Statistical Methods in Particle Physics

Lecture 3

October 29, 2012

Silvia Masciocchi, GSI Darmstadt
s.masciocchi@gsi.de

Winter Semester 2012 / 13

Outline

- Reminder:
  - Probability density function
  - Cumulative distribution function
- Functions of random variables
- Expectation values
- Covariance, correlation

Examples of probability functions:

- Binomial
- Multinomial
- Poisson
- Uniform
- Exponential
- Gaussian
  - Central limit theorem

- Chi-square
- Cauchy
- Landau

Probability density function (pdf)

Suppose outcome of experiment is **continuous** value x:

$$P(x \text{ found in } [x, x+dx]) = f(x)dx$$

→ **f(x) = probability density function (pdf)**

With:

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

Normalization
(x must be somewhere)

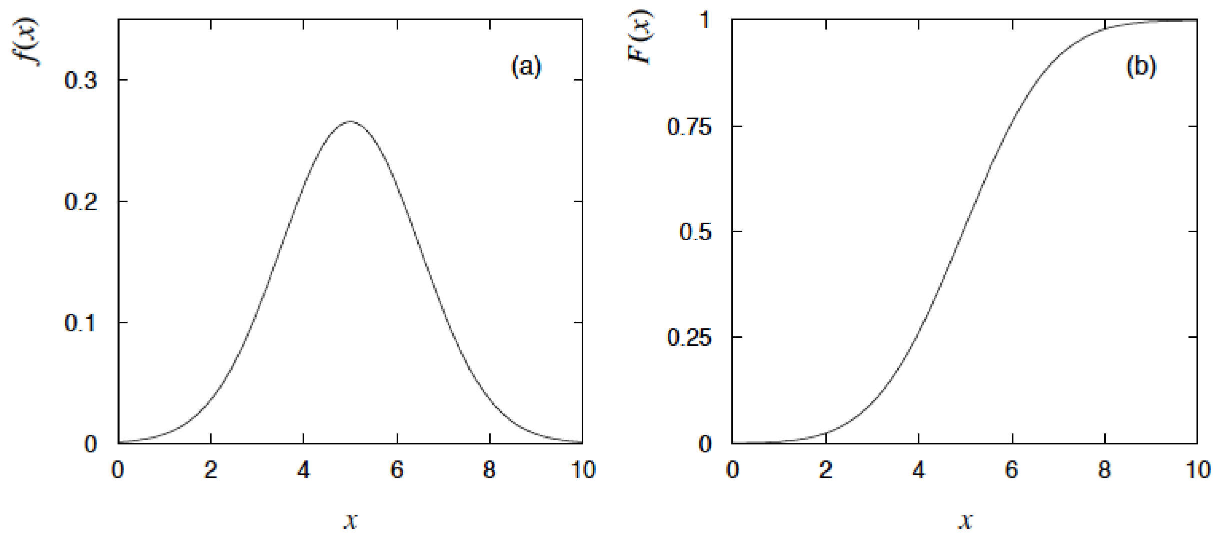
Note:

- $f(x) \geq 0$
- f(x) is NOT a probability ! It has dimension 1/x !

Cumulative distribution function (cdf)

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

Cumulative distribution function



- $F(x)$ is a continuously non-decreasing function
- $F(-\infty) = 0$, $F(\infty) = 1$
- For well behaved distributions:

$$\text{pdf}: \quad f(x) = \frac{\partial F(x)}{\partial x}$$

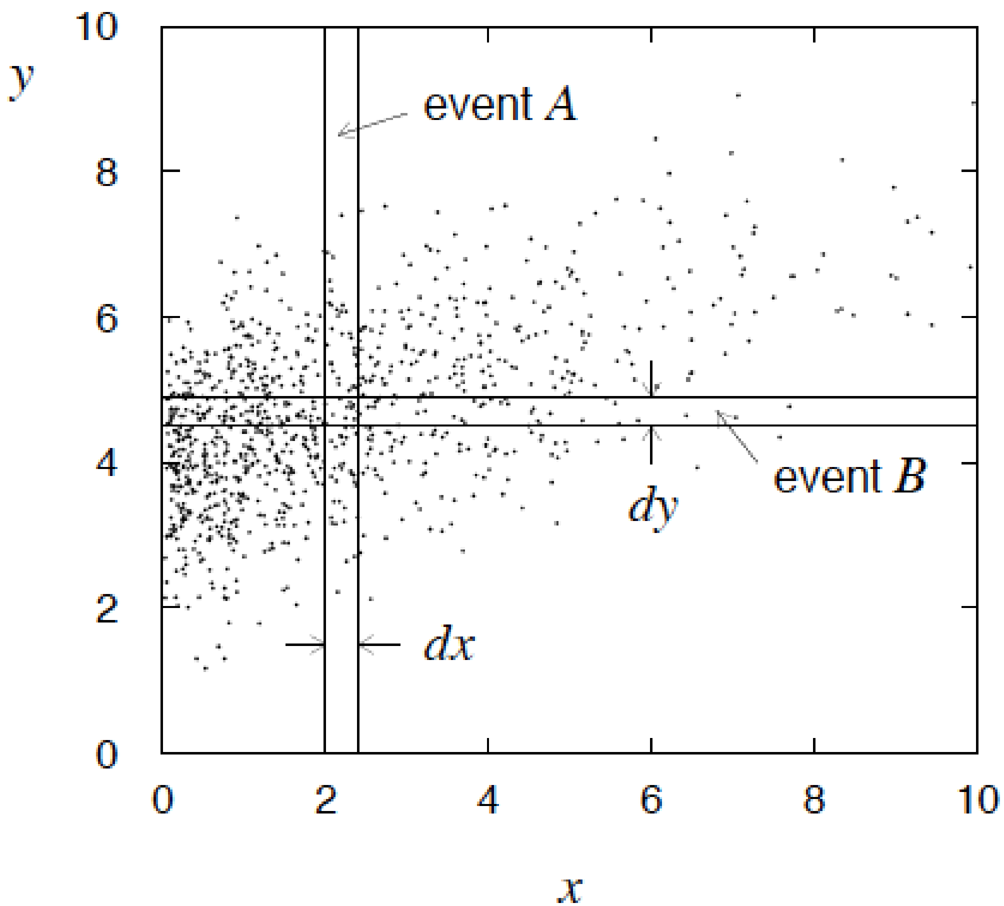
Multivariate distributions

$$f(x,y)$$

The outcome of the experiment is characterized by more than 1 quantity, e.g. by x and y

$$P(A \cap B) = f(x,y)\,dx\,dy$$

Joint pdf



Normalization:

$$\iint f(x,y)\,dx\,dy = 1$$

$$\iint \ldots \int f(x_1, x_2, \ldots x_n)\,dx_1\,dx_2 \ldots dx_n = 1$$

Marginal pdf's

From a multivariate distribution

$$f(x,y)\,dx\,dy$$

(e.g. scatter plot)

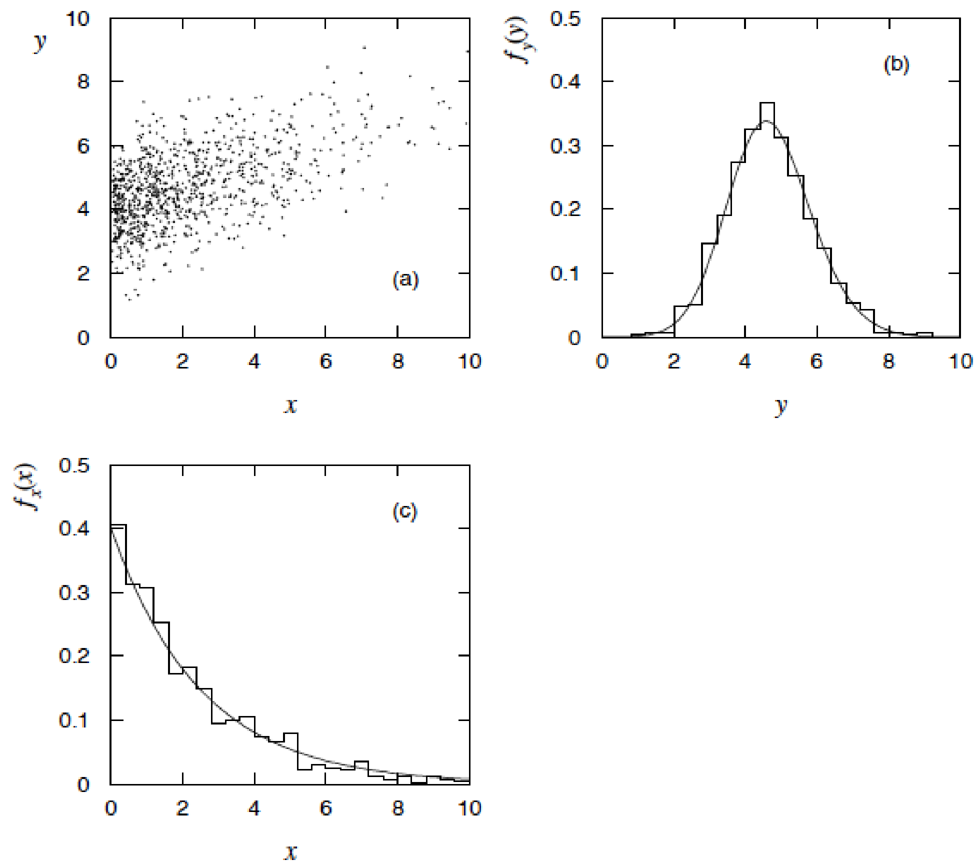
we might be in interested only in the pdf of ONE of the components (x or y, here)

→ projection of joint pdf onto individual axes

Marginal pdf

$$f_x(x) = \int f(x,y)\,dy$$

$$f_y(y) = \int f(x,y)\,dx$$



Distribution of a single variable which is part of a multivariate distribution

Functions of a random variable

A function of a random variable is itself a random variable.

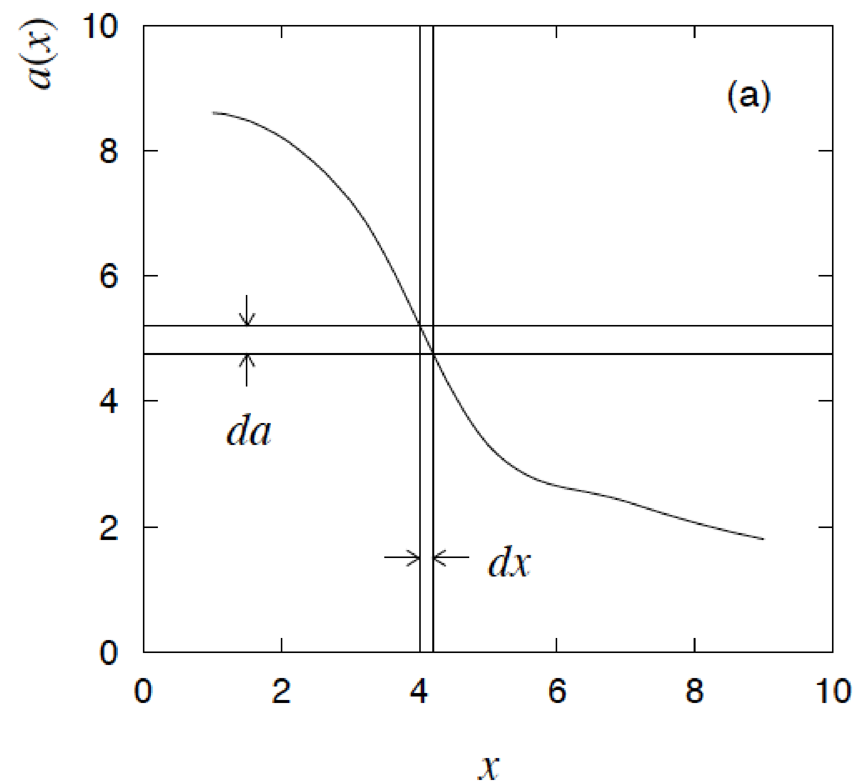
Suppose x follows a pdf f(x), consider a function a(x).

What is the pdf g(a)?

$$g(a)\,da = \int_{dS} f(x)\,dx$$

dS = region of x space for which a is in [a, a+da].

For one-variable case with unique inverse this is simply:



$$g(a)\,da = \left| \int_{x(a)}^{x(a+da)} f(x')\,dx' \right| = \int_{x(a)}^{x(a)+\left|\frac{dx}{da}\right|da} f(x')\,dx'$$

$$g(a) = f(x(a)) \left| \frac{dx}{da} \right|$$

Expectation value

Consider a continuous random variable x with pdf f (x).

Define **expectation (mean) value** as

$$E[x] = \int x f(x) dx$$

E[x] is NOT a function of x, it is rather a parameter of f(x)

Notation (often):

$$E[x] = \mu$$

~ “centre of gravity” of pdf.

For a function y(x) with pdf g(y),

$$E[y] = \int y g(y) dy = \int y(x) f(x) dx \qquad \text{(equivalent)}$$

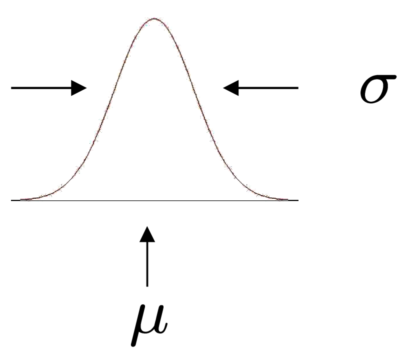
Variance and standard deviation

Variance:

$$V[x] = E[(x - \overbrace{E[x]}^{\mu})^2] = E[x^2] - \mu^2$$

Notation: $V[x] = \sigma^2$

Standard deviation: $\sigma = \sqrt{\sigma^2}$ Same dimension as x



Exercises

- Find the mean of the random variable X that has probability density function f given by:

  $f(x) = x^2 / 3$ for $-1<x<2$

- Suppose that X has the power distribution with parameter a > 1, which has density:

  $f(x) = (a - 1)x^{-a}$ for $x > 1$

  Show that:

  $$E[x] = \begin{cases} \infty, \text{if } 1 < a \leq 2 \\ \dfrac{a-1}{a-2}, \text{if } a > 2 \end{cases}$$

- Let the random variable x have the probability density function

  $$f(x) = \begin{cases} 3x^2, \text{if } 0 \leq x \leq 1 \\ 0, \text{elsewhere} \end{cases}$$

  Calculate its variance.

Covariance and correlation

Define **covariance cov[x,y]** (also use matrix notation V_{xy}) as:

$$\mathrm{cov}[x,y] = E[(x-\mu_x)(y-\mu_y)]$$

Can be written as:

$$\mathrm{cov}[x,y] = E[xy] - \mu_x \mu_y$$

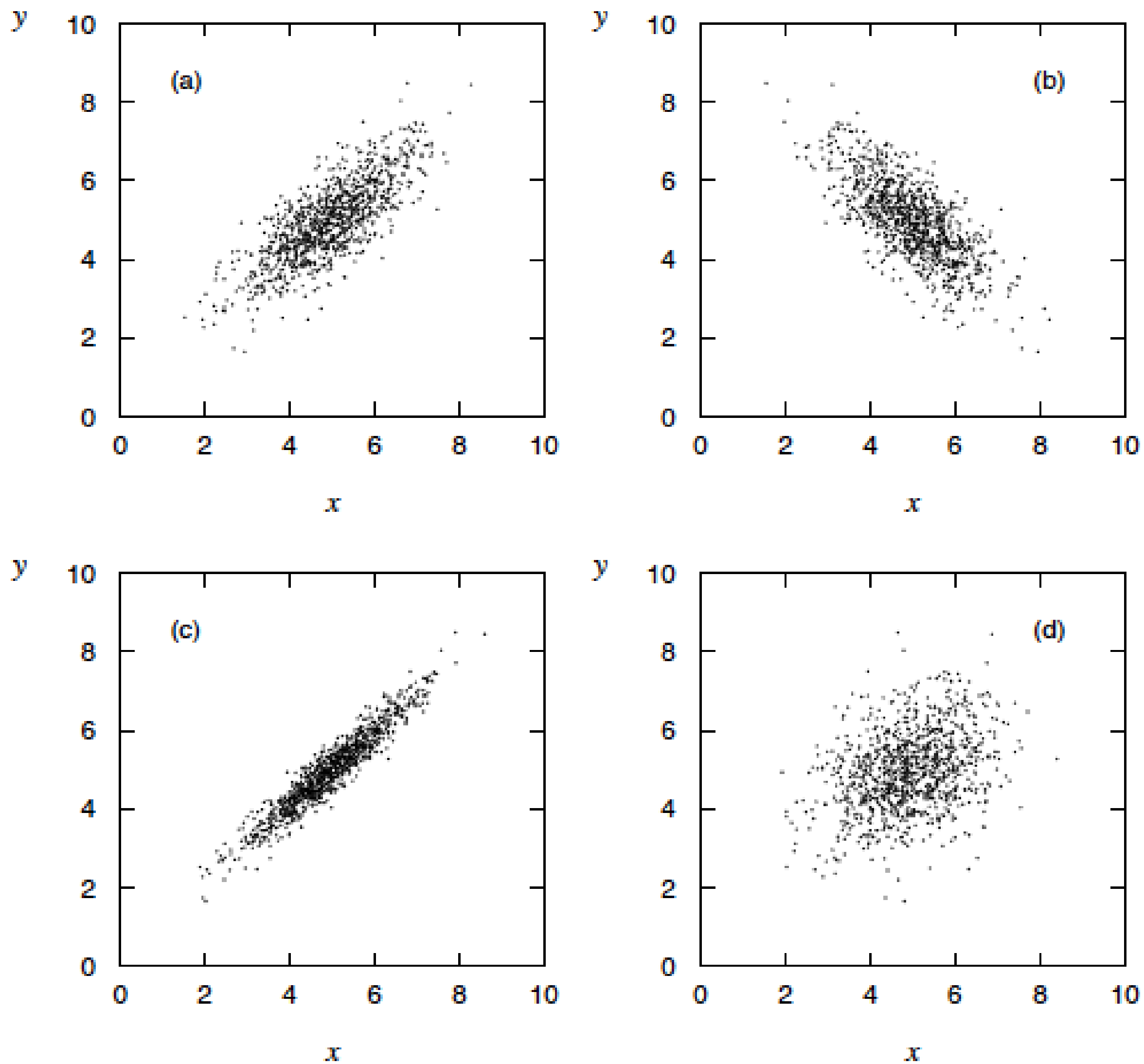
Correlation coefficient (dimensionless) defined as:

$$\rho_{xy} = \frac{\mathrm{cov}[x,y]}{\sigma_x \sigma_y}, \qquad -1 \le \rho_{xy} \le +1$$

Correlation coefficient

$\rho = 0.75$

$\rho = -0.75$



$\rho = 0.95$

$\rho = 0.25$

Independent variables

If x and y are independent, i.e. $f(x,y) = f_x(x)\, f_y(y)$, then:

$$E[xy] = \iint xy\, f(x,y)\, dx\, dy = \mu_x \mu_y$$

Therefore: $$\mathrm{cov}[x,y] = 0$$

x and y are 'uncorrelated'

Note!! The converse is NOT always true!!!

Exercise

Let [x y] be an absolutely continuous random vector with domain:

$$R_{XY} = \{(x,y): 0 \le x \le y \le 2\}$$

i.e. R_{XY} is the set of all couples (x,y) such that $0 \le y \le 2$ and $0 \le x \le y$.

Let the joint probability density function of [x y] be:

$$f(x,y) = \begin{cases} \frac{3}{8} y, & \text{if } (x,y) \in R_{XY} \\ 0 & \text{, otherwise} \end{cases}$$

Compute the covariance between X and Y.

The binomial distribution – the coin

Coin is tossed in the air → 50% heads up, 50% tails up

Meaning: if we continue tossing a coin repeatedly, the fraction of times that it lans heads up will asymptotically approach 1/2.

For each given toss, the probability cannot determine whether or not it will land heads up; it can only describe how we should expect a large number of tosses to be divided into two possibilities.

The binomial distribution – 2 coins

Suppose we toss two coins at the time

→ There are **4 different possible PERMUTATIONS** of the way in which they can land:

- Both heads up
- Both tails up
- 2 mixtures of heads and tails depending on which one is heads up

Each permutation is equally probable → the probability for any choice of them is 25% !

For the probability for the mixture of heads and tails, without differentiating between the two kinds of mixtures, we add two cases → 50%

THE SUM OF THE PROBABILITIES FOR ALL POSSIBILITIES IS ALWAYS EQUAL TO 1, BECAUSE *SOMETHING* IS BOUND TO HAPPEN

The binomial distribution – n coins

Let us extrapolate to the general case: we toss n coins in the air

(or we toss one coin n times)

P(x;n) = probability that exactly x of these coin will land heads up, without distinguishing which of the coins actually belongs to which group

X must be an integer for any physical experiment, but we can consider the probability to be smoothly varying with x as a continuous variable for mathematical purposes

The binomial distribution: Permutations and combinations

If n coins are tossed:

- 2^n different possible ways in which they can land (each coin has two possible orientations)
- Each of these possibilities is equally probable → the probability for any of these possibilities is $1/2^n$

How many of these possibilities will contribute to our observation of x coins with heads up?

Box 1 (x): heads up　　　　Box 2 (n-x): tails up

1. How many ***permutations*** of the coins result in the proper separation of the x in one box, and n-x in the other?

   Total number of choices for coins to fill the x slots in the heads box is

   $Pm(n,x) = n(n-1)(n-2) \ldots (n-x+2)(n-x+1)$

   more easily written as

$$Pm(n,x) = \frac{n!}{(n-x)!}$$

The binomial distribution: Permutations and combinations

But we care only on which is heads up or tails up, not which landed first!

We must consider contributions different only if there are different coins in the two boxes, nor if the x coins within the heads box are permuted into different time orderings!

2. the number of different **combinations** C(n,x) of the permutations results from combining the x! different ways in which x coins in the heads box can be permuted within the box.

x! → degeneracy factor of the permutations

$$C(n,x) = \frac{Pm(n,x)}{x!} = \frac{n!}{x!(n-x)!} = \binom{n}{x}$$

This is the number of different possible combinations of n items, taken x at a time

The binomial distribution - Probability

With coins: p(heads up) = p (tails up)

Then: $P(x,n)$ = all combinations x probability of each combination

$= C(n,x) \times 1/2^n$

GENERAL CASE:

n independent experiments (Bernoulli trials)

Bernoulli trial is a random experiment in which there are only two possible outcomes:

success and **failure**

p = probability of success of any given trial

$q = (1-p)$ = probability of failure

Therefore the probability of x times success (heads up) and n-x failures (tails up) is $p^x q^{n-x}$

The binomial distribution

From the definition of p and q, the probability $P_B(x;n,p)$ for observing x of the n items to be in the state with probability p is given by the **binomial distribution**

$$\boxed{P_B(x;n,p) = \binom{n}{x} p^x q^{n-x} = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}}$$

Name: the coefficients $P_B(x;n,p)$ are closely related to the binomial theorem:

$$(p+q)^n = \sum_{x=0}^{n} \left[\binom{n}{x} p^x q^{n-x} \right]$$

The (j+1)th term (corresponding to x=j) of the expansion is equal to $P_B(j;n,p)$. This proves the **NORMALIZATION** of the binomial distr.

The binomial distribution – Mean and variance

The mean of the binomial distribution is:

$$\mu = \sum_{x=0}^{n} \left[x \frac{n!}{x!(n-x)!} p^{x} (1-p)^{n-x} \right] = np$$

If we perform an experiment with n items and observe the number x of successes, after a large number of repetitions the average $\vec{x}$ of the number of successes will approach a mean value μ given by the probability for success of each item (p) times the number of items (n)

The variance is:

$$\sigma^{2} = \sum_{x=0}^{n} \left[(x-\mu)^{2} \frac{n!}{x!(n-x)!} p^{x} (1-p)^{n-x} \right] = np(1-p)$$

The binomial distribution - Types

If the probability for a single success p is equal to the probability for failure $p = q = 1/2$, then the distribution is **symmetric** about the mean μ.

The median and the most probable value are both equal to the mean.

The variance σ^2 is equal to half the mean: $\sigma^2 = \mu/2$

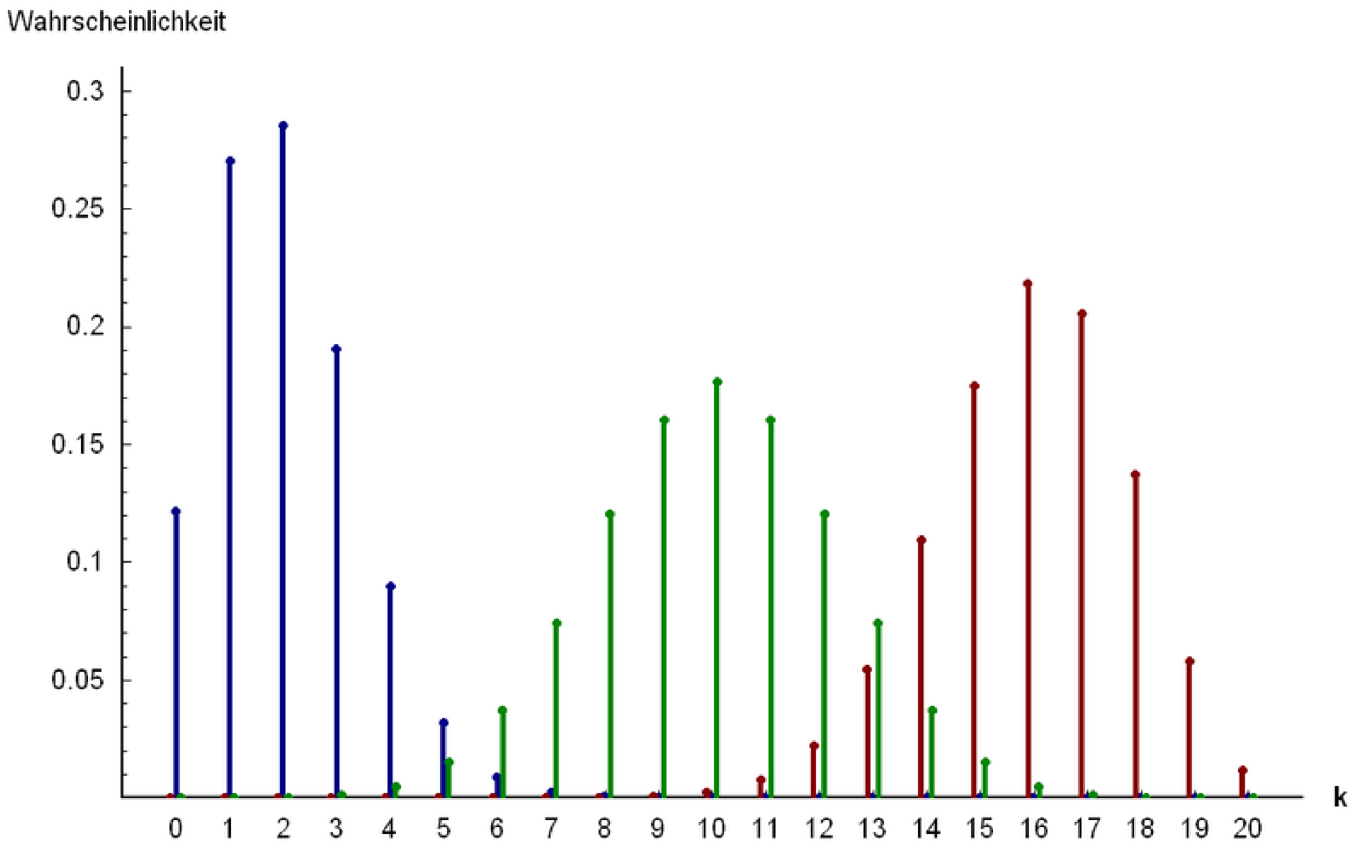
If p and q are not equal, the distribution is **asymmetric** with a smaller variance

n=20

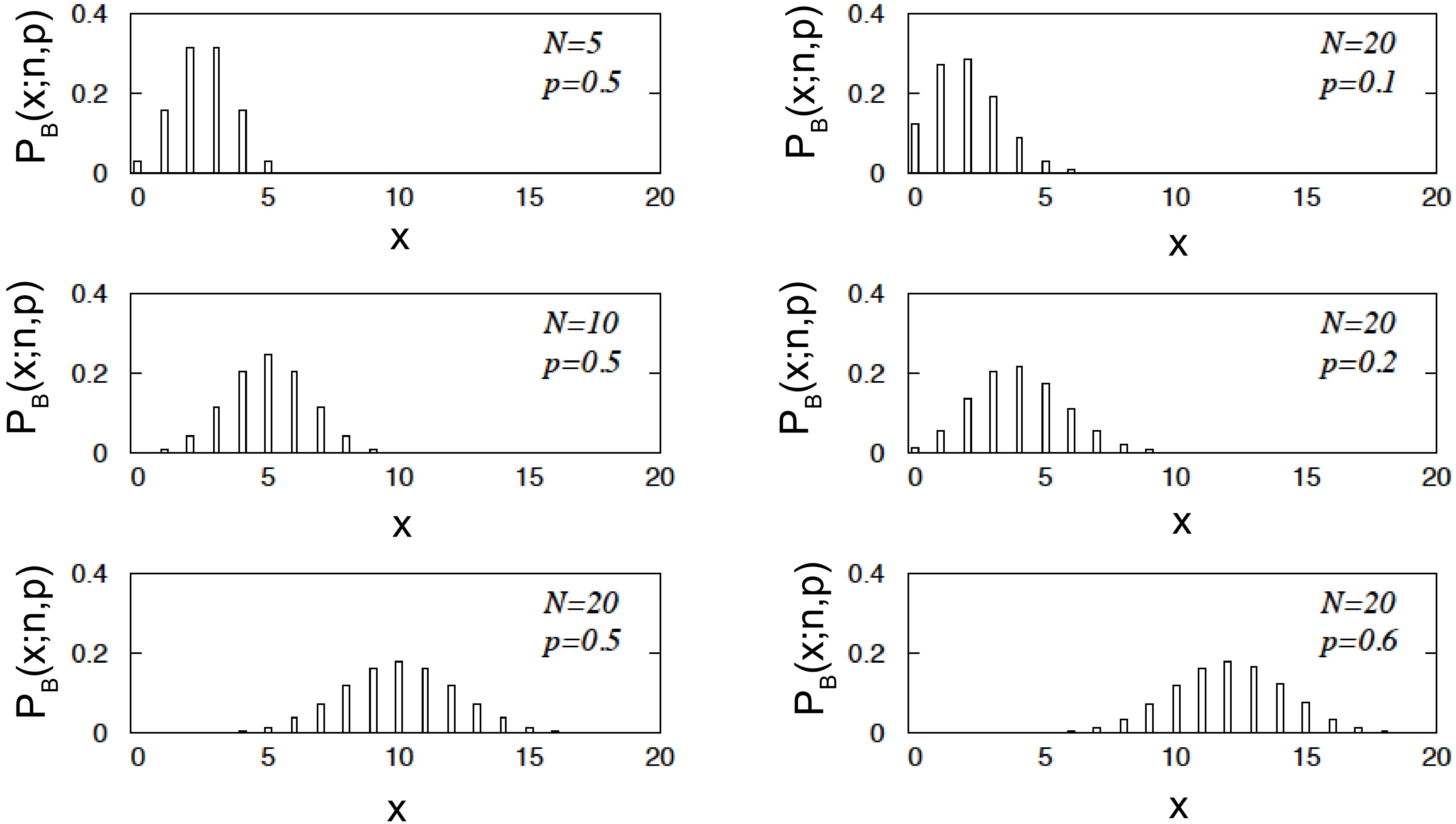
Blue: p=0.1

Green: p=0.5

Red: p=0.8



The binomial distribution - Examples



DECAYS: observe n decays of $W^{\pm}$, the number x of which are $W \to \mu\nu$ is a binomial random variable. p = **branching ratio**

Exercises

A test consists of 10 multiple choice questions with five choices for each question. As an experiment, you GUESS on each and every answer without even reading the questions.

What is the probability of getting **exactly 6** questions correct on this test?

Bits are sent over a communication channel in packets of 12. If the probability of a bit being corrupted over this channel is 0.1 and such errors are independent, what is the probability that no more than 2 bits in a packet are corrupted?

If 6 packets are sent over the channel, what is the probability that at least one packet will contain 3 or more corrupted bits?

Let X denote the number of packets containing 3 or more corrupted bits. What is the probability that X will exceed its mean by more than 2 standard deviations?

Multinomial distribution

Like binomial but now m outcomes instead of two. Probabilities are:

$$\vec{p} = (p_{1,} \ldots , p_m) \qquad \text{with} \quad \sum_{i=1}^{m} p_i = 1$$

For n trials, we want the probability to obtain:

x_1 of outcome 1,

x_2 of outcome 2,

…

x_m of outcome m

This is the multinomial distributions for $\vec{x} = (x_{1,} x_{2,.} .. x_m)$

$$P_B(\vec{x}; n, \vec{p}) = \frac{n!}{x_1! x_2! \ldots x_m!} p_1^{x_1} p_2^{x_2} \ldots p_m^{x_m}$$

Multinomial distribution

Now consider the outcome i as “success”, all others as “failure”

→ all x_i individually binomial with parameters n, p_i

Then:

$$E[x_i] = np_i \qquad V[x_i] = np_i(1-p_i)$$

One can also find the covariance to be:

$$V_{ij} = np_i(\delta_{ij} - p_j)$$

EXAMPLE: $\vec{x} = (x_1, \ldots, x_m)$ represents a **histogram** with m bins, n total entries, all entries independent

Poisson distribution

Consider a binomial x in the limit:

$n \to \infty$

$p \to 0$

$E[x] = \mu = np \to \nu$

When the average number of successes is much smaller than the possible number: $\mu \ll n$ because $p \ll 1$

Approximation of the binomial distribution

$\to$ **POISSON DISTRIBUTION**

$$P_P(x;\nu) = \frac{\nu^x}{x!} e^{-\nu} \qquad (x \geq 0)$$

$$E[x] = \nu, \qquad V[x] = \nu$$

Often in these experiments, neither the number n of possible events nor the probability p for each is known. What may be known is the the average number of events μ expected in each time interval.

Poisson distribution

Properties:

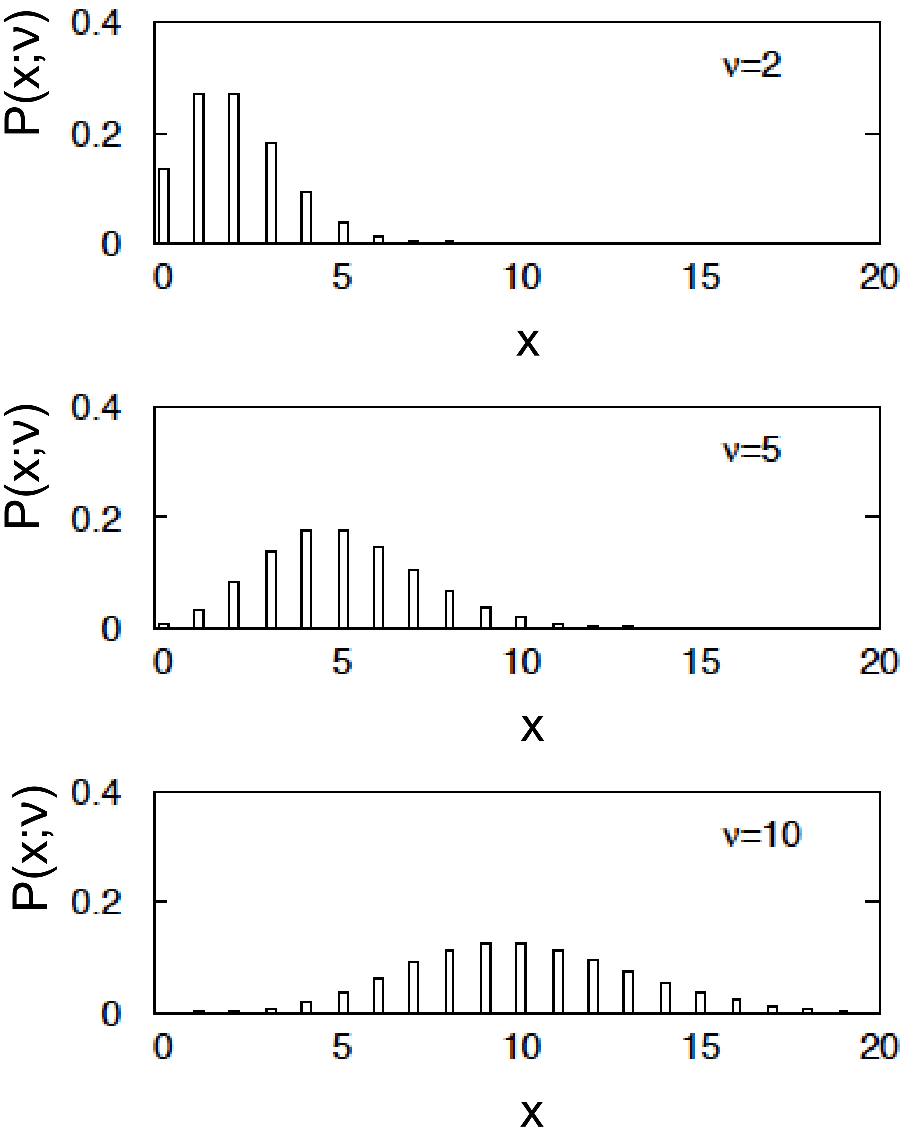
- Since this is an approximation to the binomial distribution, and p«1, the Poisson distribution is asymmetric about its mean
- It does not become 0 for x=0
- It is not defined for x<0

EXAMPLE:

NUMBER OF EVENTS FOUND

Given a cross section σ, and a fixed integrated luminosity, with

$$\nu = \sigma \int L\,dt$$



Exercise Poisson

The number of industrial injuries per working week in a particular factory is known to follow a Poisson distribution with mean 0.5

Find the probability that:

- In a particular week there will be
  - Less than 2 accidents
  - More than 2 accidents
- In a three week period there will be no accidents

The mean number of bacteria per millimetre of a liquid is known to be 4. Assuming that the number of bacteria follows a Poisson distribution, find the probability that, in 1ml of liquid, there will be

(a) no bacteria

(b) 4 bacteria

(c) less than 3 bacteria

Find the probability that

(i) in 3ml of liquid there will be less than 2 bacteria

(ii) in 0.5ml of liquid there will be more than 2 bacteria

Uniform distribution

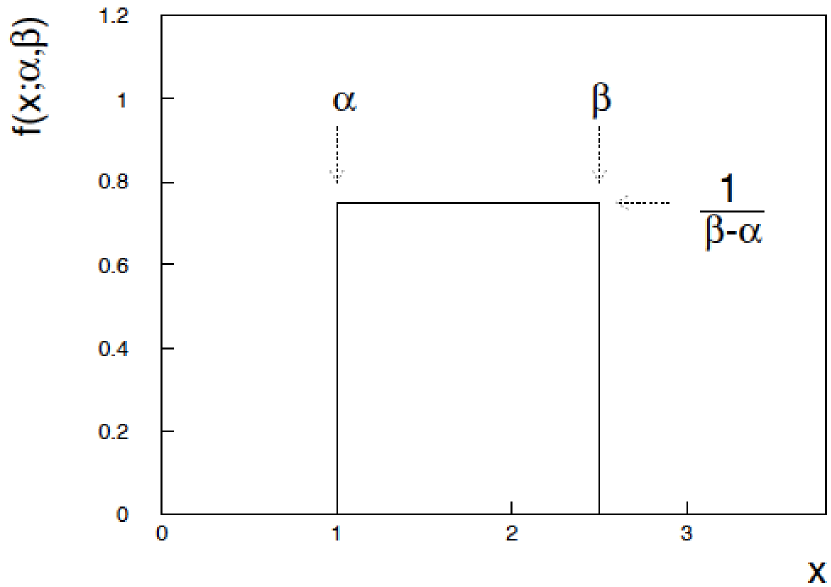
Consider a continuous random variable x, with -∞< x <∞

Uniform pdf is:

$$f(x;\alpha,\beta) = \begin{cases} \dfrac{1}{\beta-\alpha}, & \alpha \le x \le \beta \\ 0, & \text{otherwise} \end{cases}$$

$$E[x] = \frac{1}{2}(\alpha+\beta)$$

$$V[x] = \frac{1}{12}(\beta-\alpha)^2$$



EXAMPLE: for $\pi^0 \to \gamma\gamma$, $E\gamma$ is uniform is $[E_{min}, E_{max}]$, with

$$E_{min} = \frac{1}{2}E_\pi(1-\beta), \qquad E_{max} = \frac{1}{2}E_\pi(1+\beta)$$

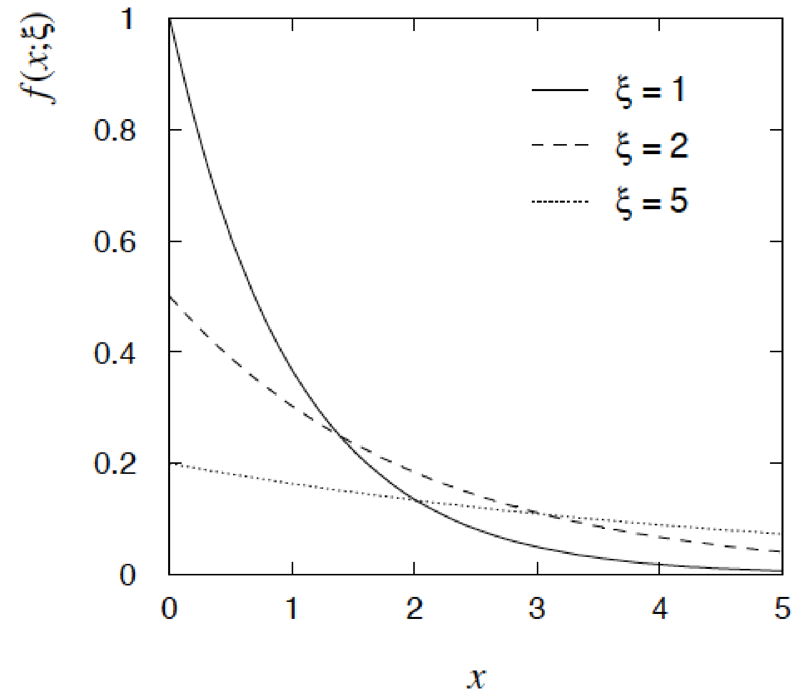
Exponential distribution

The exponential distribution for the continuous random variable x is defined by:

$$f(x;\xi) = \begin{cases} \frac{1}{\xi} e^{-x/\xi} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

$$E[x] = \xi$$

$$V[x] = \xi^2$$



EXAMPLE:

Proper decay time t of an unstable particle

$$f(t;r) = \frac{1}{\tau} e^{-t/\tau} \qquad (\tau = \text{mean lifetime})$$

Lack of memory: UNIQUE TO EXPONENTIAL

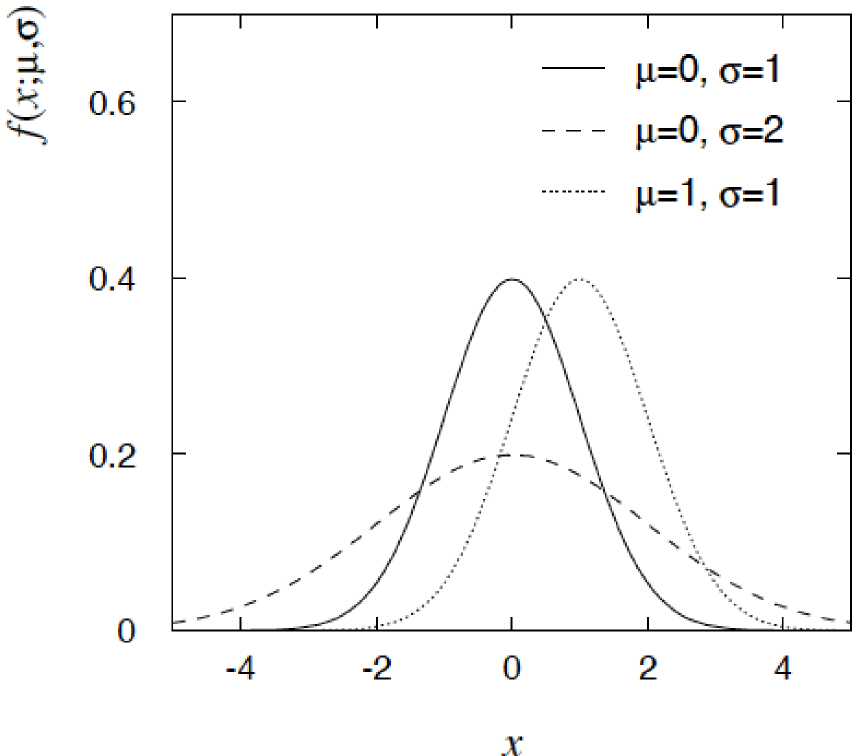
$$f(t - t_0 | t \geq t_0) = f(t)$$

Gaussian distribution

The Gaussian (normal) pdf for a continuous random variable is defined by:

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

$$E[x] = \mu$$
$$V[x] = \sigma^2$$



Special case: $\mu=0$, $\sigma^2=1$:

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \qquad \phi(x) = \int_{-\infty}^{x} \varphi(x')\, dx'$$

If y~Gaussian with μ, σ^2, then **$z=(y-\mu)/\sigma$** follows $\varphi(z)$

Central limit theorem

For n independent random variables x_i with finite variances σ_i^2, otherwise arbitrary pdf's, consider the sum:

$$y = \sum_{i=1}^{n} x_i$$

In the limit $n \to \infty$, y is a Gaussian random variable with:

$$E[y] = \sum_{i=1}^{n} \mu_i \qquad V[y] = \sum_{i=1}^{n} \sigma_i^2$$

Almost any random variable that is a sum of a large number of small contributions, follows a Gaussian distribution.

Measurement errors are often the sum of many contributions, so frequently measured values can be treated as Gaussian r.v.s

Central limit theorem - 2

For the proof, see book by Cowan e.g.

For finite n, the theorem is approximately valid to the extent that the fluctuation of the sum is not dominated by one (or few) terms

Beware of measurement errors with non-Gaussian tails !!!

Good example: velocity components v_x of air molecules

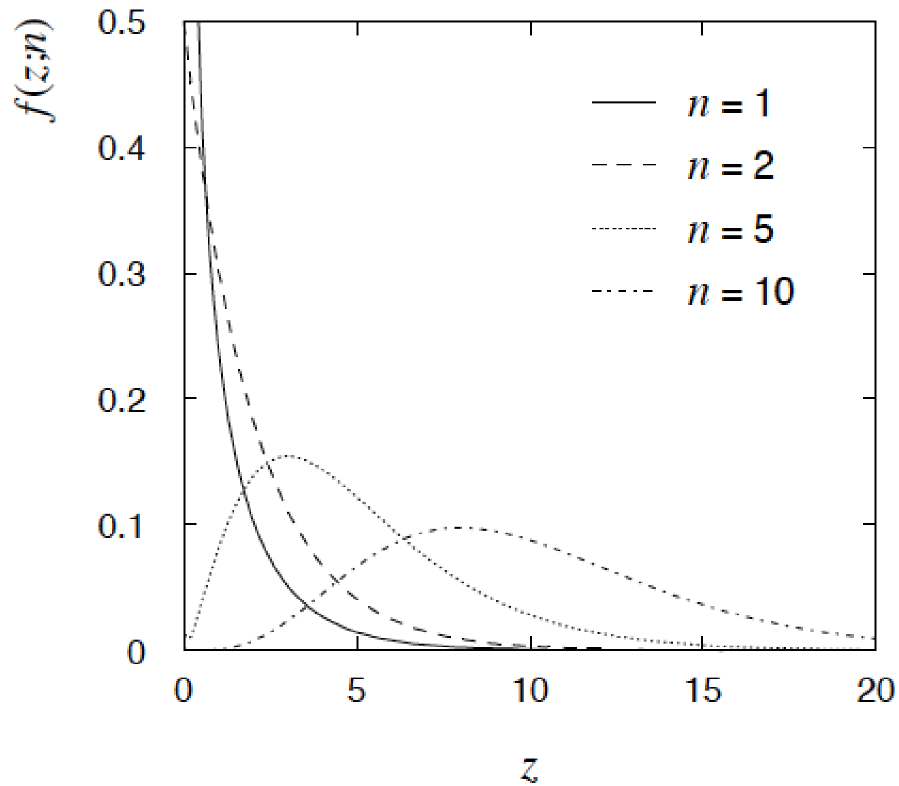
Medium good example: total deflection due to multiple Coulomb scattering (rare large angle deflections give non-Gaussian tail!!)

Bad example: energy loss of charged particle traversing thin gas layer (rare collisions make up large fraction of energy loss! → Landau pdf)

Chi-square (χ^2) distribution

The chi-square distribution for the continuous random variable z (z≥0) is defined by:

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



n = 1, 2, .. = number of degrees of freedom (dof)

E[z] = n, V[z] = 2n

EXAMPLE: goodness-of-fit test variable
especially in conjunction with method of least squares (soon!)

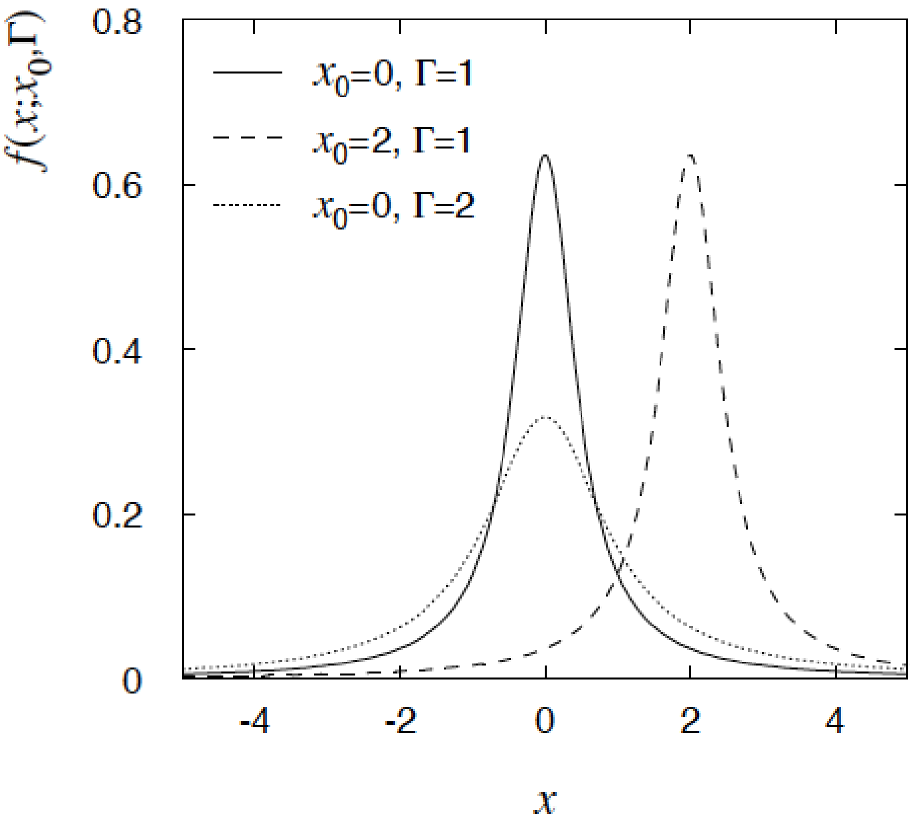
Cauchy (Breit-Wigner) distribution

The Cauchy pdf for the continuous random variable x is defined by

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

This is a special case of the Breit-Wigner pdf:

$$f(x;\Gamma,x_0) = \frac{1}{\pi}\frac{\Gamma/2}{\Gamma^2/4+(x-x_0)^2}$$



E[x] not well defined, V[x] $\rightarrow \infty$

x_0 = mode (most probable value)

Γ= full width at half maximum

EXAMPLE: mass of resonance particle, e.g. ρ, K*, Φ^0, …

Γ = decay rate (inverse of mean lifetime)

Landau distribution

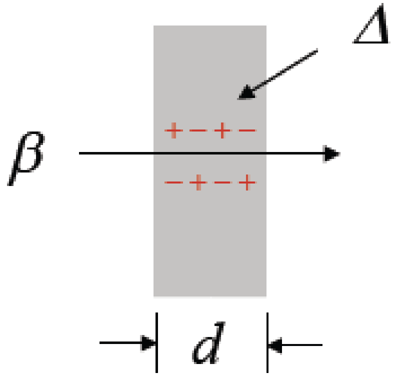
Describes the energy loss Δ of a charged particle with β=v/c traversing a layer of matter of thickness d

$$f(\Delta;\beta) = \frac{1}{\xi}\phi(\lambda)\ ,$$

$$\phi(\lambda) = \frac{1}{\pi}\int_0^{\infty} \exp(-u \ln u - \lambda u)\sin \pi u\, du\ ,$$

$$\lambda = \frac{1}{\xi}\left[\Delta - \xi\left(\ln\frac{\xi}{\epsilon'} + 1 - \gamma_{\mathrm{E}}\right)\right]\ ,$$

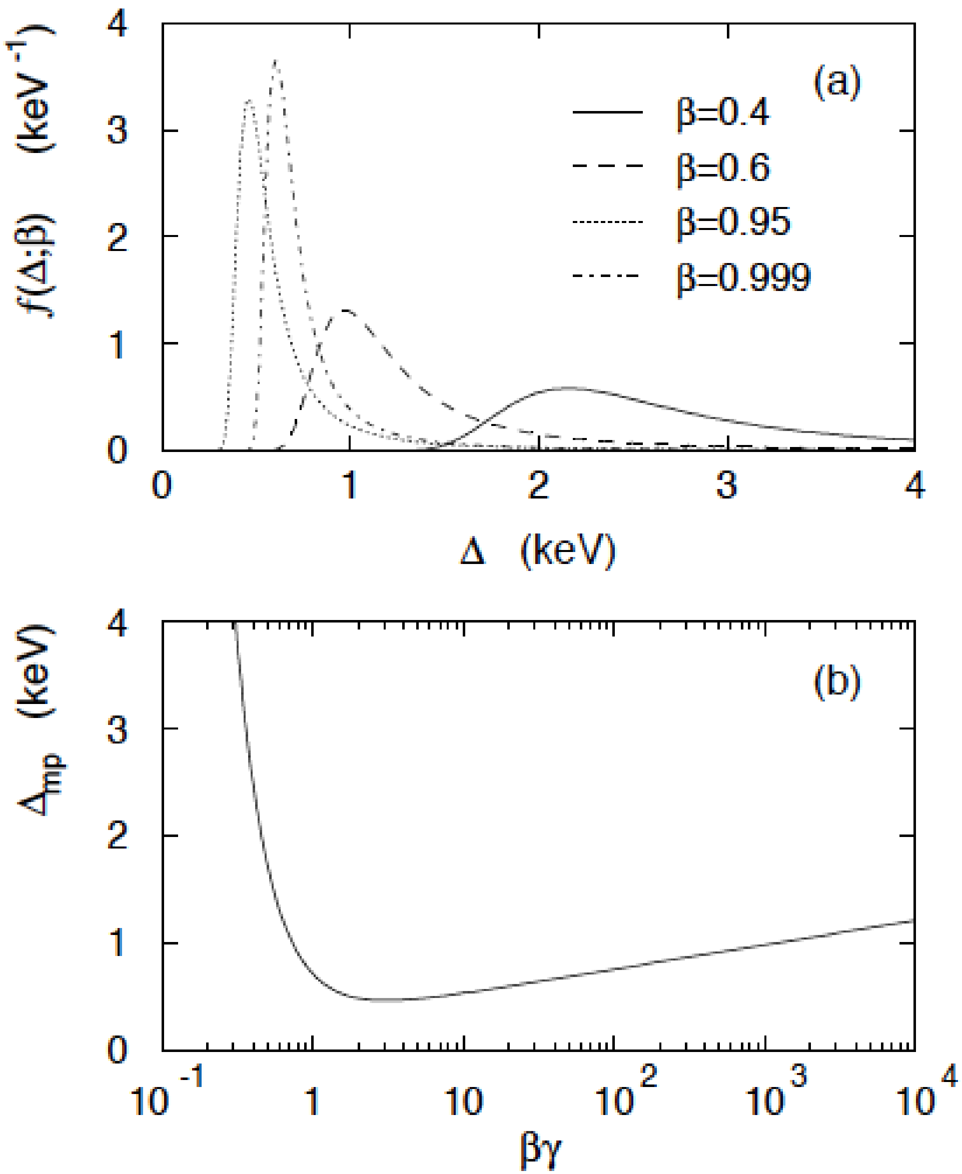
$$\xi = \frac{2\pi N_{\mathrm{A}} e^4 z^2 \rho \sum Z}{m_{\mathrm{e}} c^2 \sum A}\frac{d}{\beta^2}\ , \qquad \epsilon' = \frac{I^2 \exp\beta^2}{2 m_{\mathrm{e}} c^2 \beta^2 \gamma^2}\ .$$



L. Landau, J. Phys. USSR **8** (1944) 201; see also
W. Allison and J. Cobb, Ann. Rev. Nucl. Part. Sci. **30** (1980) 253.

Landau distribution - 2

Long Landau tail



Mode (most probable value) is sensitive to β

→ particle identification!

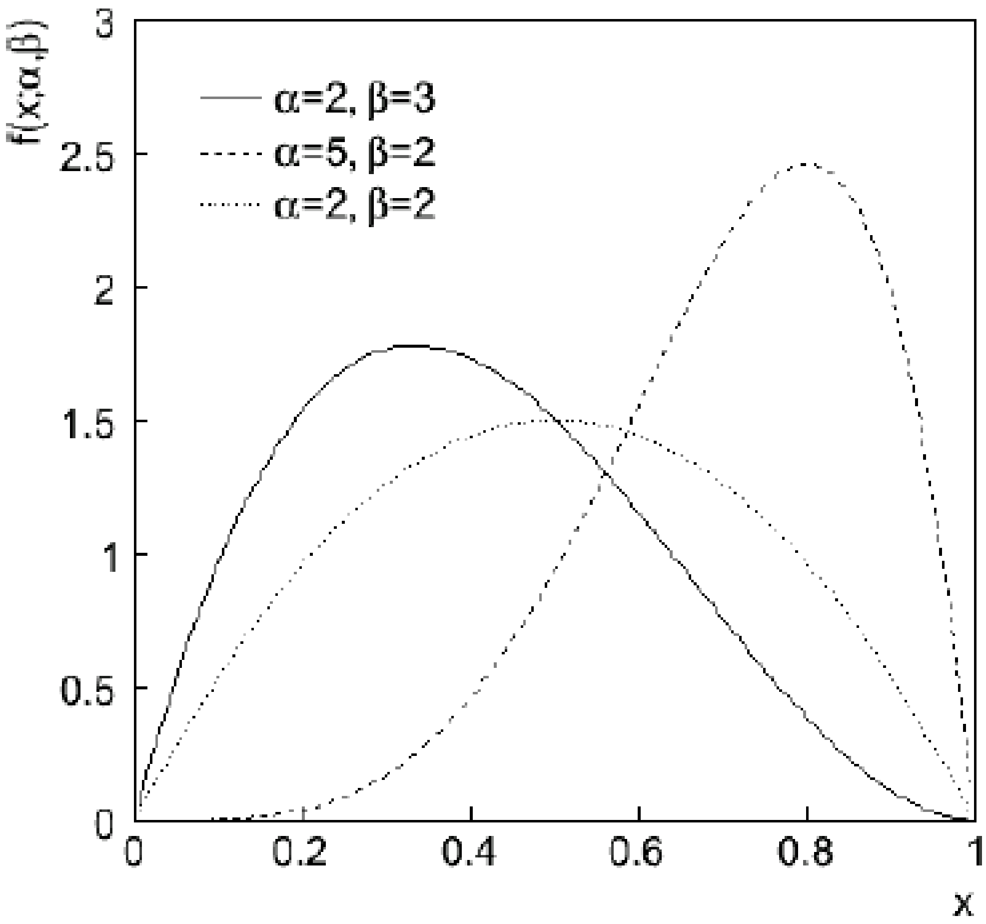
Beta distribution

You flipped 2 coins of type US $1 - George Washington:

$$f(x;\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)}\, x^{\alpha-1}(1-x)^{\beta-1}$$

$$E[x] = \frac{\alpha}{\alpha+\beta}$$

$$V[x] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$



Often used to represent pdf of continuous random variable non-zero only between finite limits

Gamma distribution

$$f(x;\alpha,\beta) = \frac{1}{\Gamma(\alpha)\,\beta^{\alpha}}\, x^{\alpha-1}\, e^{-x/\beta}$$

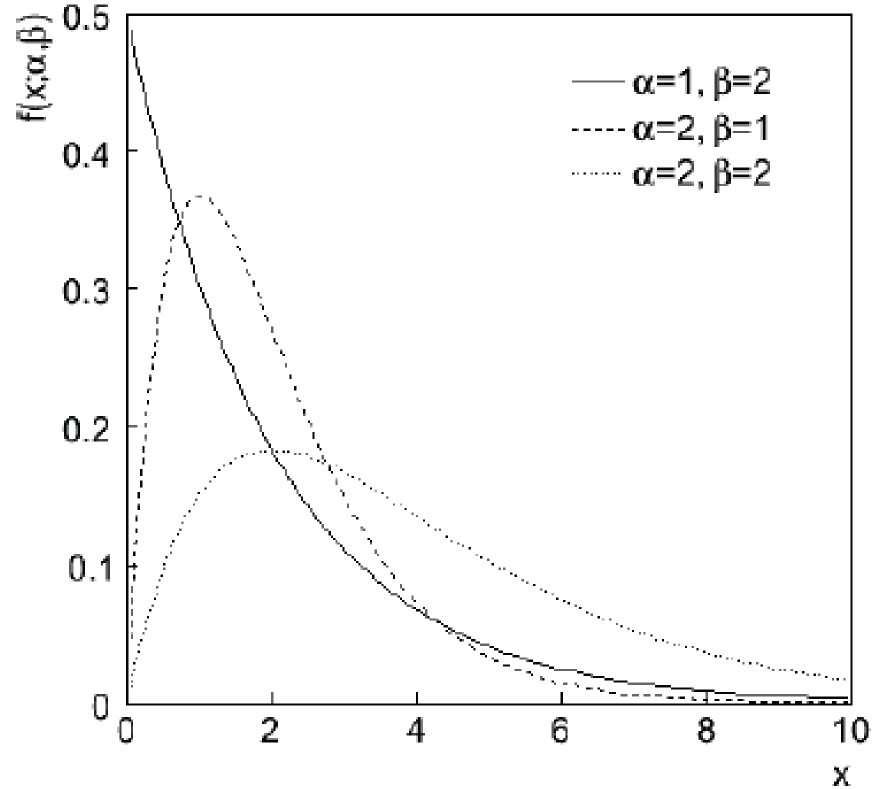
$$E[x] = \alpha\beta$$

$$V[x] = \alpha\beta^2$$

Often used to represent pdf of continuous random variable non-zero only in [0,∞]

Also similar to the gamma distribution:

- Sum of n exponential r.v.s
- Time until the nth event in Poisson process



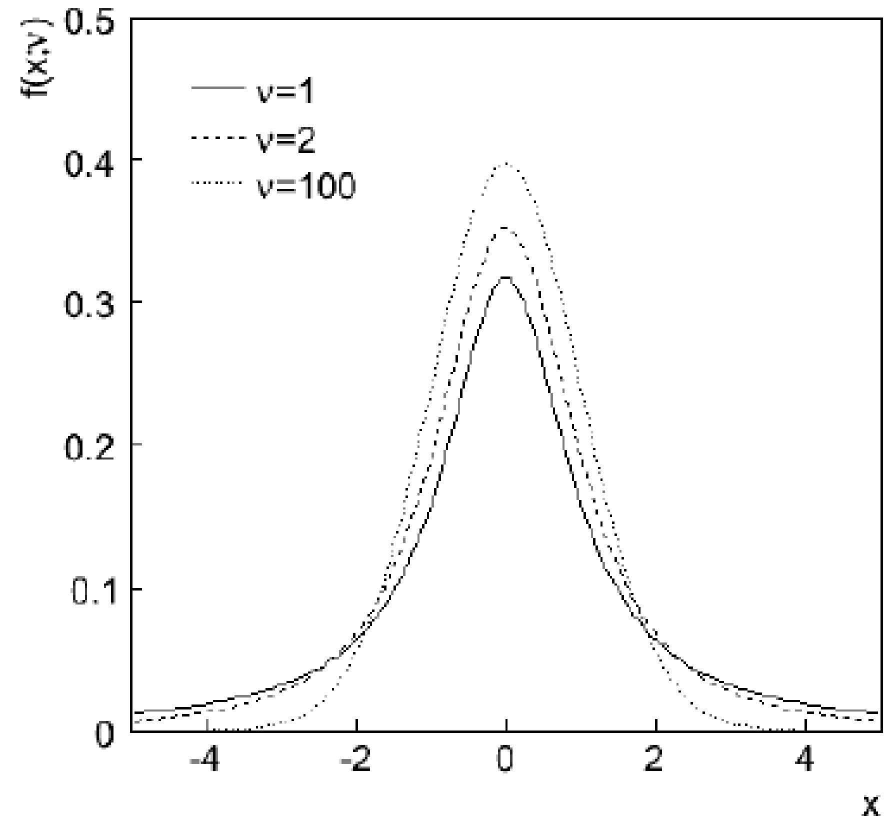
Student's t distribution

$$f(x;\nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\;\Gamma(\nu/2)}\left(1+\frac{x^2}{\nu}\right)^{(\frac{\nu+1}{2})}$$

$$E[x] = 0 \qquad (\nu > 1)$$

$$V[x] = \frac{\nu}{\nu-2} \qquad (\nu > 2)$$

- ν = number of degrees of freedom (not necessarily integer)
- ν = 1 gives Cauchy
- $\nu \rightarrow \infty$ gives Gaussian



Student's t distribution

The student's t provides a **bell-shaped pdf with adjustable tails**, ranging

- from those of a Gaussian, which fall off very quickly ($v \to \infty$, but in fact very Gauss-like for v = two dozen)
- To the very long-tailed Cauchy ($v = 1$)

Developed in 1908 by William Gosset, who worked under the pseudonym “Student” for the Guiness Brewery

Next

- Error propagation
- Monte Carlo methods
  - Transformation method
  - Integration
- Monte Carlo for particle / nuclear physics
  - Event generators
  - Detector simulation