Statistics - handout
Part-III:
Continuous random variables and distributions
-

Andras Vetier

January 13, 2017

Contents

1 Continuous random variables 2

2 Distribution function 3

3 Empirical distribution function 5

4 Density function 6

5 Uniform distributions 9

6 Beta distributions 11

7 Exponential distribution 15

8 Gamma distribution 17

9 Normal distributions 19

10 Generating a random variable with a given continuous distribution 22

11 Expected value of continuous distributions 25

12 Expected value of a function of a continuous random variable 31

13 Median 34

14 Standard deviation, etc. 37

1 Continuous random variables

In this chapter, we start studying random variables which, contrary to discrete random variables, may take any numbers from an interval of the real line. Some examples:

1. X = the amount of time someone has to wait for the bus when he or she goes to work. The waiting time may be any positive number not exceeding a certain upper limit.

2. X = the weight of a new-born baby. The weight may be any number between certain lower and upper limits.

3. X = the height of a randomly chosen Hungarian man. The height may be any number between certain lower and upper limits.

Obviously, when we observe such a random variable, that is, we check the time or measure the weight or the height,
- we get the time rounded, for example, to minutes, possibly with some decimals,
- we get the weight rounded, for example, to kilograms or grams, or to pounds, possibly with some decimals,
- we get the height rounded, for example, to centimeters, or to inches, possibly with some decimals, as well.

Using more decimals, we get more precise results. However, what we get, are always rational numbers. We emphasize that, in spite of the fact that the measurement results we get are rational numbers, the waiting time, the weight, the height themselves may be not only rational numbers but any real numbers between certain lower and upper limits.

Each specific possible value has zero probability. An important property of such random variables is that, for any fixed real number x, the probability that the random variable is equal to that specific number x is equal to 0:

$$\mathbf{P}(X = x) = 0$$

A random variable X is called to be **continuous** if for any fixed x value, the probability that the random X value is equal to the given x value is equal to 0.

2 Distribution function

The notion of the (cumulative) distribution function (often abbreviated in text-books as **c.d.f.**) plays an important role both in theory and in practice. The **(cumulative) distribution function** $F(x)$ of a random variable X is defined by

$$F(x) = \mathbf{P}(X \le x)$$

In some text-books, the definition of distribution function may be different

$$F(x) = \mathbf{P}(X < x)$$

However, for a continuous distribution this is not a real difference, because for a continuous distribution:

$$\mathbf{P}(X \le x) = \mathbf{P}(X < x)$$

Clearly, for any real number x, the event $X > x$ is the complement of the event $X \le x$, so

$$\mathbf{P}(X > x) = 1 - \mathbf{P}(X \le x) = 1 - F(x)$$

For any real numbers $a < b$, the event $a < X \le b$ is the difference of the events $X \le b$ and $X \le a$, so

$$\mathbf{P}(a < X \le b) = \mathbf{P}(X \le b) - \mathbf{P}(X \le a) = F(b) - F(a)$$

For a continuous random variable

$$\mathbf{P}(X < x) = \mathbf{P}(X \le x) = F(x)$$

and

$$\mathbf{P}(X \ge x) = \mathbf{P}(X > x) = 1 - F(x)$$

and

$$\mathbf{P}(a < X < b) = F(b) - F(a)$$

$$\mathbf{P}(a \le X < b) = F(b) - F(a)$$

$$\mathbf{P}(a < X \le b) = F(b) - F(a)$$

$$\mathbf{P}(a \le X \le b) = F(b) - F(a)$$

Characteristic properties of a distribution function:

1. Any distribution function $F(x)$ is an increasing function (not necessarily strictly increasing), that is, if $x_1 < x_2$, then $F(x_1) \le F(x_2)$.

2. $F(x)$ has a limit equal to 0 at $-\infty$: $\lim_{x \to -\infty} F(x) = 0$.

3. $F(x)$ has a limit equal to 1 at $(+\infty)$: $\lim_{x\to\infty} F(x) = 1$.

4. The distribution function of a continuous random variable is continuous.

These four properties are characteristics for distribution functions, because, on one side, they are true for the distribution function of any continuous random variable, and on the other side, if a function $F(x)$ is given which has these four properties, then it is possible to define a random variable X so that its distribution function is the given function $F(x)$.

Sometimes it is advantageous to use the so called **tail function** of a distribution:

$$T(x) = \mathbf{P}(X > x)$$

which is the "complement" of the distribution function:

$$T(x) = 1 - F(x)$$

or equivalently:

$$F(x) = 1 - T(x)$$

The tail function is obviously a decreasing function: if $x_1 < x_2$, then $T(x_1) \geq T(x_2)$.

3 Empirical distribution function

It is an important fact that the distribution function of a random variable X can be approximated from experimental results as described here. Imagine that we make N experiments for X, and we get the experimental results $X_1, X_2, \dots, X_N$. Using these experimental results, let us consider the horizontal line segments (closed from the left and open from the right) defined by the point-pairs:

$$(-\infty; 0), \quad (X_1; 0)$$
$$(X_1; \frac{1}{N}), \quad (X_2; \frac{1}{N})$$
$$(X_2; \frac{2}{N}), \quad (X_3; \frac{2}{N})$$
$$(X_3; \frac{3}{N}), \quad (X_4; \frac{3}{N})$$
$$(X_4; \frac{4}{N}), \quad (X_5; \frac{4}{N})$$
$$\vdots$$
$$(X_{N-2}; \frac{N-2}{N}), \quad (X_{N-1}; \frac{N-2}{N})$$
$$(X_{N-1}; \frac{N-1}{N}), \quad (X_N; \frac{N-1}{N})$$
$$(X_N; 1), \quad (\infty; 1)$$

These line segments constitute the graph of a function called **empirical distribution function** fitted to the experimental results $X_1, X_2, \dots, X_N$.

For technical purposes, it is more convenient to draw not only the horizontal but the vertical line segments, as well, which will yield a broken line connecting the points

$$(-\infty; 0), \quad (X_1; 0), \quad (X_1; \frac{1}{N}), \quad (X_2; \frac{1}{N}), \quad (X_2; \frac{2}{N}), \quad (X_3; \frac{2}{N}), \quad \dots$$
$$\dots, \quad (X_{N-1}; \frac{N-1}{N}), \quad (X_N; \frac{N-1}{N}), \quad (X_N; 1), \quad (\infty; 1)$$

It is convenient to think of this broken line as a representation of the graph of the **empirical distribution function** fitted to the experimental results $X_1, X_2, \dots, X_N$.

Imagine that we make now more and more experiments for X, and we get the experimental results $X_1, X_2, \dots$. Using these experimental results, we may construct the empirical distribution function or the broken line representing the graph of the empirical distribution function fitted to the experimental results $X_1, X_2, \dots, X_N$ for all N, and we get a sequence of functions or graphs. The so called **basic theorem of mathematical statistics** states that as N goes to infinity, the sequence of functions or graphs approaches uniformly to the (graph of the) theoretical distribution function of the random variable X. Obviously, the question of the precision of the approximation opens several questions, which we do not discuss here.

4 Density function

The **(probability) density function** (often abbreviated in text-books as **p.d.f.**) is the function $f(x)$ which has the property that for any interval $[a, b]$

$$\mathbf{P}(a < X < b) = \int_a^b f(x)\, dx$$

If $a = x$ and $b = x + \Delta x$, then

$$\mathbf{P}(x < X < x + \Delta x) = \int_x^{x+\Delta x} f(x)\, dx$$

For small Δx, the integral can be approximated by $f(x)\Delta x$, and we get:

$$\mathbf{P}(x < X < x + \Delta x) \approx f(x)\Delta x$$

that is

$$f(x) \approx \frac{\mathbf{P}(x < X < x + \Delta x)}{\Delta x}$$

We emphasize that the value $f(x)$ of the density function does not represent any probability value. If x is a fixed value, then $f(x)$ may be interpreted as a constant of approximate proportionality: for small Δx, the interval $[x, x + \Delta x]$ has a probability approximately equal to $f(x)\Delta x$:

$$\mathbf{P}(x < X < x + \Delta x) \approx f(x)\Delta x$$

Mechanical meaning of the density. While learning probability theory, it is useful to know about the mechanical meaning of a density function. For many mechanical problems, if the density function of a mass-distribution is given, then we are able to imagine the mass-distribution. We study one-dimensional distributions in this chapter, but since we live in a 3-dimensional world, the mechanical meaning of the density function will be first introduced in the 3-dimensional space. The reader must have learned the following arguments in mechanics.

First imagine that mass is distributed in the 3-dimensional space. If we take a point x and a small set (for example, sphere) A around x, then we may compare the amount of mass located in A to the volume of A:

$$\frac{\text{amount of mass located in } A}{\text{volume of } A}$$

This ratio is the average mass-density inside A. Now, if A is getting smaller and smaller, then the average density inside A will approach a number, which is called the mass-density at x. If $f(x, y, z)$ is the density function of mass-distribution in the space, and A is a region in the space, then the total amount of mass located in the region A is equal to the integral

$$\iiint_A f(x, y, z)\, dxdydy$$

Now imagine that mass is distributed on a surface, or specifically on a plane. If we take a point on the surface and a small set (for example, circle) A around the point on the surface, then we may compare the amount of mass located in A to the surface-area of A:

$$\frac{\text{amount of mass located in } A}{\text{surface-area of } A}$$

This ratio is the average density inside A. Now, if A is getting smaller and smaller, then the average density inside A will approach a number, which is called the mass-density at the point on the surface, or specifically on the plane. If $f(x,y)$ is the density function of mass-distribution on the plane, and A is a region in the plane, then the total amount of mass located in the region A is equal to the integral

$$\iint_A f(x,y)\ dxdy$$

Finally, imagine that mass is distributed on a curve, or specifically on a straight line. If we take a point x on the curve and a small set (for example, interval) A around x on the curve, then we may compare the amount of mass located in A to the length of A:

$$\frac{\text{amount of mass located in } A}{\text{length of } A}$$

This ratio is the average density inside A. Now, if A is getting smaller and smaller, then the average density inside A will approach a number, which is called the mass-density at x on the curve, or specifically on the straight line. If $f(x)$ is the density function of mass-distribution on the real line, and $[a,b]$ is an interval, then the total amount of mass located in the interval $[a,b]$ is equal to the integral

$$\int_a^b f(x)\ dx$$

It is clear that a probability density function corresponds to a mass distribution when the total amount of mass is equal to 1.

Characteristic properties of a probability density function:

1. $f(x) \geq 0$ for all x,
2. $\int_{-\infty}^{\infty} f(x)\ dx = 1$

These two properties are characteristic for density functions, because, on one side, they are true for all density functions of any continuous random variables, and on the other side, if a function $f(x)$ is given which has these two properties, then it is possible to define a random variable X so that its density function is the given function $f(x)$.

Relations between the distribution function and the density function. The relations between the distribution function and the density function can be given by integration and differentiation. Integrating the density function from $-\infty$ to x, we get the distribution function at x:

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

However, when $f(x) = 0$ outside an interval $[A, B]$, then to get $F(x)$ for any x between A and B, instead of integrating from $-\infty$ to x, we may integrate from A:

$$F(x) = \int_{A}^{x} f(x)\, dx \quad \text{if} \quad A < x < B$$

and

$$F(x) = \begin{cases} 0 & \text{if} \quad x < A \\ 1 & \text{if} \quad B < x \end{cases}$$

Differentiating the distribution function, we get the density function:

$$f(x) = F'(x)$$

Uniform distribution under the graph of the density function. If X is a random variable, and $f(x)$ is its density function, then we may plug the random X value into the density function. We get the random value $f(X)$. It is clear that the random point $(X, f(X))$ will always be on the graph of the density function. Now let us take a random number RND, independent of X, and uniformly distributed between 0 and 1, and let us consider the random point $(X, f(X)\text{RND})$. It is easy to see that the random point $(X, f(X)\text{RND})$ is uniformly distributed in the region under the graph of the density function.

5 Uniform distributions

1. Special case: Uniform distribution on $(0;1)$

Applications:

1. When something happens during a time interval of unit length so that it may happen in any small part of the interval with a probability equal to the length of that small part, then the time-instant when it occurs follows uniform distribution on $(0;1)$.

2. Random numbers generated by calculators or computers follow uniform distribution on $(0;1)$.

Density function:

$$f(x) = 1 \quad \text{if} \quad 0 < x < 1$$

Distribution function:

$$F(x) = x \quad \text{if} \quad 0 < x < 1$$

2. General case: Uniform distribution on $(A;B)$

Applications:

1. When something happens during the time interval (A,B) so that it may happen in any small part of the interval with a probability proportional to the length of that small part, then the time-instant when it occurs follows uniform distribution on (A,B).

2. Take a circle with radius 1, and choose a direction, that is, choose a radius of the circle. Then choose a point on the circumference of the circle at random so that none of the parts of the circle have any preference compared to other parts. Then the angle, measured in radians, determined by the fixed radius and the radius drawn to the chosen point has a uniform distribution on the interval $(0;\pi)$. If, instead of radians, the angle is measured in degrees, we get a random variable uniformly distributed between 0 and 360.

3. Random numbers generated by calculators or computers follow uniform distribution on $(0;1)$. If we multiply them by $B-A$, then we get random numbers which follow uniform distribution on $(0;B-A)$. If we add now A to them, then we get random numbers which follow uniform distribution on $(A;B)$.

Density function:

$$f(x) = \frac{1}{B - A} \quad \text{if} \quad A < x < B$$

Distribution function:

$$F(x) = \frac{x - A}{B - A} \quad \text{if} \quad A < x < B$$

Parameters: $-\infty < A < B < \infty$

6 Beta distributions

1. Special case: Beta distribution on the interval $[0;1]$

Applications:

1. If n people arrive between noon and 1pm independently of each other according to uniform distribution, and we are interested in the time instant when the kth person arrives, then this arrival time follows the beta distribution related to size n and rank k.
2. If we generate n independent random numbers by a computer, and we consider the kth smallest of the generated n values, then this random variable follows the beta distribution related to size n and rank k.

Density function:

$$f(x) = \frac{n!}{(k-1)!(n-k)!}\, x^{k-1}(1-x)^{n-k} \qquad \text{if} \quad 0 < x < 1$$

Distribution function:

$$F(x) = \sum_{i=k}^{n} \binom{n}{i} x^i (1-x)^{n-i} \qquad \text{if} \quad 0 < x < 1$$

or, equivalently,

$$F(x) = 1 - \sum_{i=0}^{k-1} \binom{n}{i} x^i (1-x)^{n-i} \qquad \text{if} \quad 0 < x < 1$$

Parameters: k and n are positive integers so that $k \leq n$. n can be called as the size, k as the rank of the distribution.

Remark In order to remember the exponents in the formula of the density function, we mention that $k-1$ is the number of random numbers before the kth smallest, and $n-k$ is the number of random numbers after the kth smallest.

Proof of the formula of the density function. Let us generate n uniformly distributed independent random points between 0 and 1. Let X be the kth smallest among them. We calculate here the density function of the random variable X. Let $0 < x < 1$, and let $\Delta x = [x_1, x_2]$ be a small interval around x. By the meaning of the density function:

$$f(x) \approx \frac{\mathbf{P}(X \in \Delta x)}{x_2 - x_1}$$

The event $X \in \Delta x$, which stands in the numerator, means that the kth smallest point is in $[x_1, x_2)$, which means that

there is	at least one	point X	in	$[x_1, x_2)$,	and
there are	$k-1$	points	in	$[0, X)$,	and
there are	$n-k$	points	in	$[X, 1]$.	

This, with a very good approximation, means that

there are	$k-1$	points	in	$[0, x_1)$,	and
there is	1	point	in	$[x_1, x_2)$,	and
there are	$n-k$	points	in	$[x_2, 1]$.	

Using the formula of the poly-hyper-geometrical distribution, we get that the probability of the event $X \in \Delta x$ is approximately equal to

$$\frac{n!}{(k-1)!\; 1!\; (n-k)!}\;\; x_1^{k-1}(x_2 - x_1)^1 (1 - x_2)^{n-k}$$

Since $1! = 1$, we may omit some unnecessary factors and exponents, and the formula simplifies to

$$\frac{n!}{(k-1)!\; (n-k)!}\;\; x^{k-1}\; (x_2 - x_1)\; (1 - x_2)^{n-k}$$

Dividing by $(x_2 - x_1)$, we get that the density function, for $0 < x < 1$, is

$$f(x) = \frac{n!}{(k-1)!\; (n-k)!}\;\; x^{k-1}\; (1 - x)^{n-k} \quad \text{if } 0 < x < 1$$

Proof of the formulas of the distribution function. The proof of the first formula is based on the fact that the kth point is on the left side of x if and only if there are k or $k+1$ or $\ldots$ n points on the left side of x. So we use the binomial distribution with parameters n and $p = x$, and summarize its kth, $(k+1)$th $\ldots$ nth terms. The second formula follows from the complementary rule of probability.

Relations between the density and the distribution functions. Since the density function is equal to the derivative of the distribution function, and the distribution function is equal to the integral of the density function, we get the equalities:

$$\frac{d}{dx}\left(\sum_{i=k}^{n} \binom{n}{i} x^i (1-x)^{n-i}\right) = \frac{n!}{(k-1)!(n-k)!}\; x^{k-1}(1-x)^{n-k}$$

and

$$\int_0^x \frac{n!}{(k-1)!(n-k)!}\; x^{k-1}(1-x)^{n-k}\; dx = \sum_{i=k}^{n} \binom{n}{i} x^i (1-x)^{n-i}$$

The first equality can be derived by simple differentiation and then simplification of the terms which cancel out each other. The second can be derived by integration by parts.

Using Excel. In Excel, the function `BETADIST` (in Hungarian: `BÉTA.ELOSZLÁS`) is associated to the distribution function of the beta distribution:

$F(x) =$ `BETADIST(x;k;n-k+1;0;1)`

We may omit the parameters 0 and 1, and write simply

$F(x) =$ `BETADIST(x;k;n-k+1)`

There is no special Excel function for the density function, so if you need the density function, then - studying the mathematical formula of the density function - you yourself may construct it using the Excel functions `COMBIN` and `POWER` (in Hungarian: `KOMBINÁCIÓ` and `HATVÁNY`). In Excel, the inverse of the distribution function `BETADIST(x;k;n-k+1;0;1)` is `BETAINV(x;k;n-k+1;0;1)` (in Hungarian: `INVERZ.BÉTA(x;k;n-k+1;0;1)`).

2. General case: Beta distribution on the interval $[A;B]$

Applications:

1. If n people arrive between the time instants A and B independently of each other according to uniform distribution, and we are interested in the time instant when the kth person arrives, then this arrival time follows the beta distribution on the interval $[A,B]$ related to size n and rank k.
2. If we have n uniformly distributed, independent random values between A and B, and we consider the kth smallest of the generated n values, then this random variable follows the beta distribution on the interval $[A,B]$ related to size n and rank k.

Density function:

$$f(x) = \frac{1}{B-A}\,\frac{n!}{(k-1)!\,(n-k)!}\,\left(\frac{x-A}{B-A}\right)^{k-1}\left(\frac{B-x}{B-A}\right)^{n-k}$$

if $A < x < B$

Distribution function:

$$F(x) = \sum_{i=k}^{n}\binom{n}{i}\left(\frac{x-A}{B-A}\right)^{i}\left(\frac{B-x}{B-A}\right)^{n-i} \quad \text{if} \quad A < x < B$$

or, equivalently

$$F(x) = 1 - \sum_{i=0}^{k-1}\binom{n}{i}\left(\frac{x-A}{B-A}\right)^{i}\left(\frac{B-x}{B-A}\right)^{n-i} \quad \text{if} \quad A < x < B$$

Parameters: k and n are positive integers so that $k \leq n$. A and B are real numbers so that $A < B$. n can be called **the size**, and k **the rank of the distribution**.

The proofs of the above formulas are similar to the special case $A = 0$, $B = 1$, and are left for the reader as an exercise.

Using Excel. In Excel, the function `BETADIST` (in Hungarian: `BÉTA.ELOSZLÁS`) is associated to the beta distribution. The distribution function of the beta distribution on the interval $[A, B]$ related to size n and rank k in Excel is:

$F(x) =$ `BETADIST(x;k;n-k+1;A;B)` $=$

There is no special Excel function for the density function. If you need an Excel formula for the density function, then - studying the mathematical formula of the density function - you yourself may construct it using the Excel functions `COMBIN` and `POWER` (in Hungarian: `KOMBINÁCIÓ` and `HATVÁNY`). In Excel, the inverse of the distribution function `BETADIST(x;k;n-k+1;A;B)` is `BETAINV(x;k;n-k+1;A;B)` (in Hungarian: `INVERZ.BÉTA(x;k;n-k+1;A;B)`).

7 Exponential distribution

Notion of the memoryless property: We say that the random life-time X of an object has the memoryless property if

$$\mathbf{P}(\ X > a + b \mid X > a) = \mathbf{P}(\ X > b\ ) \quad \text{for all positive } a \text{ and } b$$

The meaning of the memoryless property in words: if the object has already lived a units of time, then its chances to live more b units of time is equal to the chance of living b units of time for a brand-new object of this type after its birth. In other words: if the object is still living, then it has the same chances for its future as a brand-new one. We may also say that the memoryless property of an object means that the past of the object does not have an effect on the future of the object.

Example and counter-example for the the memoryless property:

1. The life-time of a mirror hanging on the wall has the memoryless property.
2. The life-time of a tire on a car dos not have the memoryless property.

Applications of the exponential distribution:

1. The life-time of objects having the memoryless property have an exponential distribution.
2. Under certain circumstances waiting times follow an exponential distribution. (We learn about these circumstances in Chapter 20 entitled "Poisson process") For example, the amount of time until the next serious accident in a big city where the traffic has the same intensity day and night continuously follows an exponential distribution.

Density function:

$$f(x) = \lambda\, e^{-\lambda x} \quad \text{if} \quad x \geq 0$$

Distribution function:

$$F(x) = 1 - e^{-\lambda x} \quad \text{if} \quad x \geq 0$$

Parameter: $\lambda > 0$

Remark. We will see later that the reciprocal of the parameter λ shows how much the theoretical average life-time (or the theoretical average waiting time) is.

Remark. In some real-life problems the memory-less property is only approximately fulfilled. In such cases, the the application of the exponential distribution is only an approximate model for the problem.

Proof of the formula of the distribution and density functions. The memoryless property

$$\mathbf{P}(\ X > a + b \mid X > a) = \mathbf{P}(\ X > b\ )$$

can be written like this:

$$\frac{\mathbf{P}(\ X > a + b\ )}{\mathbf{P}(\ X > a)} = \mathbf{P}(\ X > b\ )$$

Using the tail function $T(x) = \mathbf{P}(X > x)$, we may write the equation like this:

$$\frac{T(a+b)}{T(a)} = T(b)$$

that is

$$T(a+b) = T(a)T(b)$$

It is obvious that the exponential functions $T(x) = e^{cx}$ with an arbitrary constant c satisfy this equation. On the contrary, it can be shown that if a function is monotonous and satisfies this equation, then it must be an exponential functions of the form $T(x) = e^{cx}$. Since a tail function is monotonously decreasing, the memoryless property really implies that $T(x) = e^{cx}$ with a negative constant c, so we may write $c = -\lambda$, that is, $T(x) = e^{-\lambda x}$, which means that the distribution function is $F(x) = 1 - T(x) = 1 - e^{-\lambda x}$. Differentiating the distribution function, we get the density function: $f(x) = (F(x))' = \left(1 - e^{-\lambda x}\right)' = \lambda e^{-\lambda x}$.

Using Excel. In Excel, the function `EXPONDIST` (in Hungarian: `EXP.ELOSZLÁS`) is associated to this distribution. If the last parameter is `FALSE`, we get the density function of the exponential distribution:

$$f(x) = \lambda\, e^{-\lambda x} = \texttt{EXPONDIST(x;n;}\lambda\texttt{;FALSE)}$$

If the last parameter is `TRUE`, and the third parameter is the reciprocal of λ, we get the distribution function of the exponential distribution:

$$F(x) = 1 - e^{-\lambda x} = \texttt{EXPONDIST(x;n;}\lambda\texttt{;TRUE)}$$

You may use also the function `EXP` (in Hungarian: `KITEVŐ`) like this:

$$f(x) = \lambda\, e^{-\lambda x} = \lambda\ \texttt{EXP(-}\lambda\texttt{x)}$$

The distribution function then looks like as this:

$$F(x) = 1 - e^{-\lambda x} = 1 - \texttt{EXP(-}\lambda\texttt{x)}$$

8 Gamma distribution

Application: In a big city where the traffic has the same intensity days and nights continuously, the amount of time until the nth serious accident follows a gamma distribution.

Density function:

$$f(x) = \frac{\lambda^n x^{n-1}}{(n-1)!}\, e^{-\lambda x} \quad \text{if} \quad x \geq 0$$

Distribution function:

$$F(x) = 1 - \sum_{i=0}^{n-1} \frac{(\lambda x)^i}{i!} e^{-\lambda x} \quad \text{if} \quad x \geq 0$$

Parameters: n is a positive integer, and $\lambda > 0$

Remark. The reciprocal of the parameter λ shows how much the theoretical average waiting time for the first accident is. Thus, the reciprocal of the parameter λ multiplied by n shows how much the theoretical average waiting time for the nth accident is.

The proof of the formulas of the density and distribution functions is omitted here. The formulas can be derived after having learned about Poisson-processes in Chapter 20.

Remark. If $n = 1$, then the gamma distribution reduces to the exponential distribution.

Using Excel. In Excel, the function `GAMMADIST` (in Hungarian: `GAMMA.ELOSZLÁS`) is associated to this distribution. If the last parameter is `FALSE` and the third parameter is the reciprocal of λ (unfortunately, in the `GAMMADIST` function of Excel, the third parameter should be the reciprocal of λ, and not λ), then we get the density function of the gamma distribution with parameters n and λ:

$$f(x) = \frac{\lambda^n x^{n-1}}{(n-1)!}\, e^{-\lambda x} = \texttt{GAMMADIST(x;n;}\frac{1}{\lambda}\texttt{;FALSE)}$$

If the last parameter is `TRUE`, and the third parameter is the reciprocal of λ, then we get the distribution function of the gamma distribution with parameters λ and λ:

$$F(x) = \int_0^x \frac{\lambda^n x^{n-1}}{(n-1)!}\, e^{-\lambda x}\, dx = \texttt{GAMMADIST(x;n;}\frac{1}{\lambda}\texttt{;TRUE)}$$

Using Excel. If $n = 1$, then the Excel function GAMMADIST returns the exponential distribution. This means that, with the `FALSE` option

$$\texttt{GAMMADIST(x;1;}\frac{1}{\lambda}\texttt{;FALSE)} = \lambda\, e^{-\lambda x}$$

is the exponential density function with parameter λ, and, with the `TRUE` option

$$\texttt{GAMMADIST(x;1;}\frac{1}{\lambda}\texttt{;TRUE)} = 1 - e^{-\lambda x}$$

is the exponential distribution function with parameter λ.

9 Normal distributions

Application: If a random variable can be represented as the sum of many, independent, random quantities so that each has a small standard deviation, specifically, the sum of many, independent, small random quantities, then this random variable follows a normal distribution with some parameters μ and σ. Such random variable are, for example, the amount of electricity used by the inhabitants of a town, or the total income of a shop during a day.

1. Special case: Standard normal distribution

Density function:

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \; e^{-\frac{x^2}{2}} \quad \text{if } -\infty < x < \infty$$

Distribution function:

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \; e^{-\frac{x^2}{2}} \; dx \quad \text{if } -\infty < x < \infty$$

Remark: The usage of the Greek letters φ and Φ for the density and distribution functions of the standard normal distribution is so wide-spread as, for example, sin and cos for the sine- and cosine-functions. The symmetry of the density function about the origin φ implies the equality:

$$\int_{-\infty}^{-x} \frac{1}{\sqrt{2\pi}} \; e^{-\frac{x^2}{2}} \; dx = \int_{x}^{\infty} \frac{1}{\sqrt{2\pi}} \; e^{-\frac{x^2}{2}} \; dx$$

that is

$$\Phi(-x) = 1 - \Phi(x)$$

Since $\Phi(x)$ is a strictly increasing function of x, its inverse exists. It is denoted by $\Phi^{-1}(y)$.

Using Excel. In Excel, the function `NORMDIST` (in Hungarian: `NORM.ELOSZL`) with the special parameter values 0 and 1 corresponds to this distribution. If the last parameter is `FALSE`, then we get the density function of the normal distribution:

`NORMDIST(x;0;1;FALSE)` $= \varphi(x)$

If the last parameter is `TRUE`, then we get the distribution function of the normal distribution:

`NORMDIST(x;0;1;TRUE)` $= \Phi(x)$

Since the function $\Phi(x)$ plays a very important role, there is a special simple Excel function for $\Phi(x)$, namely, `NORMSDIST` (in Hungarian: `STNORMELOSZL`), which has only one variable and no parameters. You may remember that the letter `S` stands (in Hungarian: the letters `ST` stand) for "standard":

$$\texttt{NORMSDIST(x)} = \Phi(x)$$

In Excel, the notation for the inverse of the function Φ is `NORMSINV` or `NORMINV` (in Hungarian: `INVERZ.STNORM`, or `INVERZ.NORM`) with the special parameter values $(\mu =)$ 0 and $(\sigma =)$ 1

$$\texttt{NORMSINV(y)} = \texttt{NORMINV(y;0;1)} = \Phi^{-1}(y)$$

2. General case: Normal distribution with parameters μ and σ

Density function:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \; e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} \quad \text{if } -\infty < x < \infty$$

Distribution function:

$$F(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \; e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} \; dx \quad \text{if } -\infty < x < \infty$$

Parameters: μ may be any real number, and σ may be any positive number.

Remark. We will see later that the parameters μ and σ are the expected value and the standard deviation, respectively. The normal distribution with parameters $\mu = 0$ and $\sigma = 1$ is the standard normal distribution.

The distribution function of the normal distribution with parameters μ and σ can be expressed in terms of the standard normal distribution function:

$$F(x) = \Phi(\frac{x-\mu}{\sigma}) \qquad \text{if } -\infty < x < \infty$$

Here $\frac{x-\mu}{\sigma}$ is called the z**-value** associated with x, or the **standardized value** of x. The transformation $y = \frac{x-\mu}{\sigma}$ is called **standardization**.

If a random variable X follows a normal distribution with parameters μ and σ, then the following rules, called 1**-sigma rule**, 2**-sigma rule**, 3**-sigma rule**, are true:

$$\begin{array}{rcccl} \mathbf{P}(\mu-\sigma < X < \mu+\sigma) & \approx & 0.68 & = & 68\% \\ \mathbf{P}(\mu-2\sigma < X < \mu+2\sigma) & \approx & 0.95 & = & 95\% \\ \mathbf{P}(\mu-3\sigma < X < \mu+3\sigma) & \approx & 0.997 & = & 99.7\% \end{array}$$

These rules, in words, sound like this:

1**-sigma rule**: the value of a normally distributed random variable X falls in the interval $(\mu - \sigma, \mu + \sigma)$ with a probability 0.68.

2**-sigma rule**: the value of a normally distributed random variable X falls in the interval $(\mu - 2\sigma, \mu + 2\sigma)$ with a probability 0.95.

3**-sigma rule**: the value of a normally distributed random variable X falls in the interval $(\mu - 3\sigma, \mu + 3\sigma)$ with a probability 0.997.

Using Excel. In Excel, the function `NORMDIST` (in Hungarian: `NORM.ELOSZL`) corresponds to this distribution. If the last parameter is `FALSE`, then we get the density function of the normal distribution with parameters μ and σ:

$$\texttt{NORMDIST(x;}\mu\texttt{;}\sigma\texttt{;FALSE)} = f(x)$$

If the last parameter is `TRUE`, then we get the distribution function of the normal distribution with parameters μ and σ:

$$\texttt{NORMDIST(x;}\mu\texttt{;}\sigma\texttt{;TRUE)} = F(x)$$

In Excel, the notation for the inverse $F^{-1}(y)$ of the distribution function $F(x)$ of the normal distribution with parameters μ and σ is `NORMINV` (in Hungarian: `INVERZ.NORM`):

$$F^{-1}(y) = \texttt{NORMINV(y;}\mu\texttt{;}\sigma\texttt{)}$$

10 Generating a random variable with a given continuous distribution

It is important for a given continuous distribution that a random variable can be generated by a calculator or a computer so that its distribution is the given continuous distribution. The following is a method to define such a random variable.

Assume that a continuous distribution function $F(x)$ is given so that the function $F(x)$ is strictly increasing on an interval (A, B), and it is 0 on the left side of A, and it is 1 on the right side of B. If either $A = -\infty$ or $B = +\infty$, then we do not have to even mention them. The restriction of $F(x)$ onto the interval (A, B) has an inverse $F^{-1}(u)$ which is defined for all $u \in [0, 1]$. The way how we find a formula for $F^{-1}(u)$ is that we solve the equation

$$u = F(x)$$

for x, that is, we express x from the equation in terms of u:

$$x = F^{-1}(u)$$

We may consider the random variable X defined by

$$X = F^{-1}(\text{RND})$$

It is easy to be convinced that the distribution function of the random variable X is the given function $F(x)$.

Some examples:

1. **Uniform random variable on** $(A; B)$

   Distribution function:

   $$u = F(x) = \frac{x - A}{B - A} \quad \text{if } A < x < B$$

   Inverse of the distribution function:

   $$x = F^{-1}(u) = A + (B - A)u \quad \text{if } 0 < u < 1$$

   Simulation:

   $$X = A + (B - A)\text{RND}$$

2. **Exponential random variable with parameter** λ

   Distribution function:

   $$y = F(x) = 1 - e^{-\lambda x} \quad \text{if } x \geq 0$$

Inverse of the distribution function:

$$x = F^{-1}(u) = -\frac{\ln(1-u)}{\lambda} \quad \text{if } 0 < u < 1$$

Simulations:

$$X = -\frac{\ln(1 - \text{RND})}{\lambda}$$

Obviously, the substraction from 1 can be omitted, and

$$X = -\frac{\ln(\text{RND})}{\lambda}$$

is also exponentially distributed.

For some distributions, an explicit formula for the inverse of the distribution function is not available. In such cases, the simulation may be based on other relations. We give some examples for this case:

3. **Gamma random variable with parameters n and λ**

   Simulation:

$$X = \sum_{i=1}^{n} X_i^0 = X_1^0 + X_2^0 + ... + X_n^0$$

   where $X_1^0, X_2^0, ..., X_n^0$ are independent, exponentially distributed random variables with parameter λ.

4. **Standard normal random variable**

   Simulations:

$$X = \texttt{NORMSINV(}\text{RND}\texttt{)}$$

$$X = \texttt{NORMINV(}\text{RND}\texttt{;0;1)}$$

   Accept that the following simulations are correct, and that they are more efficient:

$$X = \left(\sum_{i=1}^{12} \text{RND}_i \right) - 6$$

   where $\text{RND}_1, \text{RND}_2, ..., \text{RND}_{12}$ are independent random variables, uniformly distributed between 0 and 1.

$$X = \sqrt{(2\ln(\text{RND}_1))}\ \cos(\text{RND}_2)$$

   where $\text{RND}_1, \text{RND}_2$ are independent random variables, uniformly distributed between 0 and 1.

5. **Normal random variable with parameters μ and σ**
   Simulations:

   $$X = \texttt{NORMINV(}\text{RND}\texttt{;}\mu;\sigma\texttt{)}$$

   $$X = \mu + \sigma X^0$$

   where X^0 is a standard normal random variable.

11 Expected value of continuous distributions

In Chapter 15 of Part II. we learned that, by performing a large number of experiments for a discrete random variable, the average of the experimental results $X_1, X_2, \ldots, X_N$ stabilizes around the expected value of X:

$$\frac{X_1 + X_2 + \ldots + X_N}{N} \approx \mathbf{E}(X)$$

The same stabilization rule is true in case of a continuous random variable. In this chapter, we define the notion of the expected value for continuous distributions, and we list the formulas of the expected values of the most important continuous distributions.

The definition of the **expected value** for continuous distributions is:

$$\mathbf{E}(X) = \int_{-\infty}^{\infty} x f(x)\, dx$$

Remark. We shall give here some motivation for the declared formula of the expected value. For this purpose, let us take a continuous random variable X, and let $X_1, X_2, \ldots, X_N$ be experimental results for X. We will show that the average of the experimental results is close to the above integral:

$$\frac{X_1 + X_2 + \ldots + X_N}{N} \approx \int_{-\infty}^{\infty} x f(x)\, dx$$

In order to show this, we choose the fixed points $\ldots, y_i, y_{i+1}, \ldots$ on the real line so that all the differences $\Delta y_i = y_{i+1} - y_i$ are small. Then we introduce a discrete random variable Y, so that the value of Y is derived from the value of X by rounding down to the closest y_i value which is on the left side of X, that is,

$$Y = y_i \quad \text{if and only if} \quad y_i \le X < y_{i+1}$$

Applying the rounding operation to each experimental result, we get the values $Y_1, Y_2, \ldots, Y_N$. Since all the differences $\Delta y_i = y_{i+1} - y_i$ are small, we have that

$$\frac{X_1 + X_2 + \ldots + X_N}{N} \approx \frac{Y_1 + Y_2 + \ldots + Y_N}{N}$$

Obviously, Y is a discrete random variable with the possible values $\ldots, y_i, \ldots$, so that the probability of y_i is

$$p_i = \int_{y_i}^{y_{i+1}} f(x)\, dx \approx f(y_i)\Delta y_i$$

and thus, the expected value of Y is

$$\sum_i y_i\, p_i = \sum_i y_i \int_{yi}^{yi+1} f(x)\, dx \approx \sum_i y_i\, f(y_i)\Delta y_i \approx \int_{-\infty}^{\infty} x f(x)\, dx$$

We know that the average of the experimental results of a discrete random variable is close to the expected value, so

$$\frac{Y_1 + Y_2 + \ldots + Y_N}{N} \approx \sum_i y_i \, p_i$$

From all these approximations we get that

$$\frac{X_1 + X_2 + \ldots + X_N}{N} \approx \int_{-\infty}^{\infty} x f(x) \, dx$$

Remark. It may happen that the expected value does not exist! If the integral

$$\int_{\infty}^{\infty} x f(x) \, dx$$

is not absolutely convergent, that is

$$\int_{\infty}^{\infty} |x| f(x) \, dx = \infty$$

then one of the following 3 cases holds:

1. Either

$$\int_{0}^{\infty} x f(x) \, dx = \infty \quad \text{and} \quad \int_{-\infty}^{0} x f(x) \, dx > -\infty$$

2. or

$$\int_{0}^{\infty} x f(x) \, dx < \infty \quad \text{and} \quad \int_{-\infty}^{0} x f(x) \, dx = -\infty$$

3. or

$$\int_{0}^{\infty} x f(x) \, dx = \infty \quad \text{and} \quad \int_{-\infty}^{0} x f(x) \, dx = -\infty$$

It can be shown that, in the first case, as N increases

$$\frac{X_1 + X_2 + \ldots + X_N}{N}$$

will become larger and larger, and it approaches ∞. This is why we may say that the expected exists, and its value is ∞. In the second case, as N increases,

$$\frac{X_1 + X_2 + \ldots + X_N}{N}$$

will become smaller and smaller, and it approaches $-\infty$. This is why we may say that the expected exists, and its value is $-\infty$. In the third case, as N increases,

$$\frac{X_1 + X_2 + \ldots + X_N}{N}$$

does not approach to any finite or infinite value. In this case we say that the expected value does not exist.

Here we give a list of the formulas of the expected values of the most important continuous distributions. The proofs are given after the list.

1. **Uniform distribution on an interval** $(A;B)$

$$\mathbf{E}(X) = \frac{A+B}{2}$$

2. **Arc-sine distribution**

$$\mathbf{E}(X) = 0$$

3. **Cauchy distribution**

   The expected value does not exist.

4. **Beta distribution related to size** n **and** k

$$\mathbf{E}(X) = \frac{k}{n+1}$$

5. **Exponential distribution with parameter** λ

$$\mathbf{E}(X) = \frac{1}{\lambda}$$

6. **Gamma distribution of order** n **with parameter** λ

$$\mathbf{E}(X) = \frac{n}{\lambda}$$

7. **Normal distribution with parameters** μ **and** σ

$$\mathbf{E}(X) = \mu$$

Proofs.

1. **Uniform distribution on an interval** $(A; B)$**.** Since the distribution is concentrated on a finite interval, the expected value exists. Since the density function is symmetrical about $\frac{A+B}{2}$, the expected value is

$$\mathbf{E}(X) = \frac{A+B}{2}$$

We may get this result by calculation, too:

$$\mathbf{E}(X) = \int_{-\infty}^{\infty} x f(x)\, dx = \int_{A}^{B} x\, \frac{1}{B-A}\, dx =$$

$$\left[\frac{x^2}{2}\right]_A^B \frac{1}{B-A}\, dx = \left[\frac{B^2-A^2}{2}\right] \frac{1}{B-A}\, dx = \frac{A+B}{2}$$

2. **Arc-sine distribution.** Since the distribution is concentrated on the interval $(-1, 1)$, the expected value exists. Since the density function is symmetrical about 0, the expected value is

$$\mathbf{E}(X) = 0$$

3. **Cauchy distribution.** Since the density function is symmetrical about 0, the 0 is a candidate for being the expected value. However, since

$$\int_0^{\infty} x f(x)\, dx = \int_0^{\infty} x\, \frac{1}{\pi} \frac{1}{1+x^2}\, dx = \left[\frac{1}{2\pi} \ln(1+x^2)\right]_0^{\infty} = \infty$$

and

$$\int_{-\infty}^{0} x f(x)\, dx = -\infty$$

the expected value does not exist.

4. **Beta distribution related to size** n **and** k

$$\mathbf{E}(X) = \int_{-\infty}^{\infty} x f(x)\, dx =$$

$$\int_0^1 x\, \frac{n!}{(k-1)!(n-k)!}\, x^{k-1}(1-x)^{n-k}\, dx =$$

$$\frac{k}{n+1} \int_0^1 \frac{(n+1)!}{k!(n-k)!}\, x^k(1-x)^{n-k}\, dx = \frac{k}{n+1}$$

In the last step, we used the fact that

$$\int_0^1 \frac{(n+1)!}{k!(n-k)!}\, x^k(1-x)^{n-k}\, dx = 1$$

which follows from the fact that

$$\frac{(n+1)!}{k!(n-k)!}\; x^k(1-x)^{n-k}$$

that is

$$\frac{(n+1)!}{k!(n-k)!}\; x^k(1-x)^{(n+1)-(k+1)}$$

is a the density function of the beta distribution related to size $n+1$ and $k+1$.

5. **Exponential distribution with parameter λ.** Using integration by parts with $u = x$, $v' = \lambda\, e^{-\lambda x}$, $u' = 1$, $v = -e^{-\lambda x}$, we get that

$$\mathbf{E}(X) = \int_{-\infty}^{\infty} x f(x)\, dx = \int_0^{\infty} x\;\lambda\; e^{-\lambda x}\; dx =$$

$$\left[x\;\left(-e^{-\lambda x}\right)\right]_0^{\infty} - \int_0^{\infty} 1\;\;\left(-e^{-\lambda x}\right)\; dx =$$

$$0 + \int_0^{\infty} e^{-\lambda x}\; dx = \int_0^{\infty} e^{-\lambda x}\; dx = \left[\frac{e^{-\lambda x}}{-\lambda}\right]_0^{\infty} = \frac{1}{\lambda}$$

6. **Gamma distribution of order n with parameter λ**

$$\mathbf{E}(X) = \int_{-\infty}^{\infty} x f(x)\, dx =$$

$$\int_0^{\infty} x\; \frac{x^{n-1}\lambda^n}{(n-1)!}\; e^{-\lambda x}\; dx =$$

$$\frac{n}{\lambda}\;\int_0^{\infty} \frac{x^n\;\lambda^{n+1}}{n!}\; e^{-\lambda x}\; dx = \frac{n}{\lambda}$$

In the last step, we used the fact that

$$\int_0^{\infty} \frac{x^n\;\lambda^{n+1}}{n!}\; e^{-\lambda x}\; dx = 1$$

This follows from the fact that

$$\frac{x^n\;\lambda^{n+1}}{n!}\; e^{-\lambda x}$$

is a density function of the gamma distribution of order $n+1$ with parameter λ.

7. **Normal distribution with parameters μ and σ.** Since the improper integrals

$$\int_0^{\infty} x f(x)\, dx = \int_0^{\infty} x\;\; \frac{1}{\sqrt{2\pi}}\;\; e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}\; dx$$

$$\int_{-\infty}^{0} xf(x)\,dx = \int_{-\infty}^{0} x\ \frac{1}{\sqrt{2\pi}}\ e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}\ dx$$

are obviously convergent, the expected value exists. Since the density function is symmetrical about μ, the expected value is

$$\mathbf{E}(X) = \mu$$

Minimal property of the expected value. If X is a continuous random variable with the density function $f(x)$, and c is a constant, then distance between X and c is $|X-c|$, the distance squared is $(X-c)^2$, the expected value of the squared distance is

$$\mathbf{E}\left((X-c)^2\right) = \int_{-\infty}^{\infty} (x-c)^2\ f(x)\,dx$$

This integral is minimal if c is the expected value of X.

Proof. The value of the integral depends on c, so the integral defines a function:

$$h(c) = \int_{-\infty}^{\infty} (x-c)^2\ f(x)\,dx$$

Expanding the square, we get:

$$h(c) = \int_{-\infty}^{\infty} x^2\ f(x)\,dx - \int_{-\infty}^{\infty} 2xc\ f(x)\,dx + \int_{-\infty}^{\infty} c^2\ f(x)\,dx =$$

$$\int_{-\infty}^{\infty} x^2\ f(x)\,dx - 2c\ \int_{-\infty}^{\infty} x\ f(x)\,dx + c^2\ \int_{-\infty}^{\infty} 1\ f(x)\,dx =$$

$$\int_{-\infty}^{\infty} x^2\ f(x)\,dx - 2c\,\mathbf{E}(X) + c^2$$

Since the integral in the last line does not depend on c, differentiating with respect to c, we get that

$$h'(c) = -2\,\mathbf{E}(X) + 2c$$

Equating the derivative to 0, we get that the minimum occurs at $c = \mathbf{E}(X)$.

12 Expected value of a function of a continuous random variable

We leared in Chapter 17 of Part II that if we make N experiments for a discrete random variable X, and we substitute the experimental results $X_1, X_2, \ldots, X_N$ into the function $y = t(x)$, and we consider the values $t(X_1), t(X_2), \ldots, t(X_N)$, then their average is is close to the expected value of $t(X)$:

$$\frac{t(X_1) + t(X_2) + \ldots + t(X_N)}{N} \approx \mathbf{E}\left(t(X)\right)$$

The same stabilization rule is true in the case of a continuous random variable. Let X be a continuous random variable, and $t(x)$ a continuous function. The expected value of the random variable $t(X)$ is calculated by the integral:

$$\mathbf{E}\left(t(X)\right) = \int_{-\infty}^{\infty} t(x) f(x)\, dx$$

Motivation of the declared formula. We give here some motivation of the declared formula of the expected value of $t(X)$. For this purpose, let us take a continuous random variable X, and a continuous function $t(x)$, and and let $X_1, X_2, \ldots, X_N$ be the experimental results for X. We will show that the average of the function values of the experimental results is close to the above integral:

$$\frac{t(X_1) + X_2 + \ldots + t(X_N)}{N} \approx \int_{-\infty}^{\infty} t(x) f(x)\, dx$$

In order to show this, we choose the fixed points $\ldots, y_i, y_{i+1}, \ldots$ on the real line so that all the differences $\Delta y_i = y_{i+1} - y_i$ are small. Then we introduce a discrete random variable, so that the value of Y is derived from the value of X by rounding down to the closest y_i value which is on the left side of X, that is,

$$Y = y_i \quad \text{if and only if} \quad y_i \le X < y_{i+1}$$

Applying the rounding operation to each experimental result, we get the values

$$Y_1, Y_2, \ldots, Y_N$$

Since all the differences $\Delta y_i = y_{i+1} - y_i$ are small, and the function $t(x)$ is continuous, we have that

$$\frac{t(X_1) + t(X_2) + \ldots + t(X_N)}{N} \approx \frac{t(Y_1) + t(Y_2) + \ldots + t(Y_N)}{N}$$

Obviously, Y is a discrete random variable with the possible values $\ldots, y_i \ldots$, so that the probability of y_i is

$$p_i = \int_{y_i}^{y_{i+1}} f(x)\, dx \approx f(y_i) \Delta y_i$$

and thus, the expected value of $t(Y)$ is

$$\sum_i t(y_i)\; p_i = \sum_i t(y_i) \int_{yi}^{yi+1} f(x)\; dx \approx \sum_i t(y_i)\; f(y_i)\Delta y_i \approx \int_{-\infty}^{\infty} t(x)f(x)\; dx$$

We know that the average of the function values of the experimental results of a discrete random variable is close to its expected value, so

$$\frac{t(Y_1) + t(Y_2) + \ldots + t(Y_N)}{N} \approx \sum_i t(y_i)\; p_i$$

From all these approximations we get that

$$\frac{t(X_1) + t(X_2) + \ldots + t(X_N)}{N} \approx \int_{-\infty}^{\infty} t(x)f(x)\; dx$$

The expected value of X^n is called the n**th moment** of X:

$$\mathbf{E}\left(X^n\right) = \int_{-\infty}^{\infty} x^n f(x)\; dx$$

specifically, the **second moment** of X is:

$$\mathbf{E}\left(X^2\right) = \int_{-\infty}^{\infty} x^2 f(x)\; dx$$

The expected value of $(X - c)^n$ is called the n**th moment** about a the point c:

$$\mathbf{E}\left((X - c)^n\right) = \int_{-\infty}^{\infty} (x - c)^n f(x)\; dx$$

specifically, the **second moment** about a point c is:

$$\mathbf{E}\left((X - c)^2\right) = \int_{-\infty}^{\infty} (x - c)^2 f(x)\; dx$$

Second moment of some continuous distributions:

1. **Uniform distribution on an interval** $(A; B)$

$$\mathbf{E}\left(X^2\right) = \frac{A^2 + AB + B^2}{3}$$

2. **Exponential distribution**

$$\mathbf{E}\left(X^2\right) = \frac{2}{\lambda^2}$$

Proofs.

1.

$$\mathbf{E}(X^2) = \int_{-\infty}^{\infty} x^2 \; f(x) \; dx = \int_A^B x^2 \; \frac{1}{B-A} \; dx =$$

$$\frac{1}{B-A} \; \left[\frac{x^3}{3}\right]_A^B = \frac{1}{B-A} \; \frac{B^3 - A^3}{3} = \frac{A^2 + AB + B^2}{3}$$

2. Using integration by parts with $u = x^2$, $v' = \lambda \; e^{-\lambda x}$, $(u^2)' = 2u$, $v = -e^{-\lambda x}$, we get that

$$\mathbf{E}(X^2) = \int_{-\infty}^{\infty} x^2 \; f(x) \; dx = \int_0^{\infty} x^2 \quad \lambda \; e^{-\lambda x} \; dx =$$

$$\left[x^2 \; \left(-e^{-\lambda x}\right)\right]_0^{\infty} - \int_0^{\infty} 2x \quad \left(-e^{-\lambda x}\right) \; dx =$$

$$0 + 2 \int_0^{\infty} x \; e^{-\lambda x} \; dx =$$

$$2 \; \frac{1}{\lambda} \; \int_0^{\infty} x \;\; \lambda \; e^{-\lambda x} \; dx =$$

Here we recognize that the integral in the last line is the expected value of the λ-parametric exponential distribution, which is equal to $\frac{1}{\lambda}$, so we get

$$2 \; \frac{1}{\lambda} \; \frac{1}{\lambda} = \frac{2}{\lambda^2}$$

as it was stated.

13 Median

In this chapter, we learn about the notion of the median, which is a kind of a "center" of a data-set or of a distribution. In the next chapter, we will learn the notion of the expected value also for continuous random variables and distributions, which is a kind of "center", too, and then we will be able to compare them.

If a data-set consists of n numbers, then we may find
the smallest of these numbers, let us denote it by z_1^*,
the second smallest, let us denote it by z_2^*,
the third smallest, let us denote it by z_3^*,
and so on,
the kth smallest, let us denote it by z_k^*,
and so on,
the nth smallest, which is actually the largest, let us denote it by z_n^*.

Using Excel. In Excel, for a data-set, the function `SMALL` (in Hungarian: `KICSI`) can be used to find the kth smallest element in an `array`:

$z_k^* =$ `SMALL(array;k)`

Now we may arrange the numbers $z_1, z_2, \ldots, z_n$ in the increasing order: $z_1^*, z_2^*, \ldots, z_n^*$. If the number n is odd, then there is a well defined center element in the list $z_1^*, z_2^*, \ldots, z_n^*$. This center element is called the **median of the data-set**. If n is even, then there are two center elements. In this case, the average of these two center elements is the **median of the data-set**.

Using Excel. In Excel, for a data-set, the function `MEDIAN` (in Hungarian: `MEDIÁN`) is used to calculate the median of a data-set:

`MEDIAN(array)`

A **median** of a continuous random variable or distribution is a value c for which it is true that both the probability of being less than c and the probability of being greater than c are equal to 0.5:

$$P\left((-\infty, c)\right) = 0.5$$

$$P\left((c, \infty)\right) = 0.5$$

that is a median is a solution to the equation

$$F(x) = 0.5$$

If the distribution function $F(x)$ is strictly increasing, which practically holds for all the important distributions, then the median can be given explicitly in terms of the inverse of the distribution function:

$$c = F^{-1}(0.5)$$

Using the density function, the median can be characterized obviously by the property

$$\int_{-\infty}^{c} f(x)\, dx = 0.5$$

or, equivalently,

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

The notion of the median can be defined for discrete distributions, too, but the definition is a little bit more complicated. The **median** of a discrete random variable or distribution is any value c for which it is true that both the probability of being less than$c$, and the probability of being greater than c are at most 0.5:

$$P\left((-\infty, c)\right) \leq 0.5$$

$$P\left((c, \infty)\right) \leq 0.5$$

In a long sequence of experiments, the median of the experimental results for a random variable stabilizes around the median of the distribution of the random variable: if $X_1, X_2, \ldots, X_N$ are experimental results for a random variable X, and N is large, then the median of the data-set $X_1, X_2, \ldots, X_N$, the so called experimental median is close to the median of the distribution of the random variable.

Minimal property of the median. If X is continuous random variable with the density function $f(x)$, and c is a constant, then the expected value of the distance between X and c is

$$\mathbf{E}(|X - c|) = \int_{-\infty}^{\infty} |x - c|\; f(x)\, dx$$

This integral is minimal if c is the median.

Proof. Let us denote the value of the integral, which depends on c, by $h(c)$

$$h(c) = \int_{-\infty}^{\infty} |x - c|\; f(x)\, dx =$$

$$\int_{-\infty}^{c} |x-c| \; f(x)\, dx + \int_{c}^{\infty} |x-c| \; f(x)\, dx =$$

$$\int_{-\infty}^{c} (x-c) \; f(x)\, dx + \int_{c}^{\infty} (c-x) \; f(x)\, dx =$$

$$\int_{-\infty}^{c} x \; f(x)\, dx - c \int_{-\infty}^{c} 1 \; f(x)\, dx + c \int_{c}^{\infty} 1 \; f(x)\, dx - \int_{c}^{\infty} x \; f(x)\, dx$$

Let us take the derivative of each term with respect to c:

$$\left(\int_{-\infty}^{c} x \; f(x)\, dx \right)' = c \; f(c)$$

$$\left(- c \int_{-\infty}^{c} \; f(x)\, dx \right)' = -1 \int_{-\infty}^{c} 1 \; f(x)\, dx - c \; f(c) = -F(x) - c \; f(c)$$

$$\left(c \int_{c}^{\infty} 1 \; f(x)\, dx \right)' = 1 \int_{c}^{\infty} 1 \; f(x)\, dx - c \; f(c) = 1 - F(x) - c \; f(c)$$

$$\left(- \int_{c}^{\infty} x \; f(x)\, dx \right)' = c \; f(c)$$

Now adding the 6 terms on the right sides, the terms $c \; f(c)$ cancel each other, and what we get is

$$h'(c) = 1 - 2F(c)$$

Since the equation

$$1 - 2F(c) = 0$$

is equivalent to the equation

$$F(c) = 0.5$$

and the solution to this equation is the median, we get that

$$h'(c) = 1 - 2F(c) = 0 \quad \text{if } \; c = \text{median}$$

$$h'(c) = 1 - 2F(c) < 0 \quad \text{if } \; c < \text{median}$$

$$h'(c) = 1 - 2F(c) > 0 \quad \text{if } \; c > \text{median}$$

which means that the minimum of $h(c)$ occurs if $c =$ median.

14 Standard deviation, etc.

Both the median and the average define a kind of "center" for a data-set, or for a distribution. It is important to have some characteristics to measure the deviation from the center for a data-set and for a distribution. In this chapter, we shall learn such characteristics.

If $z_1, z_2, \ldots, z_N$ is a data-set, consisting of numbers, then their average is a well-known characteristic of the data-set, which will be denoted by $\overline{z}_N$ or, for simplicity, by $\overline{z}$:

$$\overline{z}_N = \overline{z} = \frac{z_1 + z_2 + \ldots + z_N}{N}$$

The average shows where the center of the data-set is.

It is important for us to know how far the data are from the average. This is why we consider the distance (that is, the absolute value of the difference) between the data elements and the average:

$$|z_1 - \overline{z}| \ , \ \ |z_2 - \overline{z}| \ , \ \ \ldots \ \ , \ |z_N - \overline{z}|$$

The average of these distances is a characteristic of how far the data elements, in the average, are from their average:

$$\frac{|z_1 - \overline{z}| + |z_2 - \overline{z}| + \ldots + |z_N - \overline{z}|}{N}$$

This quantity is called the **average distance from the average**.

Using Excel. In Excel, for a data-set, the function `AVEDEV` (in Hungarian: `ÁTL.ELTÉRÉS`) calculates the average distance from the average.

If, before taking the average, instead of the absolute value, we take the square of each difference, we get another characteristic of the data-set, the average squared distance from the average, which is called the **variance** of the data-set:

$$\frac{(z_1 - \overline{z})^2 + (z_2 - \overline{z})^2 + \ldots + (z_N - \overline{z})^2}{N}$$

The square root of the variance is called the **standard deviation** of the data-set:

$$\sqrt{\frac{(z_1 - \overline{z})^2 + (z_2 - \overline{z})^2 + \ldots + (z_N - \overline{z})^2}{N}}$$

Using a calculator. Most calculators have a key to determine not only the average of a data-set, but the average distance from the average, the variance and the standard deviation, as well.

Using Excel. In Excel, for a data-set, the function VARP (in Hungarian: VARP, too) calculates the variance, and the function STDEVP (in Hungarian: SZÓRÁSP) calculates the standard deviation.

Sample variance and sample standard deviation in Excel. In Excel, the functions VAR (in Hungarian: VAR, too) and STDEV (in Hungarian: SZÓRÁS) calculate the so called **sample variance** and **sample standard deviation**. The sample variance and sample standard deviation are defined almost the same way as the variance and standard deviation, but the denominator is $N-1$ instead of N:

$$\frac{(z_1-\overline{z})^2+(z_2-\overline{z})^2+\ldots+(z_N-\overline{z})^2}{N-1}$$

$$\sqrt{\frac{(z_1-\overline{z})^2+(z_2-\overline{z})^2+\ldots+(z_N-\overline{z})^2}{N-1}}$$

The advantage of taking $N-1$ instead of N becomes clear in statistics. We will not use the functions VAR and STDEV.

Recall that if we make a large number of experiments for a random variable X, then the average of the experimental results, in most cases, stabilizes around a non-random value, the expected value of the random variable, which we denote by μ:

$$\overline{X}=\frac{X_1+X_2++X_N}{N}\approx$$

$$\mu=\begin{cases}\sum_x xp(x) & \text{in the discrete case}\\ \int_{-\infty}^{\infty} x\,p(x) & \text{in the continuous case}\end{cases}$$

The average distance from the average of the experimental results, long sequence of experiments, also stabilizes around a non-random value, which we call the **average distance from the average** of the random variable or of the distribution, which we denote by d:

$$\frac{|X_1-\overline{X}|+|X_2-\overline{X}|+\ldots+|X_N-\overline{X}|}{N}\approx$$

$$\frac{|X_1-\mu|+|X_2-\mu|+\ldots+|X_N-\mu|}{N}\approx$$

$$d=\begin{cases}\sum_x |x-\mu|\;p(x) & \text{in the discrete case}\\ \int_{-\infty}^{\infty} |x-\mu|\;f(x)\;dx & \text{in the continuous case}\end{cases}$$

The variance of the experimental results, in a long sequence of experiments, also stabilizes around a non-random value, which we call the **variance** of the random variable or of the distribution, which we denote by σ^2:

$$\frac{(X_1-\overline{X})^2+(X_2-\overline{X})^2+\ldots+(X_N-\overline{X})^2}{N}\approx$$

$$\frac{(X_1-\mu)^2+(X_2-\mu)^2+\ldots+(X_N-\mu)^2}{N} \approx$$

$$\textbf{VAR}(X)=\sigma^2=\begin{cases} \sum\limits_x (x-\mu)^2\ p(x) & \text{in the discrete case} \\ \int_{-\infty}^{\infty} (x-\mu)^2\ f(x)\ dx & \text{in the continuous case} \end{cases}$$

The standard deviation of the experimental results, which is the square root of the variance, in a long sequence of experiments, obviously stabilizes around the square root of σ^2, that is, σ, which we call the **standard deviation** of the random variable or of the distribution, which we denote by σ:

$$\sqrt{\frac{(X_1-\overline{X})^2+(X_2-\overline{X})^2+\ldots+(X_N-\overline{X})^2}{N}} \approx$$

$$\sqrt{\frac{(X_1-\mu)^2+(X_2-\mu)^2+\ldots+(X_N-\mu)^2}{N}} \approx$$

$$\textbf{SD}(X)=\sigma=\begin{cases} \sqrt{\sum\limits_x (x-\mu)^2\ p(x)} & \text{in the discrete case} \\ \sqrt{\int_{-\infty}^{\infty} (x-\mu)^2\ f(x)\ dx} & \text{in the continuous case} \end{cases}$$

These non-random values are characteristics of the random variable and of the distribution. Among these three characteristics the variance and the standard deviation play a much more important theoretical and practical role than the average distance from the average.

Mechanical meaning of the variance. The mechanical meaning of the variance is the inertia about the center, because it is calculated by the same formula as the variance:

$$\begin{cases} \sum\limits_x (x-\mu)^2\ p(x) & \text{in the discrete case} \\ \int_{-\infty}^{\infty} (x-\mu)^2\ f(x)\ dx & \text{in the continuous case} \end{cases}$$

Remark. It may seem a little bit strange that the notion of the variance and the standard deviation play a more important role than the notion of the average distance from the average. The reason is that the variance and the standard deviation satisfy a rule which is very important both for the theory and the practice. Namely, it is true that the variance of the sum of independent random variables equals the sum of the variances of the random variables, or equivalently the standard deviation of the sum of independent random variables equals to the sum of the squares of the standard deviations of the random variables. Such a general rule does not hold for the average distance from the average.

The variance is very often calculated on the basis of the following relation.

The variance equals the second moment minus the expected value squared:

$$\sum_x (x-\mu)^2\ p(x) = \sum_x x^2\ p(x) - \mu^2$$

in the discrete case, and

$$\int_{-\infty}^{\infty} (x-\mu)^2\ f(x)\ dx = \int_{-\infty}^{\infty} x^2\ f(x)\ dx - \mu^2$$

in the continuous case.

The proof of these relations is quite simple. In the continuous case:

$$\int_{-\infty}^{\infty} (x-\mu)^2\ f(x)\ dx =$$

$$\int_{-\infty}^{\infty} \left(x^2 - 2x\mu + \mu^2\right)\ f(x)\ dx =$$

$$\int_{-\infty}^{\infty} x^2\ f(x)\ dx - \int_{-\infty}^{\infty} 2x\mu\ f(x)\ dx + \int_{-\infty}^{\infty} \mu^2\ f(x)\ dx =$$

$$\int_{-\infty}^{\infty} x^2\ f(x)\ dx - 2\ \mu \int_{-\infty}^{\infty} x\ f(x)\ dx + \mu^2 \int_{-\infty}^{\infty}\ f(x)\ dx =$$

$$\int_{-\infty}^{\infty} x^2\ f(x)\ dx - 2\ \mu\ \mu + \mu^2\ 1 =$$

$$\int_{-\infty}^{\infty} x^2\ f(x)\ dx - \mu^2$$

In the discrete case, the integral is replaced by summation, $f(x)$ is replaced by $p(x)$.

Using Excel. In Excel, the variance of a discrete distribution given by numerical values can be calculated like this: if the distribution is arranged in a table-form so that the x values constitute $\texttt{array}_1$ (a row or a column) and the associated $p(x)$ values constitute $\texttt{array}_2$ (another row or column) then we may calculate the expected value μ by the

$\mu = \texttt{SUMPRODUCT(array}_1\texttt{;array}_2\texttt{)}$

command, and then we may calculate $(x-\mu)^2$ for each x, and arrange these squared distances into $\texttt{array}_3$. Then the variance is

$\sigma^2 = \texttt{SUMPRODUCT(array}_3\texttt{;array}_2\texttt{)}$

and the standard deviation is

$\sigma = \texttt{SQRT(SUMPRODUCT(array}_3\texttt{;array}_2\texttt{))}$

Steiner's equality. The second moment of a distribution about a point c is equal to the variance plus the difference between the expected value and c squared:

$$\sum_x (x-c)^2\ p(x) = \sum_x (x-\mu)^2\ p(x) + (\mu-c)^2 = \sigma^2 + (\mu-c)^2$$

in the discrete case, and

$$\int_{-\infty}^{\infty} (x-c)^2\ f(x)\ dx = \int_{-\infty}^{\infty} (x-\mu)^2\ f(x)\ dx + (\mu-c)^2 = \sigma^2 + (\mu-c)^2$$

in the continuous case.

Steiner's inequality. The second moment of a distribution about any point c is greater than the variance, and equality holds only if $c=\mu$. In other words, the second moment of a distribution about any point c is minimal, if $c=\mu$, and the minimal value is σ^2:

$$\sum_x (x-c)^2\ p(x) \geq \sigma^2$$

in the discrete case, and

$$\int_{-\infty}^{\infty} (x-c)^2\ f(x)\ dx \geq \sigma^2$$

in the continuous case. Equality holds if and only if $c=\mu$.

The proof of Steiner's equality, for the continuous case:

$$\int_{-\infty}^{\infty} (x-c)^2\ f(x)\ dx =$$

$$\int_{-\infty}^{\infty} \left(\ (x-\mu)+(\mu-c)\ \right)^2\ f(x)\ dx =$$

$$\int_{-\infty}^{\infty} \left(\ (x-\mu)^2 + 2(x-\mu)(\mu-c) + (\mu-c)^2\ \right)\ f(x)\ dx =$$

$$\int_{-\infty}^{\infty} (x-\mu)^2\ f(x)\ dx + \int_{-\infty}^{\infty} 2(x-\mu)(\mu-c) f(x)\ dx + \int_{-\infty}^{\infty} (\mu-c)^2 f(x)\ dx =$$

$$\sigma^2 + 2(\mu-c) \int_{-\infty}^{\infty} (x-\mu) f(x)\ dx + (\mu-c)^2 \int_{-\infty}^{\infty} f(x)\ dx =$$

$$\sigma^2 + 2(\mu-c)\ 0 + (\mu-c)^2\ 1 =$$

$$\sigma^2 + 0 + (\mu-c)^2 =$$

$$\sigma^2 + (\mu-c)^2$$

In the above argument, we used the fact that

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

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

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

$$\mu - \mu\, 1 = 0$$

In the discrete case, the integral is replaced by summation, $f(x)$ is replaced by $p(x)$. The Steiner's inequality is an obvious consequence of the Steiner's equality.

Steiner's equality in mechanics. Steiner's equality in mechanics is well-known: the inertia about a point c is equal to the inertia about the center of mass plus inertia about the point c as if the total amount of mass were in the center of mass.

Steiner's inequality in mechanics. Steiner's inequality in mechanics is well-known: the inertia about a point c which is different from the center of mass is greater than the inertia about the center of mass.

Variance and standard deviation of some distributions:

1. **Uniform distribution**

   The second moment of the uniform distribution on an interval $(A;B)$ is

   $$\frac{A^2+AB+B^2}{3}$$

   The expected value of the uniform distribution is

   $$\frac{A+B}{2}$$

   So, the variance is

   $$\mathbf{VAR}(X) = \left(\frac{A^2+AB+B^2}{3}\right) - \left(\frac{A+B}{2}\right)^2 = \frac{(B-A)^2}{12}$$

   Thus, the standard deviation of the uniform distribution is

   $$\mathbf{SD} = \frac{(B-A)}{\sqrt{12}}$$

2. **Exponential distribution**

   The second moment of the exponential distribution is

   $$\mathbf{E}\left(X^2\right) = \frac{2}{\lambda^2}$$

   The expected value of the exponential distribution is

   $$\mathbf{E}\left(X\right) = \frac{1}{\lambda}$$

   So, the variance is

   $$\mathbf{VAR}(X) = \left(\frac{2}{\lambda^2}\right) - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}$$

   Thus, the standard deviation of the exponential distribution is

   $$\mathbf{SD} = \frac{1}{\lambda}$$

3. **Normal distributions**

   We know that the expected value of the normal distribution with parameters μ and σ is μ. The variance of the normal distribution with parameters μ and σ is

   $$\mathbf{VAR}(X) = \sigma^2$$

   Thus, the standard deviation of the normal distribution is

   $$\mathbf{SD} = \sigma$$

   This is why the normal distribution with parameters μ and σ is also called the normal distribution with expected value μ and standard deviation σ.