Learning goals of the week:

- covariance and correlation
- recall the basics of combinatorial calculus
- be able to work with discrete/continuous distributions
- recall the most common pdfs

Week 3

Multidimensional pdf’s

More than one random variable

The output of a measurement can be more than one variable:

Example: the direction of a particle in terms of two angles $(x_1, x_2) = (\vartheta, \phi)$;
the four-vectors of a particle $(x_1, x_2, x_3, x_4) = (pT, \vartheta, \phi, m)$

When performing the measurement we are performing a sampling of a 2D (or generally nD) joint pdf and the probability becomes

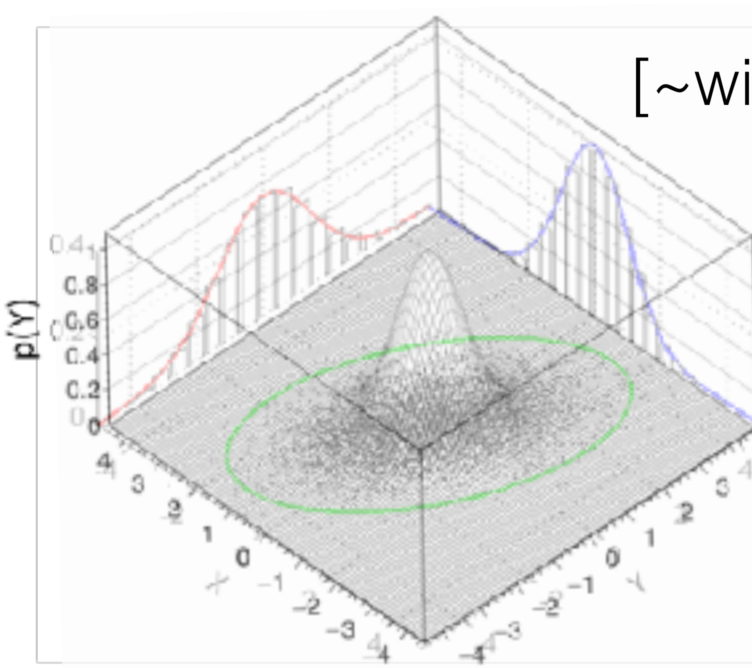
$$P(x_1 < X_1 < x_1 + \mathrm{d}x_1\,,\, x_2 < X_2 < x_2 + \mathrm{d}x_2) = f(x_1, x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2$$

$$P(a < X_1 < b\,,\, c < X_2 < d) = \int_a^b \mathrm{d}x_1 \int_c^d \mathrm{d}x_2\, f(x_1, x_2)$$

Marginal pdf: it's the pdf describing x_1 independently of the value of x_2; the 2D probability can be "marginalised / projected" by integrating over one variable:

$$f_1(x_1) = \int_{-\infty}^{+\infty} f(x_1, x_2)\,\mathrm{d}x_2$$

$$f_2(x_2) = \int_{-\infty}^{+\infty} f(x_1, x_2)\,\mathrm{d}x_1$$

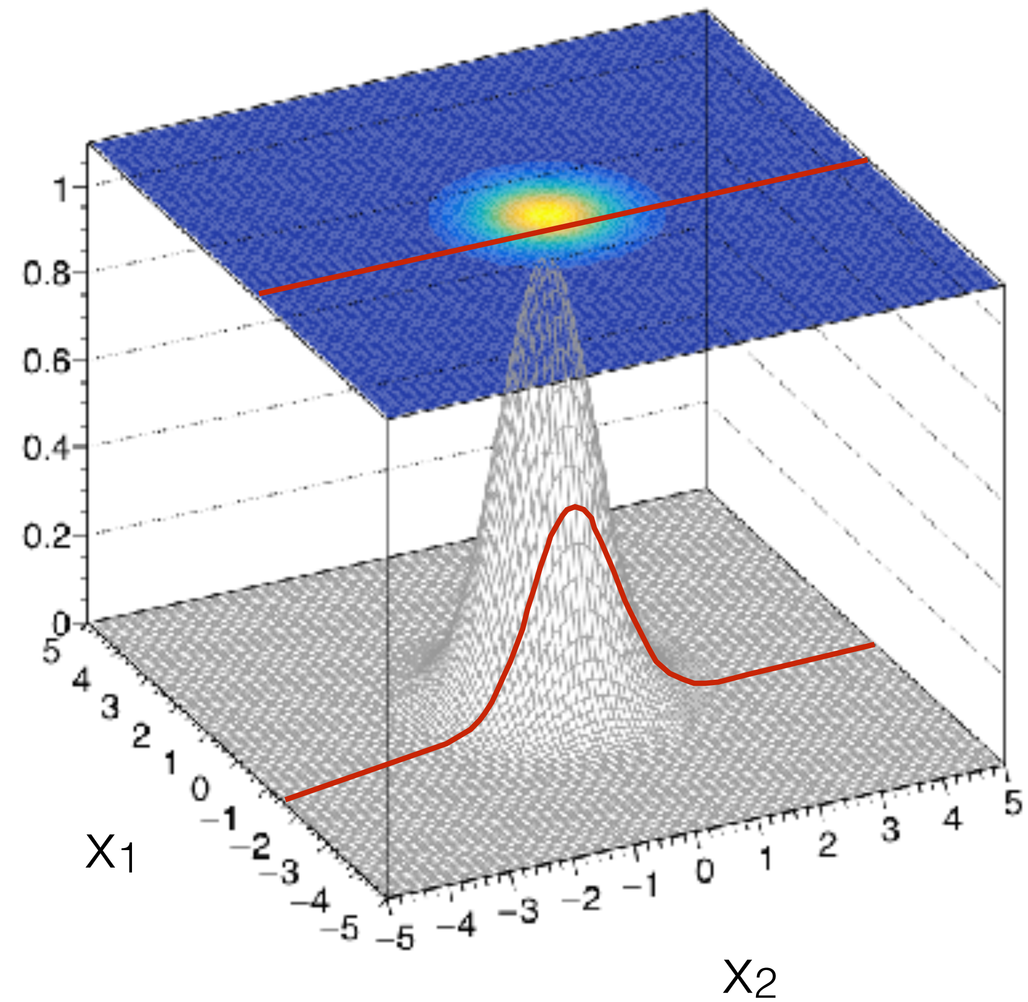


[~wiki]

More than one random variable

Conditional pdf: it's the pdf of x_2 for a fixed value of x_1: $f(x_2|x_1)$

$$P(a < X_2 < b \mid X_1 = x_1) = \int_a^b f(x_2 \mid x_1)\,\mathrm{d}x_2$$



In plain english:

Marginal pdf ignores the values of the other variable(s)

Conditional pdf fix the value of the other variable(s)

Correlated measurements

Example:

If you spend 42 hours each week at the university, the probability that at a randomly chosen moment your head is at the university is 1/4.
Similarly, the probability that your feet are at the university is 1/4. [Metzger]

What is the probability that both your head and your feet are at the university ?

Correlated measurements

Example:

If you spend 42 hours each week at the university, the probability that at a randomly chosen moment your head is at the university is 1/4.
Similarly, the probability that your feet are at the university is 1/4. [Metzger]

What is the probability that both your head and your feet are at the university ?

1/4, not 1/16 ! The location of your head and your feet are highly correlated…

Covariance

A way to quantify the dependence between two variables is given by the covariance

$$cov(x_1, x_2) = <(x_1 - <x_1>) \cdot (x_2 - <x_2>)> = <x_1 x_2> - <x_1><x_2>$$

If the two variables are independent then $<x_1 x_2> = <x_1><x_2>$ and the covariance is zero

Some properties:

- cov(x,a) = 0
- cov(x,x) = V(x)
- cov(x,y) = cov(y,x)
- cov(ax, by) = ab cov(x,y)
- cov(x+a,y+b) = cov(x,y)
- cov(x,y) has units

Covariance matrix:

when you have N variables the covariance is a [NxN] symmetric real matrix

Variance with two variables

The general formula for the variance of two random variables is:

$$V[ax + by] = a^2V[x] + b^2V[y] + 2abCov(x, y)$$

which, in case of uncorrelated variables becomes simply:

$$V\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} V(X_i)$$

Correlation

The correlation coefficient (linear correlation) is obtained normalizing the covariance by the standard deviation of the two variables

$$\rho_{x_1 x_2} = \frac{cov(x_1, x_2)}{\sqrt{V(x_1)V(x_2)}}$$

NB: contrary to the covariance, the correlation coefficient is dimensionless

Correlation matrix:
when you have N variables the covariance is a [NxN] symmetric real matrix with “1” on the diagonal

Sample covariance and correlation

Given a sample of size n: $(x_1,y_1)\ldots(x_n,y_n)$, the sample covariance or empirical covariance estimate of the covariance is

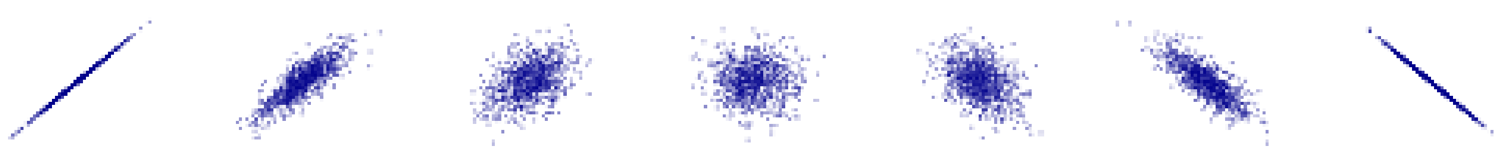
$$s_{xy} = \frac{1}{n-1}\sum_i (x_i - \bar{x})(y_i - \bar{y})$$

and the sample correlation or empirical correlation estimate of ρ_{xy}

$$r_{xy} = \frac{s_{xy}}{s_x s_y}$$

The (linear) correlation coefficient tells how well a set of measurement of two variables supports the hypothesis that the two are linearly dependent.

Exercise: Which one has the largest correlation ? positive / negative ?



(scatter plot)

Sample covariance and correlation

Given a sample of size n $(x_1,y_1)\ldots(x_n,y_n)$ the sample covariance or empirical covariance estimate of the covariance is

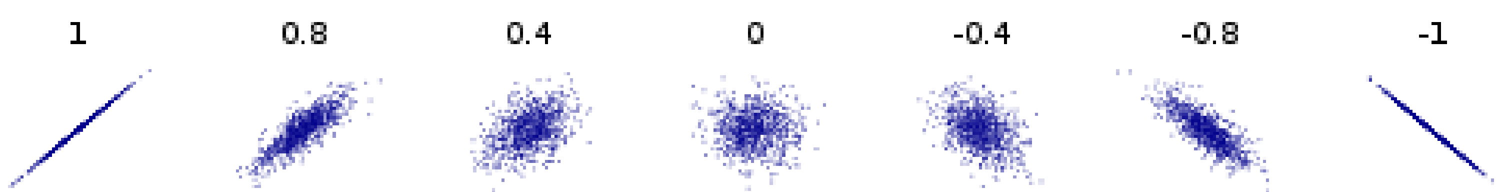
$$s_{xy} = \frac{1}{n-1}\sum_i (x_i - \bar{x})(y_i - \bar{y})$$

and the sample correlation or empirical correlation estimate of ρ_{xy}

$$r_{xy} = \frac{s_{xy}}{s_x s_y}$$

The (linear) correlation coefficient tells how well a set of measurement of two variables supports the hypothesis that the two are linearly dependent.

Exercise: Which one has the largest correlation ? positive / negative ?



(scatter plot)

Sample covariance and correlation

Given a sample of size n $(x_1,y_1)\ldots(x_n,y_n)$ the sample covariance or empirical covariance estimate of the covariance is

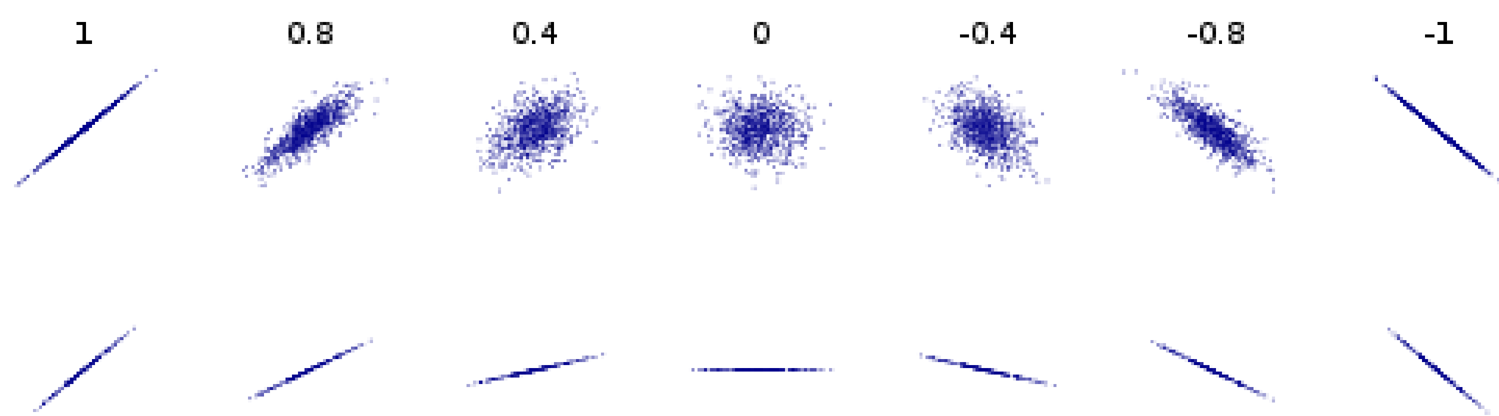
$$s_{xy} = \frac{1}{n-1}\sum_i (x_i - \bar{x})(y_i - \bar{y})$$

and the sample correlation or empirical correlation estimate of ρ_{xy}

$$r_{xy} = \frac{s_{xy}}{s_x s_y}$$

The (linear) correlation coefficient tells how well a set of measurement of two variables supports the hypothesis that the two are linearly dependent.

Exercise: Which one has the largest correlation ? positive / negative ?



(scatter plot)

Sample covariance and correlation

Given a sample of size n $(x_1,y_1)\ldots(x_n,y_n)$ the sample covariance or empirical covariance estimate of the covariance is

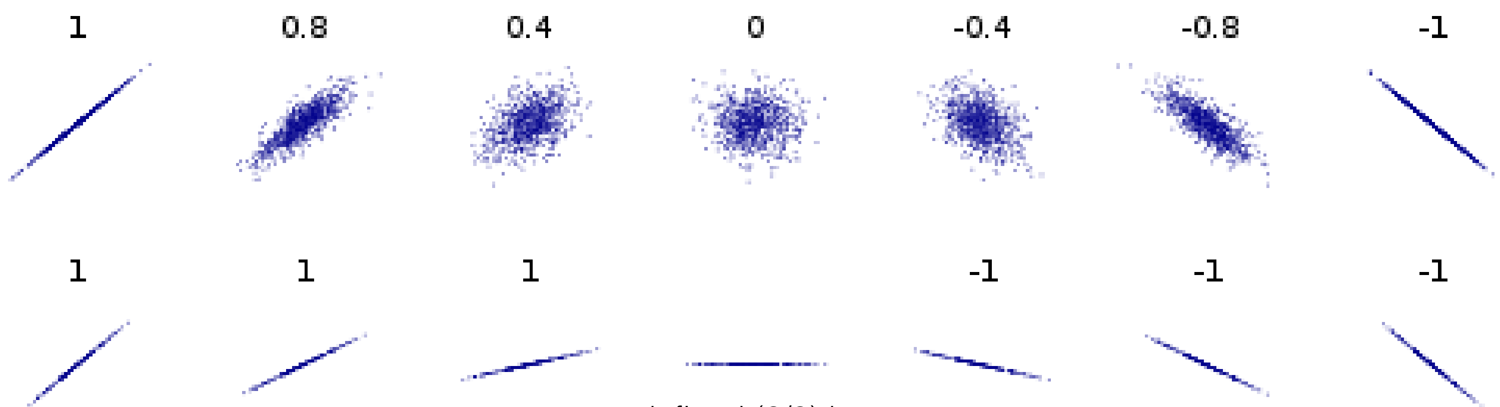
$$s_{xy} = \frac{1}{n-1}\sum_i (x_i - \bar{x})(y_i - \bar{y})$$

and the sample correlation or empirical correlation estimate of ρ_{xy}

$$r_{xy} = \frac{s_{xy}}{s_x s_y}$$

The (linear) correlation coefficient tells how well a set of measurement of two variables supports the hypothesis that the two are linearly dependent.

Exercise: Which one has the largest correlation ? positive / negative ?

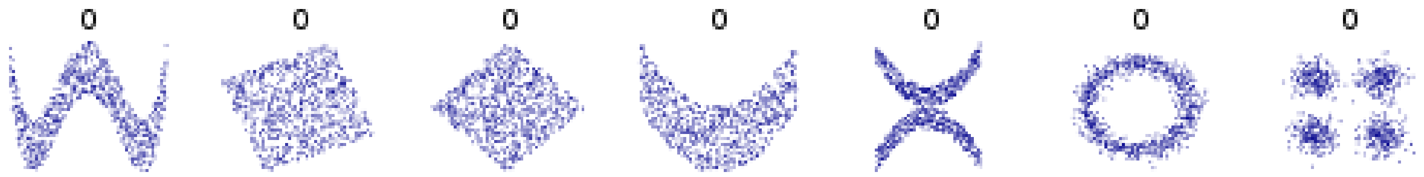


(scatter plot)

undefined (0/0) but not very interesting, as y is a constant for each value of x

Sample covariance and correlation

Exercise: Which one has the largest correlation ? positive / negative ?



The linear correlation coefficient for all these samples is zero, but it doesn't mean that the two variables are independent !

Error propagation (x,y)

Suppose you have two variables: x and y.
The uncertainty on a function of the two variables can be computed again from the Taylor expansion:

$$f(x,y) \approx f(x_0,y_0) + \left(\frac{\partial f}{\partial x}\right)_{x_0,y_0} \cdot (x - x_0) + \left(\frac{\partial f}{\partial y}\right)_{x_0,y_0} \cdot (y - y_0)$$

$$\sigma_f^2 \quad = \quad \left(\frac{\partial f}{\partial x}\right)^2 \sigma_x^2 + \left(\frac{\partial f}{\partial y}\right)^2 \sigma_y^2 + 2\frac{\partial f}{\partial x}\frac{\partial f}{\partial y} \cdot cov(x,y)$$

Which can be extended to several variables as

$$\sigma_f^2 = \sum_j \left(\frac{\partial f}{\partial x_j}\right)^2 \cdot \sigma_{x_j}^2 + \sum_j \sum_{k \neq j} \left(\frac{\partial f}{\partial x_j}\right)\left(\frac{\partial f}{\partial x_k}\right) \cdot cov(x_j, x_k)$$

Correlation is not causation

When looking at data, don't fool yourself !

"The first principle is that you must not fool yourself and you are the easiest person to fool."

R.P.Feynman

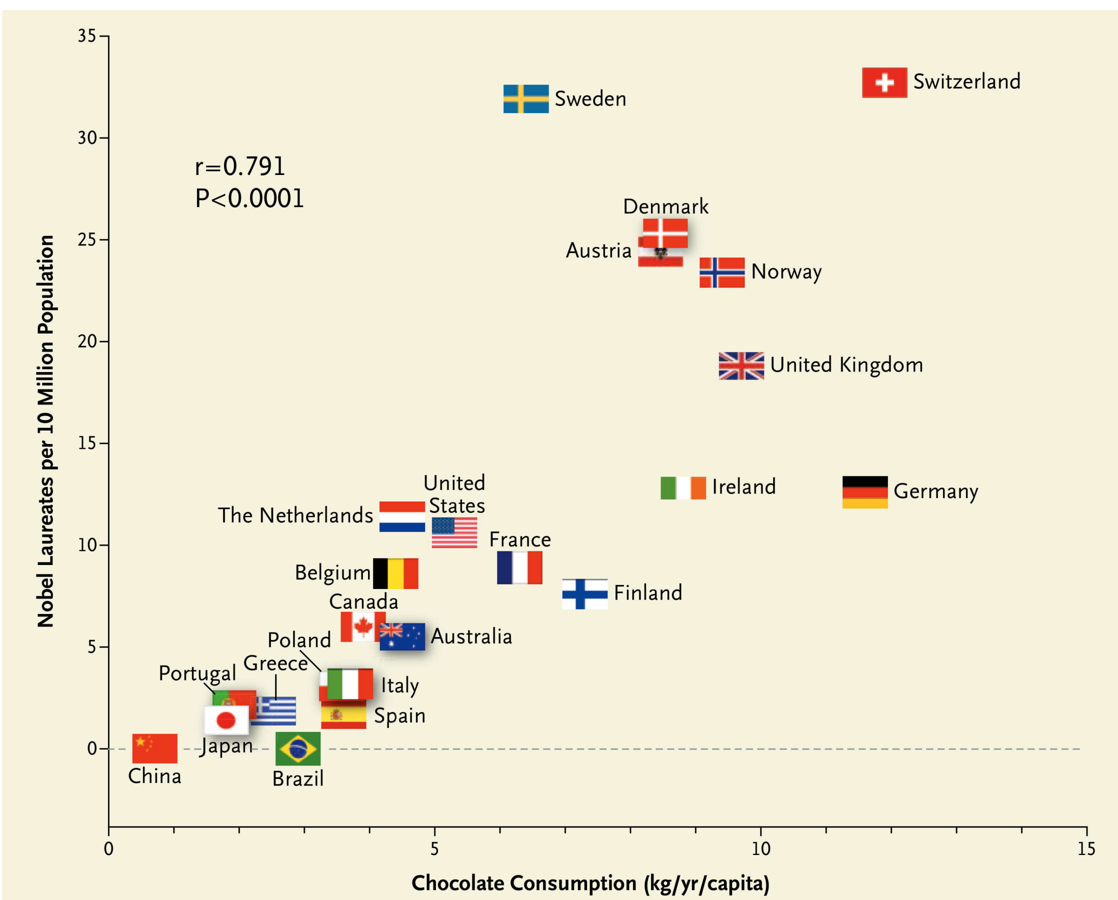
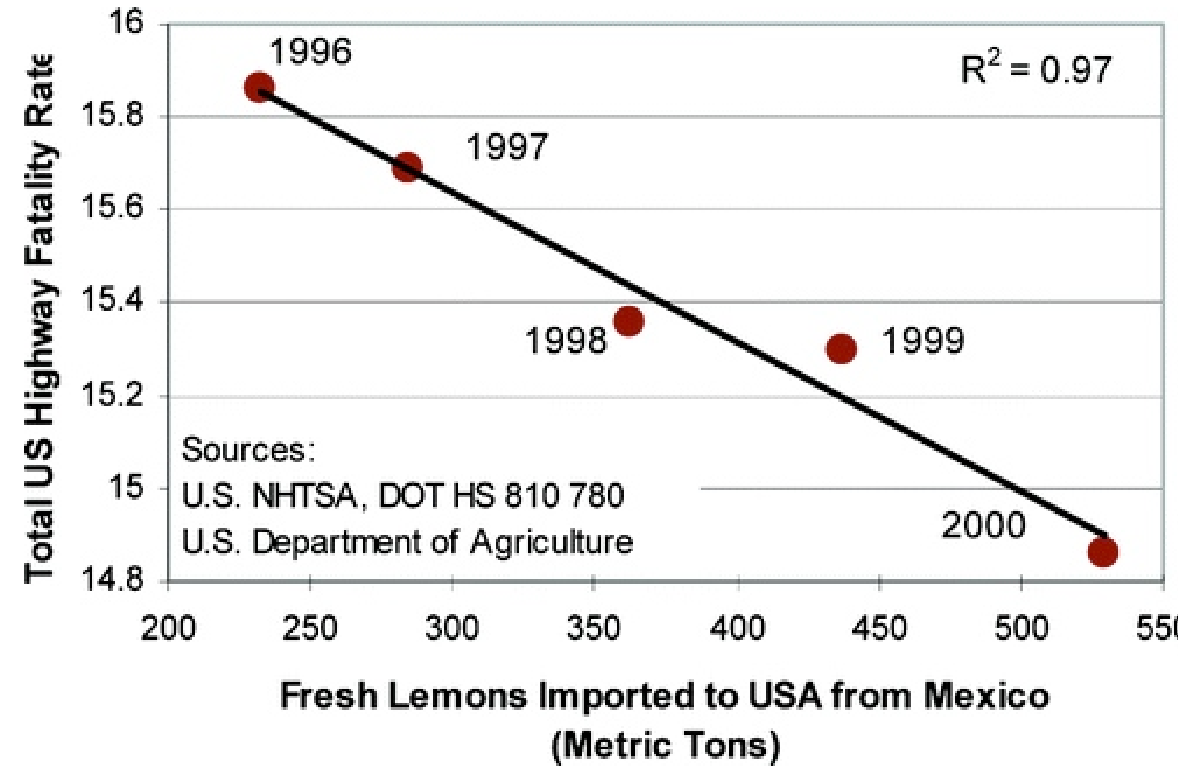
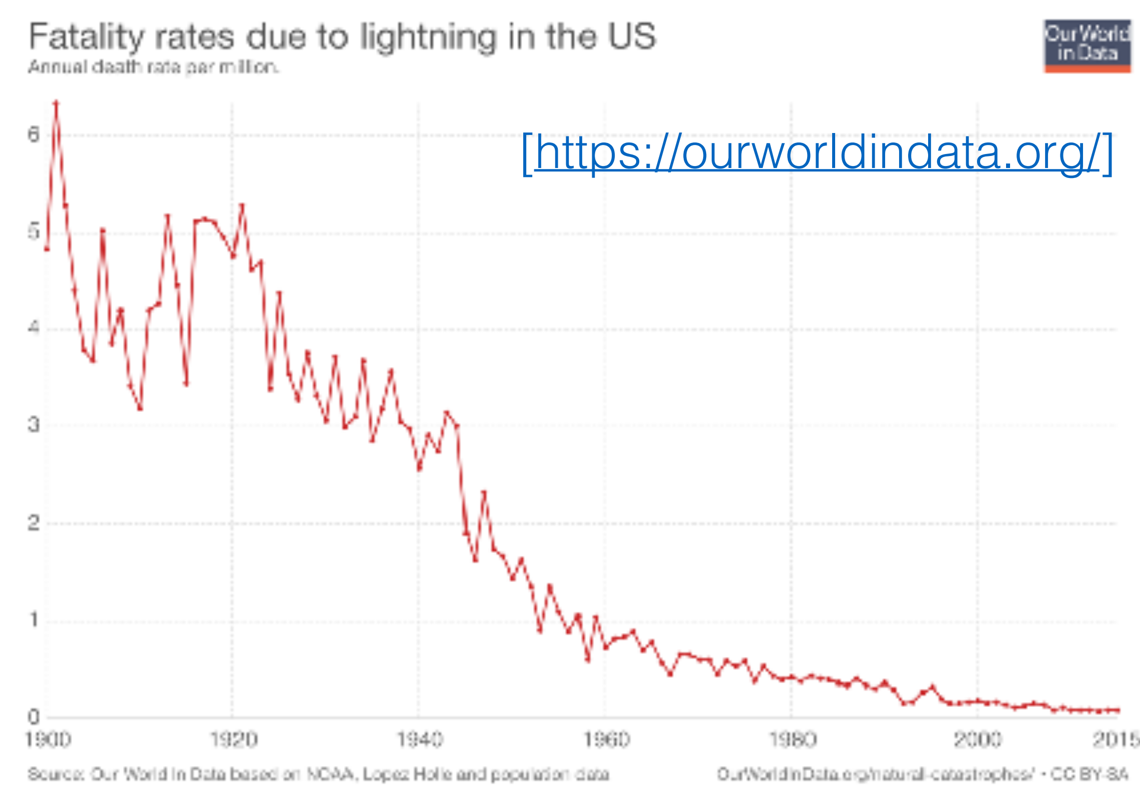


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.



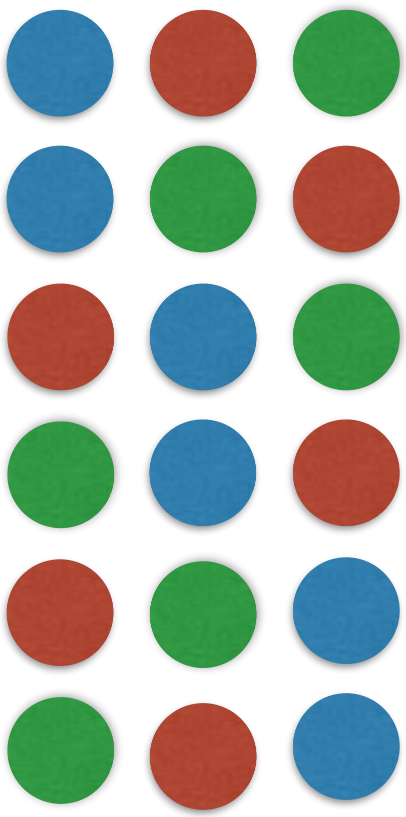


[https://ourworldindata.org/]

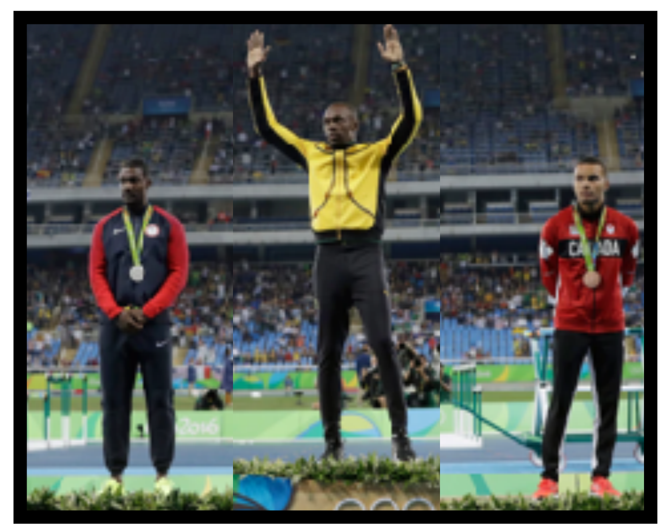
Combinatorial calculus

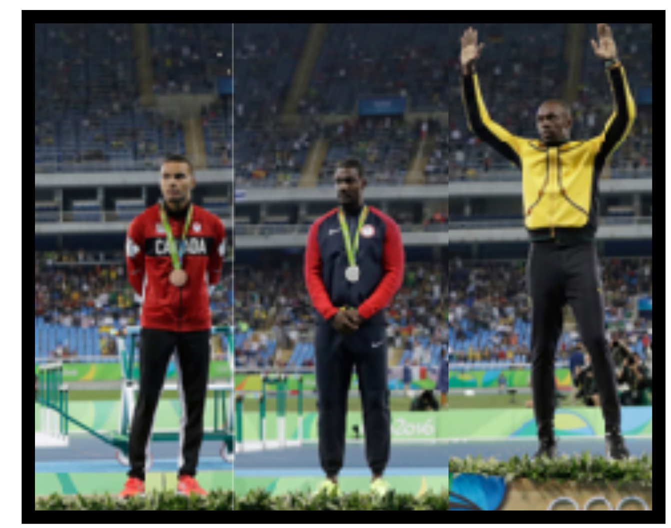
Combinatorial calculus

When talking about sequences of objects we call:
- permutations if we care about the order
- combinations if we don’t care about the order



Rio 2016 - 100m : Bolt, Gatlin, DeGrasse





These are all different permutations but equivalent combinations

Permutations with repetition

Pick r times from a set of n objects and put them back each time.
The number of possible permutations is —>
You always have all n objects to choose from

$$n^r$$

Example:
2 bit pick 4 times

0	0	0	0
1	0	0	0
0	1	0	0
1	1	0	0
0	0	1	0
1	0	1	0
0	1	1	0
1	1	1	0
0	0	0	1
1	0	0	1
0	1	0	1
1	1	0	1
0	0	1	1
1	0	1	1
0	1	1	1
1	1	1	1

2^4= 16 permutations

Example:
lock with 3 digits 0..9

10^3 = 1000 permutations

Example:
Mastermind: 6 colored pegs, 4 holes

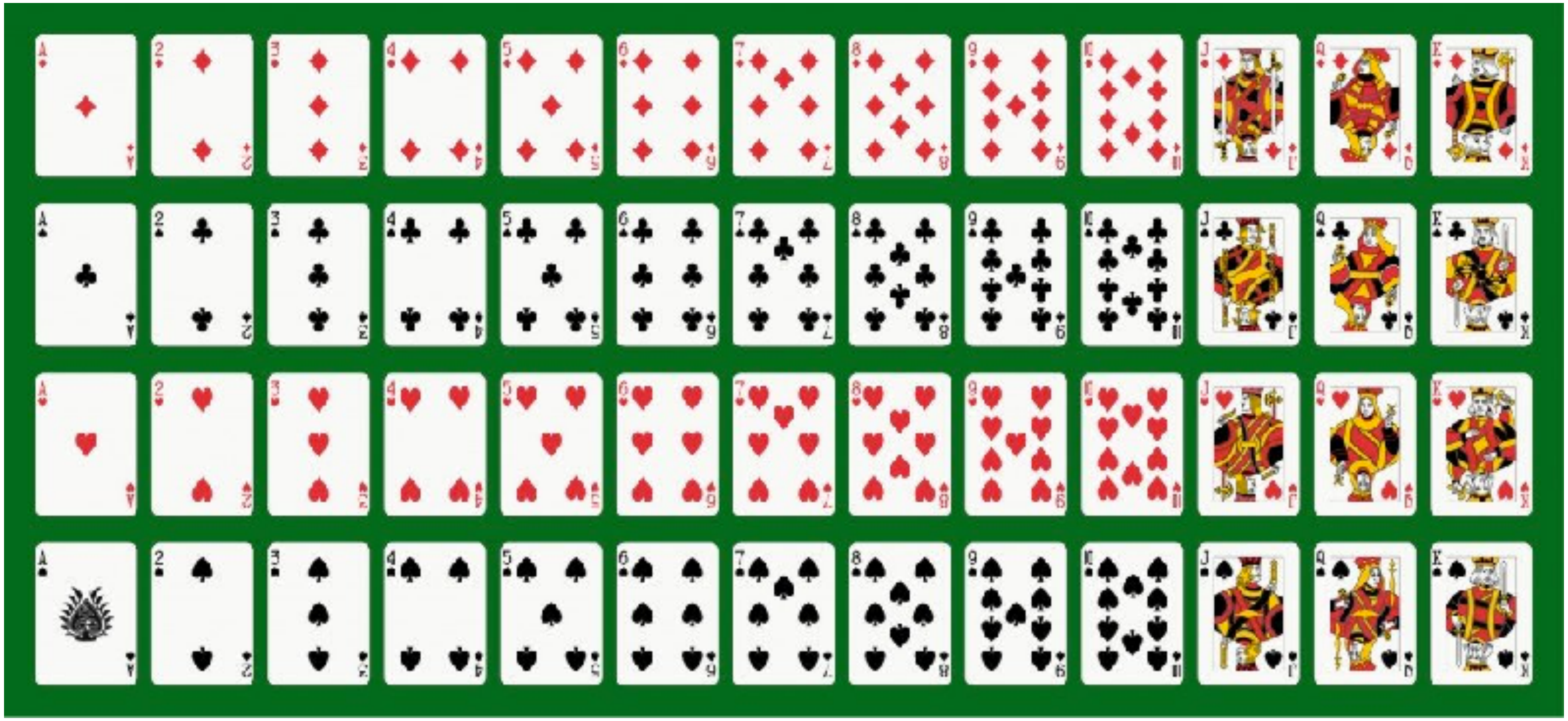
6^4= 1296 permutations

Permutations without repetition

Pick r times from a set of n objects and don't put them back.
The number of possible permutations is —>
Every time you pick one you have one less to choose from.

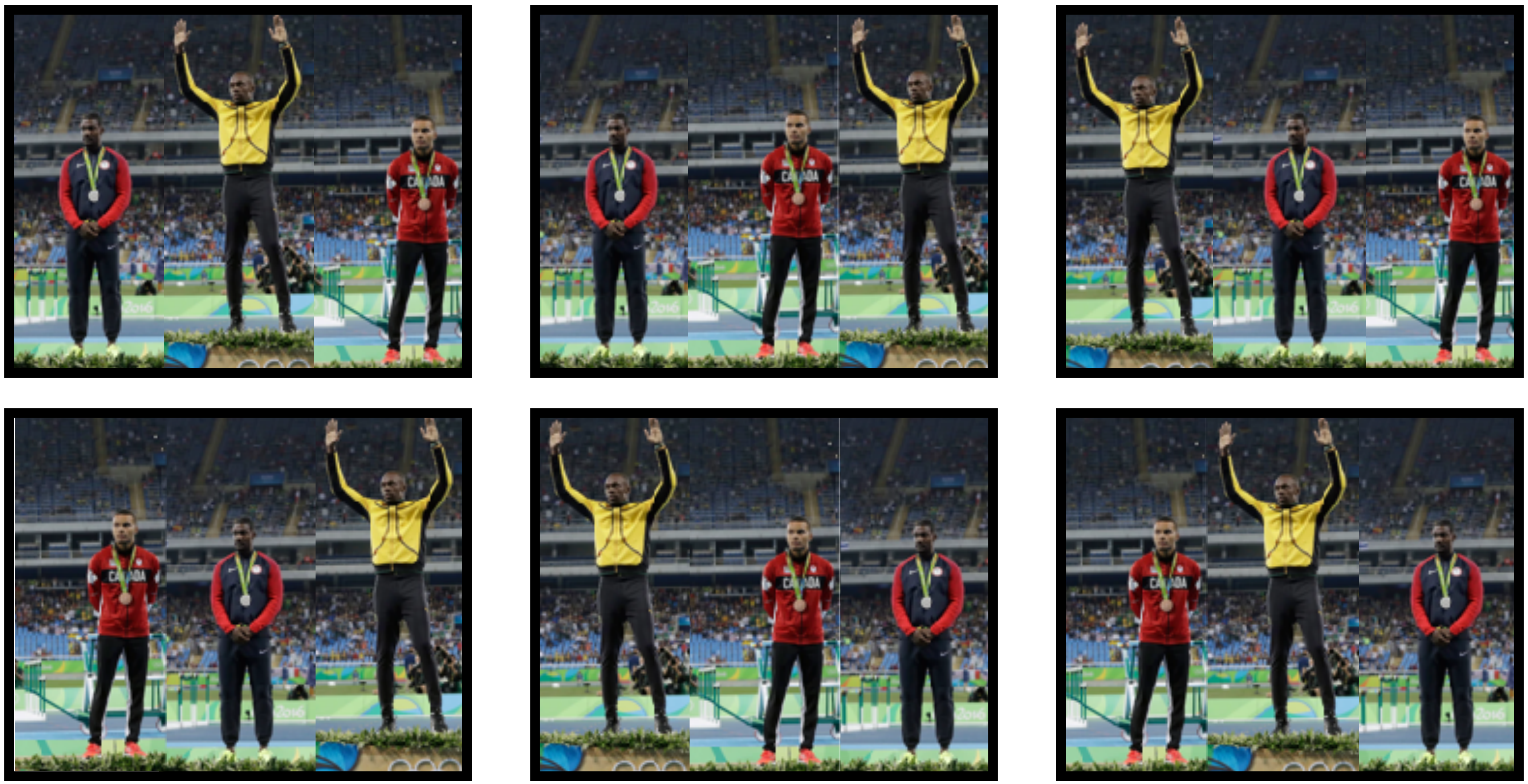
$$\frac{n!}{(n-r)!}$$

Example: 4 picks in a deck of 52 cards
first pick choose from 52, second from 51, third from 50, fourth from 49
52· 51· 50· 49 = 52! / (52-4)!



Permutations without repetition

Example: 3 athletes, 3 podium positions = 3! = 6 permutations



Combinations without repetition

Pick r times from a set of n objects and don't put them back, ignoring the order
The number of possible combinations is —>
Every time you pick one you have one less to choose from, and you need to divide by the number of ordered sequences r!

$$\frac{n!}{(n-r)!}\frac{1}{r!} = \binom{n}{r}$$

(read "n pick r")

Example: Lotto: 6 numbers extracted without putting them back from the set 1..90, the order is irrelevant.

The chances to win are $1/\binom{n}{r}$ ~ 10^{-9}

Combinations with repetition

Pick r times from a set of n objects and put them back, ignoring the order
The number of possible combinations is —>

$$\frac{(n+r-1)!}{(n-1)!r!} = \binom{n+r-1}{r}$$

Example: take n scoops of icecream. You can repeat them and the order is not important

Combinations with repetition

Pick r times from a set of n objects and put them back ignoring the order.

Derivation:
start from an example: take 3 objects from a set of 5 (a,b,c,d,e)
e.g. a a b,   a b c,   c c c

Think about the sequences of objects as classes and put a separator in between them
aab —> a a | b | | | (the last three are empty classes corresponding to c, d, e)
abc —> a | b | c | |
ccc —> | | c c c | |

We can also drop the letters and replace all symbols with "x"
e.g.: aab —> a a | b | | | —> x x | x | | |

The problem is now how many ways we can place r = 3 "x" and n = 5-1 = 4 "|"

This is the same as the combination w/o repetition " N pick R ", where in this case
N = n-1 + r (sum of all "|" and "x")
R = r

$$\binom{n-1+r}{r}$$

Discrete pdfs

Bernoulli trials

A Bernoulli trial is an experiment with only two outcomes: 0 / 1, pass / fail, yes / no and the success probability p (and failure probability q =1-p) is constant

The random variable is discrete and can take only the values r ∈{0,1}
The pdf is trivially the probability for an experiment to succeed or fail

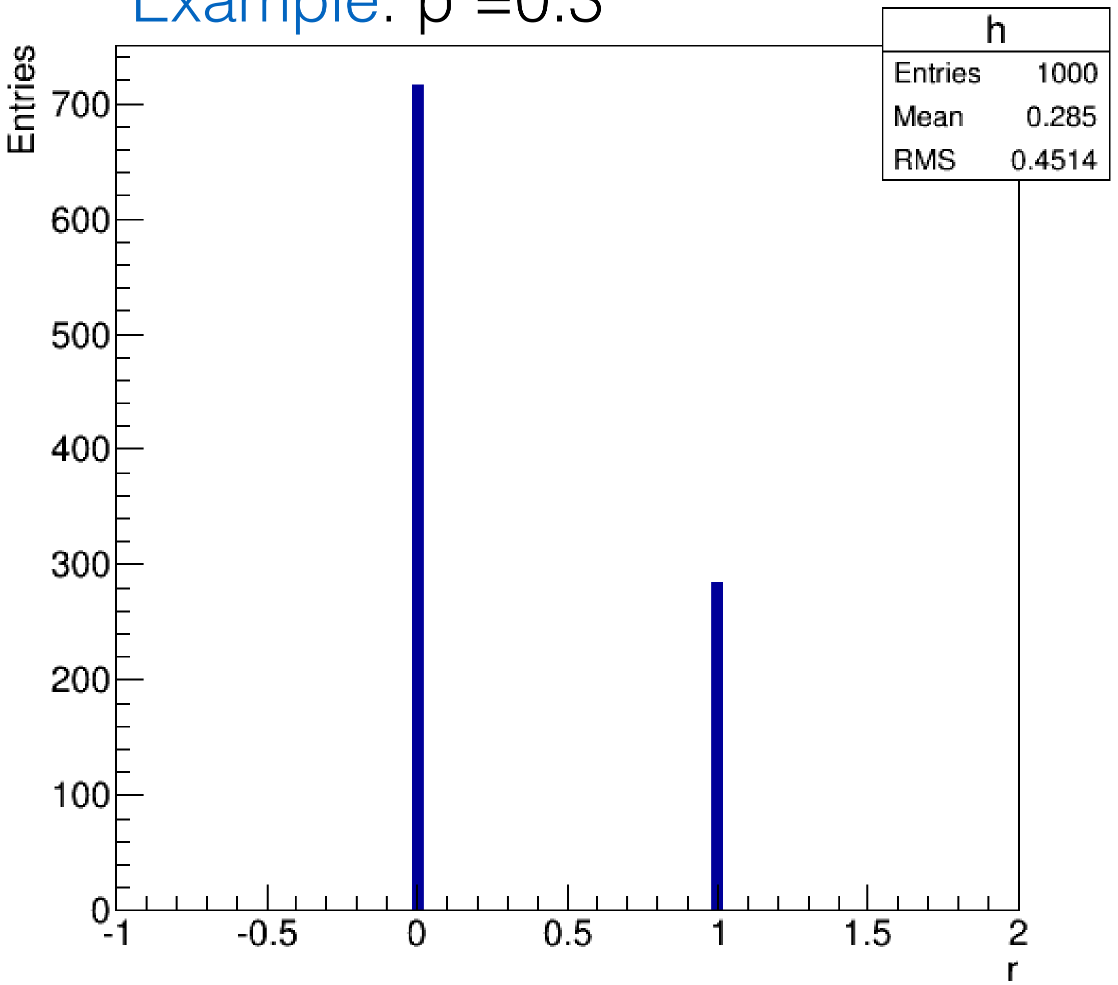
$$f(r;p) = p^r \; q^{(1-r)}$$

The first two moments are:

$$\mu = p$$

$$V(r) = p(1-p)$$

Example: p =0.3



Binomial distribution

Given n Bernoulli trials with success probability p, the binomial distribution gives the probability to observe r successes and n-r failures (independently of the order).

The random variable is r ∈{0,n}, the probability for r successes (for Bernoulli trials n = 1)

$$P(r;n,p) = \binom{n}{r} p^r (1-p)^{n-r}$$

The first two moments are:

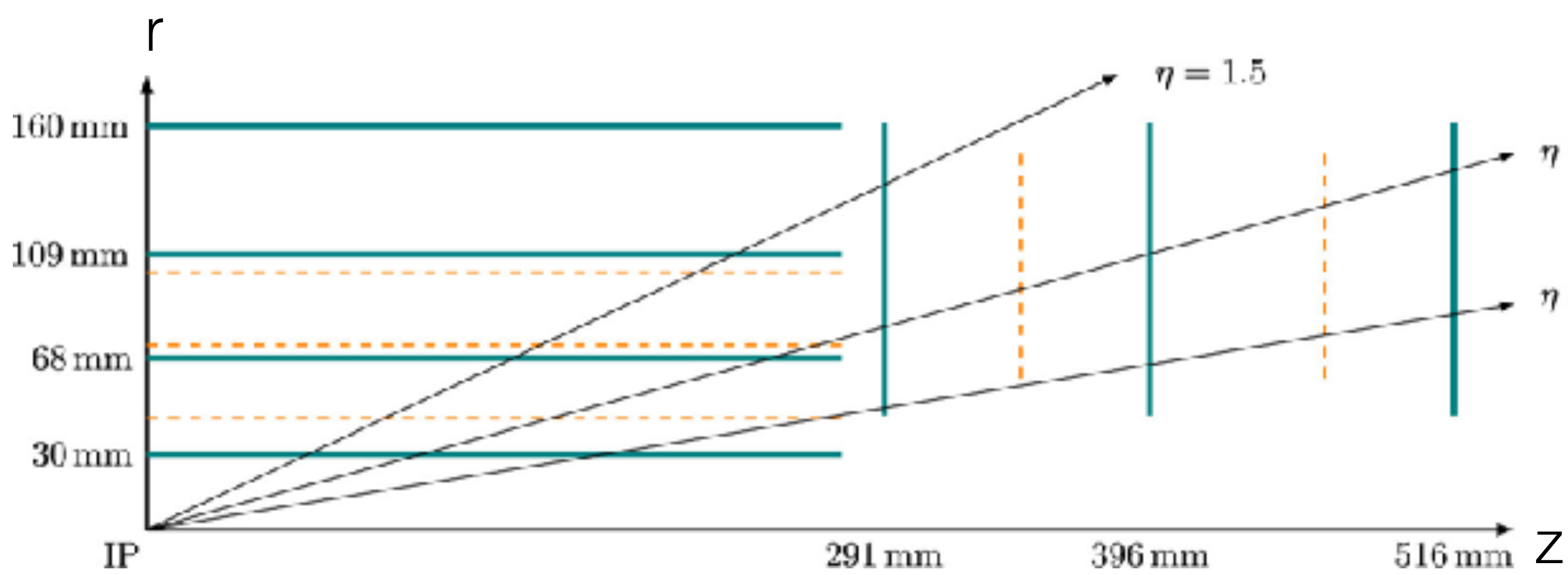
$$< r >= \sum_{r=0}^{n} r \cdot P(r) = np$$

$$V(r) = np(1-p)$$

Reproductive property: if X is binomially distributed as P(X;n,p) and Y is binomially distributed as P(Y; m,p) then X+Y is binomially distributed as P(X+Y; n+m, p).
NB: p is the same for X and Y

Binomial distribution

Example: a pixel detector has four layers and an efficiency per layer to detect a charged particle of 99%. What is the probability to reconstruct a track with 2 or 3 or 4 hits ? and what is the probability to reconstruct a track with at least 3 hits ?



r-ϕ

P(r = 2; n = 4, p=0.99) = 0.00 (0.0006)
P(r = 3; n = 4, p=0.99) = 0.038
P(r = 4; n = 4, p=0.99) = 0.96

P(r ≥ 2; n = 4, p=0.88) = sum of the above > 99%

Multinomial distribution

We can generalize the binomial distribution by considering k-types of outcomes instead of only 0 / 1. (Think about a measurement giving you the k-th bin in an histogram). Assume we have r_i measurements of type i with $i \in \{0,k\}$ (number of entries in bin i) with the total number of measurements $n = r_1+r_2+\ldots+r_k$ and probability p_i for each type of outcome (probability to enter in bin i).
The multinomial pdf is :

$$P(r_1, ..., r_k; n, p_1, ..., p_k) = \left(\frac{n!}{r_1!...r_k!}\right) p_1^{r_1} ... p_k^{r_k}$$

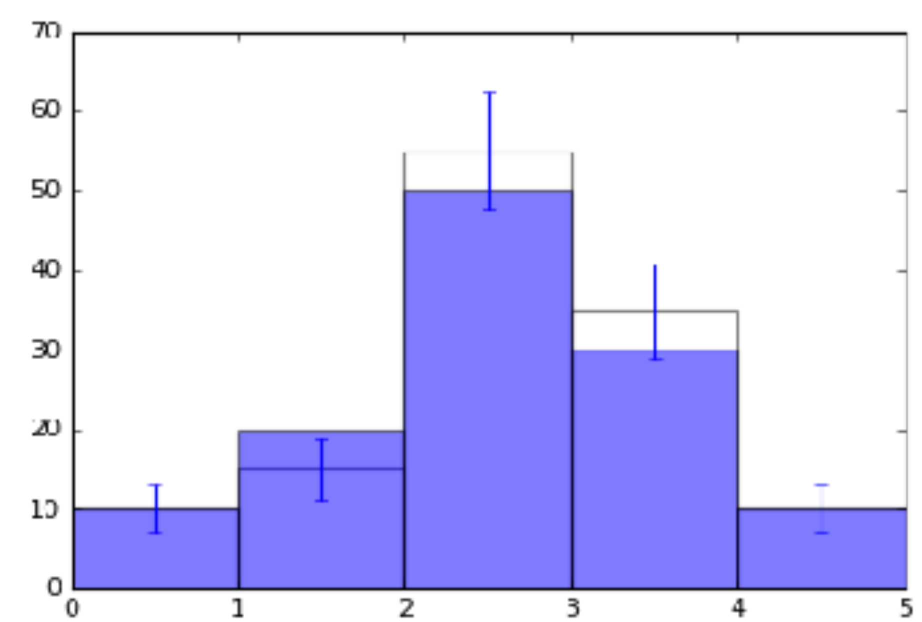
The first two moments are:

$$< r_i >= np_i$$

$$V(r_i) = np_i(1 - p_i)$$

Example: the multinomial distribution describes the probability to have r_i entries in bin i for an histogram with n bins

n = 100
p = [0.1, 0.2, 0.5, 0.3, 0.1]
r = [10, 15, 55, 35, 10]



Multinomial distribution

The covariance is:

$$cov(r_i, r_j) = -np_ip_j$$

and the correlation coefficient:

$$\rho_{ij} = \frac{cov(r_i, r_j)}{\sigma_i \sigma_j} = -\sqrt{\frac{p_i}{1-p_i}\frac{p_j}{1-p_j}}$$

NB: the correlation among bins comes from the fact that the total number of entries $n = r_1 + \ldots + r_k$ is fixed, i.e. $r_i = n - r_1 - \ldots - r_{i1} - r_{i+1} - \ldots - r_k$.
If n is not fixed, i.e. n is another random variable, the bin entries are uncorrelated and instead of having a multinomial we will have a Poisson for each bin.

Poisson distribution

The Poisson distribution is used when we know the number of outcomes but we don't know the number of trials. The probability to observe r events when we expect λ is:

$$P(r;\lambda) = \frac{\lambda^r e^{-\lambda}}{r!}$$

Example: with a radioactive source we measure the number of decays but we don't have the total number of nuclei (or the non decayed nuclei)

The Poisson distribution can be obtained as the limit of the binomial distribution when the number of independent trials is high, the success probability is small and constant and their product is constant (average rate) pn = λ (d*erivation on the next slide*).

First two moments:

$<r> = \lambda$

$V(r) = \lambda$ <— this is the origin of the n ±√n uncertainty in counting experiments
The number of observed counts fluctuates around its mean with this s.d.

Reproductive property: if X is Poisson distributed as $P(r;\lambda_r)$ and Y is Poisson distributed as $P(s;\lambda_s)$ then X+Y is Poisson distributed as $P(r+s;\lambda_r+\lambda_s)$.

Example: number of muons from two different sources (e.g. pions and kaons decays)

Poisson distribution

The distribution can be obtained as a limit of the binomial: let λ be the probability to observe a radioactive decay in a period T of time. Now divide the period T in n time intervals $\Delta T = T/n$ small enough that the probability to observe two decays in an interval is negligible. The probability to observe a decay in ΔT is then λ/n, while the probability to observe r decays in the period T is given by the binomial probability to observe r events in n trials each of which has a probability λ/n.

$$P\left(r; n, \frac{\lambda}{n}\right) = \frac{n!}{(n-r)!}\frac{1}{r!}\left(\frac{\lambda}{n}\right)^r\left(1-\frac{\lambda}{n}\right)^{n-r} \tag{2.1.16}$$

Under the assumption that $n >> r$ then:

$$\frac{n!}{(n-r)!} = n(n-1)(n-2)\ldots(n-r+1) \sim n^r \tag{2.1.17}$$

and

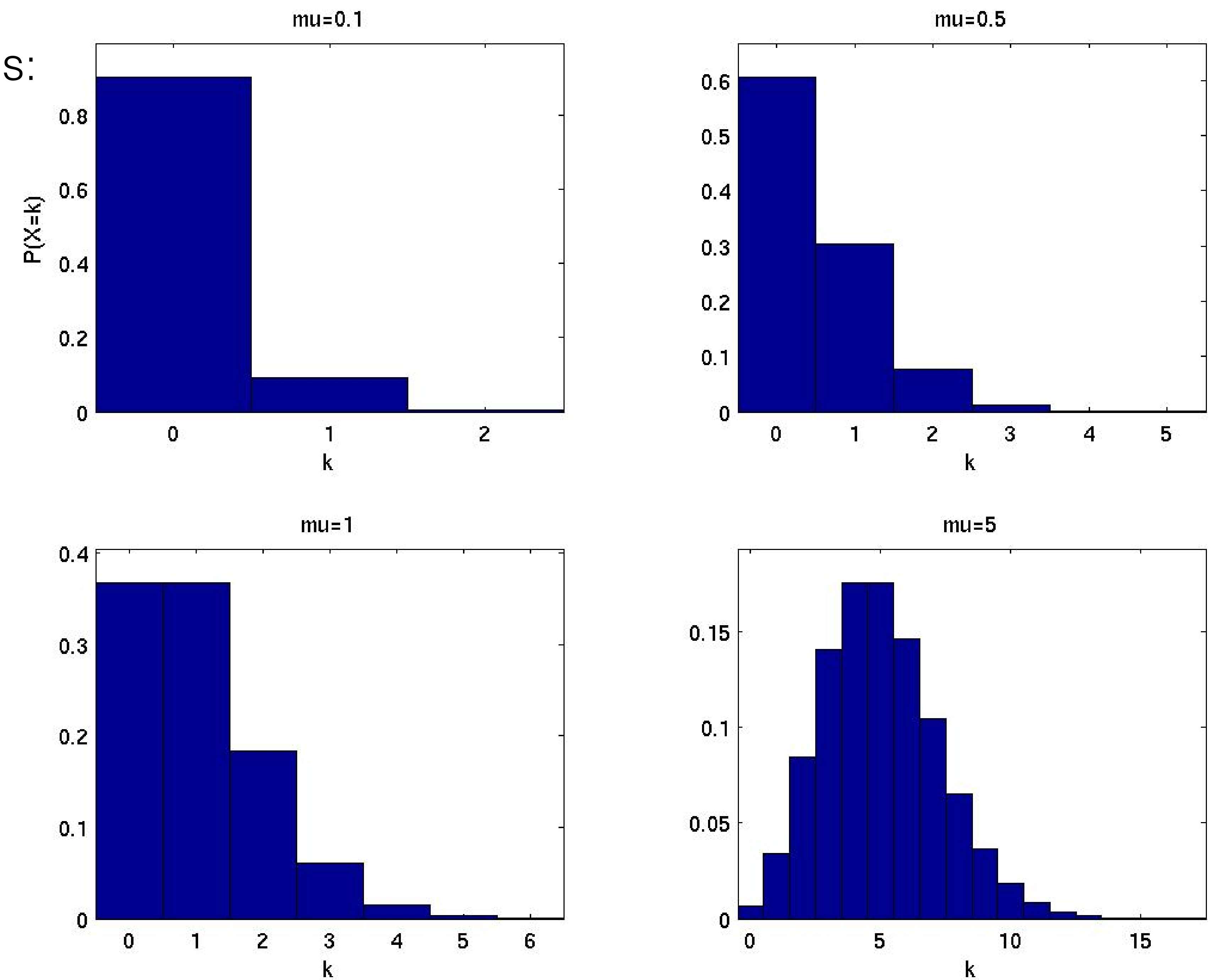
$$\left(1-\frac{\lambda}{n}\right)^{n-r} \sim \left(1-\frac{\lambda}{n}\right)^{n} \to e^{-\lambda} \qquad \text{for} \qquad n\to\infty \tag{2.1.18}$$

and replacing this in the binomial expression in Eq. 2.1.16 we obtain the Poisson p.d.f:

$$P(r;\lambda) = \frac{\lambda^r e^{-\lambda}}{r!}. \tag{2.1.19}$$

Poisson distribution

Some examples:



When λ < 1 the most probable value of the distribution is 0
When λ is integer then both λ=n and λ=n-1 are equally probable

Poisson distribution

Example: An historical example [1] is the number of deadly horse accidents in the Prussian army. The fatal incidents were registered over twenty years in ten different cavalry corps. There was a total of 122 fatal incidents, and therefore the expectation value per corps per year is given by $\lambda = 122/200 = 0.61$. The probability that no soldier is killed per year and corps is $P(0; 0.61) = e^{-0.61} \cdot 0.61^0/0! = 0.5434$. To get the total events (of no incidents) in one year and per corps, we have to multiply with the number of observed cases (here 200), which yields $200 \cdot 0.5434 = 108.7$. The total statistics of the Prussian cavalry is summarized in Tab. 2.1.1, in agreement with the Poisson expectation. □

Fatal incidents per corps and year	Reported incidents	Poisson distribution
0	109	108.7
1	65	66.3
2	22	20.2
3	3	4.1
4	1	0.6

Table 2.1.1: The total statistics of deadly accidents of Prussian soldiers.

Poisson distribution

The Poisson distribution is used to describe counting experiments, provided that the assumption “number of trials is high, the success probability is small and constant” is fulfilled.

Examples described by a Poisson distribution:

- number of interactions caused by an intense beam of particles on a thin target
- number of entries in an histogram for events taken in a given time period (integral not fixed)

Examples not described by a Poisson distribution:

- decay of a small amount of radioactive material in a time interval comparable to its lifetime
- number of interactions caused by an intense beam of particles on a thick target

Poisson distribution

The Poisson p.d.f. requires that the events be independent.

Example: Consider the case of a counter with a dead time of 1 μsec.
In this case for high fluxes, the number of particles detected in some time interval will not be Poisson distributed because the detection of a particle is not independent of the detection of other particles.

- If the particle flux is low, the chance of a second particle within the dead time is small and it can be neglected —> Poisson distributed
- If the flux is high the dead time cannot be —> not Poisson distributed

Continuous pdfs

Uniform distribution

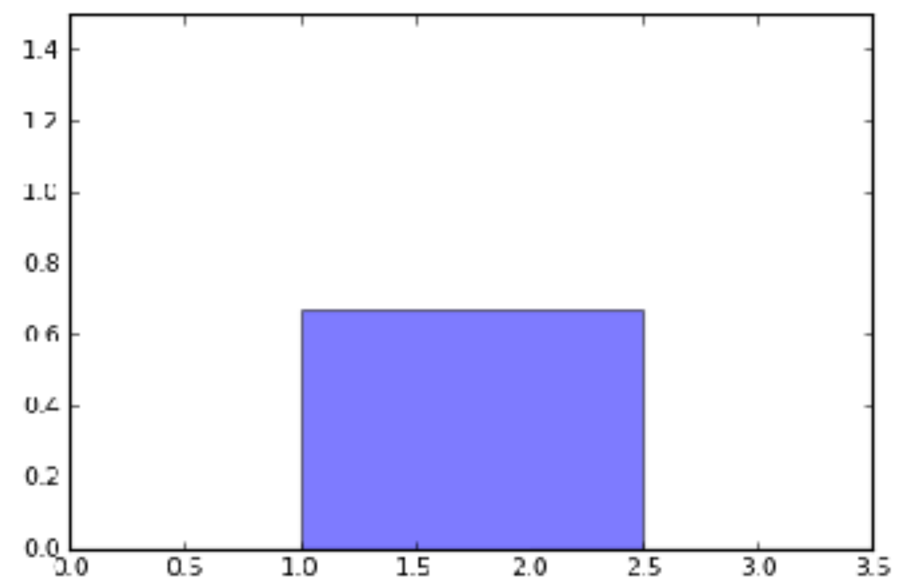
The uniform pdf is

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } a \leq x \leq b \\ 0 & \text{else.} \end{cases}$$

The probability to obtain a random value between a and b is constant

First two moments:

$$\begin{aligned} < x > &= \int_a^b \frac{x}{b-a} dx = \frac{1}{2}(a+b) \\ Var(x) &= \frac{1}{12}(b-a)^2 \end{aligned}$$

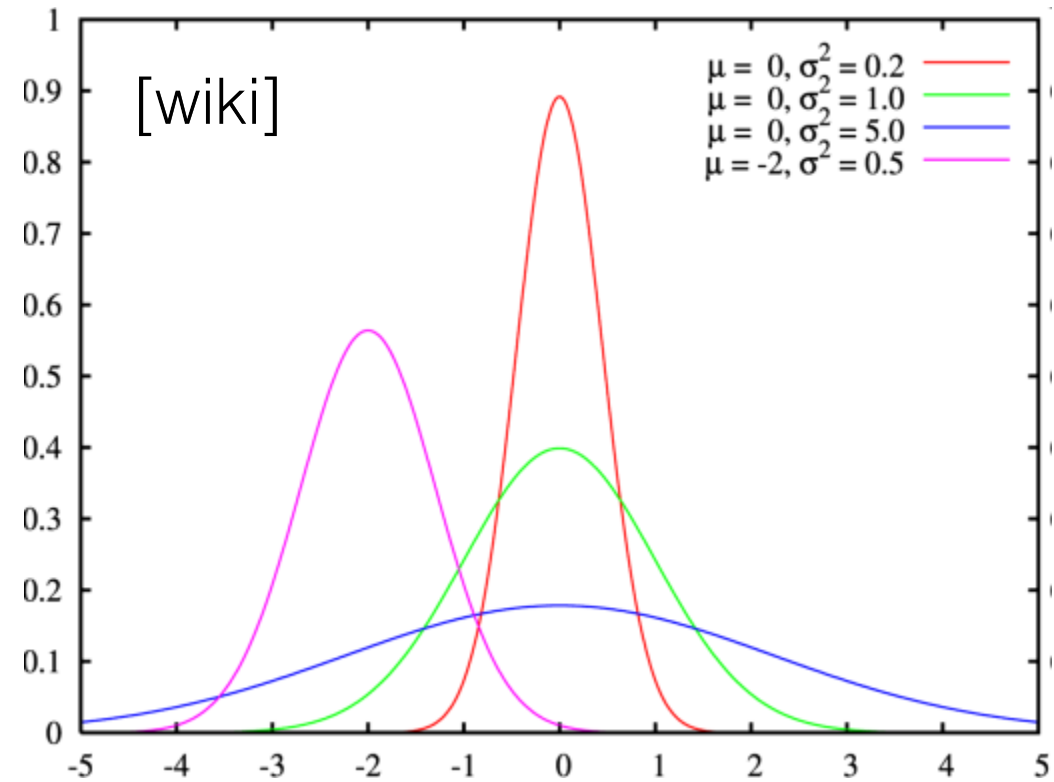


Example: what is the resolution of a single silicon strip detector of width 1 mm ? You have a binary signal: if a charge particle hits the strip you have a signal, if it misses the signal you don't. Take as resolution the 1/√12 = 290 um.

Gaussian distribution

The gaussian distribution is how you expect your measurements to be distributed when they are affected by a large number of additive 'noise' components.

$$f(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



The first two moments are:

$$\int_{-\infty}^{+\infty} xP(x;\mu,\sigma)dx = \mu$$

$$\int_{-\infty}^{+\infty} (x-\mu)^2 P(x;\mu,\sigma)dx = \sigma^2$$

Basic properties:

- symmetric about μ
- σ characterize the width
- inflection point happens at x = σ
- the maximum is at x = μ ; G(μ, σ) = 1/√2π σ
- FWHM = 2 σ√2ln2 = 2.355 σ

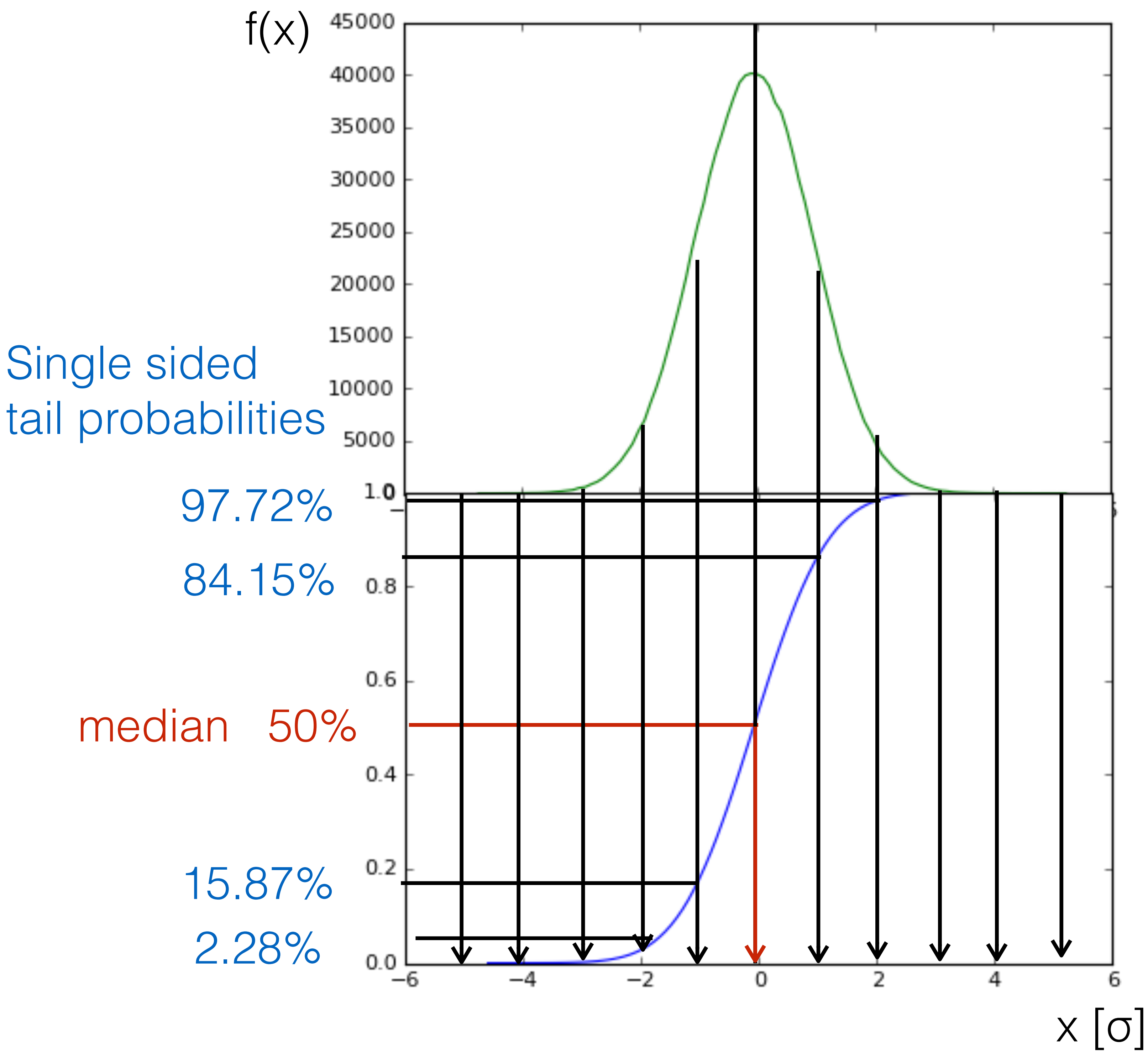
The "standard distribution" G(μ=0,σ=1) can be obtained by changing variable

$$z = \frac{x-\mu}{\sigma} \quad N(0,1) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

Reproductive property: If X and Y are two independent r.v.'s distributed as $f(x; \mu_x, \sigma_x)$ and $f(y; \mu_y, \sigma_y)$ then Z = X+Y is distributed as $f(z;\mu_z,\sigma_z)$ with $\mu_z = \mu_x + \mu_y$ and $\sigma_z = \sigma_x + \sigma_y$.

Cumulative of the Gaussian

$$\Phi(x) = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{x} e^{-(t-\mu)^2/2\sigma^2} dt$$



Double sided tail probability

$P(|x-\mu| \leq 1\sigma) = 68.27\%$
$P(|x-\mu| \leq 2\sigma) = 95.45\%$
$P(|x-\mu| \leq 3\sigma) = 99.73\%$

$90\% = P(|x-\mu| \leq 1.645\sigma)$
$95\% = P(|x-\mu| \leq 1.960\sigma)$
$99\% = P(|x-\mu| \leq 2.576\sigma)$
$99.9\% = P(|x-\mu| \leq 3.290\sigma)$

Multidimensional Gaussian

$$f(x,y;\mu_u,\sigma_x,\mu_y,\sigma_y) = \frac{1}{\sqrt{2\pi}\sigma_x}\frac{1}{\sqrt{2\pi}\sigma_y} e^{-\frac{(x-\mu_x)^2}{2\sigma_x^2}} e^{-\frac{(x-\mu_y)^2}{2\sigma_y^2}}$$

Properties:

- $\langle x \rangle = \mu_x$ and $\langle y \rangle = \mu_y$
- $V(x) = \sigma_x$ and $V(y) = \sigma_y$

Or in general with N *correlated* variables

$$f(\vec{x};\vec{\mu},V) = \frac{1}{(2\pi)^{N/2}|V|^{1/2}} \exp\left(-\frac{1}{2}(\vec{x}-\vec{\mu})^T V^{-1}(\vec{x}-\vec{\mu})\right)$$

$\vec{x}$ and $\vec{\mu}$ are column vectors with the components
$\vec{x}^T$and $\vec{\mu}^T$are the corresponding row vectors
V is the covariance matrix (|V | its the determinant)

Properties:

- $\langle x_i \rangle = \mu_i$
- $V(x_i) = V_{ii}$
- $\mathrm{cov}(x_i,x_j)=V_{ij}$

Exponential distribution

Example: unstable particle/nucleus lifetime

$$f(t) = \frac{1}{\tau}e^{-\frac{t}{\tau}} \qquad f(ct) = \frac{1}{c\tau}e^{-\frac{ct}{c\tau}}$$

t: proper decay time   ct: proper decay length

Example: charge of a short-circuited capacitor

$$Q = Q_0 e^{-\lambda t}$$

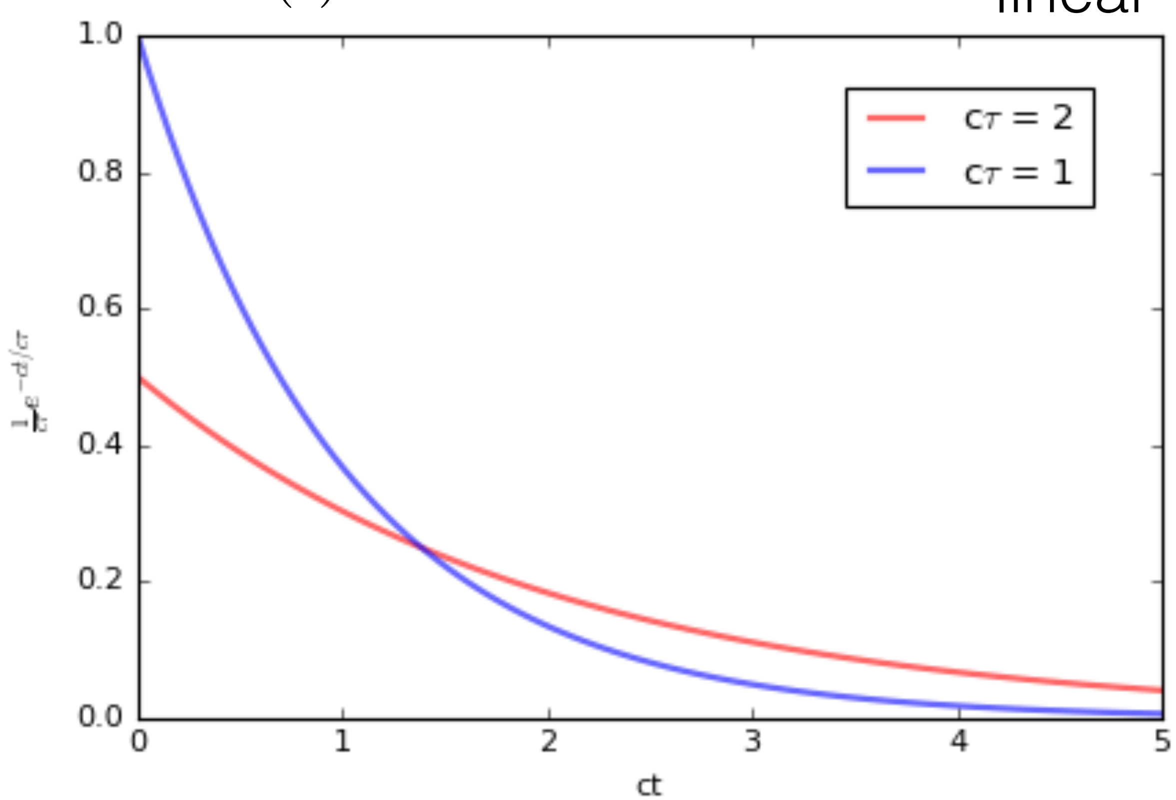
where $\lambda = 1/(RC)$.

Properties:

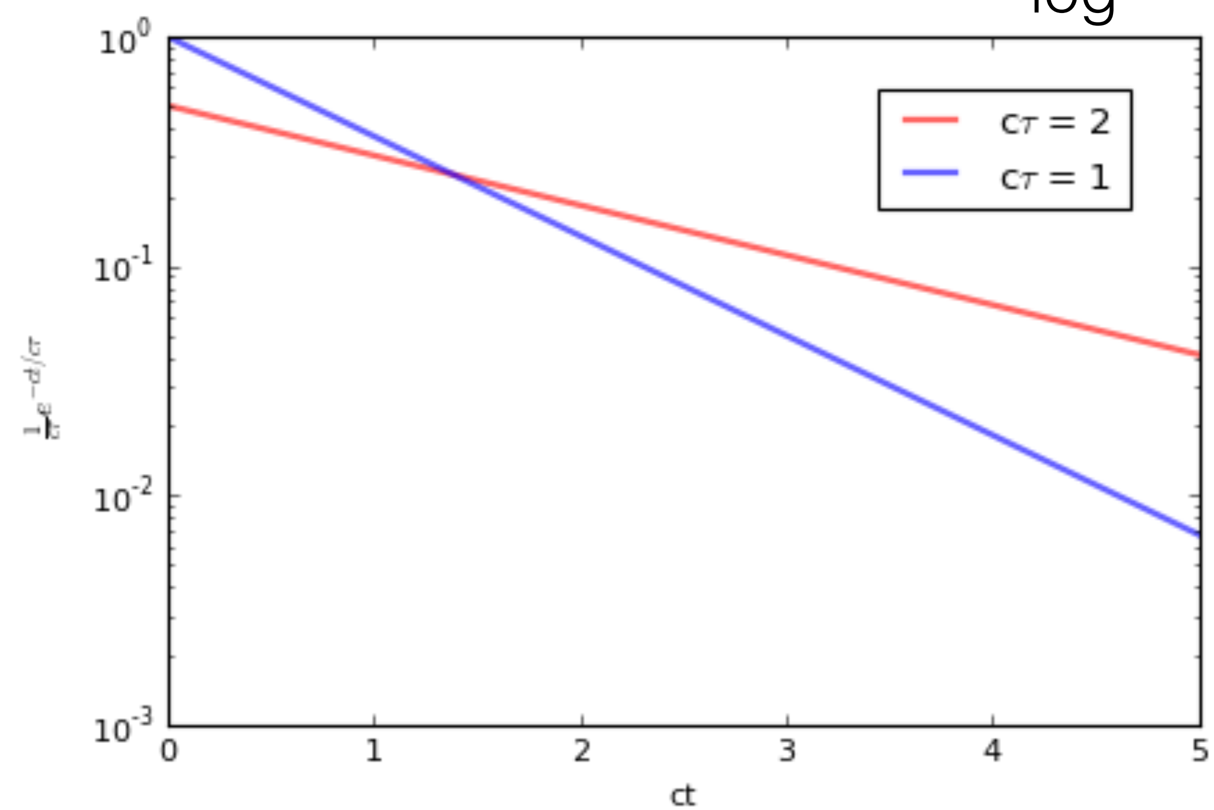
$$\langle \tau \rangle = \frac{1}{\tau}\int_0^{\infty} t e^{-t/\tau} dt = \tau$$

$$Var(t) = \tau^2$$

linear



log



Proper decay time (length) of an unstable particle.
(the steepest the distribution the shorter the lifetime)

LogNormal distribution

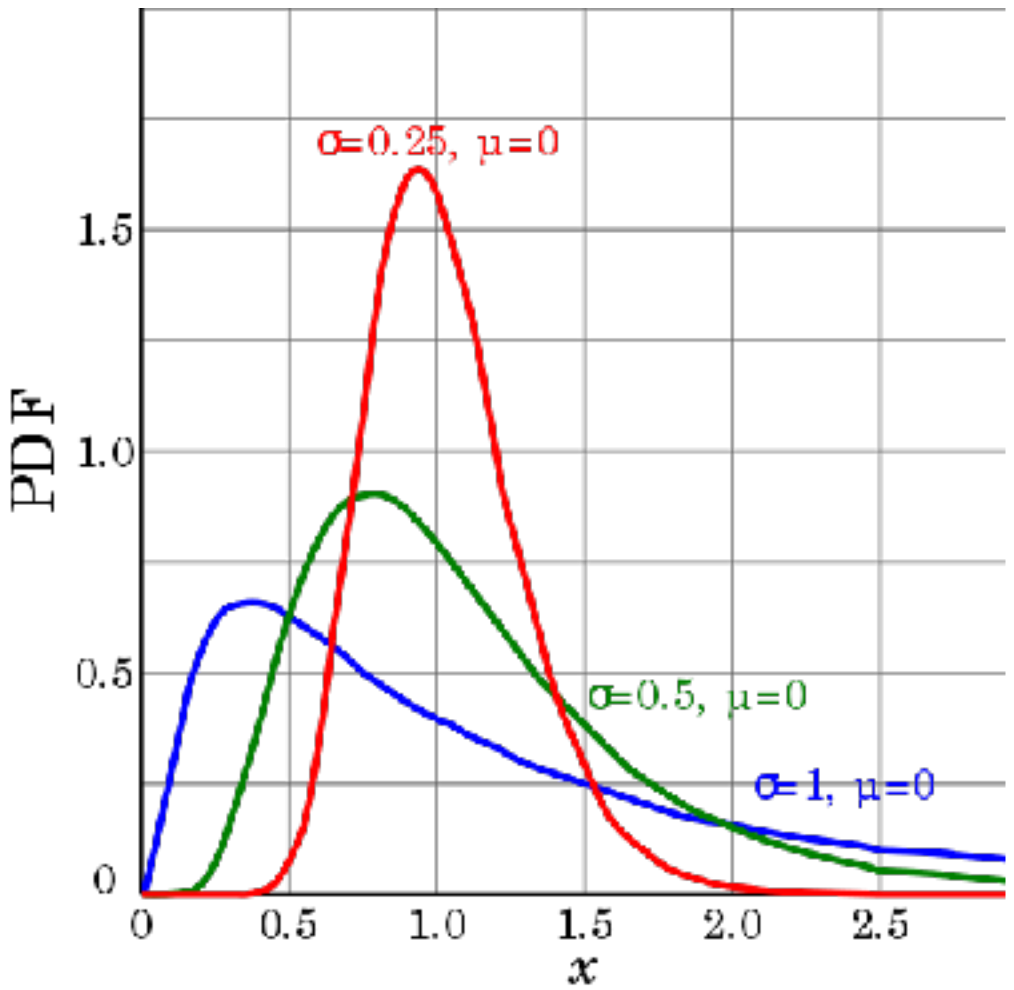
$$f(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}}\frac{1}{x}e^{-(\ln x-\mu)^2/2\sigma^2}$$

If y obeys a normal distribution with μ, then x = exp(y) obeys a log-normal distribution (y = ln x)

Properties:

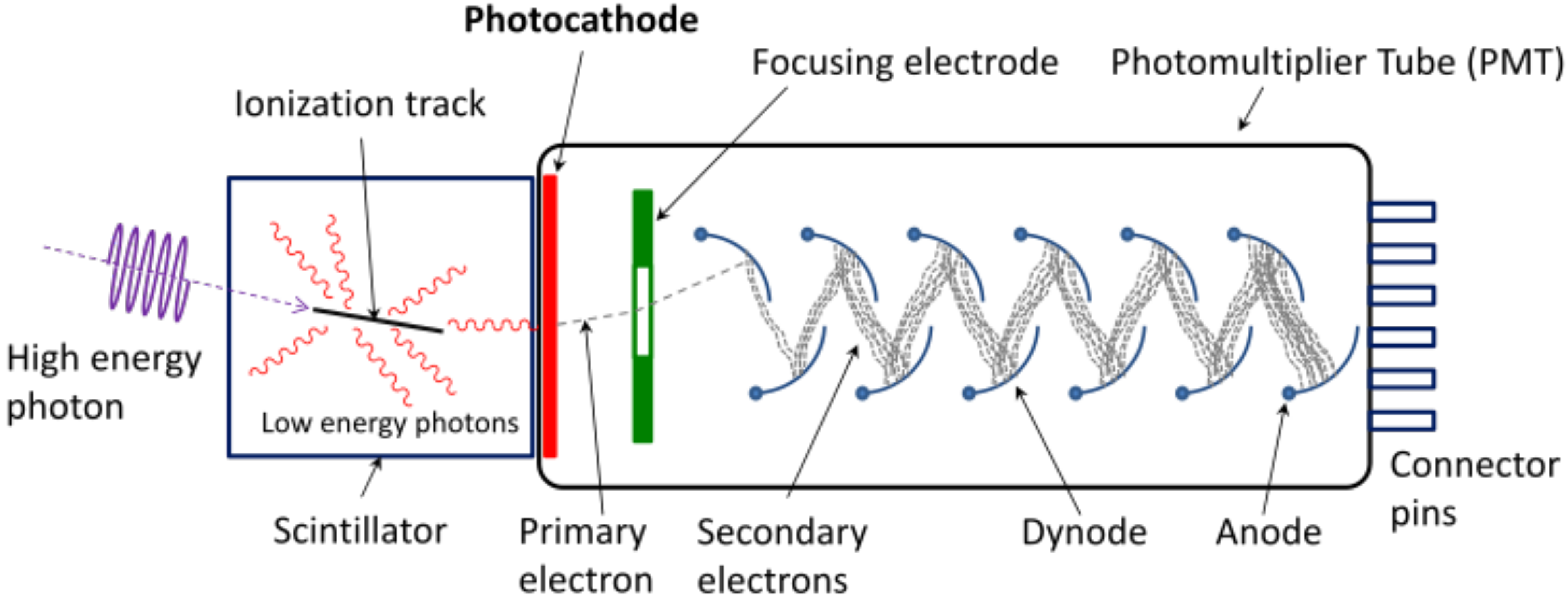
$$\begin{aligned} < x > &= e^{(\mu+\frac{1}{2}\sigma^2)} \\ Var(x) &= e^{(2\mu+\sigma^2)}(e^{\sigma^2}-1) \end{aligned}$$

The log-normal distribution is used to model the response of a system where the resolution is given by the product of the effect of several sources (as opposed to the gaussian where you have additive contributions)



LogNormal distribution

Example Consider the signal of a photomultiplier (PMT), which converts light signals into electric signals. Each photon hitting the photo-cathode emits an electron, which gets accelerated by an electric field generated by an electrode (dynode) behind. The electron hits the dynode and emits other secondary electrons which gets accelerated to the next dynode. This process if repeated several times (as many as the number of dynodes in the PMT). At every stage the number of secondary electrons emitted depends on the voltage applied.



If the amplification per step is a_i, then the number of electrons after the k^{th} step,

$n_k = \Pi_{i=0}^{k} a_i$ is approximately log-normal distributed.

χ^2 (chi-squared) distribution

The chi-squared pdf is the joint probability of the product of n gaussians

$$f(\mathbf{x};\mu,\sigma) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left[-\frac{1}{2}\left(\frac{x_i-\mu_i}{\sigma_i}\right)^2\right]$$

$$= \exp\left[-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i-\mu_i}{\sigma_i}\right)^2\right] \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma_i}$$

We define $$\chi^2(n) = \sum_{i=1}^{n}\left(\frac{x_i-\mu_i}{\sigma_i}\right)^2$$ as the chi-squared variable with n-degrees of freedom

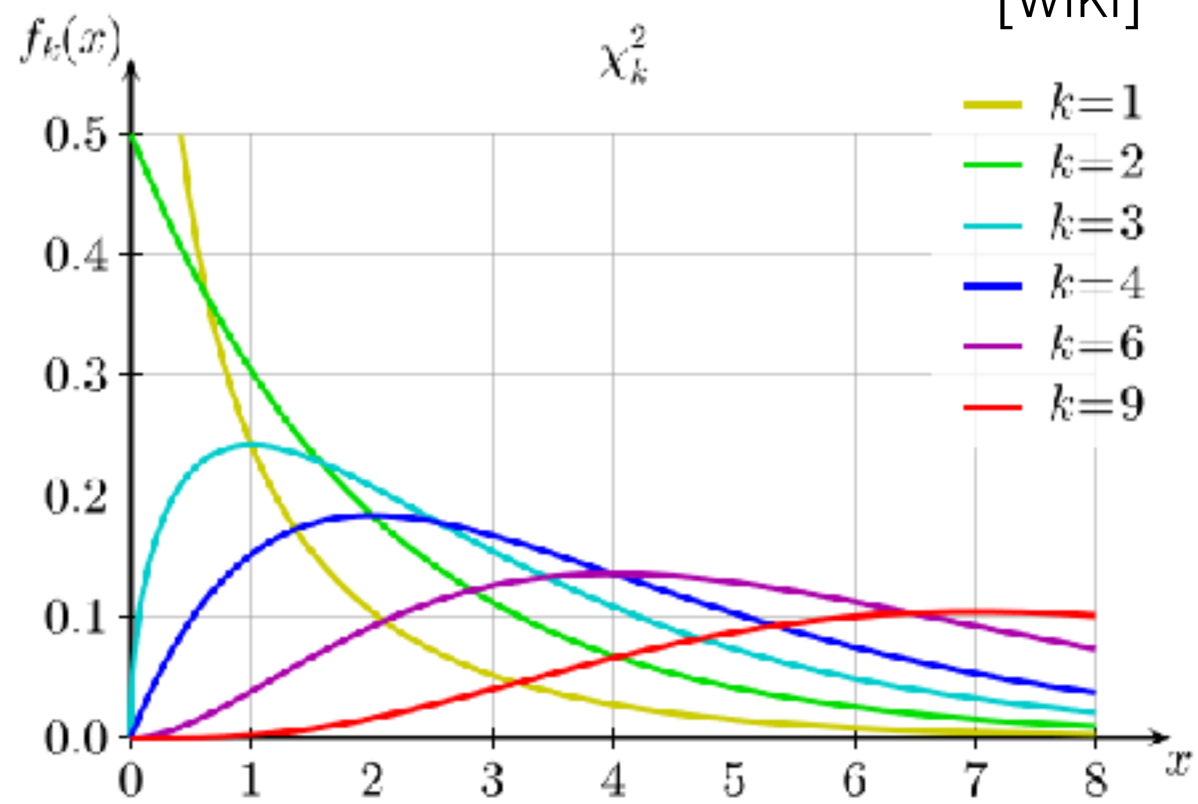
The pdf is:

$$\chi^2(n) = f(\chi^2;n) = \frac{(\chi^2)^{n/2-1}e^{-\chi^2/2}}{\Gamma(n/2)2^{n/2}}$$

This will be particularly relevant when discussing least-squares fits

χ^2 (chi-squared) distribution

[wiki]



Properties:

- mean=n
- variance = 2n
- mode=n-2 for n≥2 and 0 for n≤2

We define: $\chi^2(n)/n$ as reduced χ^2. (we will encounter this when talking about goodness of fit)

For $n \to \infty$, $\chi^2(n) \to \text{Gaussian}(\chi^2; n, 2n)$

Typically we can approximate the χ^2 distribution to a gaussian for $n \geq 30$.

Student's t-distribution

The Student's t distribution can be used to estimate the mean of a normally distributed parent distribution when the sample size is small and parent distribution standard deviation is unknown and the one evaluated from the sample s is used.

Take n samples from a gaussian parent and compute:

mean $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ variance $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$

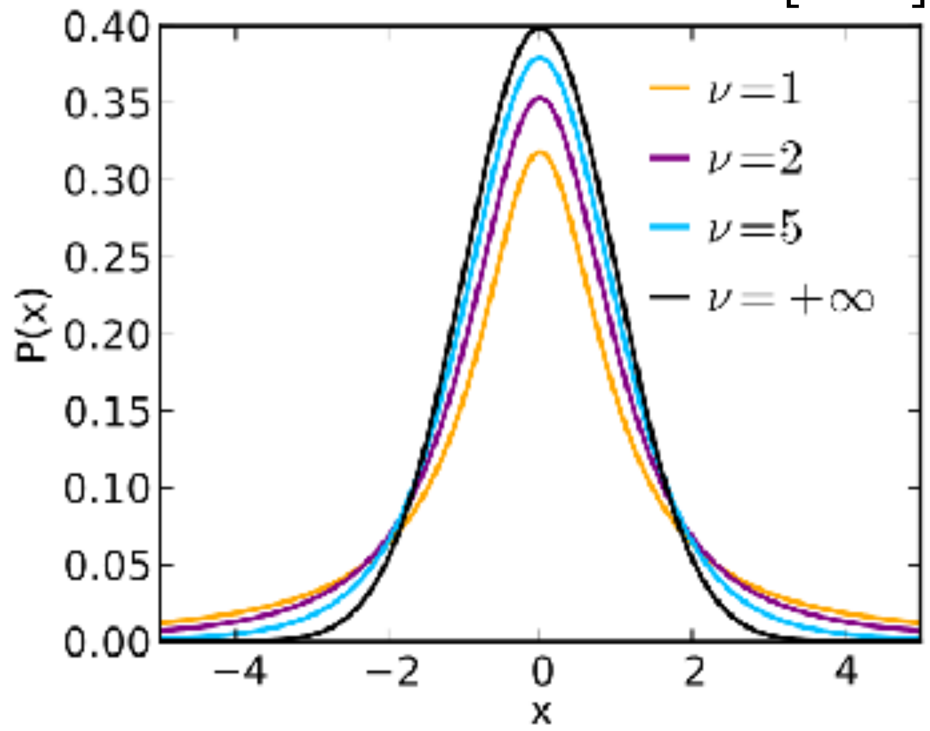
the random variable $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$

is gaussian distributed with mean 0 and variance 1.

the random variable $$\frac{\bar{X} - \mu}{S/\sqrt{n}}$$

has a student's distribution with n-1 degrees of freedom

[wiki]



Cauchy / Lorentz / Breit-Wigner

$$f(x; x_0, \gamma) = \frac{1}{\pi\gamma \left[1 + \left(\frac{x - x_0}{\gamma}\right)^2\right]} = \frac{1}{\pi\gamma} \left[\frac{\gamma^2}{(x - x_0)^2 + \gamma^2}\right]$$

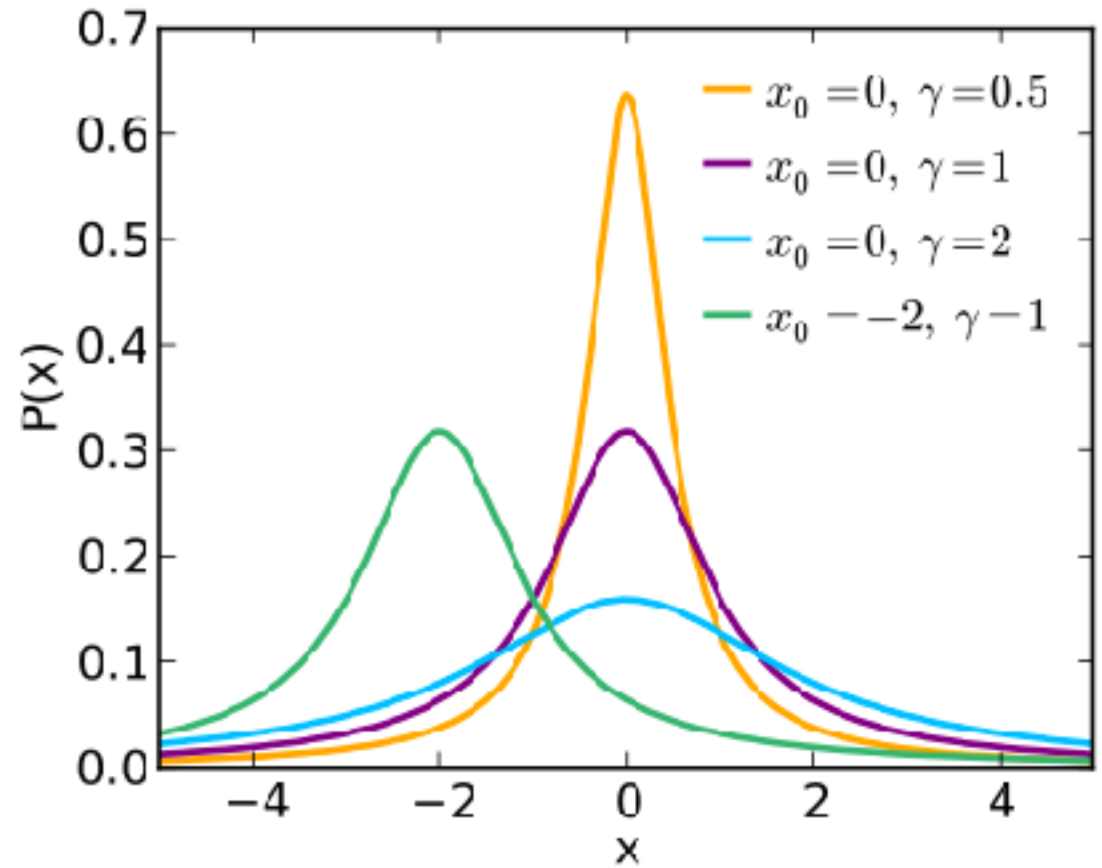
Or as standard Cauchy distribution:

$$f(x) = \frac{1}{\pi} \frac{1}{1 + x^2}.$$

Nor the mean nor the variance are defined !
(divergent integrals)

In HEP is usually written as (Breit-Wigner):

$$f(m; M, \Gamma) = \frac{1}{2\pi} \frac{\Gamma}{(m - M)^2 + (\Gamma/2)^2}$$



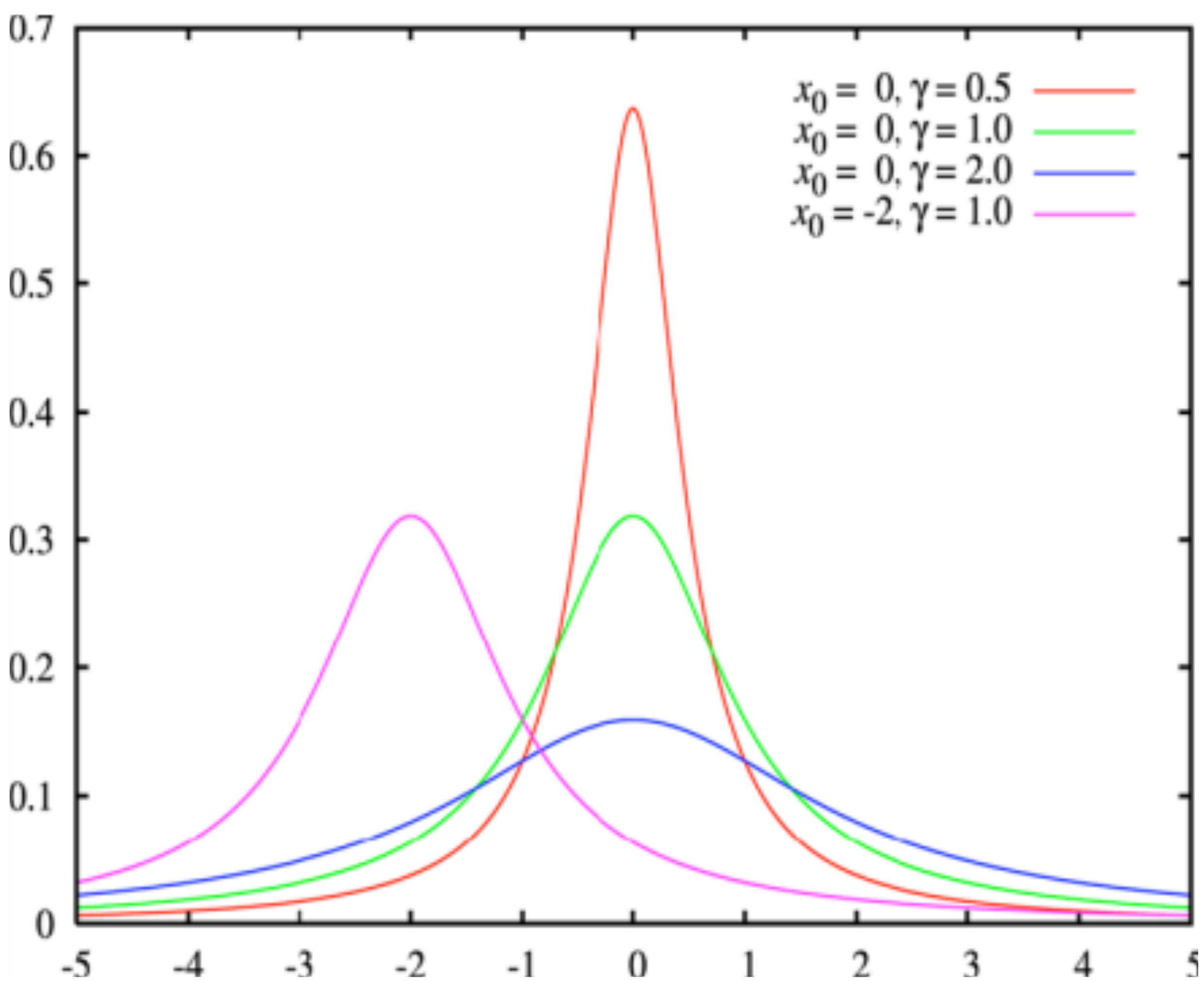
which is used to describe the cross section near a resonance of mass M and width Γ.
The Breit-Wigner is the Fourier transform of the wave function of an unstable particle:

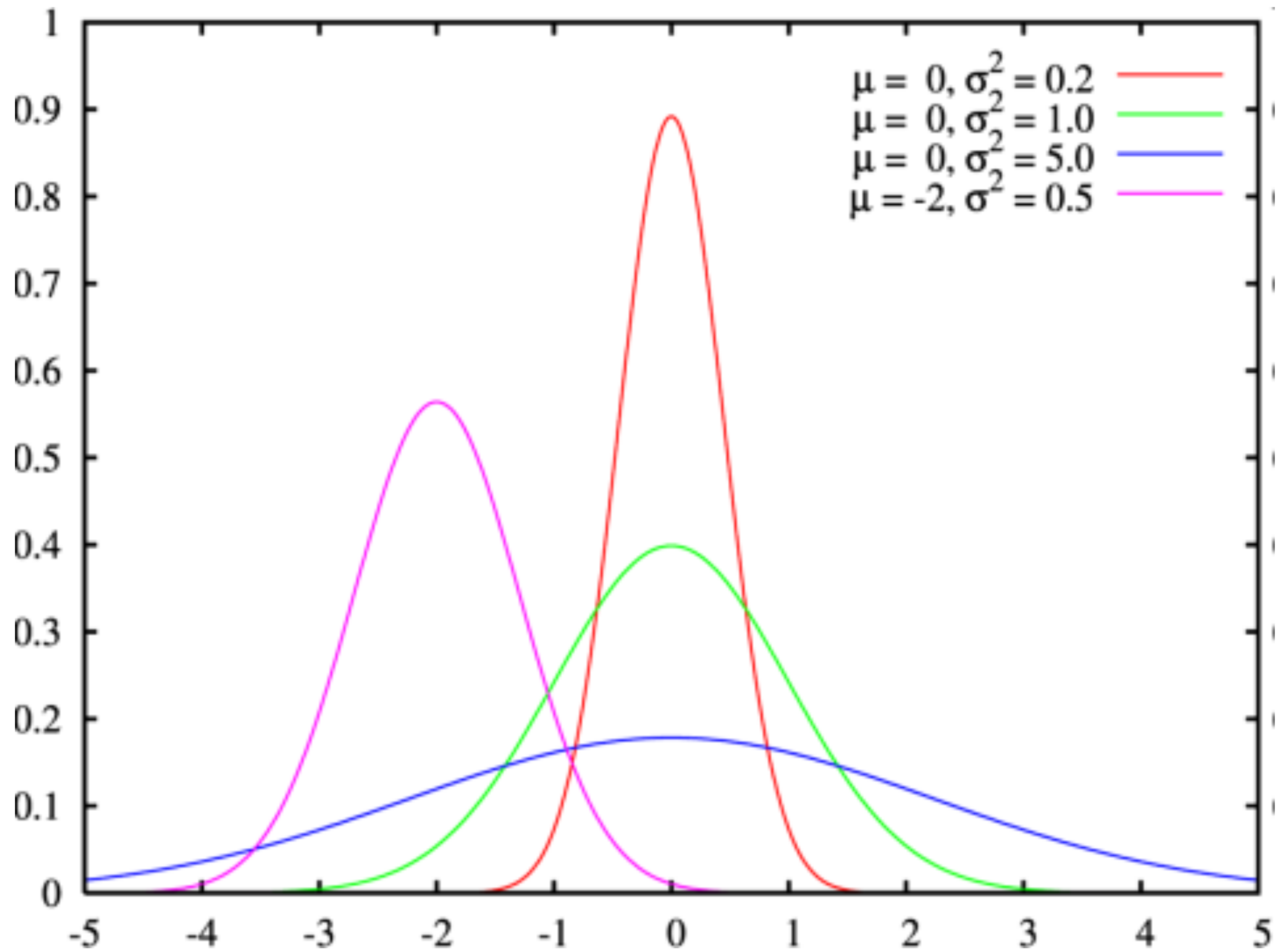
$$\psi(t) \propto e^{-iE_i t/\hbar} e^{-\Gamma t/2}$$

$$\phi(\omega) \propto \int_0^\infty \psi(t) e^{i\omega t} dt = \frac{i}{(\omega - \omega_0 + i\frac{\Gamma}{2})}$$ which squares gives: $$|\phi(\omega)|^2 = \frac{1}{(\omega - \omega_0)^2 + \frac{\Gamma^2}{4}}$$

Cauchy / Lorentz / Breit-Wigner

Compared to a gaussian:

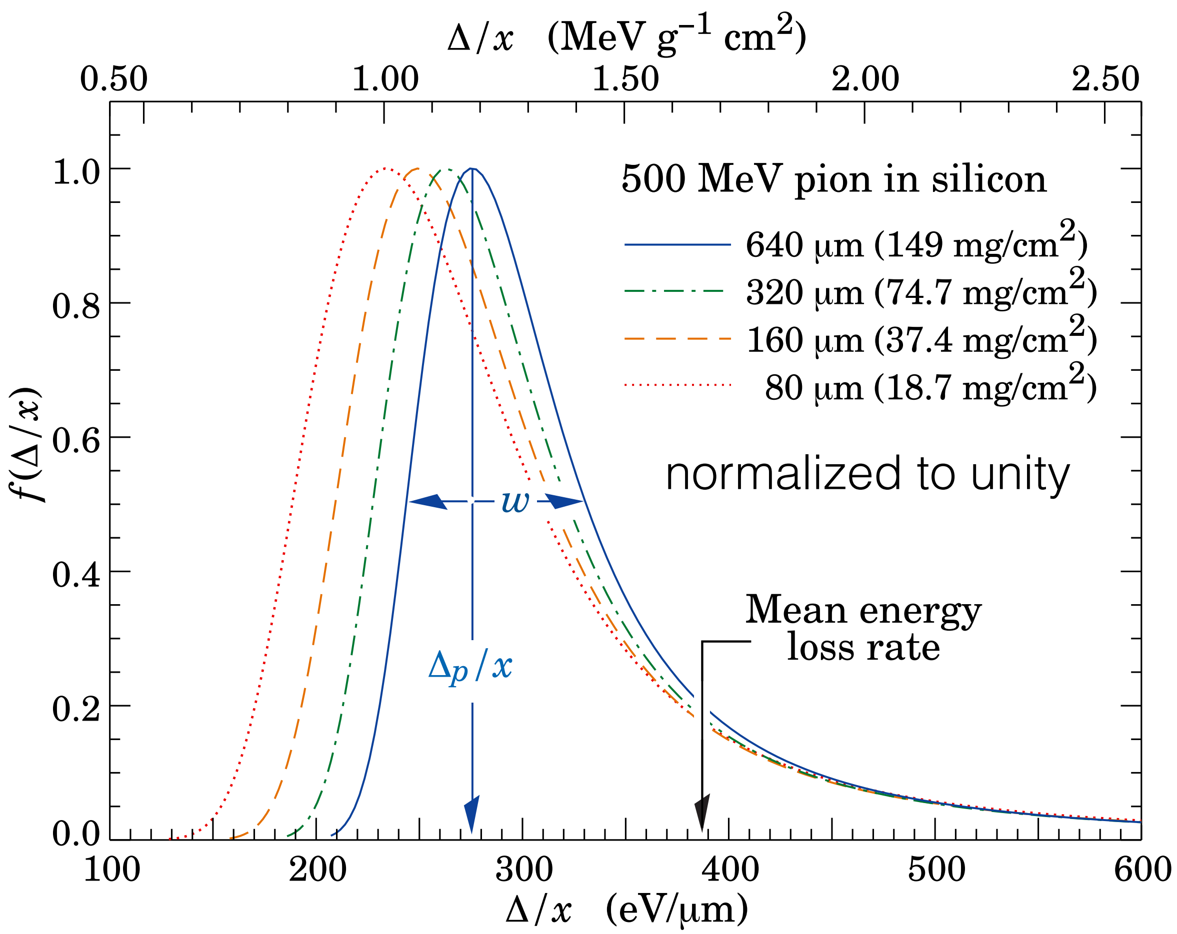




Landau distribution

[PDG: passage of particles through matter]

Used to describe the distribution of the energy loss of a charged particle (by ionisation) passing through a thin layer of matter.



$$p(x) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \exp(s \log s + xs) ds \qquad (c > 0)$$

Conveniently approximated by

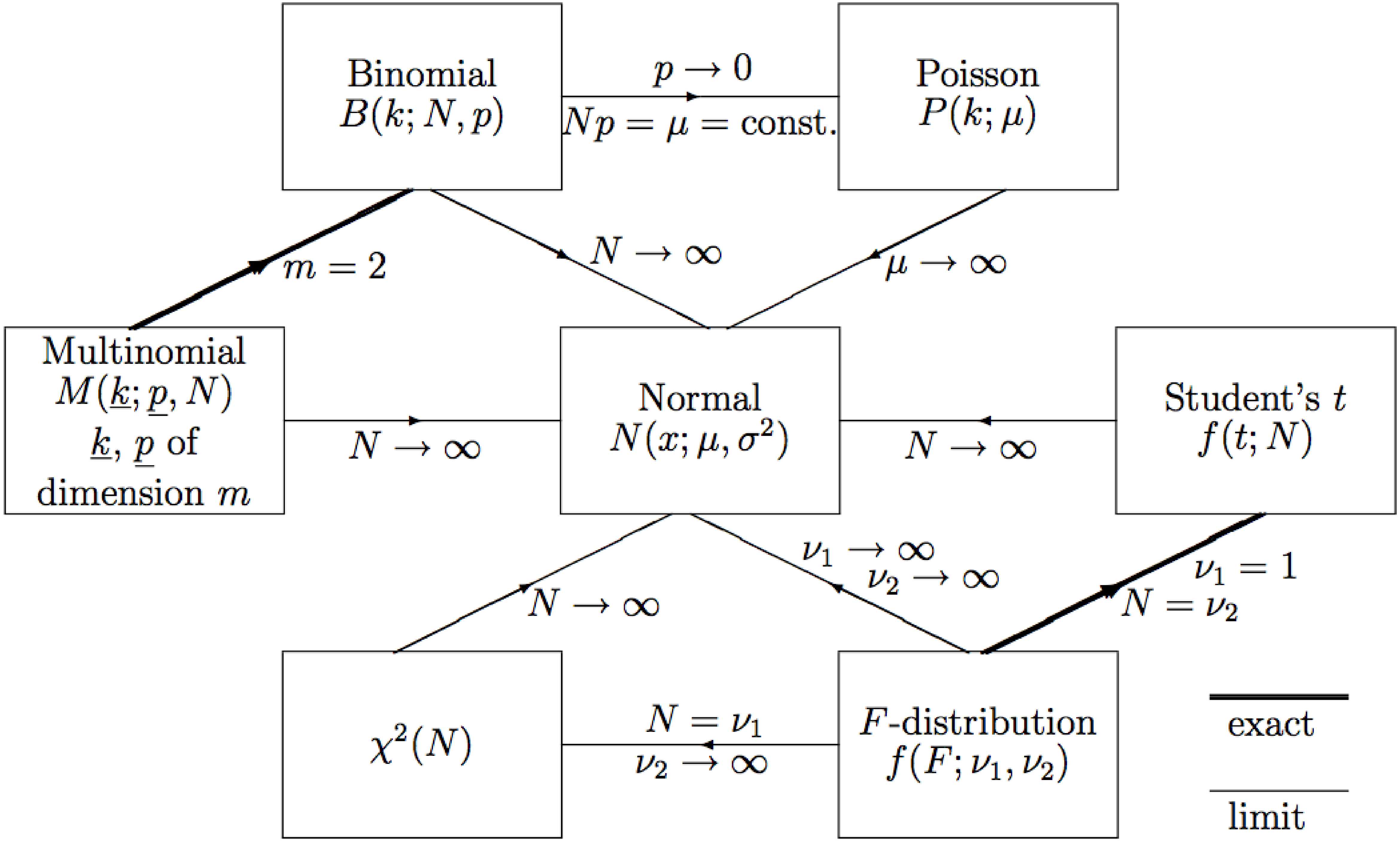
$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(x+e^{-x}\right)}$$

$x = E - E_{mp}$
E_{mp} = most probable energy deposited

The long tail towards large energies models the large energy loss fluctuations in thin layers.
The mean and the variance of the distribution are not defined.

Limits

[Metzger]



Bibliography

Covariance and correlation

Taylor: Chapter 9

Continuous pdfs:

any of: Taylor, Lyons, Cowan

Backup

F distribution

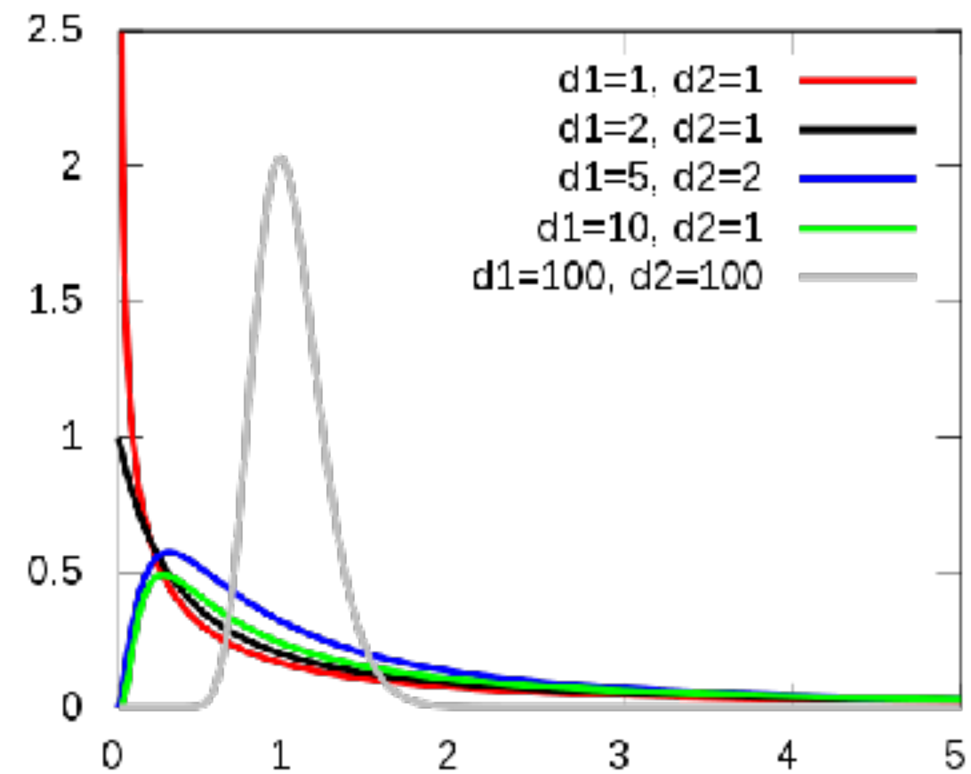
The F-distribution describes the ratio between two random variables χ^2_1 and χ^2_2, distributed as a χ^2 with ν_1 and ν_2 degrees of freedom.

$$F = \frac{\chi_1^2/\nu_1}{\chi_2^2/\nu_2}$$

The corresponding pdf is:

$$f(F) = \left(\frac{n_1}{n_2}\right)^{n_1/2} \cdot \frac{\Gamma((n_1+n_2)/2)}{\Gamma(n_1/2)\Gamma(n_2/2)} \cdot F^{(n_1-2)/2}\left(1+\frac{n_1}{n_2}F\right)^{-(n_1+n_2)/2}$$

[wiki]



We will encounter this distribution when discussing the hypothesis test for the variance of two samples.
F can be re-written as the ratio of two variances (where for convenience the largest one is at the numerator $F \geq 1$).

Random walk

Example: taken form Feynman Lectures on Physics Ch. 6 Vol 1

notebook https://gitlab.ethz.ch/mdonega/STAMET_FS18/blob/master/notebooks/randomWalk.ipynb