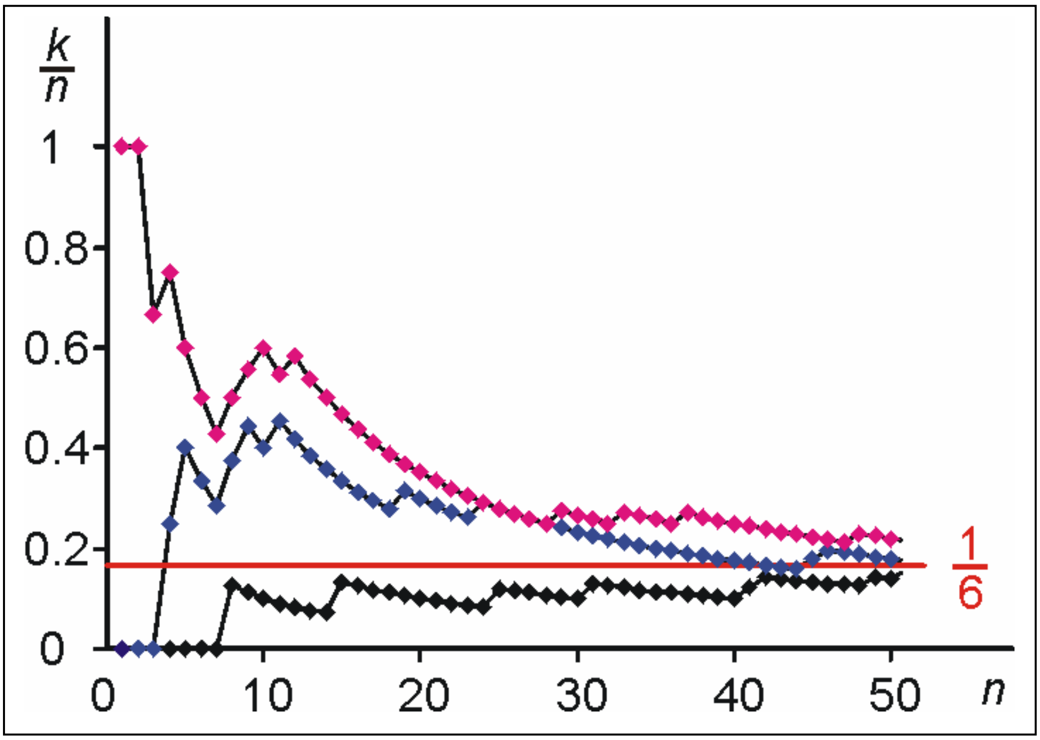
Elements of probability calculus

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 die 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.

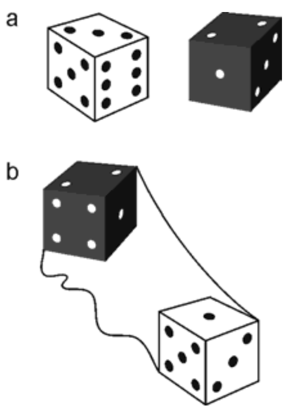
(It is an empirical fact it can not be proven by logical sequence.)
(Karl Pearson 1857-1936)
We assign a number to the event: **probability**.

Properties of probability:

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

Independence

1000 rolls



a

	⚀	⚁	⚂	⚃	⚄	⚅
⚀	30	25	30	29	28	25
⚁	24	27	31	27	24	27
⚂	28	30	39	32	24	29
⚃	28	28	22	26	27	33
⚄	27	24	26	21	31	27
⚅	30	25	32	30	29	25

b

	⚀	⚁	⚂	⚃	⚄	⚅
⚀	40	41	46	12	9	21
⚁	51	38	37	13	22	15
⚂	42	49	52	8	20	17
⚃	8	10	15	36	52	44
⚄	11	16	9	45	39	35
⚅	10	17	8	43	41	28

Conditional probability

The probability that "the result of the black die is 1" (event A) if "the result of the white die is 1" (event B),
$P(A|B)$: the probability of event A is conditioned on the prior occurrence of event B.

If $P(A|B) = P(A)$ then event A is statistically **independent** of event B

If $P(A \cap B) \equiv P(AB)$ is the probability of occurrence of A and B, then

$$P(A|B)\,P(B) = P(A \cap B) \qquad \text{(rule of multiplication)}$$

Independence (equivalent equation): $P(A)P(B) = P(A \cap B)$.

Problem:

After a die roll, are the next two events independent or not? The result is smaller than 3 (event A), the result is even (event B).

Use the previous equation!

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 by the parameters of distribution, if they exist.

In general 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: $x_j = 2, 3, \ldots, 12$.

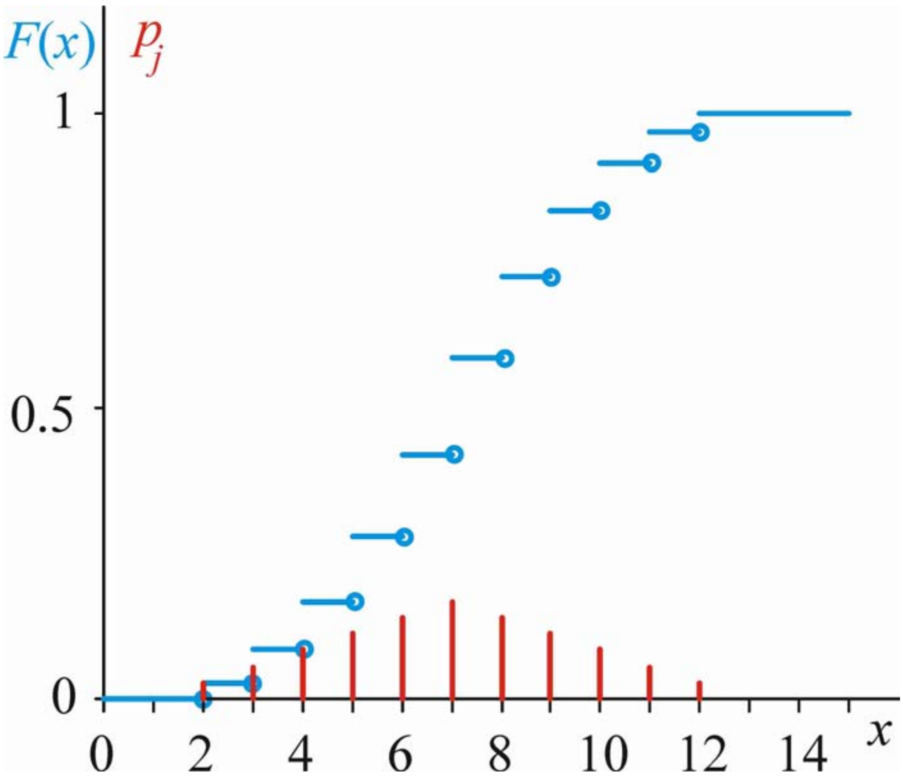
The „result" of the roll is one of the possible values.

Characterization (by):

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

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

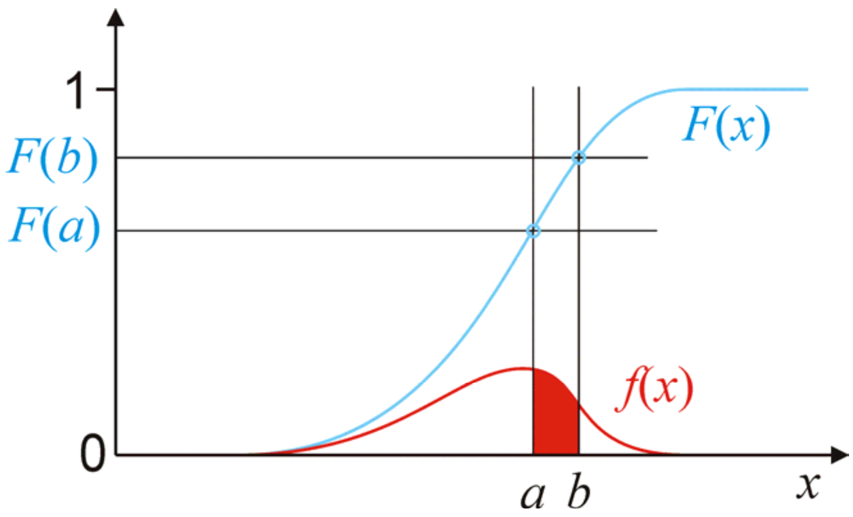
$$p_j = p(\xi = x_j)$$



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?

1a. **expected value** [$M(\xi)$]

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

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

(roll of a pair of dice)

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

If we have only a **few data** we can not see the characteristics of **data set**.
Numerical characteristics of quantitative data
(can be determined in every case)

Where is the **"middle"** of data set with n elements?

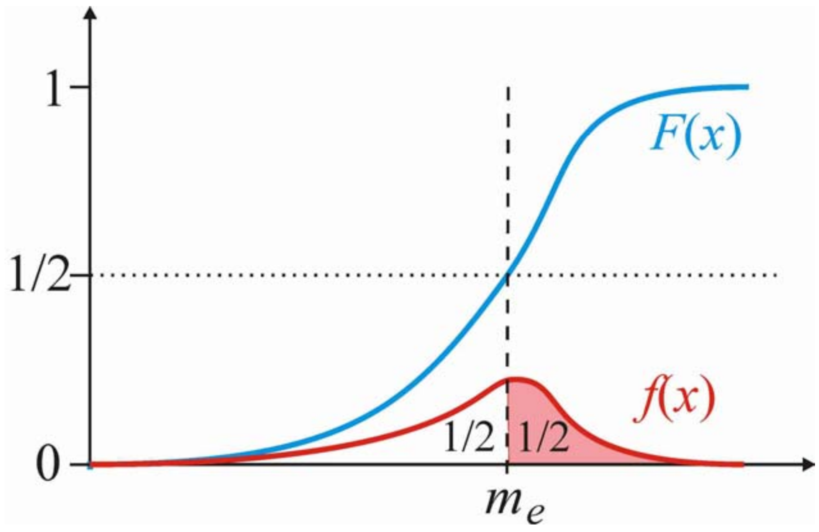
1b. **mean** (arithmetical average)

$$x_{\text{mean}} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{\sum_{j=1}^{m} w_j x_j}{\sum_{j=1}^{m} w_j}$$

It is sensitive to the extreme values!

2a. **median** (m_e)

$F(m_e) = 1/2$



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

2b. **median** (x_{median}) of data set

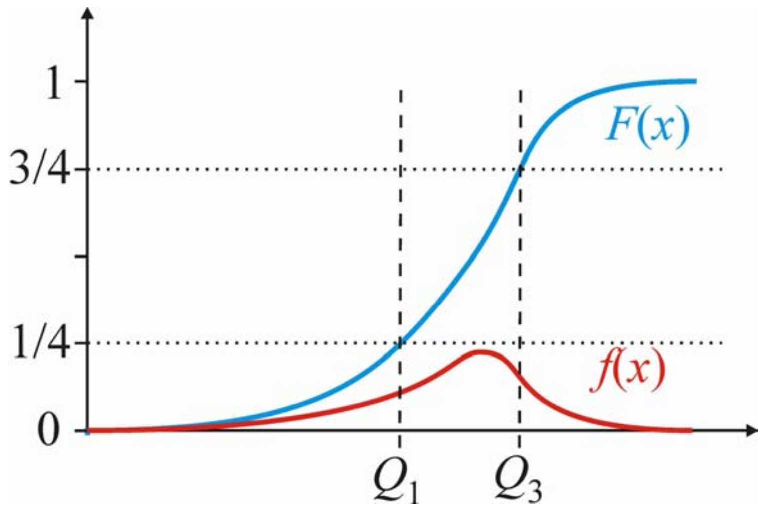
We order the data according to their magnitudes, and look for the middle or middles.

3a. **quantiles**

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

$F(Q_1) = 1/4$

$F(Q_3) = 3/4$



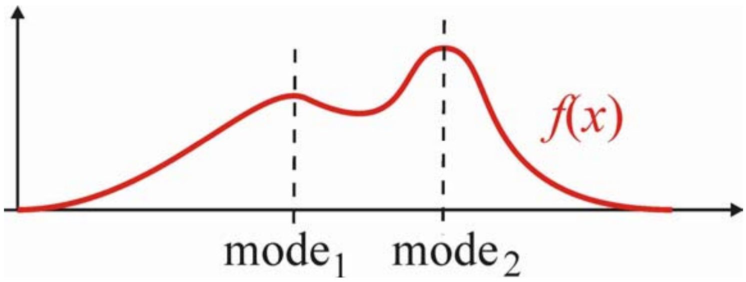
3b. for a set of data; e.g.: What income makes a person become a member of the “upper ten thousand”.

First we order the data by magnitude again.

E.g. lower quartile, middle quartile = median, upper quartile

4a. **mode(s)**

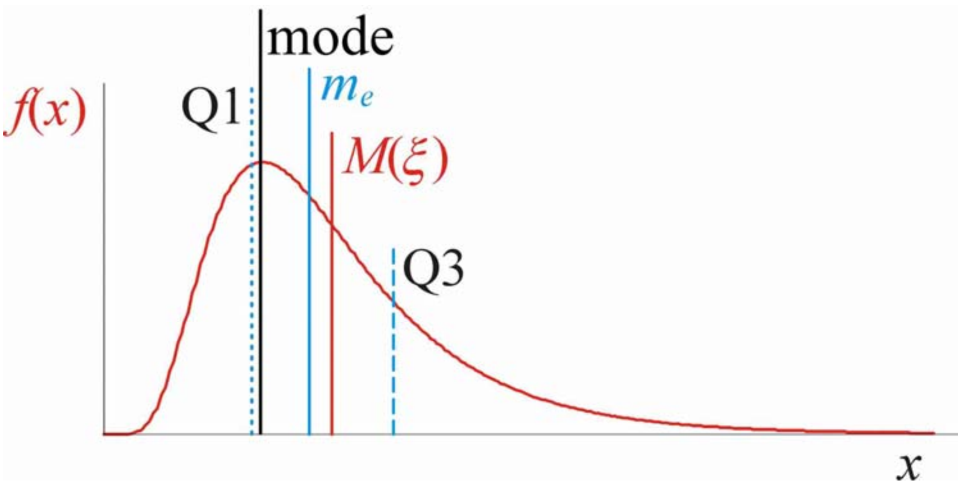
most probable value(s),
local maxima of the probability density function



4b. if the data set has identical data, the one which has the most copy called as **mode**. (But more mode also may exist in the same data set.) („mode” → fashionable)

They are not sensitive to the extreme values!

Relation of the numerical parameters of the „middle”:



How large is the **spread** of the distribution?

1. **variance**

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

Characteristics of measures of spread of data set

0. **range**

 the difference of the biggest and the smallest elements of the data set

1. **variance** (s_x^2)

 average of the squared deviation of the data from the mean

$$\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2 .$$

2. **standard deviation**

 of the data set is given by the formula

$$s_x=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

Further characteristics can also be quantified (skewness, kurtosis).

Some properties of expected value

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

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

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

Some properties of variance

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

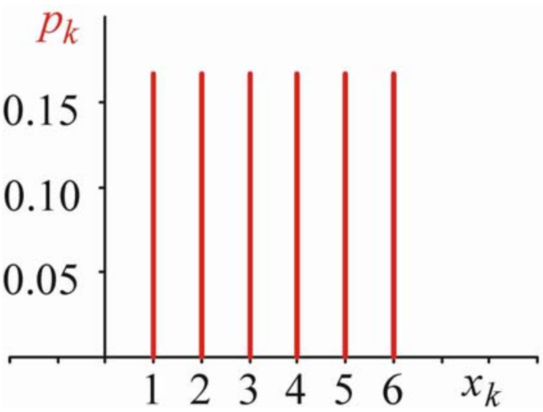
if ξ and η are independent random variables, than
$D^2(\xi+\eta)=D^2(\xi)+D^2(\eta)$

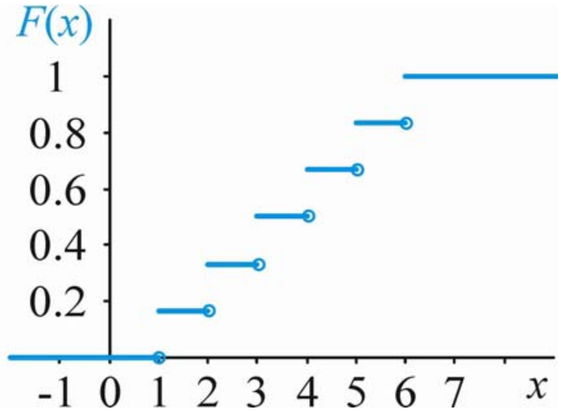
Some remarkable (model) **distributions**

1. Discrete probability distributions

Uniform distribution

In a specific case
Example: dice; probability of an outcome p = 1/6.
Possible values: 1, 2, 3, 4, 5, 6.





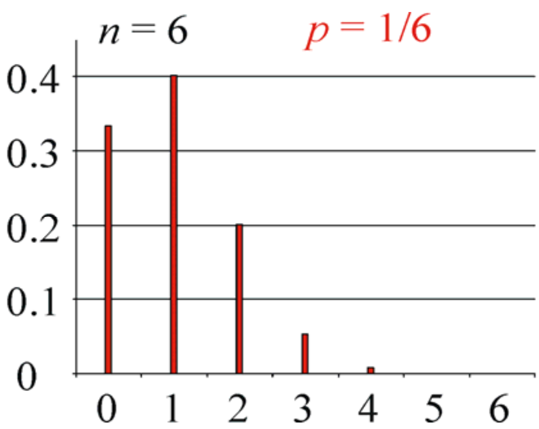
Binomial distribution (Bernoulli-distribution)

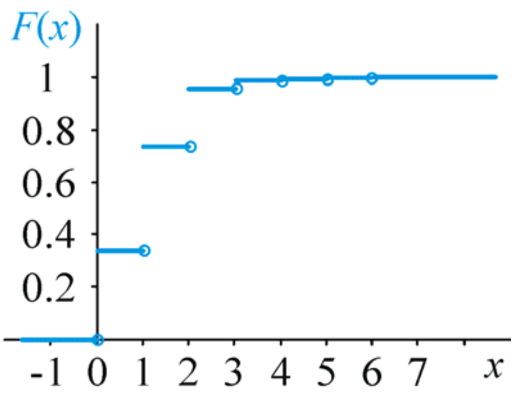
alternative p, $(1–p)$
n trials $P(\xi = k) = B(n, k)$

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

Example: dice, 6 rolls, $n = 6$ ($p = 1/6$)
What is the probability that we get never ($k = 0$), ones, twice (k-times) a result of 6?

$M(\xi) = np$,
$D^2(\xi) = np(1–p)$

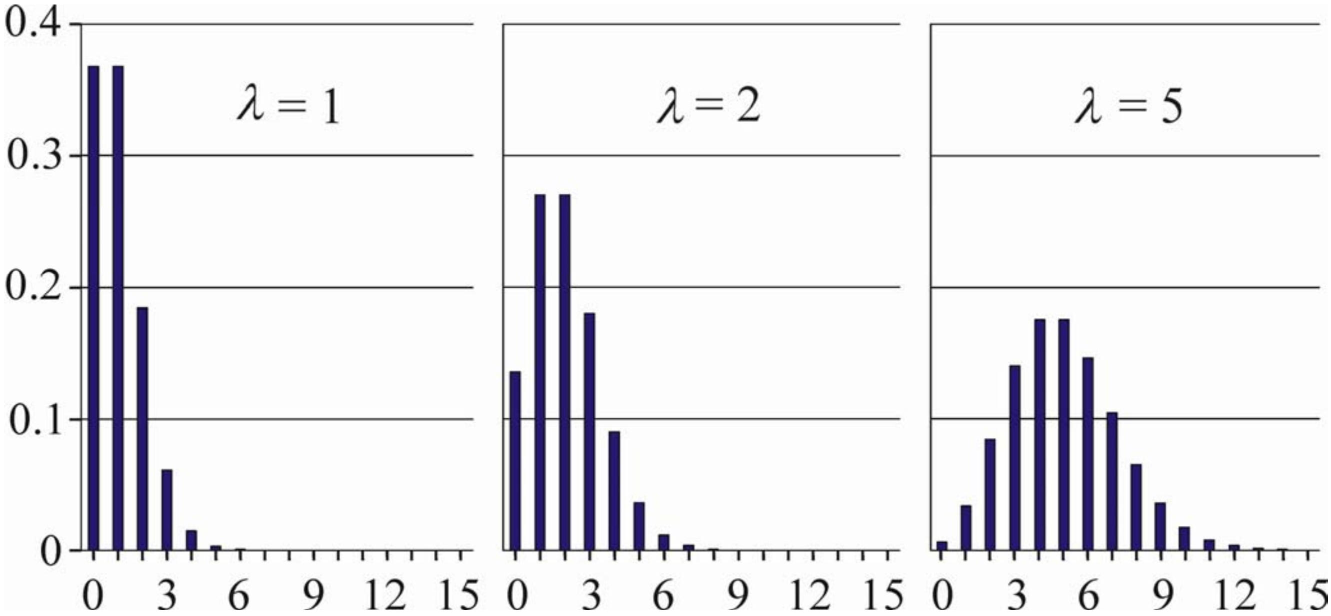




End of lecture

Poisson-distribution

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

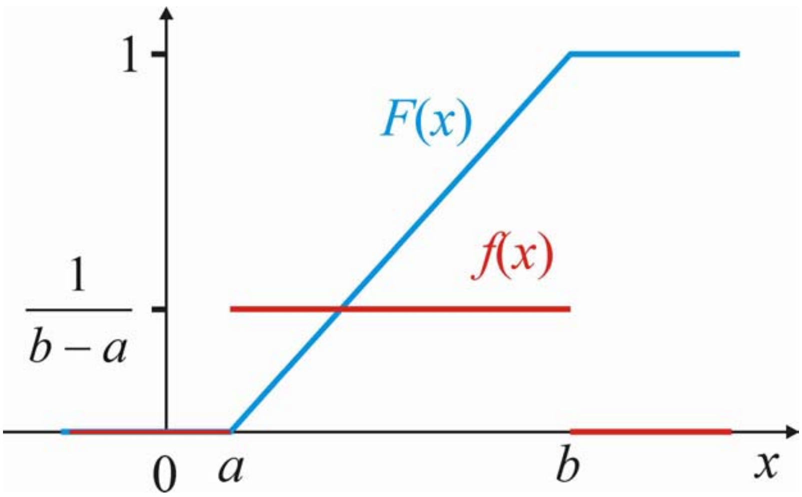


Examples: number of particles in a given volume
number of decayed atoms in a radioactive substance during a given time interval

2. Continuous probability distributions

Uniform distribution

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

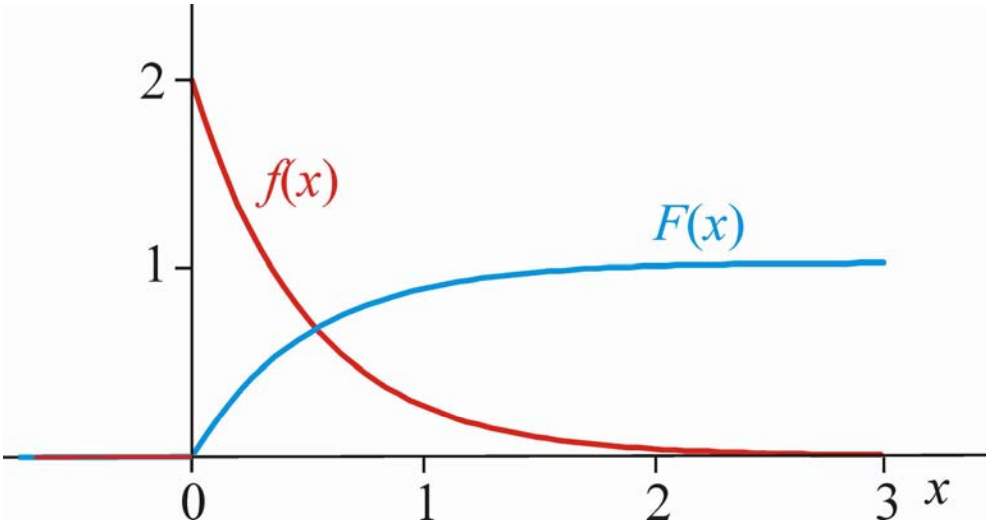


Example: the density or temperature of air in a room

Exponential distribution

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

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

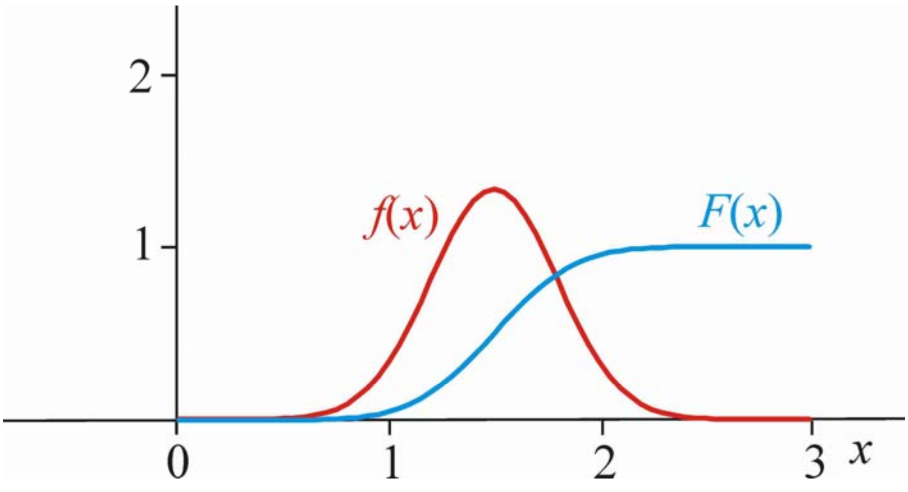


Example: lifetime of the individual atoms in the course of radioactive decay

Normal distribution (Gaussian distribution)

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

$D^2(\xi) = \sigma^2$ $N(1.5;0.3)$



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 [N(μ;σ)] $\rightarrow z$ [N(0;1)] $z = \dfrac{x-\mu}{\sigma}$

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

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 changes, than it should be a random variable having normal distribution with a good approximation.

You may try it!