Cumulative distribution function (CDF) $F(y)$

$$F(y) = p(x \leq y)$$

Expectation value:

$$\langle x \rangle = \int y dF(x)$$

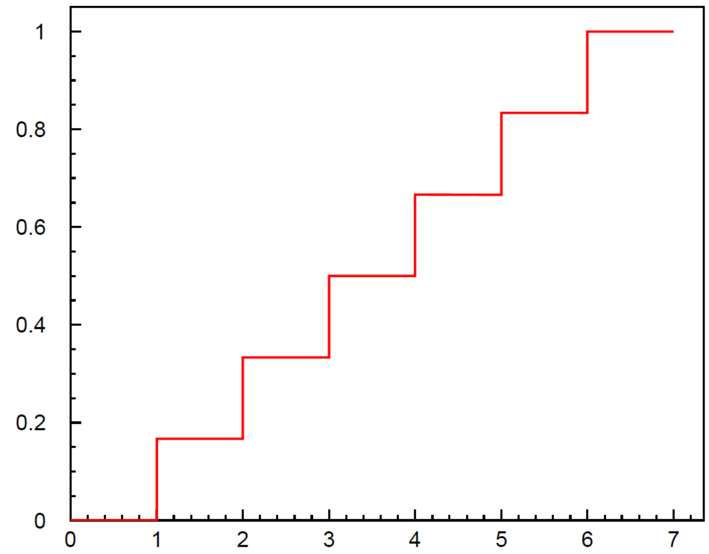
or more general for a function

$$\langle g(x) \rangle = \int g(x) dF(x)$$

Basic terminology

For a finite number of events, *F(y)* becomes a step function

i.e. throwing a dice:



For the expectation value we get:

$$\langle g(x) \rangle = \sum_i g(x_i) p_i$$

Basic terminology

Variance and standard deviation/standard error

The mean deviation from the mean vanishes:

$$\langle x - \langle x \rangle \rangle = \int x dF(x) - \int x dF(x) = 0;$$

For the quadratic deviation this is no longer the case, we define the variance by;

$$Var(x) = \langle (x - \langle x \rangle)^2 \rangle = \int dF(x)(x - \langle x \rangle)^2 = \langle x^2 \rangle - \langle x \rangle^2.$$

The square root of the variance is a measure of the dispersion of the random variable.

$$\Delta x = \sqrt{Var(x)}$$

It is called standard deviation or standard error.

Basic terminology

Probability distribution function (PDF)

Extend concepts to continous random variables

Define probability distribution function as probabiliy to find the random variable between x and $x + dx$

For differentiable $F(x)$ we have $f(x) = \frac{dF(x)}{dx}$

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

$$\langle g(x) \rangle = \int g(x)dF(x) = \int g(x)f(x)dx$$

$$\langle (g(x) - \langle g(x) \rangle)^2 \rangle = \int (g(x) - \langle g(x) \rangle)^2 dF(x)$$

Examples for probability distribution functions

1. Uniform distribution in $[a,b]$

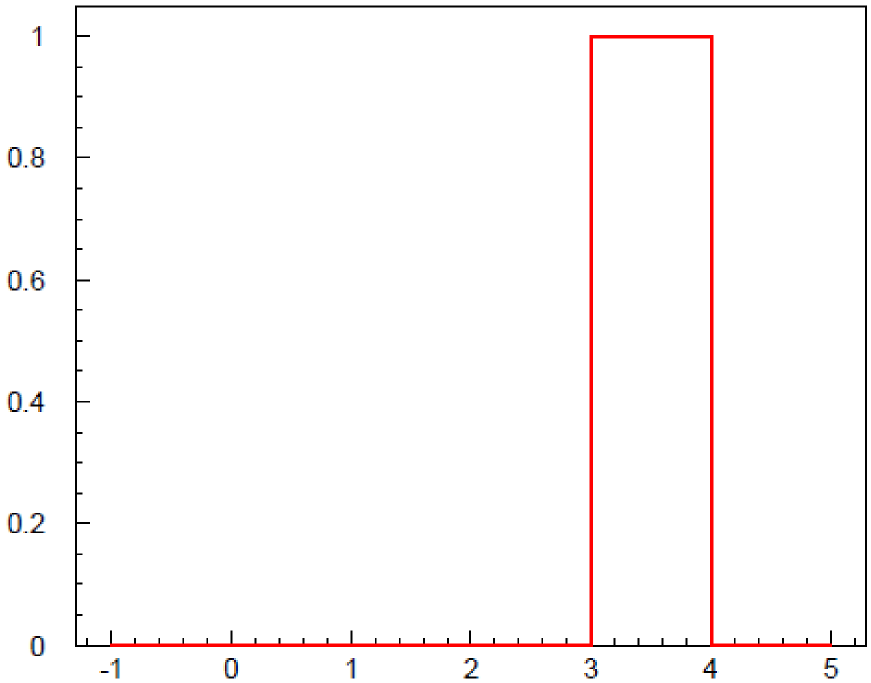
$$f(x,a,b) = \frac{1}{b-a} \quad (18)$$

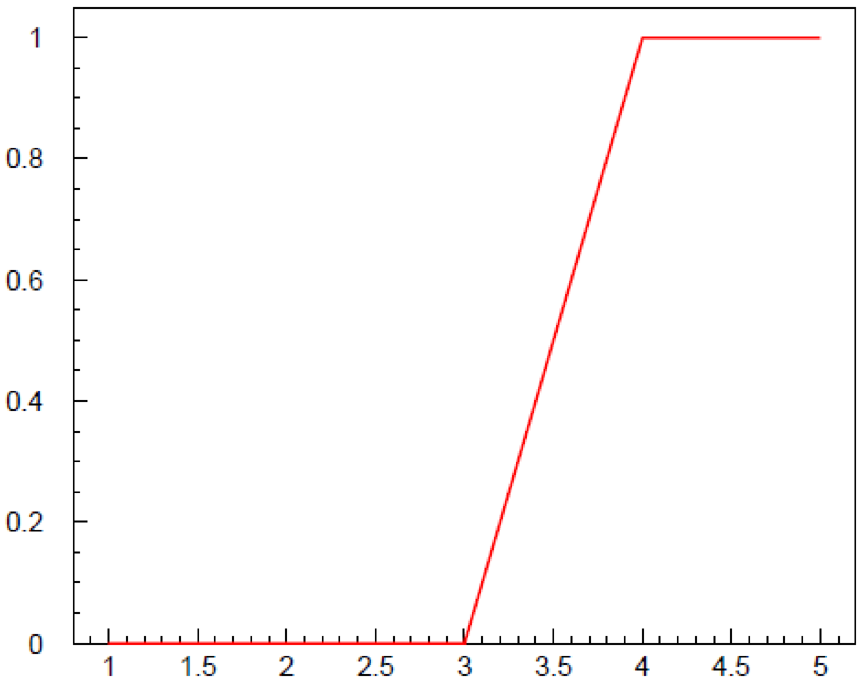
$$F(x,a,b) = \begin{cases} 0 & \text{for } x < a \\ \frac{y-a}{b-a} & \text{for } a < x \leq b \\ 1 & \text{for } x > b \end{cases} \quad (19)$$

$$\langle x \rangle = \frac{1}{2}(b-a) \quad (20)$$

$$\mathrm{Var}(x) = \frac{(b-a)^2}{12} \quad (21)$$

Uniform distribution





$f(x)$ $F(x)$

Gaussian distribution

2. Gaussian distribution / Normal distribution

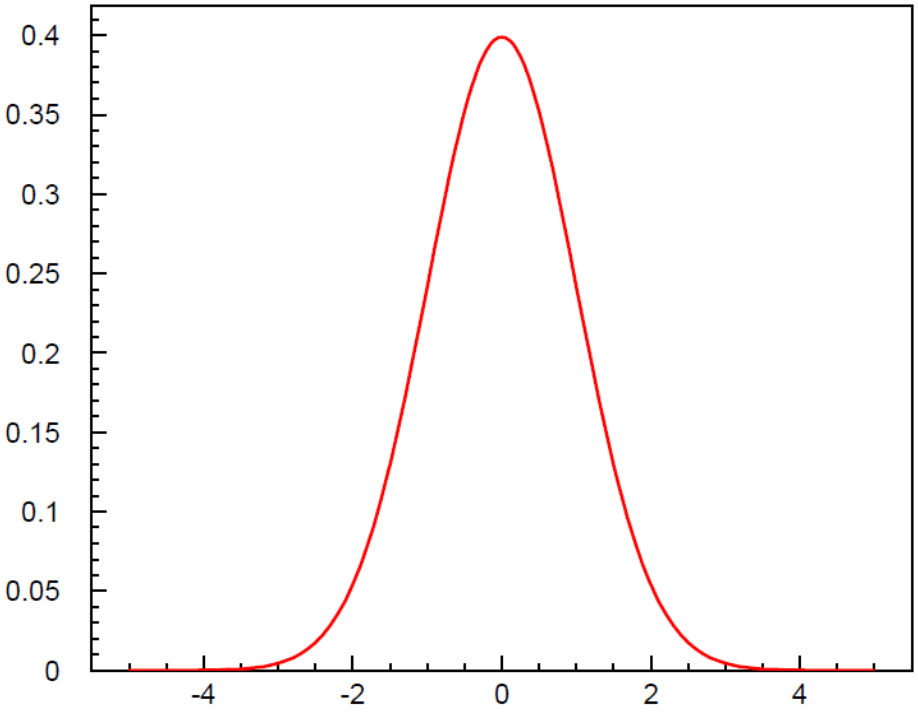
$$f(x,\overline{x},\sigma) = \frac{1}{\sqrt{2\pi}\sigma}\exp(-\frac{(x-\overline{x})^2}{2\sigma^2}) \tag{22}$$

$$F(x,0,1) = \frac{1}{2}(1+\mathrm{erf}(x/\sqrt{2})) \tag{23}$$

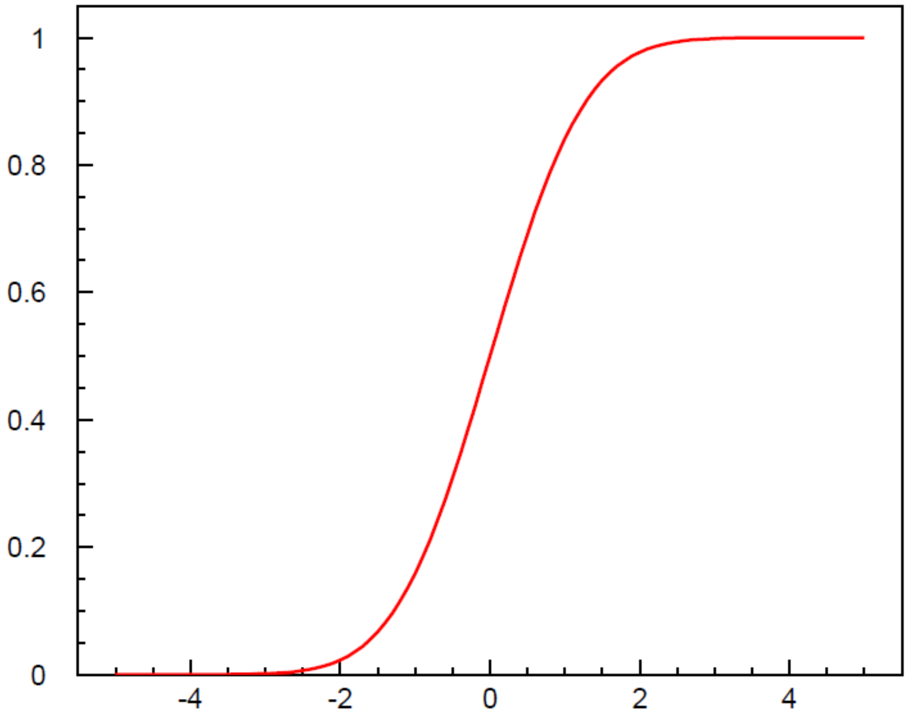
$$\langle x \rangle = \overline{x} \tag{24}$$

$$\mathrm{Var}(x) = \sigma^2 \tag{25}$$

Gaussian distribution



$f(x)$



$F(x)$

Poisson and χ^2 -distribution

4. Poisson distribution

$$f(n,\nu) = \frac{\nu^n e^{-\nu}}{n!} \tag{29}$$

$$\langle n \rangle = \nu \tag{30}$$

$$\mathrm{Var}(n) = \nu \tag{31}$$

5. χ^2-distribution

$$f(z,n) = \frac{z^{n/2-1} e^{-z/2}}{2^{n/2}\Gamma(n/2)} \tag{32}$$

$$\langle z \rangle = n \tag{33}$$

$$\mathrm{Var}(z) = 2n \tag{34}$$

(35)

Basic terminology

Composite random events

Event consisting of several elementary random events

Important case: $f(x) + g(y)$

What can we say about expectation value and variance ?

Expectation value:

$$\langle (\lambda_1 f(x) + \lambda_2 g(y)) \rangle = \int dxdy \rho(x,y)(f(x)+g(y))$$

$$= \lambda_1 \int dx \rho_x(x) f(x) + \lambda_2 \int dy \rho_y(y) g(y) = \lambda_1 \langle f \rangle + \lambda_2 \langle g \rangle$$

Basic terminology

$$
\begin{aligned}
Var(\lambda_1 f(x) + \lambda_2 g(y)) &= \langle (\lambda_1 f(x) + \lambda_2 g(y) - \langle \lambda_1 f(x) + \lambda_2 g(y) \rangle)^2 \rangle \\
&= \langle (\lambda_1 f(x) + \lambda_2 g(y) - \lambda_1 \langle f(x) \rangle + \lambda_2 \langle g(y) \rangle)^2 \rangle \\
&= \lambda_1^2 Var(f(x)) + \lambda_2^2 Var(g(x)) \\
&+ 2\lambda_1\lambda_2 (\langle f(x) g(y) \rangle - \langle f(x) \rangle \langle g(y) \rangle)
\end{aligned}
$$

→ variance of the sum can be larger or smaller than sum of indvidual variances

Covariance

$$cov(f(x), g(y)) = \langle f(x)g(y) \rangle - \langle f(x) \rangle \langle g(y) \rangle$$

Correlation

$$corr(f(x), g(y)) = \frac{cov(f(x), g(y))}{\sqrt{Var(f(x))Var(g(y))}}$$

$$-1 \leq corr(f(x), g(y)) \leq 1$$

If x and y are independent we have

$$\rho(x,y) = \rho_x(x)\rho_y(y).$$

$$\langle f(x)g(y)\rangle = \int dxdy\rho_x(x)\rho_y(y)f(x)g(y) = \langle f(x)\rangle \; \langle g(y)\rangle$$

$$cov(f(x),g(y)) = corr(f(x),g(y)) = 0.$$

Note that the opposite is in general not true!

From $cov = corr = 0$ we cannot conclude that x and y are independent

The central limit theorem

Consider sum x of N independent random variables s_i: $x = \sum_{i=1}^{N} s_i.$

We get:

$$\begin{aligned} \langle x \rangle &= \langle \sum_{i=1}^{N} s_i \rangle = \langle s_1 + s_2 + \ldots + s_N \rangle \\ &= \int ds_1 w_1(s_1) \ldots \int ds_N w_N(s_N)(s_1 + s_2 + \ldots + s_N) = \sum_{i=1}^{N} \langle s_i \rangle \end{aligned}$$

$$(\Delta x)^2 = \mathrm{Var}(x) = \mathrm{Var}(\sum s_i) = \sum \mathrm{Var}(s_i) = \sum (\Delta s_i)^2$$

and thus $$\frac{\Delta x}{\langle x \rangle} = \frac{\sqrt{\sum \mathrm{Var}(s_i)}}{\sum \langle s_i \rangle}$$

The central limit theorem

Assuming

$$\sum_i \mathrm{Var}(s_i) = O(N), \quad \sum \langle s_i \rangle = O(s_i)$$

we get the famous $1/\sqrt{N}$ rule:

$$\frac{\Delta x}{\langle x \rangle} = O(1/\sqrt{N}).$$

For the special case that all s_i are due to the same pdf we get

$$\frac{\Delta x}{\langle x \rangle} = \frac{\sqrt{N \mathrm{Var}(s)}}{N \langle s \rangle} = \frac{\sqrt{\mathrm{Var}(s)}}{\langle s \rangle} \frac{1}{\sqrt{N}}$$

The central limit theorem

Let us now study the distribution of x if we repeat the sum many times over different si, for simplicity we assume that all si follow the same distribution and that the higher moments exist:

$$\langle s^i \rangle = \int ds s^n w(s)$$

The probability distribution function is than given by

$$p(x) = \int ds_1 w(s) \int ds_2 w(s_2) \cdots \int ds_N w(s_N) \delta(x - \sum_{i=1}^{N} s_i).$$

Rewriting the delta function using

$$\delta(y) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dk exp(-iky)$$

The central limit theorem

we get:

$$
\begin{aligned}
p(x) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} dk \int ds_1 w(s) \int ds_2 w(s_2) \cdots \int ds_N w(s_N) exp(-ik(x - \sum_{i=1}^{N} s_i)) \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} dk exp(-ikx) \int ds_1 exp(-iks_1) w(s) \int ds_2 exp(-iks_2) w(s_2) \cdots \\
&\quad \int ds_N exp(-iks_N) w(s_N) \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} dk exp(-ikx) W(k)^N \qquad (61)
\end{aligned}
$$

The central limit theorem

To calculate $W(k)^N$ we expand

$$W(k) = \int ds \exp(-iks) w(s) = 1 + ik\langle s \rangle - \frac{1}{2} k^2 \langle s^2 \rangle + \dots$$

use

$$\ln(1+y) = y - y^2/2 + \dots$$

and obtain:

$$\begin{aligned} \ln\left([W(k)]^N\right) &= N \ln(1 + ik\langle s \rangle - \frac{1}{2} k^2 \langle s^2 \rangle + \dots) \\ &= N(ik\langle s \rangle - \frac{1}{2} k^2 \mathrm{Var}(s) + \dots) \end{aligned}$$

The central limit theorem

If we drop higher order terms we get

$$W(k)^N = \exp(N(ik\langle s\rangle - \frac{1}{2}k^2\mathrm{Var}(s))) = \exp(ik\langle x\rangle - \frac{1}{2}k^2\mathrm{Var}(x)))$$

and thus

$$\begin{aligned} p(x) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp(-ikx)\exp(ik\langle x\rangle - \frac{1}{2}k^2\mathrm{Var}(x))) \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp(ik(\langle x\rangle - x) - \frac{1}{2}k^2\mathrm{Var}(x))) \\ &= \frac{1}{2\pi\Delta x}\exp(-\frac{(x-\langle x\rangle)^2}{2\Delta x^2}) \end{aligned}$$

The End

Frequentist versus Bayesian probability

$$\frac{Hits}{Hits+miss}$$

$$P(\text{theory}|\text{data}) \propto P(\text{data}|\text{theory})P(\text{theory})$$