Elements of probability calculus

Fundamental concepts

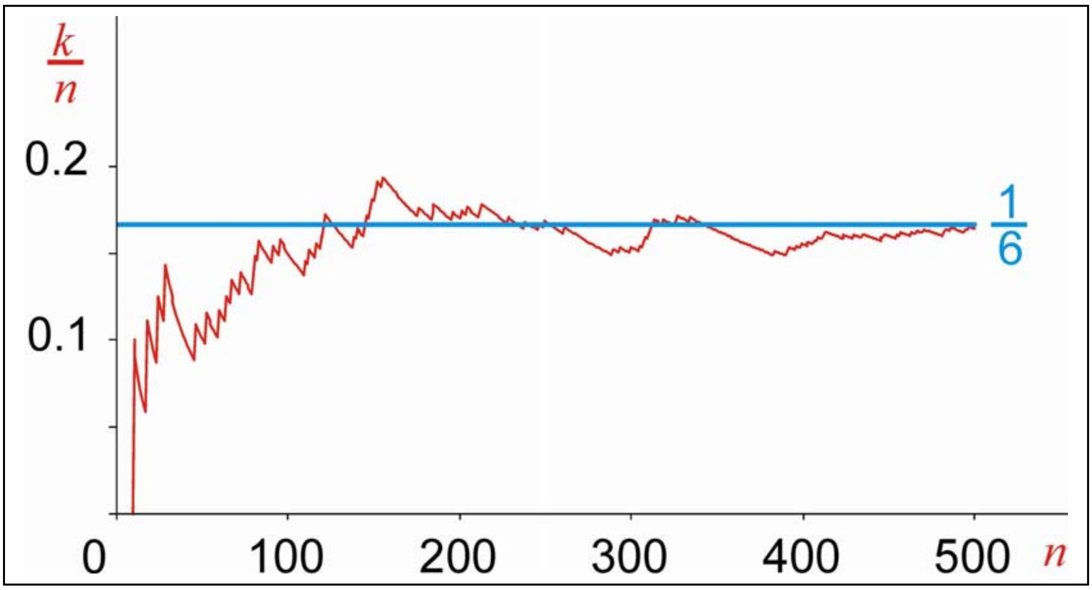
Phenomenon: all the things which are repeatable in essence at identical conditions, in connection with them we can do observations we can make "experiments".
E.g. medical examination | toss of a coin | waiting for a tram

Observation: we give what we are interested in connection with the phenomenon and how we can detect or measure it.
E.g. color of skin | falling time of the coin | how many passengers

Event: a statement which comes true or not.
E.g. yellow | between 0,5 s and 1,5 s | 10 passengers

Relative frequency of the event in the series of trials:
k/n, where k is the absolute frequency of the occurrence of the event and n is the number of experiments.

E.g. Phenomenon: a dice is rolled.
Observation: what is the outcome.
Event: the result is 6.



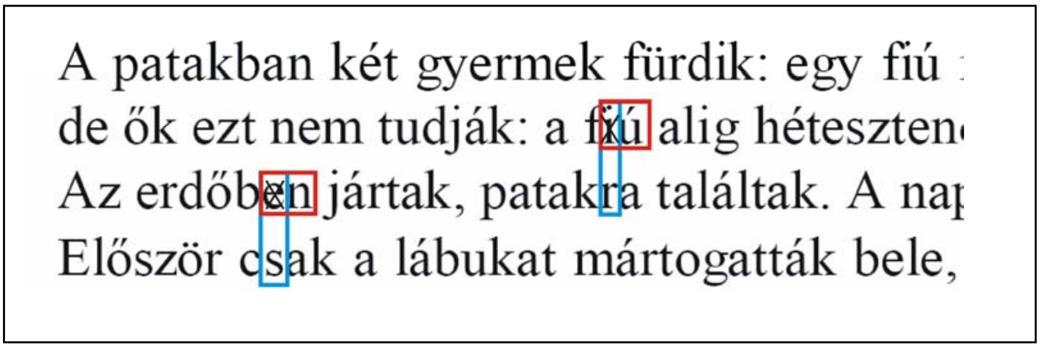
Law of large numbers (for the relative frequencies):
As the n (number of die rolls) increases, the relative frequency, k/n becomes stable around a certain value. This value is independent of the actual series of trials.
(It is an empirical fact it can not be proven by logical sequence.)
(Karl Pearson 1857-1936)
We assign a number to the event: **probability**.

Properties of probability:

1. The probability of an event [$P(A)$] is always: $0 \leq P(A) \leq 1$.
2. The probability of certain event is: $P(sure) = 1$,
   the probability of impossible event is: $P(impossible) = 0$.
3. The probability of the union of two mutually exclusive events (e.g. A and B) is: $P(A+B) = P(A) + P(B)$.

Joint observation

Example:
We choose a letter randomly from a Hungarian text and observe its neighbor to the right and the one stands underneath.
(We do not take into consideration the empty spaces.)


A patakban két gyermek fürdik: egy fiú
de ők ezt nem tudják: a fiú alig héteszten
Az erdőben jártak, patakra találtak. A na
Először csak a lábukat mártogatták bele,

A event: ”the second letter is a vowel”
B event: “the first letter is a vowel”
$\overline{A}$ event (complement of A): “the second letter is a consonant”
$\overline{B}$ event (complement of B): “the first letter is a consonant”

The result of 100 repeated trials:

Absolute frequencies of neighbor letters

	B	$\overline{B}$	total
A	**6**	34	40
$\overline{A}$	36	24	60
total	42	58	100

Absolute frequencies of letters one under the other

	B	$\overline{B}$	total
A	**16**	23	39
$\overline{A}$	26	35	61
total	42	58	100

Calculate the conditional relative frequencies of the cyan area.

	B
A	**0.14**
$\overline{A}$	0.86
total	1

	B
A	**0.38**
$\overline{A}$	0.62
total	1

Based on this we may introduce the **conditional probability.** The probability that ”the second letter is a vowel” if “the first letter is a vowel”, $P(A|B)$: the probability of event A is conditioned on the prior occurrence of event B.

$P(AB)$: probability of occurrence of A and B.
This is ≈0,06 for neighbor letters, and ≈0,16 for the letters one under the other.
$P(B)$: probability of occurrence of B, this is ≈0,42.

Thus:

$$P(A|B) = \frac{P(AB)}{P(B)} .$$ (rule of multiplication)

The event A is statistically **independent** of event B if

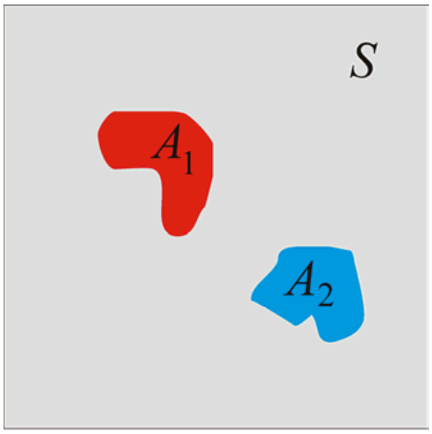
$$P(A|B) = P(A),$$

that is equivalent to equation:

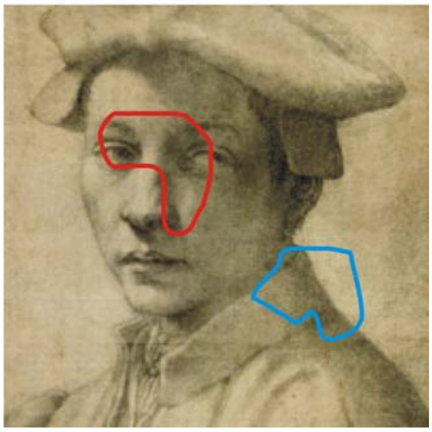
$$P(AB) = P(A)P(B).$$

Concept of distribution in every day meaning

Let us take an example: graphics.



We need a sheet of paper, S basic set, where the graphite is distributed.
We assign numbers to its subsets A (e.g. based on the amount (mass, weight) of graphite on the given set).



These numbers are additive:

$$P(A) = \sum_{k=1}^{n} P(A_k),$$

if all the A_k have no intersection.

Normalization: let the total amount of graphite be unit (mass or weight),

$$P(S) = 1.$$

Random variable

We observe a quantitative thing in connection with a phenomenon.

1. We give what and how to “measure”.
2. Random variable is characterized by its distribution or the parameters of distribution, if they exist.

Generally we do not know these parameters.

Practically all the “change” which based on any observation and we may assign numbers are of these kind. Its value depends on circumstances what we are not able to take into account, thus depends on “chance”.

Characterization of discrete random variable

E.g. roll of a pair of (independent) dice with 36 possible outcomes.

Let $\xi = i + k$ be the random variable
$i = 1, 2, 3, 4, 5, 6$ and $k = 1, 2, 3, 4, 5, 6$, thus
ξ may have 11 different values:

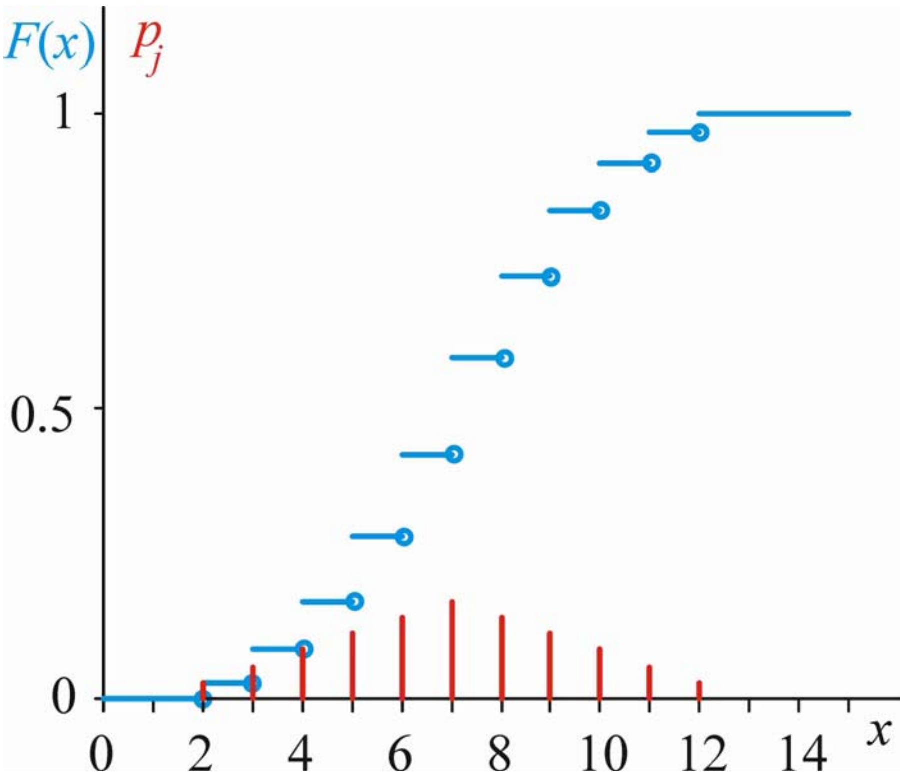
The possible values are: $x_j = 2, 3, \ldots, 12$.

The „result" of the roll is one of the possible values.

Characterization (by):
Distribution function [$F(x)$] and **Probabilities** [p_j]

$$F(x) = p(\xi < x) = \sum_{x_j < x} p(\xi = x_j)$$

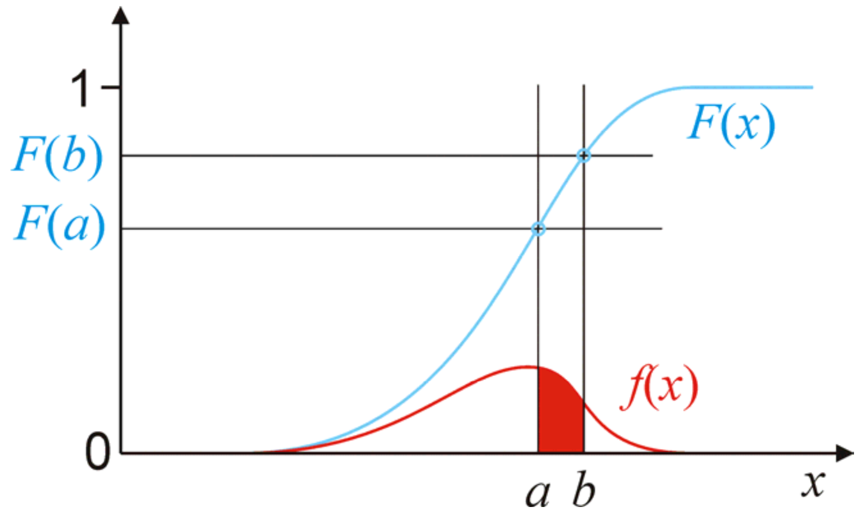
$$p_j = p(\xi = x_j)$$



x_j	p_j
2	1/36
3	2/36
4	3/36
5	4/36
6	5/36
7	6/36
8	5/36
9	4/36
10	3/36
11	2/36
12	1/36

Characterization of continuous random variable

Cumulative distribution function [$F(x)$] and Probability density function [$f(x)$]



$$F(b) - F(a) =$$
$$= p(a < \xi < b) =$$
$$= \int_a^b f(x)dx =$$
= [red area]

Numerical parameters of a random variable
or rather its distribution.

Where is the "middle" of distribution?

expected value [$M(\xi)$] ()

Discrete case: $$M(\xi) = \sum_i x_i p_i$$

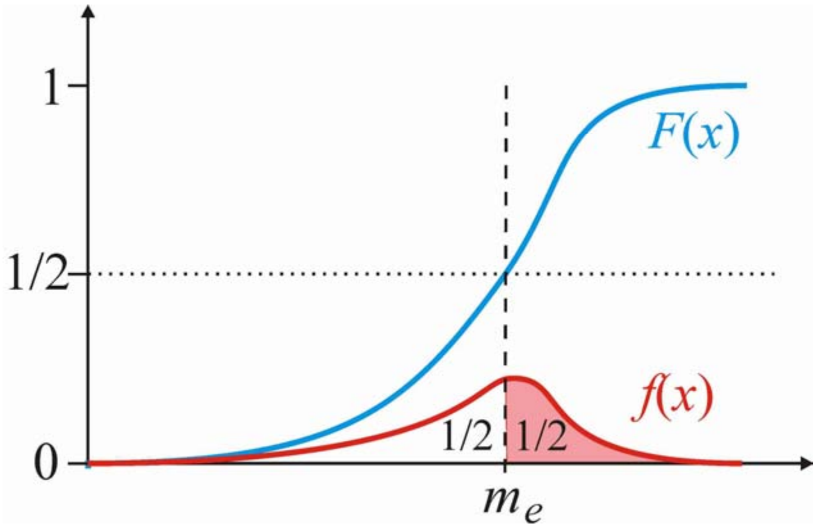
Continuous case: $$M(\xi) = \int_{-\infty}^{\infty} x f(x) dx$$

x_i	p_i	$x_i\, p_i$
2	1/36	2/36
3	2/36	6/36
4	3/36	12/36
5	4/36	20/36
6	5/36	30/36
7	6/36	42/36
8	5/36	40/36
9	4/36	36/36
10	3/36	30/36
11	2/36	22/36
12	1/36	12/36
		252/36 = 7

demonstration: location of center of mass

median (m_e)

$F(m_e) = 1/2$



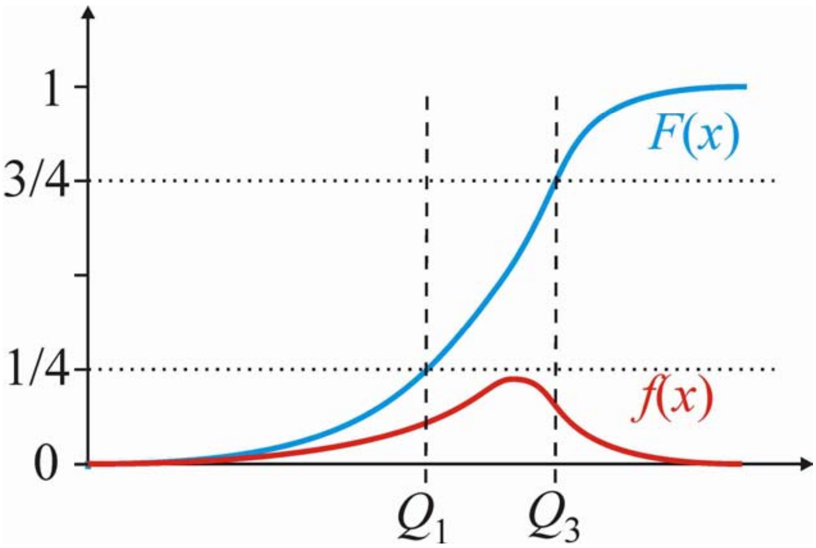
demonstration: quantile of two uniform probability (1/2) mass (weight) or rather area.

quantiles

other ratio of probability or mass (weight), or rather ratio of area (Q_1 lower, Q_3 upper quartile)

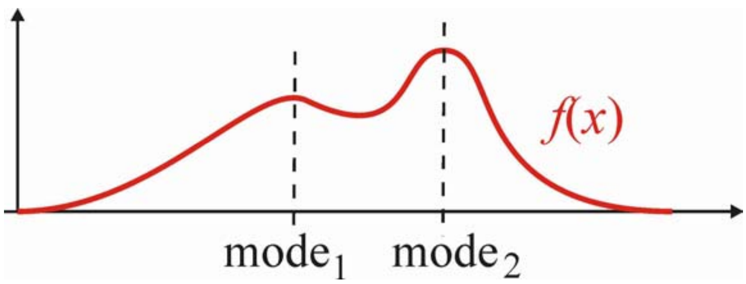
$F(Q_1) = 1/4$

$F(Q_3) = 3/4$

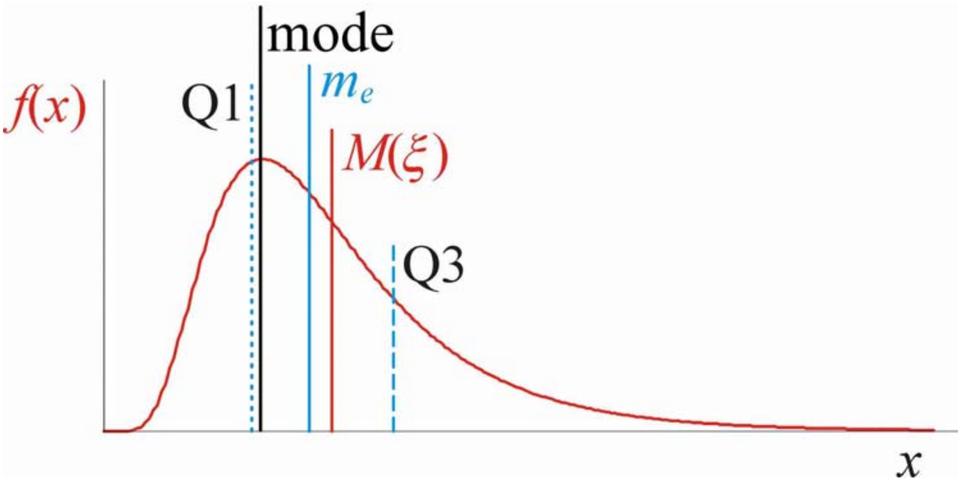


mode(s)

most probable value(s),
local maxima of the probability density function



Relation of the numerical parameters of the „middle”:



How large is the spread of the distribution?

variance

$$D^2(\xi) = M[(\xi - M(\xi))^2]$$

Further general parameters:

Moments **Central moments**

$M(\xi^k)$ $M[(\xi - M(\xi))^k]$ $k = 1, 2, 3, 4, \ldots$

Skewness or asymmetry:

It can be characterized by the 3rd central moment.

Peakedness or kurtosis:

It can be characterized by the 4th central moment.

Some properties of expected value

$$M(k\xi) = kM(\xi)$$

$$M(\xi + \eta) = M(\xi) + M(\eta)$$

if ξ and η independent random variables, than
$M(\xi\eta) = M(\xi)M(\eta)$,

Some properties of variance

$D^2(a\xi + b) = a^2 D^2(\xi)$

if ξ and η independent random variables, than
$D^2(\xi + \eta) = D^2(\xi) + D^2(\eta)$

As a consequence
if $D(\xi_i) = \sigma$, $i = 1, 2, \ldots, n$, than $D^2(\xi_1 + \xi_2 + \ldots + \xi_n) = n\sigma^2$

Some discrete probability distributions

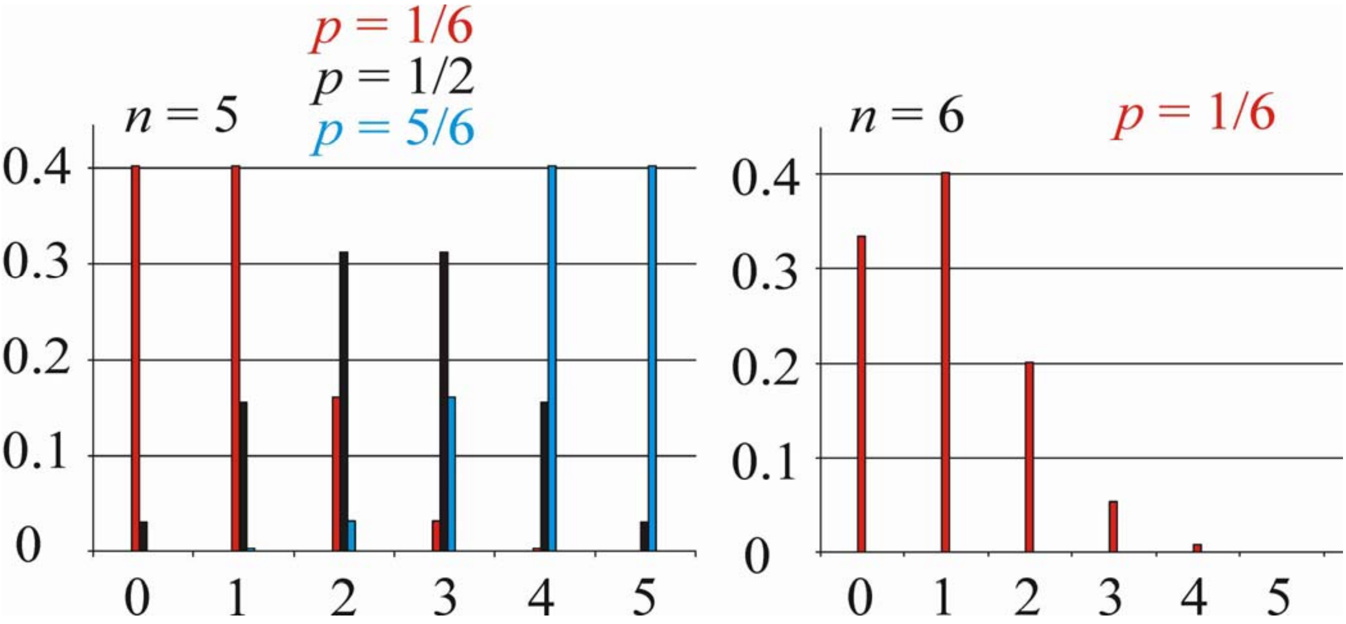
Binomial distribution (Bernoulli-distribution)

alternative p, $(1-p)$
n trials $P(\xi = k) = B(n, k)$

$M(\xi) = np$,
$D^2(\xi) = np(1-p)$

Example: dice, 6 rolls, $n = 6$
How many times we get 6?

k	P
0	0.33
1	0.4
2	0.2
3	0.05
4	0.008
5	0.0006
6	0.00002



Poisson-distribution

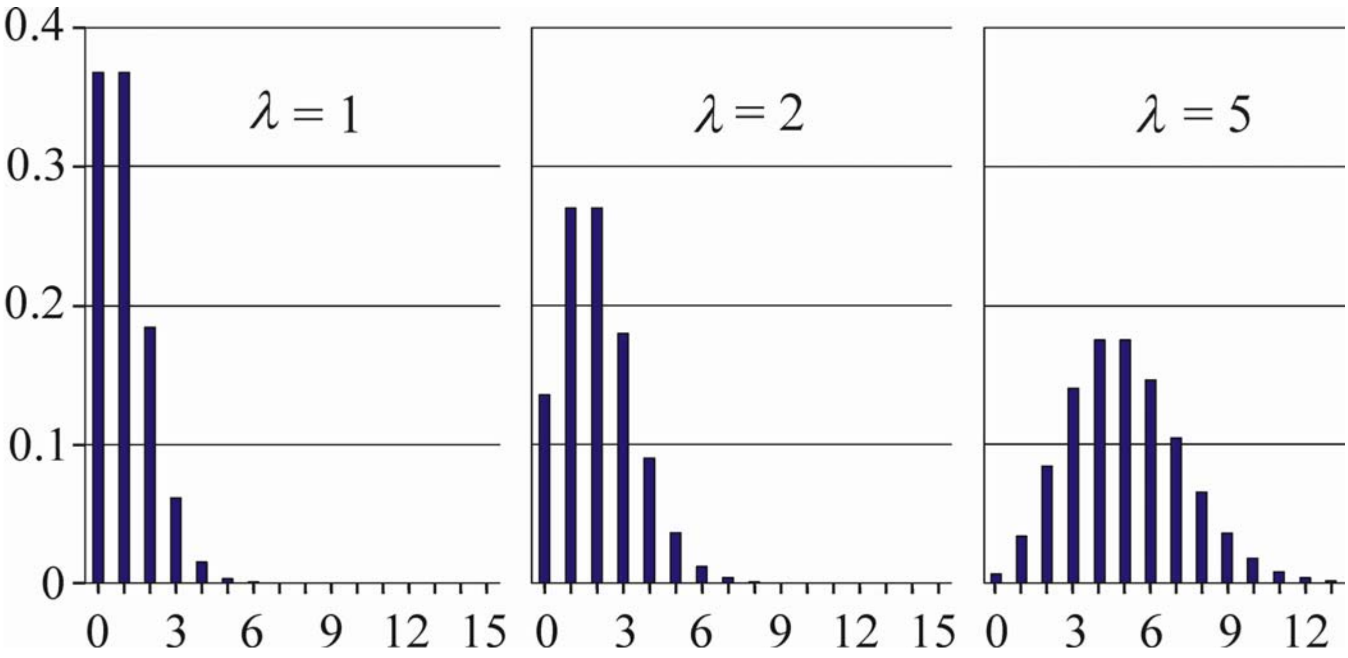
$M(\xi) = \lambda,\qquad D^2(\xi) = \lambda$

Examples:

number of calls in a telephone central during a certain time interval

number of particles in a given volume

number of decayed atoms in a radioactive substance during a given time interval

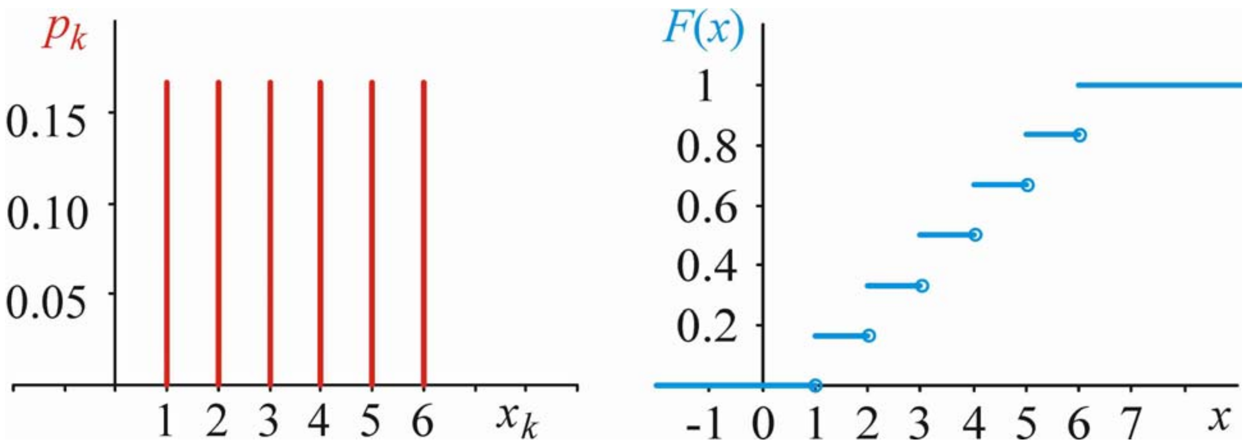


Uniform distribution

In a specific case

Example: dice; probability of an outcome $p = 1/6$.

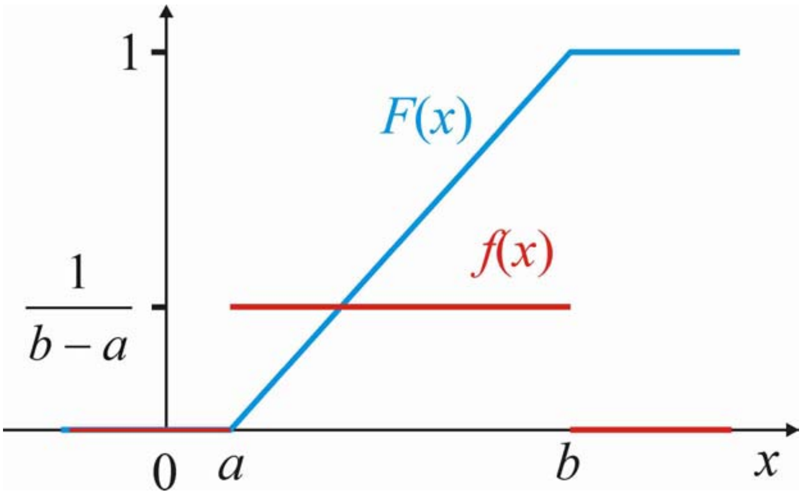
Possible values: 1, 2, 3, 4, 5, 6.



Some continuous probability distributions

Uniform distribution

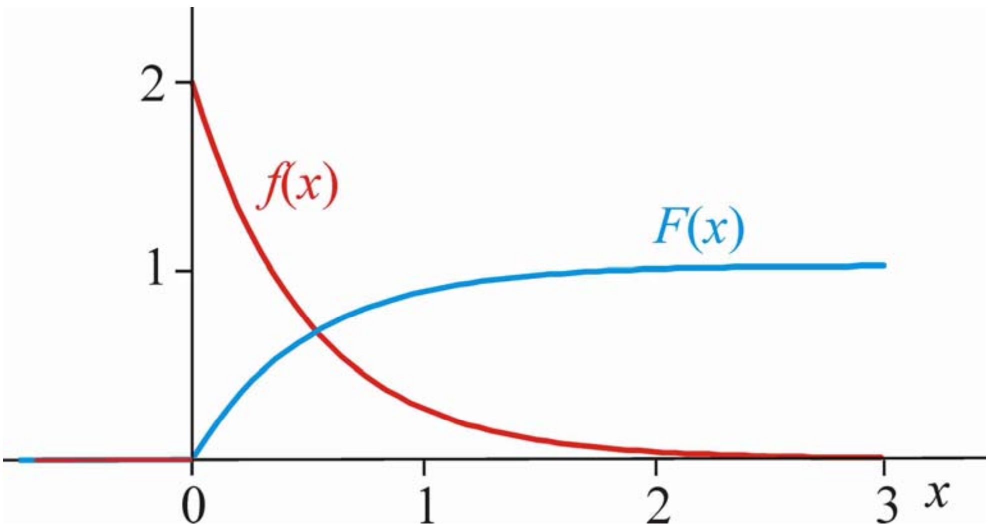
$M(\xi) = (a + b)/2$
$D^2(\xi) = (b - a)^2/12$



Example: the density or temperature of air in a room

Exponential distribution

$M(\xi) = 1/\lambda$, $(\lambda = 2)$
$D^2(\xi) = 1/\lambda^2$

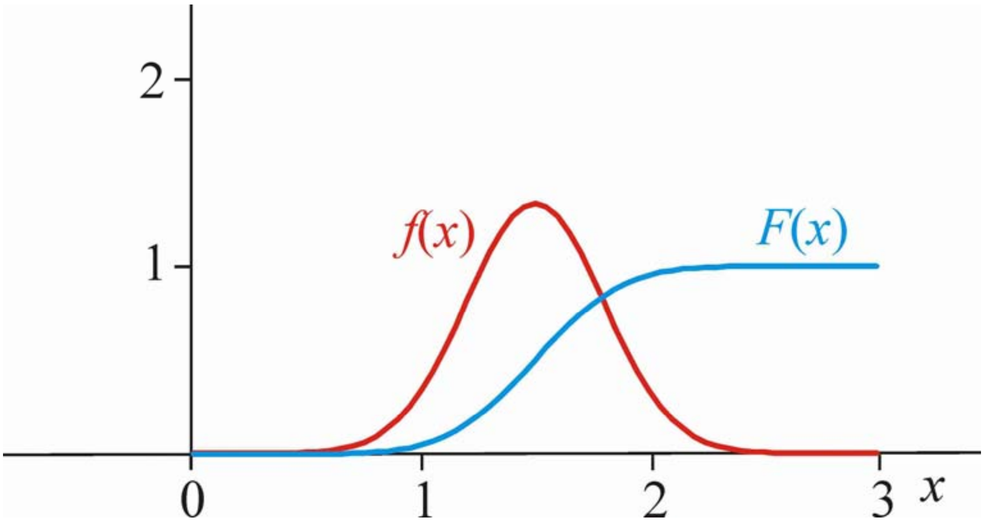


Examples: waiting times
lifetime of the individual atoms in the course of radioactive decay.

Normal distribution (Gaussian distribution)

$M(\xi) = \mu$, $\qquad N(\mu;\sigma)$

$D^2(\xi) = \sigma^2$ $\qquad N(1,5;0,3)$

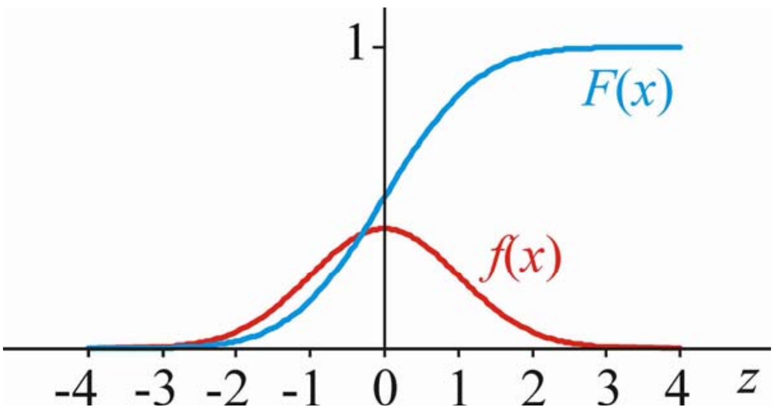


Examples:
The height of men in Hungary, given in cm: $N(171;7)$
Diastolic blood pressure of schoolboys, given in Hgmm: $N(58;8)$

Standard normal distribution

$M(\xi) = 0$

$D^2(\xi) = 1$

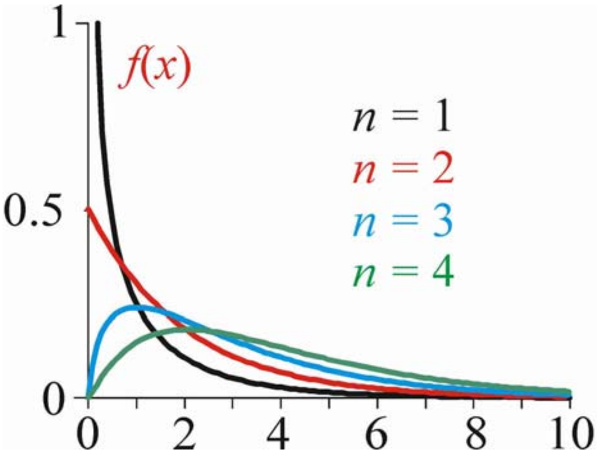


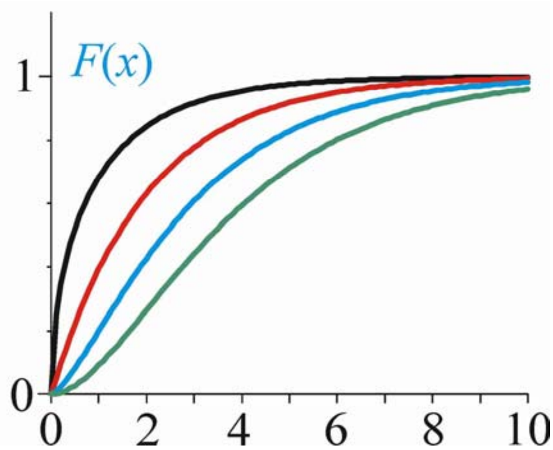
Transformation: $x\ [\mathrm{N}(\mu;\sigma)] \rightarrow z\ [\mathrm{N}(0;1)]$ $\qquad z = \dfrac{x-\mu}{\sigma}$

Both the χ^2-distribution and the t-distribution are results of the transformations of variables having standard normal distribution (ξ_n).

χ^2- distribution

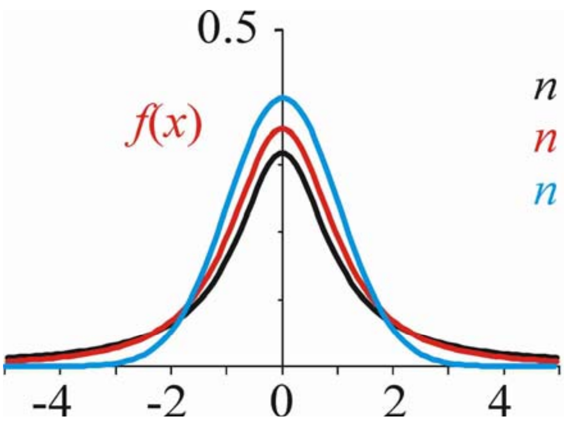
$M(\eta_n) = n$ $\qquad$ $D^2(\eta_n) = 2n$ $\qquad$ $\eta_n = \xi_1^2 + \ldots + \xi_n^2$

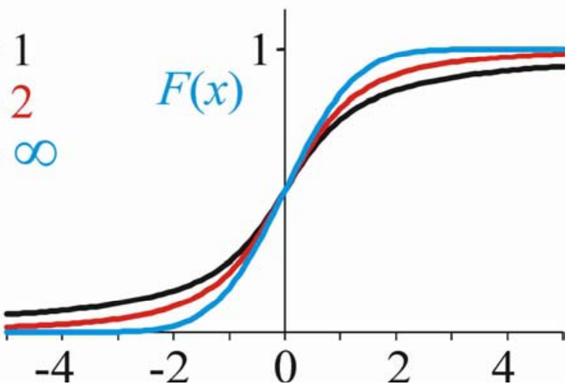




t-distribution

$M(\zeta_n) = 0$ $\qquad$ $D^2(\zeta_n) = n/(n-2)$ $\qquad$ $\zeta_n = \dfrac{\sqrt{n}\xi}{\sqrt{\eta_n}}$



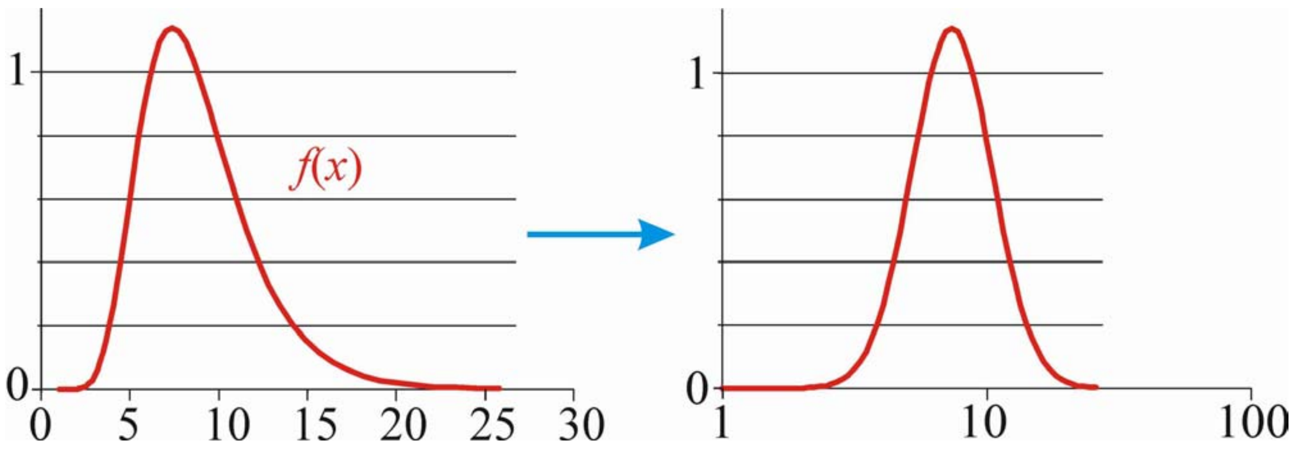


In these distributions n is the so-called **degree of freedom.** This parameter is connected to the number of elements of the dataset studied.

At the limes of $n \to \infty$ the t-distribution and the standard normal distribution are identical.

Lognormal distribution

ξ has lognormal distribution, if $\varphi = \ln\xi$ has normal distribution



Why the normal distribution is a favoured one?

Central limit theorem

If a random variable is a result of a sum of several small independent change, than it should be a random variable having normal distribution with a good approximation.

You may try it!