# Illustrative Probability Theory

**by István Szalkai, Veszprém,**
University of Pannonia, Veszprém, Hungary

**The original "Hungarian" deck of cards, from 1835**

# Note:

This is a very short summary for better understanding.
N and R denote the sets of natural and real numbers,
□ denotes the end of theorems, proofs, remarks, etc.,
in quotation marks ("...") we also give the Hungarian terms /sometimes interchanged/.
Further materials can be found on my webpage http://math.uni-pannon.hu/~szalkai in the Section "Valószínűségszámítás".

**dr. Szalkai István,** szalkai@almos.uni-pannon.hu

Veszprém, 2018.11.01.

# Content:

# 0. Prerequisites

## Elementary combintorics and counting techniques.

Recall and repeat your knowledge about combinatorics from secondary school: permutations, variations, combinations, factorials, the binomial coefficients ("binomiális együtthatók")

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

and their basic properties, the Pascal triangle, Newton's binomial theorem. The above formula is defined for all natural numbers $n,k \in \mathrm{N}$ . For $n<k$ the binomial coefficients has value 0. It is read as "*n* **choose** *k*" but in Hungarian "*n* alatt *k*".

You may read https://en.wikipedia.org/wiki/Binomial_theorem , and you can also use my booklet (in Hungarian) on my webpage http://math.uni-pannon.hu/~szalkai/Komb-elemei.pdf

*You must practice elementary counting problems, since problems of this type are unbelievable hard!*

You certainly know the **die** (plural: dice, "kocka,kockák") having 6 faces ("lap") and dots on it. Moreover, you must be familiar also with the *mathematical* background of the Hungarian and French cards, please read the following subsection carefully. However, you are forbidden (at least before your successful exam) to enter into any gambling with dice or cards !!! !!!

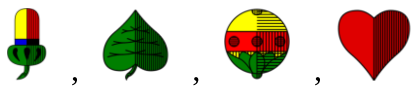
## The decks of "Hungarian" and "French" cards

("Magyar és Francia kártyák")

Most of the foreign literature calls the "Magyar" cards to "German" cards [**m1**]. However, some years ago it was proved by historicans, that this set of figures was invented and produced first in Hungary in 1835 by **József Schneider** (see front and back cover). More history is included at the end of this subsection.

*Mathematically* both **decks**/sets of cards ("kártyapaklik") contain of four **suits**/colors ("színek"), 8 and 13 **figures**/characters ("figura"), respectively, in each suit, so they can be arranged in a matrix form (see back cover). So, the Magyar kártya (deck) contains of 4*8=**32** cards while there are 4*13=**52** cards in the French set. The names of the suits is (mathematically) not so important, some new edition uses simply red, yellow, green, blue ... colors.

The real Magyar suits are:



and their names and French equivalents are, in order ([**m2**]):

le gland = **acorns** ("**makk**") corresponds to trèfle = clovers = clubs = ♣ ("treff"),
la feuille = **leaves** ("**zöld**") corresponds to pique = pikes = spades = ♠ ("pikk"),
le grelot = **bells** ("**tök**") corresponds to carreau = tiles = diamonds = ♦ ("káró"),
le cœur = **hearts** ("**piros**") corresponds to cœur = hearts = ♥ ("kőr").

The German names are *Eichel*, *Grün* oder *Blatt* oder *Laub*, *Schelle* oder *Schell*, *Herz*.

The names of the Magyar **figures** are: VII, VIII, IX, X, **alsó** ("under=inferior=sergeant"), **felső** ("over=superior=officer"), **király** ("king"), **ász** vagy **disznó** ("ass=pig").

The figures on the cards (see back cover) are famous swiss heros from the 14th century, Friedrich Schiller wrote a famous drama about the story in 1804. The drama's first Hungarian performance was in 1833 and became shortly popular. Because, in the early 19th century the passive resistance in Hungary, against the suppression of the Austrian Empire (the Habsburgs) was strengthening. The Swiss managed a successful uprising against the same Habsburgs, the portrait of their leader, **Tell Vilmos** ("William Tell") [**m3**] can be found on the card "makk felső" (search for it). The swiss characters from the drama, instead of Hngarian heros, were chosen to avoid *censorship* at that time of the Hungarian opposition to Habsburg rule. The story, after all, was about a successful revolt against the Habsburgs.

We have to add that the interesting story of Tell Vilmos is possibly a legend, modern scientific historians proved it, though there many early middle age similar legends (e.g. in Dutchland) and some decade ago serious punishment were taken in Switzerland to persons who denied the existence of Guglielm Tell. However, the successful uprising against the Habsburgs between 1308 and 1315 is a valid historical fact (Battle of Morgarten, "Morgarteni csata" [**m4**]).

Interesting also, that the Magyar kártya is sometimes called Swiss cards ("Svájci kártya") due to the nationality of the characters but this deck of cards is not used in Switzerland.

For Hungarian national card games see (after your exam!)

**Ulti** https://en.wikipedia.org/wiki/Ulti ,
**Snapszer** https://en.wikipedia.org/wiki/Sixty-six_(card_game) ,
**Huszonegyes** https://hu.wikipedia.org/wiki/Huszonegyes ,
similar to https://en.wikipedia.org/wiki/Blackjack ,
**Zsírozás** (hetes) https://en.wikipedia.org/wiki/Sedma ,
**Makaó** https://en.wikipedia.org/wiki/Macau_(card_game) .

## References:

[**m1**] https://en.wikipedia.org/wiki/German_playing_cards
[**m2**] jeu de cartes allemand
[**m3**] https://en.wikipedia.org/wiki/William_Tell
[**m4**] https://en.wikipedia.org/wiki/Battle_of_Morgarten

# 1. Events and the sample space

("Események és az eseménytér")

## Basic notions, definitions

**1.0. Definitions**
**Experiment** ("kísérlet"): active or passive observing a phenomenom.
**Deterministic (=determined) experiment:** the outcome (result) is *uniquely determined* by the preliminary conditions (the same conditions => the same result).
**Stochastic (=random,** "véletlen"**) experiment:** the outcome is *not* determined by the conditions: repeating the experiment under the same conditions usually we get (randomly) another results.
**Examples:** throwing a die or more dice, coins, picking cards from a deck, measuring any physical quantity (temperature, speed, weight, etc.), life-time of a unit or an animal or of people, lottery, etc.

**1.1. Definitions**
**Event** ("esemény")**:** a (precisely described) outcome of an experiment.
**Elementary event** ("elemi esemény")**:** can not be splitted to smaller exents.
**Compound** (or complex, "összetett") **events** are build from elementary events.
**Sample space** ("eseménytér")**:** the *set* of all elementary events, it is usually denoted by $\Omega$ (other books use H or T or other letter). $\Omega$ is also called **ground set** ("alaphalmaz").

**Examples:** *elementary events* are: "rolling one die I got 3" , "rolling three dice I got 3,2,1", "I picked the red king card" , "the temperature is (exactly) 23°C" , . . .
The sample spaces in the above examples are: $\Omega_{die}=\{1,2,3,4,5,6\}$,
$\Omega_{3dice} = \{(1,1,1),(1,1,2),(1,2,1),...,(3,2,1),...,(6,6,6)\}$, $\Omega_{cards}$={the cards of the deck},
$\Omega_{temp}$ = R (set of all real numbers),
*Compound events* are, for example: "rolling one die I got an odd number", "I picked a king card", "rolling three dice I got equal points", "the temperature is between 23 and 25 °C " .

**1.2. Warning:** in probability theory you are *not* allowed to say "three unique dice" or similar, since in the nature all objects (dice, coins, etc.) are different and we want to study the nature! That is why, for example the sample set $\Omega_{3dice}$ contains of $6^3$ elements, i.e. $|\Omega_{3dice}| = 6^3$ .

Observe, that elementary events are subsets of $\Omega$ while compound events are *subsets* of it.
The elementary events in the above examples are: $\omega_{die}$ = "3" $\in \Omega_{die}$ ,
$\omega_{3dice} = (3,2,1) \in \Omega_{3dice}$ , $\omega_{cards}$ = "red king" $\in \Omega_{cards}$ , $\omega_{temp} = 23 \in \Omega_{temp}$ , . . . ,
the above compound events are: $A_{die} = \{1,3,5\} \subset \Omega_{die}$ ,
$A_{3die} = \{ (1,1,1), (2,2,2), (3,3,3), (4,4,4), (5,5,5), (6,6,6) \} \subset \Omega_{die}$ ,
$A_{cards}$ = { spade king , heart king , diamond king , club king } $\subset \Omega_{cards}$ ,
$A_{temp} = [23,25] \subset \Omega_{temp}$ , . . . ,

Now, the general (abstract) mathematical definition is the following:

**1.3. Definition** Any nonempty set $\Omega$ is called a **sample space**, any subset A of $\Omega$, i.e. $A \subseteq \Omega$ is called an **event** and any element $\omega \in \Omega$ (or $x \in \Omega$) is called an **elementary event** in $\Omega$.

Note, that any elementary event $\omega \in \Omega$ can (must) be identified to the one element subset (singleton,"egyerlemű halmaz") $\{\omega\} \subseteq \Omega$ . □

**1.4. Note** that the result (outcome, "végeredmény") of an experiment ("kísérlet") is always an element ("elem") $\omega \in \Omega$. We say that the event A is **satisfied/occured** ("bekövetkezett") if $\omega \in A$ . □

**1.5. Definition** The set of all events is the *set of all subsets of* $\Omega$ , which is called the **power set** ("hatványhalmaz") of $\Omega$, and is denoted by ***P***$(\Omega)$, i.e. $P(\Omega) := \{A : A \subseteq \Omega\}$ . □

The notions below are obvious but we have to think over their mathematical background. Moreover, later we give generalizations of them.

**1.6. Definitions**

A **certain** or **sure event** ("biztos esemény") must occur in every case, i.e. it must contain all element $\omega \in \Omega$. Clearly there is only one such subset of $\Omega$ : $\Omega$ himself (the ground set). So, $\Omega$ is the only **certain event.**

An **impossible event** ("lehetetlen esemény") must *not* occur in any case, i.e. it must not contain any element $\omega \in \Omega$. Clearly there is only one such subset of $\Omega$ : the **empty set** $\varnothing$ ("üres halmaz"). So, $\varnothing$ is the only **impossible event**.

**Excluding** ("kizáró") events A and B may not occur at the same time, i.e. for any outcome $\omega \in \Omega$ one of them must not occur, that is either $\omega \notin A$ or $\omega \notin B$. This means that excluding events must be disjoint ("diszjunkt") sets: $A \cap B = \varnothing$ . Clearly disjoint sets always represent excluding events. (See also page 13.)

In which case can we say that the event **A implies B** ("A maga után vonja B-t"), or **B follows from A** ("B következik A-ból") ? Clearly, for any outcome of the experiment, i.e. for each $\omega \in \Omega$, in the case $\omega \in A$ we must also have $\omega \in B$. This is exactly when $A \subseteq B$ , i.e. A is a **subset** ("részhalmaz") of B. (See also page 13.) □

From now on please keep using the mathematical dictionary ("szótár") in the Appendix ("Függelék") for better understanding.

## Operations with events (algebra of the events)

("műveletek eseményekkel, eseményalgebra")

Please keep in mind that the (so called) *events* in probability theory are, in fact, subsets of a given ground set $\Omega$, we are always allowed to talk about (sub)*sets*, *union* ("únió"), *intersection* ("metszet") and *complement* ("komplementer ") instead of the following new terms. See also the Appendix.

**1.7. Definition:** The "new" operations on events are, for eny events $A, B \subseteq \Omega$ :

| | | |
|---|---|---|
| **sum** or **addition** ("összeg") | **A+B :=** | $A \cup B$ = union, |
| **product** ("szorzat") | **A·B :=** | $A \cap B$ = intersection, |

**difference** or **substraction** ("különbség") **A-B** = A\B = difference,
**negation** ("tagadás") of A $\overline{A}$ = $\overline{A}$ = complement.

That is:

the event **A+B** occurs exatly when *either* A or B (at least one of them) occurs,
the event **A·B** occurs exatly when *both* A or B occur,
the event **A-B** occurs exatly when A occurs *but* B does not,
the event $\overline{A}$ occurs exatly when A does *not* occur. □

**1.8. Notes: i)** the difference can be expressed as **A-B =** A∩$\overline{B}$ = A·$\overline{B}$, so we need ∪, ∩, ‾ only .

**ii)** Do not mix the above difference A-B with the **symmetric difference** ("szimmetrikus differencia/különbség") AΔB := (A\B)∪(B\A) . □

In this summary we use mainly the traditional set theoretical terms and notaions. We advice to the Reader to use and practice both (set- and probability theoretical) variants all the time.

## The properties of the operations

On the basis of the above remarks we have to repeat the properties of the well known set theoretical operations. We advice to the Readers to translate all these equalities to the new terminology and symbols.

**Boole-algebrák axiómái**

$$\mathfrak{B} = (X, \cup, \cap, \overline{\phantom{A}}, \emptyset, I)$$

| | | |
|---|---|---|
| kommutativitás | $A \cup B = B \cup A$ | (BA1) |
| | $A \cap B = B \cap A$ | (BA2) |
| asszociativitás | $A \cup (B \cup C) = (A \cup B) \cup C$ | (BA3) |
| | $A \cap (B \cap C) = (A \cap B) \cap C$ | (BA4) |
| disztributivitás | $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ | (BA5) |
| | $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ | (BA6) |
| elnyelési tulajdonságok | $A \cup (A \cap B) = A$ | (BA7) |
| | $A \cap (A \cup B) = A$ | (BA8) |
| $\emptyset$ és $I$ tulajdonságai | $A \cup \overline{A} = I$ | (BA9) |
| | $A \cap \overline{A} = \emptyset$ | (BA10) |
| | $A \cup \emptyset = A$ | (BA11) |
| | $A \cap \emptyset = \emptyset$ | (BA12) |
| | $A \cup I = I$ | (BA13) |
| | $A \cap I = A$ | (BA14) |

**The axioms of Boolean algebras**

**1.9. Notes: i)** The term **Boolean algebra** is a general notion: it includes not only the set $\cup$, $\cap$, $\overline{\phantom{A}}$ and probability $+$, $\cdot$ $\overline{\phantom{A}}$ operations, but the logical $\vee$, $\wedge$, $\neg$, the number theoretical **scm** , **gcd** ("lkkt, lnko"), N/$x$ operations, and many more.
**ii)** The above axioms have many consequences, for example the well known **De Morgan-rules**:

$$\overline{A\cup B}=\overline{A}\cap\overline{B} \quad \text{és} \quad \overline{A\cap B}=\overline{A}\cup\overline{B} \ .$$ □

## 2. The relative frequency and the probability

("A relatív gyakoriság és a valószínűség")

**2.0. Definition.** Fix an experiment and a possible event A . Repeat this experiment $n$ many times and denote $k$ the number of occurences of A during these $n$ many experiments (clearly $k\le n$). In this case $k$ is called the (**absolute**) **frequency** ("gyakoriság") of A , while the proportion $k/n$ is the **relative frequency** ("relatív gyakoriság") of A . □

*Practical experiences* show that after fixing an experiment and a possible event A , the relative frequency $k/n$ is very close to a fixed, theoretical number, say $p$ , which number is a characteristic of A . Moreover the higher is $n$ the closer is $k/n$ to $p$ . This does not contradict to the (again practical) phenomenon that $k/n$ always may have large fluctuations around $p$ , even for large $n$ . This phenomenon will be proved and explained by **Bernoulli**'s Theorem in Section 10 *Law of large numbers* .

This theoretical number $p$ is called the **probability** ("valószínűség") of the event of A and is denoted by **P(A)**. However, this is only a *naive* definition, the precise mathematical definition follows below.

**2.1. Definition:** The axioms of the **probability** ("valószínűség") by **Kolmogorov.**
Any P is a **probability** (-measure, "mérték") on the sample set $\Omega$ if :
**(o)** P : $P(\Omega)\to\mathbf{R}$ is a function, i.e. $P(A)\in\mathbf{R}$ is a real number for any $A\subseteq\Omega$,
**(i)** $0\le P(A)\le 1$ for any $A\subseteq\Omega$ ,
**(ii)** $P(\emptyset)=0$ , $P(\Omega)=1$ ,
**(iii)** $P\left(\bigcup_{i=1}^{M}A_i\right)=\sum_{i=1}^{M}P(A_i)$ and $P\left(\bigcup_{i=1}^{\infty}A_i\right)=\sum_{i=1}^{\infty}P(A_i)$
if the events pairwise ("páronként") exclude each other, i.e. if $A_i\cap A_j=\emptyset$ for $i\ne j$ . □

**2.2. Corollaries** of the axioms:
**i)** $P(A\cup B)=P(A)+P(B)-P(A\cap B)$ for any $A,B\subseteq\Omega$ ,
**ii)** $P(A\cup B)=P(A)+P(B)$ (**additivity**, "additivitás") *only if* $A\cap B=\emptyset$ or $A\cap B$
that is, *if* A and B exclude each other,
**iii)** $P(\overline{A})=1-P(A)$ ,

**iv)** $P(A \setminus B) = P(A) - P(A \cap B)$ for any A,B$\subseteq\Omega$ ,
**v)** $P(A \setminus B) = P(A) - P(B)$ *only if* $B \subset A$ , i.e. B implies A ,
**vi)** $P(B) \leq P(A)$ *for* $B \subset A$ (**monotonicity**, "monotonitás"),
**vii)** $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(B \cap C) - P(A \cap C) + P(A \cap B \cap C)$

(**logical sieve**, "logikai szitaformula") for any A,B$\subseteq\Omega$ . □

**2.3. Remark:** Observe that all the above axioms and properties are very similar to the properties of the **area** ("terület") of planar regions (figures) and to the properties of the **volume** ("térfogat") of 3D bodies. This is really so, and is not surprising, since all these notions (probability, area, volume, etc.) are **measures** ("mértékek") which, in some view of point, measure the size of the set A. So, we suggest to the Readers to substitute (in her/his mind) area $T_A$ instead of *P*(A) for easier understanding!

In the following we *extend* Definition 1.6.

**2.4. Definition:** For any events (subsets) A,B$\subseteq\Omega$ we say:
A is a **certain event** if P(A)=1 ,
A is an **impossible event** if P(A)=0 ,
A and B **excluding each other** if P(A$\cap$B)=0 . □

**2.5. Remark:** In everyday speech the words **chance** ("esély") and **probability** ("valószínűség") are synonyms, however in some probability and statistical theories these words mean completely different quantities: if *p* is the probability and *q*=1-*p* , then the chance is *p*/*q* . □

# 3. Calculating the probability

Now we introduce only the two simplest ways of calculating the probability. Keep in mind, that the main purpose of all the present summary and semester is to calculate the probability.

## 3.1. Combinatorial (classic) random field

("kombinatorikai/klasszikus valószínűségi mező")

If $\Omega$ is a finite set and each $\omega\in\Omega$ elementary event has equal probability (e.g. rolling a fair die or pulling a card from a deck, etc.), then

- first we have $P(\{\omega\}) = \frac{1}{n}$ for each $\omega\in\Omega$ , where $n=|\Omega|$ is the size of $\Omega$ ,
- second, $P(A) := \frac{|A|}{|\Omega|} = \frac{satisfactory}{total} = \left(\frac{\text{"kedvező"}}{\text{"összes"}}\right)$ for all A$\subseteq\Omega$ . □

Combinatorial problems, in general, are difficult since counting |A| and $|\Omega|$ are not so easy. This means that you have to practice a lot of combinatorial problems.

## Geometric probability field

("geometriai valószínűségi mező"),

In the case when $\Omega$ is, or can be represented with, such a subset of the real line or the plane or the space, where each elementary event {ω} has "the *same*" probability, **then** for every event / subset $A \subseteq \Omega$ we have

$$P(A) = \frac{\mu(A)}{\mu(\Omega)} \qquad (*)$$

where μ(A) and $\mu(\Omega)$ denote the **length** ("hossz"), **area** ("terület") or **volume** ("térfogat") of the 1-, 2- or 3- dimensional sets A and $\Omega$ . □

**3.3 Remarks:** Having the "*same probability*" is hard to check both in the reality and in the theory. For example, if we shoot to a target $\Omega$, the probability of hitting a specific geometrical point $\omega \in \Omega$ is zero. On the other hand, the formula (*) suggest that P(A) must depend on the area of A and not the placement of it. For example, if shooting to the target $\Omega$ we must ensure that our shots spread out **uniformly** ("egyenletesen") on *all the parts* of $\Omega$ , which is not the case for an olympic champion. Besides, each shot ω *must hit* the target $\Omega$ since each ω is an element of the ground set $\Omega$. In each application these assumption must be checked!

Do not confuse the *geometrical probability* with the *geometrical distribution* (see Section 7). □

**3.4. Examples:** - *waiting for the bus* if I just accidently go out to the bus stop,
- *target throwing* (supposing I shoot on the target randomly, I am not a professional target thrower , and all my shots hit the target) .

There are many problems and examples of type *Rendez-vous in the library* ("Randevú a könyvtárban"), in which $\Omega$ is, in fact, not present in the problem, it is only a model for the solution.

# 4. Conditional probability, independence of events

("Feltételes valószínűség, események függetlensége")

## The conditional probability

Suppose that something has been happened before the event A, denote this former event by B . How much effect the event B may have to A ? An extensive analyzation of the relative frequencies leds us to the following mathematical defintion:

**4.1. Definition:** If the event B is not impossible, ie P(B)>0 , then the probability of the occurence of A, *supposing* ("feltéve") that B has already been occured, is :

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \quad . \qquad (**)$$

P(A|B) is called the **conditional** ("feltételes") probability of A, where B is the **condition** ("feltétel"). The (previous) probability P(A) is called **unconditional** ("feltétel nélküli") probability. □

**4.2. Remarks:** The *conditional probability* satisfies all the axioms and properties of the (unconditional) probability in the case the condition B is fixed.

This means that, the formulas, listed in 2.1 (o)-(iii) and 2.2 i)- vii) remain true, if instead of *P*(...) everywhere we write *P*(...|B) . □

Naturally arises the following question: In *what measure* and in what *direction* does B have **effect** ("befolyás") to A ? That is, we have to compare P(A|B) to P(A). This will be examined in this Section later.

After the multiplication of the equality (**) in 4.1 we obtain the following simple but important relation:

**4.3.Theorem** of **multiplication** ("Szorzástétel")**:** $P(A \cap B) = P(A \mid B)P(B)$ . □

**4.4. Definition:** The events $B_1, B_2, ..., B_n \subseteq \Omega$ form a **complete system of events** ("teljes eseményrendszer"), if they pairwise exclude each other and their union is the certain event, i.e. in formulas:

$$B_i \cap B_j = \emptyset \qquad \text{for any } i \neq j \ ,$$

and

$$B_1 \cup B_2 \cup ... \cup B_n = \Omega$$

(or, in more general: $P(B_i \cap B_j)=0$ and $P(B_1 \cup B_2 \cup ... \cup B_n)=1$.)

In other branches of mathematics, a set system $\{B_1, B_2, ..., B_n\}$ with the above properties is also called **partition** or **division** ("partíció / felosztás"). See also the illustration left below. □

**4.5. Theorem** of the **complete probability** ("teljes valószínűség tétele"):
Suppose that $\{B_1, B_2, ..., B_n\}$ forms a complete system of events and $P(B_i)>0$ for each $i \leq n$ . Then for every event $A \subseteq \Omega$ we have

$$P(A) = \sum_{i=1}^{n} P(A \mid B_i) \cdot P(B_i)$$

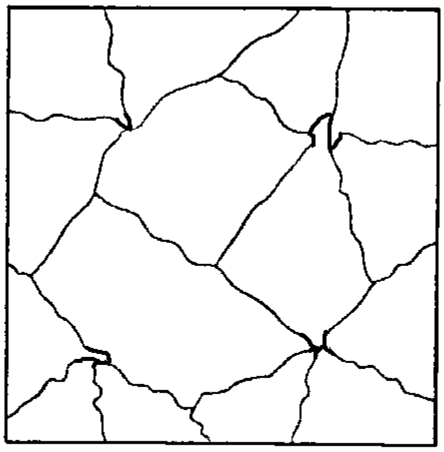
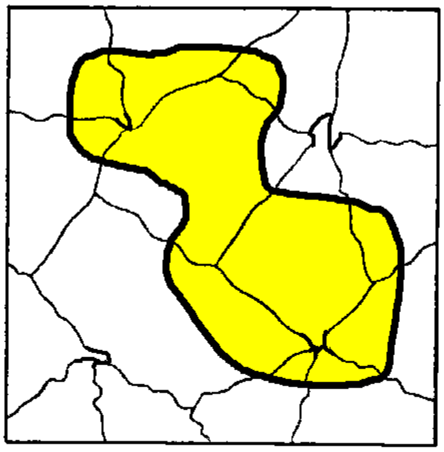
**Proof:** Using the *Theorem of multiplication* the above formula gives

$$P(A) = \sum_{i=1}^{n} P(A \cap B_i)$$

which clearly holds, since

$$A = \bigcup_{i=1}^{n} (A \cap B_i) \ . \qquad □$$

The following picture on the right illustrates the above ideas (think again on the area instead of P) :

**Complete system of events** (partition) **Complete probability**

**4.6. Example:** In a factory the goods are produced in 3 shifts. The 40% of the goods is produced in the I. shift, the 35% of them in the II. shift, and the 25% of them in the III. shift. The probability of the waste products in the I. shift is 0.05, in the II. shift is 0.06, in the III. shift is 0.07. If we choose a good randomly, how much is the probability of choosing a waste product?

**Solution:** Let $B_1$, $B_2$, $B_3$ denote the events that the good was produced in the shift I,...,III, and let W be the event that the good is a **waste** ("selejt") one. The conditions of the example say $P(B_1)=0.4$, $P(B_2)=0.35$, $P(B_3)=0.25$ (checking: $P(B_1)+P(B_2)+P(B_3)=0.4+0.35+0.25=1$) .

Further $P(W|B_1)=0.05$, $P(W|B_2)=0.06$ and $P(W|B_1)=0.07$. Now, using the Theorem of the Complete Probability we have:

$$P(W) = P(W|B_1)\cdot P(B_1) + P(W|B_2)\cdot P(B_2) + P(W|B_3)\cdot P(B_3) =$$
$$= 0.05*0.4 + 0.06*0.35 + 0.07*0.25 = \underline{0.0585} .$$ □

**Inverse question:** If the randomly chosen product is waste, what is the probability that the I. or II. or III. shift produced it? Who we have to blame for the waste product with the highest probability? For example, III. shift produced waste products with the highest probability, but on the contrary, they make the less many products. The answer is in the following theorem.

**4.7. Bayes theorem (Inversion theorem**, "megfordítási tétel"**):**

For every event $A,B\subseteq\Omega$ , assuming $P(A)>0$ and $P(B)>0$ we have

$$P(B\,|\,A) = \frac{P(A\,|\,B)\cdot P(B)}{P(A)} \quad .$$ □

**Proof:** The theorem follows from the Theorem of multiplication:

$$P(B|A)\cdot P(A) = P(B\cap A) = P(A\cap B) = P(A|B)\cdot P(B)$$ and divide by P(A). □

**Continuation of Example 4.6**

$P(B_1|W) = P(W|B_1)\cdot P(B_1)/P(W) = 0.05*0.4\ /0.0585 \approx 0.341\,880$ ,

$P(B_2|W) = P(W|B_2)\cdot P(B_2)/P(W) = 0.06*0.35/0.0585 \approx 0.358\,974$ ,

$P(B_3|W) = P(W|B_3)\cdot P(B_3)/P(W) = 0.07*0.25/0.0585 \approx 0.299\,145$ ,

which means that the largest amount of waste products was produced in the 2nd shift.

(Check: $P(B_1|W)+P(B_2|W)+P(B_3|W) = 1$ .)

**Clearly** P(A|B) = P(B|A) = 0 if the events A and B *exclude* each other (see page 6).
Similarly P(B|A) = 1 if A *implies* B (see page 6).

## The independence of events

("Események függetlensége")

We have already mentioned the natural question: in what extent and in which direction the occurence of B does have an influence ("hatás") for the occurence of A ? Obviously we have three main cases:

P(A|B) < P(A) means that B **weakens** ("gyengíti") A ,
P(A|B) > P(A) means that B **strengthens** ("erősíti") A ,
P(A|B) = P(A) means that B does **not** have influence on A .

**4.8. Special cases:** We now rethink the notions in Definition 1.6. and 2.4. on the basis of the formula in 4.1.:

If B⊆A then $P(A\,|\,B)=\dfrac{P(A\cap B)}{P(B)}=\dfrac{P(B)}{P(B)}=1$ , so B really implies A .

If B∩A=∅ then $P(A\,|\,B)=\dfrac{P(A\cap B)}{P(B)}=\dfrac{0}{P(B)}=0$ , so B really excludes A . □

**4.9.Remark:** It is an obvious requirement, that the events A and B can be **independent** ("függetlenek") only if none of them has any effect to the other, i.e.

P(A|B) = P(A) *and* P(B|A) = P(B) .

A short calculation (using the Theorem of multiplication) shows, that the above two equalities (together) are equivalent to the below one.

**4.10. Definition:** The events A and B are **independent from each other** ("függetlenek egymástól") if and only if

$$P(A\cap B)=P(A)\cdot P(B)\,.$$ □

Let us emphasize, that the above equality can *not* be applied for *any* events A,B⊆Ω but only (very) special ones !

Additionally, we can use the above equality in our practice in two directions.

**First,** if we can verify (in some physical or other way) that the two events A and B are really independent (e.g. two dice has no effect to each other), then we can use the above equality to determine P(A∩B) /i.e. reality ⇒ calculation/ .

**Second,** if our calculations (with a pocket calculator) justify the above equality, then no doubt: A and B must be considered to be independent /i.e. calculations ⇒ reality/ !

**4.11. Statement:** If $0<P(A)<1$, $0<P(B)<1$ and $A$ and $B$ are independent from each other, then the pairs of events $\overline{A}$ and $B$ , $A$ and $\overline{B}$ , $\overline{A}$ and $\overline{B}$ are also independent from each other. □

# 5. Random variables and their characteristics

("Valószínűségi változók és jellemzőik")

In most of the experiments we are detecting not only the occurence of an event (red, missing, frozen, exploided, etc.) but we are measuring some quantity. However, measuring the *same* quantity (e.g. the mass of a chocholate bar) several times, we get different data, in general, the alterations show random changes. The notion of (random) measuring is defined below.

## Random variables

**5.1. Definition:** The *functions* $\xi$ , which assign real numbers to elementary events, are called **random variables** ("valószínűségi változók"), **r.v.** ("v.v.") for short.
In formulae: $\xi$ can be any function $\xi : \Omega \to R$ . □
To memorize: an r.v. is the measured result of the experiment.

$Im(\xi)$ denotes the **image** or **range** ("képhalmaz/értékkészlet"), i.e. the set of possible measuring outcomes/results ("eredmények") of the r.v. $\xi$ .

It is essential to learn that r.v. have two essentially different types.

**5.2. Definition: i)** The r.v. $\xi$ is called **discrete** (*separated*, "diszkrét/elkülönült") if it may have finite or countable/enumerable ("megszámlálható/felsorolható") many possible values, in other words its range can be written in form $Im(\xi)= \{x_1, x_2, ... , x_n, ... \}$ where $x_i \in R$ are the possible outcomes (of the measuring).
**ii)** $\xi$ is called **continuous** ("folytonos") if its range contains an interval: $Im(\xi) \supseteq (a,b)$ . □
By Cantor's theorem no interval (a,b) is countable.

In learning probability theory it is very important to distinguish the above two types of r.v. Though, in the roots, they are the same phenomenon, but in the practice they have very different properties and formulas.

**5.2. Definition:** The **distribution** ("eloszlás") of a *discrete* r.v. $\xi$ is the set of the probabilities $\{p_1, p_2, ... , p_n, ... \}$ where $p_i :=P(\xi=x_i)$ for i=1,2,... and $Im(\xi)= \{x_1, x_2, ... , x_n, ... \}$.□

Keep in mind that *no* continuous r.v. has distribution in the above sense.

**5.4. Statement:** Any sequence of real numbers $\{p_1, p_2, ... , p_n, ... \}$ is a distribution of a discrete r.v. if and only if it fulfills the following axioms (fundamental properties):
*(i)* $0 \le p_i \le 1$ ,
*(ii)* $p_1+p_2+...+p_n+... = 1$ . □

## The distribution function

The following construction is valid both for discrete and continuous r.v. Both the construction and its properties must be obvious.

**5.5. Definiton:** For any r.v. $\xi$ (either discrete or continuous) the (**cumulative**) **distribution function** ("kumulatív/összegzési eloszlásfüggvény") is:

$$F : R \to R \quad \text{where} \quad F(x) := P(\xi < x) \quad . \qquad \square$$

**5.6. Theorem:** The basic properties (axioms) of the distribution function are:

0) $F : \mathrm{R} \to \mathrm{R}$ and Dom($F$)=R ,
1) $0 \le F(x) \le 1$ for $x \in \mathrm{R}$ ,
2) $F(x)$ is monotone increasing, i.e. $x_1 < x_2$ implies $F(x_1) < F(x_2)$ ,
3) $F(x)$ is continous from left, i.e. $\lim\limits_{x \to x_0 -} F(x) = x_0$ ("filled circles are on the right"),
4) $\lim\limits_{x \to \infty} F(x) = 1$ and $\lim\limits_{x \to -\infty} F(x) = 0$ . □

Observe, that distribution functions of discrete r.v. are always build of horizontal (straight) line segments placed on increasing heights, so called **step functions** ("lépcsős függvények").

What's the *distribution function* used for ? With its help we can give quick answers for the following, frequently asked questions (FAQ), both for discrete and continuous r.v. :

**5.7. Theorem:**

**i)** $P(\xi < a) = F(a)$ ,

**ii)** $P(\xi \ge a) = 1 - F(a)$ ,

**iii)** $P(\xi \le a) = F(a) + P(\xi = a)$ ,

**iv)** $P(\xi > a) = 1 - F(a) - P(\xi = a)$ ,

**v)** $P(a \le \xi < b) = F(b) - F(a)$ ,

**vi)** $P(a \le \xi \le b) = F(b) - F(a) + P(\xi = b)$ ,

**vii)** $P(a < \xi < b) = F(b) - F(a) - P(\xi = a)$ ,

**viii)** $P(a < \xi \le b) = F(b) - F(a) + P(\xi = b) - P(\xi = a)$ ,

**ix)** $P(\xi = a) = \lim\limits_{x \to a+} F(x) - F(a)$ . □

## The density function

Definiton 5.2 should be replaced by the following, matematically more accurate requirement.

**5.8. Definiton:** The r.v. $\xi$ is called **continous** ("folytonos") if there exist a function $f : \mathrm{R} \to \mathrm{R}$ for which (except at most finite many values $t \in \mathrm{R}$):

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

In this case $f(x)$ is called a **density function** ("sűrűségfüggvény") for the r.v. $\xi$ .

The r.v. $\xi$ and $\eta$ have the **same distribution** ("azonos eloszlásúak") if their density functions are equal: $f_\xi(x) = f_\eta(x)$ for all $x \in \mathrm{R}$ at most finitely many exception. □

Let us emphasize that only *continuous* r.v. possesses a density function! The naming *density function* will be explained in Remark 5.15.

**5.9. Theorem:** The basic properties (axioms) of the density function are:

0) $f: \mathrm{R} \to \mathrm{R}$ 1) $f(x) \geq 0$ 2) $\int_{-\infty}^{\infty} f(x) = 1$ . □

**5.10. Theorem:** If $f: R \to R$ is a continous function (except at mot finitely many points) satisfying the three properties of Theorem 9, then there is a r.v. $\xi$ for which $f(x)$ is a density function. □

Let us see first the connection between the *distibution* and *density* functions, $F(x)$ and $f(x)$ in detail.

**5.11. Theorem:** *(i)* If $\xi$ is continous then $F(x)$ is continous in each point $x \in \mathrm{R}$ .
*(ii)* If $F(x)$ is continous and it is continously differentiable (except finite many points), then there is a continous r.v. ξ that has distribution function exactly $F(x)$ .
*(iii)* $F'(x) = f(x)$ for all the points $x \in \mathrm{R}$ where the derivative $F'(x)$ does exist. □

**5.12. Theorem:** $P(\xi = x) = 0$ for all the points $x \in \mathrm{R}$ where $F(x)$ is continuous. □

What's the $f(x)$ *density function* used for ? With its help we can give quick answers for the following, frequently asked questions (FAQ), but only for continuous r.v. :

**5.13. Theorem:** For any continuous r.v. $\xi$

$$P(a \leq \xi < b) = P(a \leq \xi \leq b) = P(a < \xi \leq b) = P(a < \xi < b) = F(b) - F(a) \ . \qquad \square$$

Now we summarize the most important formulas (FAQ) of this Section.

**5.14. Theorem: "Typical" questions** and answers (FAQ) ("tipikus kérdések és válaszok"):

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

$P(a\leq\xi) = \int_{a}^{\infty} f(x)dx = 1-F(a) = 1- P(\xi<a)$ ,

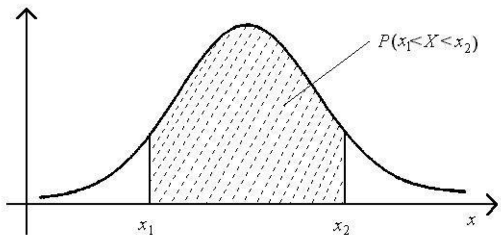
$P(a\leq\xi<b) = \int_{a}^{b} f(x)dx = F(b)-F(a)$ (Newton-Leibniz rule ("szabály"))

$P(\xi=b) = 0$ (for continuous r.v. ξ)

$P(\xi\approx c) = P(|\xi-c|<\varepsilon) = P(c-\varepsilon<\xi<c+\varepsilon) = F(c+\varepsilon)-F(c-\varepsilon)$ . □

**5.15. Remark:** Now we can answer why *f(x)* is called density function.
We must think on the horizontal real line (*x* axis) as the scale of the analogue measuring device ξ , and the height of the function *f(x)* at the point $a=x_0$ is proportionate to how many times the measuring ξ resulted $x_0$ , i.e. $f(x_0) \approx P(\xi=x_0)$. More precisely:

$$\lim_{\Delta a \to 0} \frac{P(a \leq \xi < a + \Delta a)}{\Delta a} = f(a) \ .$$

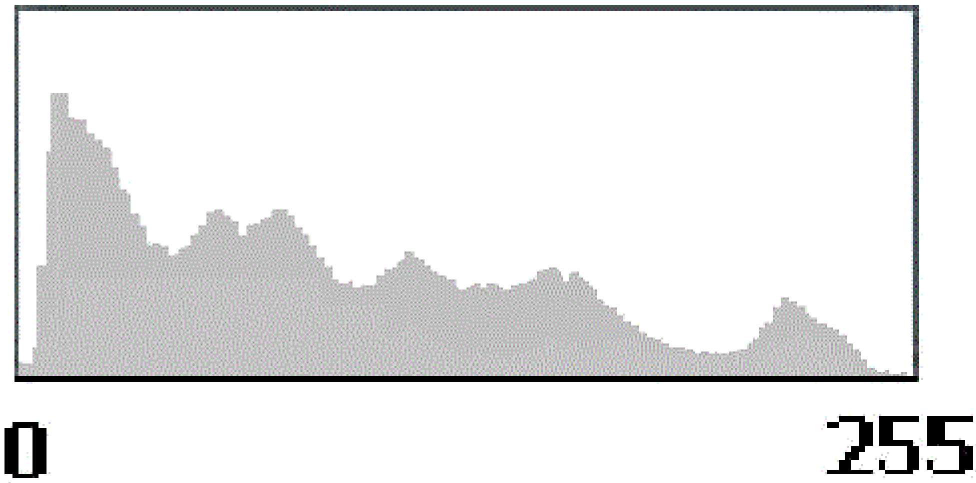



$$P(x_1 < X < x_2) = \int_{x_1}^{x_2} f(x)dx.$$

**The density function**

Recall, that *discrete* r.v. do not have density functions, instead they have **histograms / column diagrams** ("Hisztogram/oszlopdiagram") which, however do not exist for continuous r.v.

A histogram is a collection of bars at each point $x_i \in \mathrm{Im}(\xi)$ of height $p_i = P(\xi = x_i)$ . If $\xi$ may have many possible values $x_i \in \mathrm{Im}(\xi)$ , which are dense on the $x$ axis, then histograms look very *similar* to density functions. (This connection can be justified using some more complicated mathematics.)




**Histogram (column-diagram)**

## Independence of random variables

**5.16. Definition:** The *arbitrary* random **variables** $\xi$ and $\eta$ are called **independent** ("függetlenek") if for any numbers $x,y \in \mathbb{R}$ we have

$$P(\xi < x \quad and \quad \eta < y) = P(\xi < x) \cdot P(\eta < y) \quad . \qquad \square$$

**5.17. Theorem:** The *discrete* r.v. ξ and η with Im(ξ)={ $x_1, x_2,....$ } and Im(η)= { $y_1, y_2....$ } are independent if and only if for any *i,j*∈N indices the following holds:

$$P(\xi = x_i \ , \eta = y_j) \ = \ P(\xi = x_i) \cdot P(\eta = y_j) \ . \qquad \square$$

Clearly the delimiters "... *and* ..." and "... , ..." in the above expressions have the same meaning.

**5.18. Theorem:** The *continuous* r.v. ξ and η with density functions $f(x)$, $g(y)$ resp. are independent if and only if for any *x,y*∈R real numbers the following holds for the second order partial differences:

$$\frac{\partial^2 P(\xi < x \ , \eta < y)}{\partial x \, \partial y} = f(x) \cdot g(y) \qquad \square$$

# 6. Expected value, variance and dispersion

("Várható érték, szórásnégyzet és szórás")

In general we make measurements several times for more accurate results, calculate **arithmetical means/averages** ("számtani közép/átlag"), count weights for repeated results. Let us see these habits in the theory.

## The mean or expected value

**6.1. Definition:** Let the *discrete* r.v. $\xi$ have the range (possible outcomes) $\{ x_1, .... \}$ with the distribution {$p_1$,...}. Then the **mean** ("átlag") or **expected value** ("várható érték") of $\xi$ is defined as follows:

- if $\xi$ has a finite range (*n* many), then

$$M(\xi) = E(\xi) = \sum_{i=1}^{n} x_i p_i \ ,$$

- if $\xi$ has an infinite range, then

$$M(\xi) = E(\xi) = \sum_{i=1}^{\infty} x_i p_i$$

assuming $\sum_{i=1}^{\infty} |x_i| \cdot p_i < \infty$ . □

**6.2. Definition:** For the *continuous r.v.* $\xi$ with density function *f*(*x*) the **mean** ("átlag") or **expected value** ("várható érték") of $\xi$ is defined as follows:

$$M(\xi) = E(\xi) = \int_{-\infty}^{\infty} x \cdot f(x) \, dx$$

assuming $\int_{-\infty}^{\infty} |x| \cdot f(x) dx < \infty$ . □

**6.3. Remarks:** *(i)* The older notation **E(ξ)** refers to the term ("szakkifejezés") **expected value** ("várt/várható érték"), but in our days the better notation **M(ξ)** for the **mean** ("átlag") is also in use. One can "expect" that after many measurements the results are very close, around the (arithmetical) mean. However, rolling a fair dice the mean is 3.5, and I do not think it is worth expecting to roll exactly 3.5 scores !

In (everyday) statistics we often hear about "*expected life time at birth*". How can we expect the life time of a newborn baby looking at her/him? Of course it is an average again.

*(ii)* The **empirical** (=practical (greek), "gyakorlati") observation, *that measuring results are very close to the mean* are justified in rigorous mathematical tools in Theorem 10.4 **Weak Law of Large Numbers** ("Nagy számok gyenge törvénye ") of **Chebyshev** ("Csebisev"). □

**6.4. Theorem: The properties of the mean:**

o) $M(\xi)=M(\eta)$ if $\xi$ and $\eta$ have the same distribution,

i) $M(\xi)=c$ if $\xi=c$ (constant, "állandó"), i.e. the measuring device is sticked ("beragadt a mérőműszer"),

ii) $M(a\xi+b)=aM(\xi)+b$ for any fixed real numbers $a,b\in$R (e.g. $^{o}$C and $^{o}$F),

iii) $M(\xi+\eta)=M(\xi)+M(\eta)$ holds for *any* r.v. $\xi$ and $\eta$ ,

iv) $a\le M(\xi)\le b$ if $a\le\xi\le b$ , i.e. the measuring is bounded ("korlátos"),

v) $0\le M(\xi)$ if $0\le\xi$ ,

vi) $M(\xi\eta)=M(\xi)\cdot M(\eta)$ holds *only for independent* r.v. $\xi$ and $\eta$ .

vii) If $\xi_1,\xi_2,...\xi_n$ are pairwise independent r.v., having the same distribution, then

$$M\left(\sum_{i=1}^{n}\xi_i\right)=n\cdot M(\xi_1) \quad \text{and} \quad M\left(\frac{\sum_{i=1}^{n}\xi_i}{n}\right)=M(\xi_1)$$

viii) For any continuous function $g:R\to R$ , and discrete r.v. $\xi$

$$M(g(\xi))=\sum_{i=1}^{\infty}g(x_i)p_i \ ,$$

and if $\xi$ is continuous r.v., then

$$M(g(\xi))=\int_{-\infty}^{\infty}g(x)f(x)dx \ .$$

ix) Especially for $g(x)=x^2$ we have $M(\xi^2)=\sum_{i=1}^{\infty}x_i^2p_i$ and $M(\xi^2)=\int_{-\infty}^{\infty}x^2f(x)dx$ . □

Clearly vii) above means that measuring several times and calculating averages results the same average. Compare this observation to iv) of Theorem 6.10.

## The variance and the dispersion

Many times we have experienced, that merely different datasets may result the same mean (e.g. marks in classrooms, fair and the {3,3,3,4,4,4} dice, etc.). We will measure the "*distance*" or the "*spread*" ("szóródás") of the dataset below.

The formula $M(\xi - M(\xi))$ gives always 0, the other formula $M(|\xi - M(\xi)|)$ has (mathematically) bad properties and is hard to compute. Now, the formula $(\xi - M(\xi))^2$ makes small differences even smaller, the big differences even larger, and has (mathematically) good properties, so this is our choice.

**6.6. Definition:** For any (discrete or continuous) r.v. $\xi$ the **variance** ("szórásnégyzet") of $\xi$ is

$$v(\xi) := D^2(\xi) := M\left((\xi - M(\xi))^2\right)$$

assuming this expression (infinite sum or improper integral) is finite,

the **dispersion** ("szórás") or **standard deviation** of $\xi$ is : $D(\xi) = \sqrt{M\left((\xi - M(\xi))^2\right)}$ .

The symbols $\sigma(\xi)$ and $s(\xi)$ are also in use for the dispersion. □

**6.7. Statement:** Clearly the dispersion is the square root of the variance.
By Theorem 6.4.v) the quantity under the square root is always nonegative. □

**6.8. Statement:** If $M(\xi^2)$ is finite, then the variance of $\xi$ does exist, and we have:

$$D^2(\xi) = M(\xi^2) - M^2(\xi) \quad \text{and} \quad D(\xi) = \sqrt{M(\xi^2) - M^2(\xi)} \ .$$

**Proof:** By Theorem 6.4. :

$$D^2(\xi) = M\left((\xi - M(\xi))^2\right) = M\left(\xi^2 - 2\xi \cdot M(\xi) + M^2(\xi)\right) = M(\xi^2) - 2M(\xi)M(\xi) + M^2(\xi) = \\ = M(\xi^2) - M^2(\xi)$$ □

**6.9. Conclusion:** By Theorem 6.4. ix) we have the following formulae:

for *discrete* r.v. $\xi$ $$D^2(\xi) = \sum_{i=1}^{\infty} x_i^2 p_i - \left(\sum_{i=1}^{\infty} x_i p_i\right)^2 \ ,$$

for *continuous* r.v. $\xi$ $$D^2(\xi) = \int_{-\infty}^{\infty} x^2 f(x)dx - \left(\int_{-\infty}^{\infty} xf(x)dx\right)^2 \ .$$ □

In practice and in examples we calculate $D^2(\xi)$ and $D(\xi)$ by the formulae of 6.9. instead of 6.6.

**6.10. Theorem: The main properties of the variance and the dispersion:**

i) $D^2(\xi) = D(\xi) = 0$ if and only if $\xi = c$ (constant, "állandó"), i.e. the device is sticked, ("beragadt a mérőműszer"),

ii) $D^2(a\xi + b) = a^2 \cdot D^2(\xi)$ and $D(a\xi + b) = |a| \cdot D(\xi)$ for any fixed real numbers $a,b \in$ R,

iii) $D^2(\xi + \eta) = D^2(\xi) + D^2(\eta)$ and $D(\xi + \eta) = \sqrt{D^2(\xi) + D^2(\eta)}$ holds *only for independent* r.v. $\xi$ and $\eta$ .

iv) If $\xi_1, \xi_2, ... \xi_n$ are pairwise independent r.v., having the same distribution, then

$$D^2\left(\sum_{i=1}^{n} \xi_i\right) = n \cdot D^2(\xi_1) \quad , \quad D^2\left(\frac{\sum_{i=1}^{n} \xi_i}{n}\right) = \frac{D^2(\xi_1)}{n} \quad ,$$

$$D\left(\sum_{i=1}^{n}\xi_i\right)=\sqrt{n}\cdot D(\xi_1) \quad , \quad . \; D\left(\frac{\sum_{i=1}^{n}\xi_i}{n}\right)=\frac{D(\xi_1)}{\sqrt{n}}$$ □

Recall our remark concerning i) of Theorem 6.4. Now, iv) above adds, that measuring several times and calculating averages results *less* dispersion. For example, $n$=10 times *more* measurings makes $\sqrt{10}\approx 3.16$ many times *less* dispersion.

**6.11. Definition:** The **mode** ("módusz") of a *discrete* r.v. ξ is the value $\mathbf{x_k}\in\text{Im}(\xi)$ (or the values $\mathbf{x_{k1}}, \ldots \mathbf{x_{ks}}$) with highest probability(ies), that is for which the probability(ies) $p_k$= P($x_k$) or $p_{k1}$=P($x_{k1}$),...,$p_{ks}$=P($x_{ks}$) are maximal.

The **mode** of a *continuous* r.v. ξ is/are the local maximum place(s) $x\in$R of the density function $f(x)$ . □

**6.12. Definition:** The **median** ("medián") of a *discrete* r.v. ξ is the (unique) real number $m\in$R with the following property:

- if there is *no* number $x\in$R satisfying F($x$) = ½ , then we let $m$ to be the *smallest* number for which F($m$) > ½ ,
- if there *are* number(s) $x\in$R satisfying F($x$) = ½ , then the set of these numbers $x$ forms an interval (since F($x$) is a step function), so we let $m$ to be the centre ("középpont") i.e. the mean of this interval.

The **median** of a *continuous* r.v. ξ is the solution of the equality F($x$)= ½ , or the centre of the interval of the solution set. □

Informally, we get the *median* if we order the data in increasing order and choose the one in the *middle* of this row ("medián=középen áll") .

In the following three Sections we discuss some kinds of r.v. which often arise in practice and in theory, and now we give useful formulae for them. The Table after Section 9 collects the main results.

# 7. Special discrete random variables

In what follows Im(ξ)={$x_1$,...,$x_n$} or Im(ξ)={$x_1$,...,$x_n$,...} and we use the abbreviation for any natural number $k\in$N :

$$p_k := P(\xi = x_k) \; .$$

## Discrete uniform random variables

**7.1. Definition:** ξ is **discrete uniform r.v.** ("diszkrét egyenletes v.v.) if Im(ξ)={$x_1$,...,$x_n$} is an arbitrary finite set and ξ results each $x_i$ with the same probability, i.e. $p_i$ =P(ξ=$x_i$) = 1/$n$ for each $i$=1,...,$n$ . □

**7.2. Examples:** rolling with a fair dice, the last digit of the sum in a shop, we choose a random number between 1 and $n$ , we draw randomly one from the lottery numbers $\{x_1,...,x_n\}$ , we roll a pencil (prism with base of regular $n$-gon) having $n$ identical sides, etc.

**7.3. Statement:**

$$M(\xi) = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x_1 + x_2 + ... + x_n}{n} \quad \text{and} \quad D(\xi) = \sqrt{M(\xi^2) - M^2(\xi)} = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n} - \left(\frac{\sum_{i=1}^{n} x_i}{n}\right)^2} \qquad \square$$

(This is only the defintion, checking is an easy homework.)

## Hypergeometrical random variables

**7.4. Definition**: Let $N,S,n \in \mathrm{N}$ be fixed arbitrary natural numbers, $N \geq 2$, $0 \leq S \leq N$ , $1 \leq n \leq S$ , $0 \leq N - S \leq n$ . Then $\xi$ is a **hypergeometrical r.v.** ("hipergeometrikus eloszlású v.v.") *with the parameters* $S,N,n$ if the possible values of $\xi$ are $\mathrm{Im}(\xi) = \{0,1,\ldots,n\}$ and

$$p_k = P(\xi = k) = \frac{\binom{S}{k}\binom{N-S}{n-k}}{\binom{N}{n}} \quad . \qquad \square$$

where $\binom{\cdots}{\cdots}$ denote the binomial coefficients.

**7.5. Statement:** The experiments of the below type, which we call **sampling without repetition** ("visszatevés/ismétlés nélküli mintavételek"), are hypergeometrical r.v.: Suppose that we have $N$ many elements in a set (box), S many of them differs from the others (e.g. waste and good products, red and white balls, etc.). We choose $n$ many elements randomly from this set (e.g. blindly), *without* putting back any of them, i.e. we may draw these $n$ many elements at the same time. Now let us count the number of the elements among the drawn elements, differing from the others (from the S -many) and let $\xi$ denote this number. Now $\xi$ is a hypergeometric r.v. (Justifying the above statement is an easy homework, worth calculating it.) □

**7.6. Examples:**

**7.7. Theorem:** $M(\xi) = n \cdot \frac{S}{N}$ $(= np)$ and $D(\xi) = \sqrt{np(1-p)\left(1 - \frac{n-1}{N-1}\right)}$ where $p = \frac{S}{N}$

The term $\sqrt{1 - \frac{n-1}{N-1}}$ is called the **correcting term** ("korrekciós tényező"). □

**7.8. Theorem:** If $n$ and $k$ are constants, $S \to \infty$ , $N \to \infty$ , $\frac{S}{N} \to p$ =constant, then

$$\frac{\binom{S}{k}\binom{N-S}{n-k}}{\binom{N}{n}} \to \binom{n}{k} p^k (1-p)^{n-k} \qquad \square$$

**7.9. Explanation:** The above theorem says, that for large $N$ hypergeometric r. variables are close to the binomial distributions (see next subsection), which describes sampling *with* repetition. This corresponds to the empirical (practical) and theoretical experience that for large datasets ($N$) the repetiting or not (with or without) makes no big difference in sampling.

## Binomial or Bernoulli random variables

**7.10. Definition:** $\xi$ is a **binomial** ("binomiális") or **Bernoulli r.v.** *with the parameters* $n,p$ if $n\in\mathrm{N}$ , $n\neq 0$ is a fixed any nonzero natural number, $0<p<1$ is any fixed real number, the possible values of $\xi$ are (!): $\mathrm{Im}(\xi) = \{x_0,x_1,...,x_n\} = \{0,1,2\ldots,n\}$ and $\xi$ has the following distribution: for any k=0,1,2…,$n$ we have

$$p_k = \mathrm{P}(\xi = \mathrm{k}) = \binom{\mathrm{n}}{\mathrm{k}} p^k (1-p)^{n-k}$$

where $\binom{\mathrm{n}}{\mathrm{k}}$ is the binomial coefficient. □

**7.11. Theorem:** The experiments of the below type, which we call **sampling with repetition** ("visszatevéses/ismétléses mintavételek"), are binomial r.v.: Fix an experiment ($\Omega$) and an event $A\subseteq\Omega$ , fix further an arbitrary natural number $n\in\mathrm{N}$ , $n\neq 0$ , and let $p:=\mathrm{P}(A)$. Now repeat the experiment ($\Omega$) $n$ -many times, under completely the same conditions and independently from each other, and count to $\xi$ how many times the event $A$ occured during these $n$ -many experiments. In this case $\xi$ will be a binomial r.v. with the parameters $n,p$ .

Checking this statement is any easy and useful homework. □

Let us emphasize, that the conditions required above ("under completely the same conditions and independently from each other") are hard to fulfil in the practice in general, so they can be only approximated by binomial r.v.-s.

**7.12. Examples:**

a) Roll a dice (either fair or unfair) 10 times and let $\xi$ be the number of times we had 6. Now $\xi$ is a binomial r.v. with parameters $n$=10, $p$=1/6 (if the dice is fair). If we roll 10 fair dices at the same time and we count the dices showing 6 - we get the same r.v. as $\xi$ before.

b) Roll 7 times a coin (either fair or unfair) and let $\xi$ be the number of heads we have. Now $\xi$ is a binomial r.v. with parameters $n$=7, $p$=0.5 (if the coin is fair). The same r.v. results when rolling 7 coins simultaneously.

c) *Sampling with repetitions* ("*with putting back*", "ismétléses/visszatevéses mintavétel"): Let a box (set) contain of $N$ many objects (elements) total, from which S many differ from the others (e.g. are waste). Pick $n$ many times one object from the box, i.e. one at a time, randomly, remember to its type (waste or not), and before the next drawing put it back to the box. The numbers $N$, S and $n$ are fixed. Let $\xi$ denote the number of waste elements we have drawn during the above process. Then $\xi$ has the binomial distribution with parameters $n$ and $p$=S/$N$ .

d) Choose 5 cards from the Hungarian deck with repetitions, $\xi$ is the number of hearts we have drawn. Now $\xi$ has the binomial distribution with parameters 5 and $p$= 8/32 = 1/4 .

**7.13. Statement:** If $\xi$ is binomial r.v. then $M(\xi) = np$ and $D(\xi) = \sqrt{np(1-p)}$ . □

**7.14. Statement:** If $\xi$ is binomial r.v. then the mode(s) of $\xi$ is(are) $m$ ($m_1$ , $m_2$), where

$m = [(n+1)p]$ if $(n+1)p$ is not an integer,

$m_1 = (n+1)p$ and $m_2 = (n+1)p-1$, if $(n+1)p$ is an integer,

where $[x]$ denotes the **integer part** or **rounding downwards** ("egész rész, lefelé kerekítés, csonkítás") of the real number $x$ . □

**7.15. Theorem:** If $\xi_1$ is a binomial r.v. with the parameters $n_1$ , $p$ , and if $\xi_2$ is binomial r.v. with the parameters $n_2$ and (the same!) $p$ , and moreover $\xi_1$ and $\xi_2$ are independent, then their sum $\xi_1 + \xi_2$ is also a binomial r.v. with the parameters $n = n_1 + n_2$ and $p$ . □

**7.16. Theorem:** If $n \to \infty$ and $p \to 0$ such that $np = \lambda$ = constant, then we have the limit

$$p_k = \binom{n}{k} p^k (1-p)^{n-k} \to \frac{\lambda^k}{k!} e^{-\lambda} \ .$$ □

The latter result, among others, helps us to calculate correctly $p_k$ . Since in general $n$ and $k$ are large numbers, so the binomial coefficient is much larger, while, in the meantime $p$ and even $p^k$ is extremaly small. Though, in theory, it is not a problem, but in practice, when we have to calculate concrete numerical values either with a pocket calculator or with a computer, we will have a significant error when calculating $p_k$ . So, instead of $p_k$ we calculate the quantity on the right side. Poisson distributions are based on the above Theorem.

## Poisson random variables

**7.17. Definition:** The r.v. $\xi$ has **Poisson** distribution ("Poisson eloszlás") with the *parameter* $\lambda > 0$ , if it has possible values (outcomes) $Im(\xi) = \mathrm{N} = \{0,1,2,\ldots,n,...\}$ , the set of *all* natural numbers and the distribution is (where $e \approx 2{,}71828...$ is the Euler-number and $k \in \mathrm{N}$) :

$$p_k = P(\xi = k) = \frac{\lambda^k}{k!} e^{-\lambda} \ .$$ □

**7.18. Applications:** As we mentioned at the end of the previous Section, one application is the approximation of binomial distributions, when $n$ is large, $p$ is small and $np$ is about $\lambda$ . We present an example of this application at the end of Section 10.

In the practice the following types of problems (experiments) have Poisson distributions. Fix in advance a "*physical set*": a region either in the plane (area) or in the space (volume) or an interval in 1 dimension, mainly a time-interval, we *call this* physical set simply a **"volume V"** ("V térfogat"). Suppose further that inside this set there are, or there may be occur many, independent phenomena ("jelenség"), but the probability of having more of them decreases as $1/n^2$ . Then a mathematical theorem ensures that the *number* of the phenomenons has Poisson distribution. This description of Poisson r.v. may throw some light to the relationship between binomial and Poisson r.v. and Theorem 7.16. □

**7.19. Examples: 1-**dimensional: the number of clients / telephone calls / meteors / bricks falling to my head / ... in a *given* (i.e. fixed) time intervallum; the number of misprints in a given number of pages of a given book,

**2-**dimensional: the number of errors in a given area of paper / textile,

**3-**dimensional: the number of raisins in a given volume of cake, the number of errors in a given amount (volume) of a material, number of pebbles in a volume of clay, number of stars in a part of the space, number of dust/pollen particles in a volume of air, etc.

general: number of errors in a material (tube, board, volume),

**7.20. Theorem:** If $\xi$ is Poisson r.v., $\lambda>0$ then $M(\xi)=\lambda$ és $D(\xi)=\sqrt{\lambda}$ . □

**7.21. Theorem:** The mode(s) of $\xi$ is(are):

$[\lambda]$ if $\lambda$ is *not* an integer,

$\lambda$ and $\lambda-1$ if $\lambda$ is an integer. □

**7.22. Theorem:** If $\xi_1$ is a Poisson r.v. with parameter $\lambda_1$ , $\xi_2$ is a Poisson r.v. with parameter $\lambda_2$ , further $\xi_1$ and $\xi_2$ are independent, then the r.v $\xi=\xi_1+\xi_2$ is also a Poisson r.v. with the parameter $\lambda=\lambda_1+\lambda_2$ . □

**Corollary:** For any Poisson r.v. $\xi$ on the volume V with parameter $\lambda$ and for any $t\in R^+$ the **restriction/extension** ("leszűkítés/kiterjesztés") of $\xi$ to the volume V/t results again a Poisson r.v. with parameter $\lambda/t$ .

This phenomenon is called that the Poisson r.v. can be **unboundedly divided** ("korlátlanul osztható"). □

## Geometrical random variables

**7.23. Definition:** The r.v. $\xi$ has **geometrical distribution** ("geometriai/mértani") with the parameter $p$ ($0<p<1$) if it has possible values (outcomes) $Im(\xi) = N\backslash\{0\} = \{1,2,\dots,n,...\}$ , the set of *all* natural numbers except 0, and the distribution is ($k\in N$):

$$p_k = P(\xi=k) = p\cdot(1-p)^{k-1} \ . \quad □$$

**Remark:** Writing the widely used shortening $q:=1-p$ we get $p_k=p\cdot q^{k-1}$, a geometrical sequence, and this formula explains why these r.v. are called geometrical.

**7.24. Theorem:** In the practice the following types of problems are geometric ones. Let us fix an experiment ($\Omega$) and an event $A\subseteq\Omega$ . Let us repeat the experiment one or several times, independently from each other and under the same conditions, *until* the event *A* first occurs, when we stop repeating the experiment. Let $\xi$ denote the *number* of experiments we made, including the last (successfull) one ("*untill A*"). Then $\xi$ is a geometrical r.v. with parameter $p = P(A)$, supposing $0<p<1$. □

The proof of this theorem is an easy excercise, left to homework.

Let us emphasize, that geometrical r.v. occur when we do *not know* in advance the number of the neccessary experiments when starting, while, on the contrary, at binomial and hipergeometric r.v. we had to decide and *fix* the number of the experiments before starting!

**7.25. Examples:** the jug goes to the fountain until it brokes (Hungarian proverb: "addig jár a korsó a kútra, amíg el nem törik"),
we make exams until we pass,
we hit the nut until it brokes,
we shot to the enemy until we succeed,
the duell concludes until the first drop of blood,
we (try to) jump the stream until our clothes remain dry,
the girls in the harem pass before Sindbad until he chooses, ... .

**Remark**, that in the real life, in most of the above examples the requirements "*independently from each other and under the same conditions*" are not fulfilled. Think on your 2nd and 3rd exams: we make more excercies, get more nervous, crib better, or the nuts get tired on many previous hits, or Sindbad gets tired/excited, etc. In these cases either we talk about an approximation of the real problem, or, we think an idealized "theoretical" or "school" ("elméleti/iskolai") version of the problem.

See also Theorem 8.14. □

**7.26. Theorem:** If $\xi$ is a geometrical r.v. with parameter $p$, then

$$M(\xi) = \frac{1}{p} \quad \text{and} \quad D(\xi) = \frac{\sqrt{1-p}}{p} \quad .$$ □

In practical applications we often meet the question "*how many experiments are needed with 90% success* ?"

**7.27. Statement:** $P(\xi \le k) = 1-(1-p)^k$ .

**Proof:** (this easy calculation is worth reading carefully, $q=1-p$):

$P(\xi \le k) = P(\xi=1) + P(\xi=2) + ... + P(\xi=k) = p + p\cdot q^1 + p\cdot q^2 + ... + p\cdot q^{k-1} = p\cdot \dfrac{q^k-1}{q-1} = p\cdot \dfrac{q^k-1}{-p}$

$= 1-q^k$ . □

**7.28. Theorem:** If $\xi$ is geometrical r.v. then

$$P(\xi > m+n \mid \xi > m) = P(\xi > n) \; ,$$

i.e. the previous experiments do not have any effect to the forthcoming ones, this is called $\xi$ is **not turning younger** ("nem fiatalodó").

**Proof:** $P(\xi > k) = \sum_{i=k+1}^{\infty} P(\xi = i) = \sum_{i=k+1}^{\infty} p(1-p)^{i-1} = p(1-p)^k \dfrac{1}{1-(1-p)} = (1-p)^k$ .

$m+n > m$ implies $P(\xi > m+n \cap \xi > m) = P(\xi > m+n) = (1-p)^{m+n}$ ,

so $P(\xi > m+n \mid \xi > m) = \dfrac{P(\xi > m+n)}{P(\xi > m)} = (1-p)^n$ . □

# 8. Special continous random variables

## Continuous uniform random variables

**8.1. Definition** For any fixed real numbers $a,b\in\mathrm{R}$ , $a<b$ the r.v. $\xi$ has a **continuous uniform** ("folytonos egyenletes") distribution with parameters $a,b$ (on the interval $[a,b]$) if its density function is

$$f(x)=\begin{cases} c & if \quad a\le x\le b \\ 0 & otherwise \end{cases} .$$ □

This subsection contains *easy* statements with easy or obvious proofs, the calculations are useful for everybody.

**8.2. Statement:** $c=\dfrac{1}{b-a}$ and $F(x)=\begin{cases} 0 & if \quad x\le a \\ \dfrac{x-a}{b-a} & if \quad a\le x\le b \\ 1 & if \quad x\ge b \end{cases}$ □

**8.3. Statement:** $M(\xi)=\dfrac{a+b}{2}$ and $D(\xi)=\dfrac{b-a}{\sqrt{12}}$ . □

**8.4. Theorem:** In the practice the following type of problems are uniform continuous r.v.: choose randomly, independently from each effect, "*uniformly*" a real number $x$ in the interval $[a,b]$ . □

The above and the forthcoming theorems and statements justify the strong connection between the uniform continous r.v. and the geometrical probability fields (see second half of Section 3)! □

**8.5. Examples:** rolling a cylindrical pen or bottle on which point of its surface it stops,
cutting randomly a ribbon/paper strip / wooden stick into two parts,
I go to the bus stop randomly (I do not kow the time table) but the buses come regularly and exactly in each 15th minute,
we choose a random number in [0,1] with a fixed 4th digit, etc.

**8.6. Statement:** $P(c<\xi<d)=\dfrac{d-c}{b-a}$ is proportional to the length of the sub-interval. □

## Exponential random variables

**8.7. Definition:** The r.v. $\xi$ has **exponential distribution** ("exponenciális eloszlású") with the parameter $\lambda\in\mathrm{R}^+$, $\lambda>0$ if its density function is

$$f(x)=\begin{cases} \lambda e^{-\lambda x} & if \quad x\ge 0 \\ 0 & otherwise \end{cases} .$$ □

**Remark:** In the practice the following random variables (measurings) have exponential distributions: life times of machines, tools, non living constructions and objects (like bulbs, wheels, computer hardware, etc.), or the length of working periods (e.g. telephone calls, repairing bikes). However, this is not a theoretical theorem, but a lot of statistical experiments (measurings) and hypothesis were justified by the rigorous method (with theoretical proof) *goodness of fit* ("illeszkedésvizsgálat"). This method is not included in this semester. □

**8.8. Theorem:** For each exponential r.v. $\xi$ with parameter $\lambda$ the distribution function is

$$F(x)=\begin{cases}0 & if \quad x<0 \\ 1-e^{-\lambda x} & if \quad x\geq 0\end{cases}.$$ □

**8.9. Theorem:** $$M(\xi)=D(\xi)=\frac{1}{\lambda}\ .$$ □

**8.10. Theorem:** If $\xi$ is exponential r.v. and $x\geq 0$, $y\geq 0$ are arbitrary positive real numbers, then

$$P(\xi\geq x+y\,|\,\xi\geq y)=P(\xi\geq x)\ .$$

This equality is called **everyoung / not turning to older** property ("örökifjú, nem öregedő").□

Justifying the above theorem is quite trivial, worth to do as a homework.

The above equality says (look more carefully) that the requirement: " $\xi$ works *x* long from now (*y*), i.e. until *x*+*y* " is independent of *y* , moreover *y* can be eliminated (or say, *y*=0), i.e. the object can be considered as a totally new: without any assumption (the unconditional probability on the right side). Compare this theorem also to Theorem 7.28.

The following theorem describes the importance ot the everyoung property in the theory of the exponential r.v.

**8.11. Theorem:** If $\xi$ is continuous, F(0)=0, F(*x*)<1 , F has derivative for each nonnegative real number *x* , $\lim_{0+} F'(x)=\lambda>0$ and $\xi$ is everyoung, then $\xi$ must be exponential r.v. □

**8.12. Theorem:** The connection between exponential and *Poisson* r.v. :
Let $\xi_1,\xi_2,....$ be independent exponential r.v. with the *same* parameter $\lambda_\xi$ , let *T*>0 be a fixed real number and denote $\eta$ the (new) random variable: the number of the objects which went wrong until the time point *T* . Of course we change each bad object immediately it went wrong. In mathematical form: let Im($\eta$)=N and

$\eta=0$ if $\xi_1>T$ and let $\eta=1$ if $\xi_1\leq T$ but $\xi_1+\xi_2>T$ ,

and in general:

$$\eta=k \quad \text{if} \quad \sum_{i=1}^{k}\xi_i\leq T \quad \text{but} \quad \sum_{i=1}^{k+1}\xi_i>T\ .$$

In this case $\eta$ is a Poisson r.v. with parameter $\lambda_\eta=\lambda_\xi\cdot T$ . □

**8.13. Example:** The life time of a (fixed type) bulb is exponential r.v. $\xi$ with average life time 1000 hours. When the bulb goes wrong we change it immediately to the same type. What probability we need to change at least 2 bulbs during 2500 hours, i.e. P($\eta\geq$2)=?

**Solution:** M(ξ)=1000, Theorem 8.9 implies $\lambda_\xi = {}^1/_{1000}$ , by the above theorem $\lambda_\eta = {}^{2500}/_{1000} = 2.5$ , P(η≥2) = 1-P(η<2) = 1-P(η=1)-P(η=1) = 1 - $e^{-2.5}$ · **(**$(2.5)^0/0! + (2.5)^1/1!$**)** ≈ <u>0.7127</u> .

**8.14. Theorem:** The connection between exponential and *geometrical* r.v.:
If $\xi$ is exponential r.v. with parameter $\lambda$ then $\eta = [\xi] + 1$ is geometrical r.v. with parameter $p = 1 - e^{-\lambda}$ , $q$=1-$p$=$e^{-\lambda}$ .
**Proof:** The values of $\eta$ are $Im(\eta)$={1,2,3,...} and

$$P(\eta = k) = P(k-1 \le \xi < k) = F(k+1) - F(k) = 1 - e^{-\lambda k} - (1 - e^{-\lambda(k-1)}) = e^{-\lambda(k-1)}(1 - e^{-\lambda}) = q^{k-1} \cdot p \ . \quad \square$$

**Explanation:** $\eta$ denotes, e.g. that "how many *integer* hours does the bulb work", i.e. how many hours have I to wait (repeating) to destroy the bulb (until it goes wrong). □

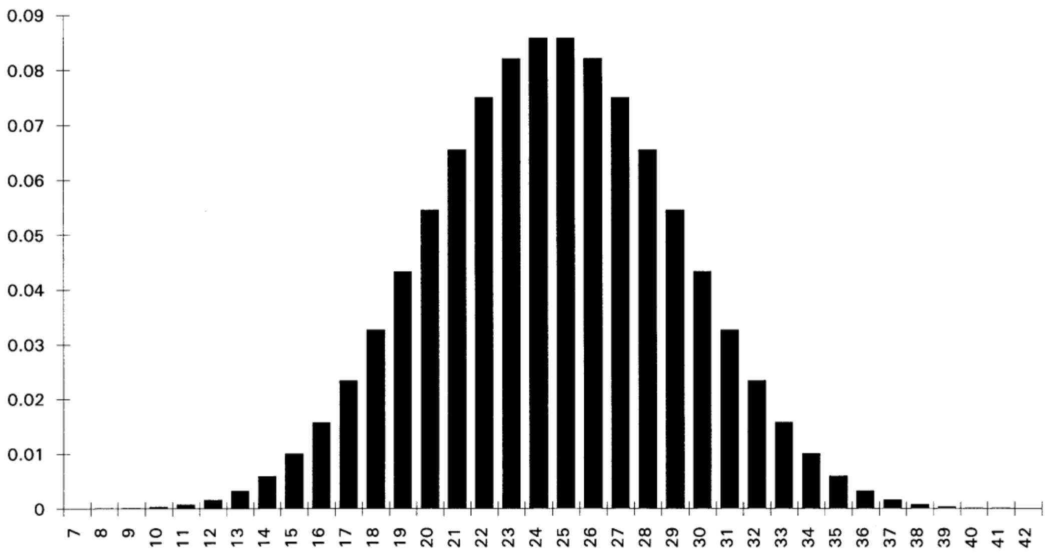
**8.15. Example:** The telephone company *A* counts the time *continuously* and the unit price is 20 HUF/minute. Company *B* makes bill for each minute you started, 15 HUF/minute. E.g. if you phoned for 2 min 21 sec = 2 $^{21}/_{60}$ = 2.35 min, you should pay 2.35·20= 47 HUF to company *A*, or 3·15=45 HUF to company *B*. What company you have to choose if your wife's calling length is an exponential r.v. with average 2 min ?
**Solution:** Denote $\xi$ the length of the call, $\theta_1$ and $\theta_2$ the prices for company *A* and *B* respectively. Then

$$M(\theta_1) = 20M(\xi) = 40 \text{ HUF} \quad \text{and} \quad M(\theta_2) = 15 \cdot (M[\xi] + 1) = 15 \cdot \frac{1}{1 - e^{-0.5}} = 38.1 \text{ HUF} . \quad \square$$

# 9. The normal distribution

**9.1. Introduction:** Having investigated carefully many physical and other practical quantities measuring statistical large tables, Gauss oberved that the histograms (column diagrams) of these tables can be approximated by the graph of the transformations of the function az $e^{-x^2}$ . These r.v. are called having normal or Gaussian distribution. In general, those quantities are normal, which arise as the sum of many (1000 or more) +/- effects which are very small each, e.g. height or weight or volume of a pencil, a man, woman, animal, volume of the rain, temperature, voltage or power of a battery, etc.

There are many experiments illustrating this assumptions, you can find some on my webpage: sum of several dices (see also Figure below), Galton board, etc. These formulas (assumptions) are justified in the *Central Limit Theorem* 10.5 (see in next Section) and the method *goodness of fit* , not included in this semester.




**Sums of 7 dices**
http://math.uni-pannon.hu/~szalkai/7kocka.gif

**9.2. Definition:** $\xi$ is a **standard normal r.v.** if its density function is

$$f(x)=\frac{1}{\sqrt{2\pi}}e^{\frac{-x^2}{2}} \qquad (x\in\mathrm{R})\ . \qquad \square$$

**9.3. Theorem:** $M(\xi)=0$ and $D(\xi)=1$ . □

**9.4. Remark:** It is worth studying the formulas and the graphs of the function $e^{-x^2}$ and its linear transformations!

By Liouville's theorem there is *no* formula for the primitive function of $e^{-x^2}$ . This is why we have to use table for $F(x)=\Phi(x)$, a small table is included at the end of this booklet.

**9.5. Notation:** $$\Phi(x):=F(x)=\int_{-\infty}^{x}f(t)dt \qquad (x\in\mathrm{R})\ . \qquad \square$$

Clearly $\Phi(x)$ is the distribution function $F(x)$ for the standard normal r.v.

**9.6. Statement:** $\Phi(-x)=1-\Phi(x)$ for any $x\in\mathrm{R}$ . □

**Corollary:** $-\xi$ is also a standard normal r.v.

**Proof:** $F(x)=P(-\xi<x)=P(-x<\xi)=1-\Phi(-x)=1-(1-\Phi(x))=\Phi(x)$ . □

**9.7. Definition:** Let $\xi$ standard normal r.v. and let $m,\sigma\in\mathrm{R}$ be any real numbers, $\sigma>0$ . Then the r.v. $\eta=\sigma\cdot\xi+m$ is called **normal r.v. with parameters** $m,\sigma$ and denote this fact by $\eta\sim N(m,\sigma)$ . □

The below theorem gives an alternative to the above definition for normal r.v.:

**9.8. Theorem:** The density function of $\eta$ is $$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{\frac{-(x-m)^2}{2\sigma^2}} \qquad (x\in\mathrm{R})\ . \qquad \square$$

**9.9. Statement:** $M(\eta) = m$ and $D(\eta) = \sigma$ . □

**9.10. Theorem:** $\eta \sim N(m,\sigma)$ implies $F_\eta(x) = F_{m,\sigma}(x) = \Phi\left(\frac{x-m}{\sigma}\right)$ . □

**9.11. Theorem:** "*k -times σ (sigma) rule*" ("*k*-szor σ szabály"):

For $\xi \sim N(m,\sigma)$ and $k \in R^+$ $P(m - k\sigma < \xi < m + k\sigma) = 2\Phi(k) - 1$ . □

**Meaning:** The probability that the values of $\xi$ are at most $k\sigma$ far from the average *m* is exactly $2 \cdot \Phi(k)-1$ (look at the table).
The $k*\sigma$ *rule* can be easily checked by 9.10. and 9.6., this calculation is an easy homework.
Below we give some special cases of the $k*\sigma$ *rule* :

**9.12. Theorem:** Special cases:

$k=1 \Rightarrow P(m - \sigma < \xi < m + \sigma) = 0.68$ ,

$k=2 \Rightarrow P(m - 2\sigma < \xi < m + 2\sigma) = 0.95$ ,

$k=3 \Rightarrow P(m - 3\sigma < \xi < m + 3\sigma) = 0.997$ . □

**9.13. Theorem:** If $\xi \sim N(m,\sigma)$ and $\eta = a\xi + b$ , $a \neq 0$ then $\eta \sim N(am + b\, ,\, |a|\sigma)$ . □

**9.14. Theorem:** If $\xi \sim N(m_1,\sigma_1)$ and $\eta \sim N(m_2,\sigma_2)$ are independent r.v. then

$$\xi + \eta \sim N(m_1 + m_2\, ,\, \sqrt{\sigma_1^2 + \sigma_2^2}\,).$$ □

**Remark:** We already know from the general Theorems 6.4. and 6.10. that the mean and dispersion of $\xi+\eta$ are $m_1+m_2$ and $\sqrt{\sigma_1^2 + \sigma_2^2}$ . The above Theorem adds that the *sum* of normal distributions is *again a normal* one. □

The below theorem is often used in practical applications and follows from the above one:

**9.15. Theorem:** If $\xi_1, \xi_2, ....\xi_n \sim N(m,\sigma)$ are independent normal r.v. with the same distribution, then

$$\sum_{i=1}^{n} \xi_i \sim N(nm, \sigma\sqrt{n}) \quad \text{and} \quad \frac{\sum_{i=1}^{n} \xi_i}{n} \sim N\left(m, \frac{\sigma}{\sqrt{n}}\right) .$$ □

**Remark:** We have already mentioned several times, that in the practice (not only in the laboratory) we have to measure anything several times and calculating average, instead of a single measuring. This method decreases the dispersion and makes the measuring more precise. It is useful to know, that both the sum, difference and the average of the measurements are also *normal distribution* r.v. □

**9.16. Example:** The weight of an adult is a normal r.v. with mean 75 kg and dispersion 15kg, children at school have weight in mean 35kg and dispersion 6kg. If we consider the two r.v. independent, then what is the probability of

**a)** that an adult is heavier than a child,

**b)** the weight of an adult and a child together is between 80kg and 140 kg?

**9.17. Example:** We plan an elevator (lift) for 8 person, the weight of the people is a r.v. with mean 75kg and dispersion 15kg, can be considered to be independent. How strong we have to plan the elevator that would be able to lift 8 person with 99% probability?

**Solution:** $\xi_i \sim N(75,15)$, $i$=1,...,8, $m$=75, $\sigma$=15 and let $\eta := \sum_{i=1}^{8} \xi_i$ . By Theorem 9.15 we have $\eta \sim N(8\cdot m, \sqrt{8}\cdot\sigma) = N(8\cdot 75, \sqrt{8}\cdot 15) = N(600, 42.43)$. The question is to find an $x \in R$ such that $P\left(\sum_{i=1}^{8} \xi_i < x\right) = P(\eta < x) = 0.99$. Since $P(\eta<x) = F_\eta(x) = \Phi\left(\frac{x-600}{42.43}\right) = 0.99$ and from the table of $\Phi$ we have $\Phi(2.32)=0.99$, so $\frac{x-600}{42.43} = 2.32$ and so $x$=698.5 ~ 700 kg . □

## Other random variables, derived from normal distributions

The following Theorems will be used in mathematical Statistics.

**9.18. Theorem:** If $\xi \sim N(0,1)$ then the distribution and density functions of $\eta = \xi^2$ are:

$$F_\eta(x) = \begin{cases} 0 & \text{if} \quad x \le 0 \\ 2\Phi(\sqrt{x}) - 1 & \text{if} \quad x > 0 \end{cases}, \quad f_\eta(x) = \begin{cases} \frac{1}{\sqrt{2\pi x}} e^{-\frac{x}{2}} & \text{if} \ x > 0 \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad M(\eta)=1 .$$

**Proof:** $F_\eta = P(\eta<x) = P(\xi^2<x) = P(-\sqrt{x}<\xi<\sqrt{x}) = \Phi(\sqrt{x})-\Phi(-\sqrt{x}) = \Phi(\sqrt{x})-(1-\Phi(\sqrt{x})) = 2\Phi(\sqrt{x})-1$, $f_\eta = F'_\eta = \ldots$ . □

**9.19. Theorem:** For any $n \in N$ and independent r.v. $\xi_1,\ldots,\xi_n \sim N(0,1)$ the density functions of $\eta_n = (\xi_1)^2+\ldots+(\xi_n)^2$ is $f_\eta(x) = cx^{\frac{n}{2}-1} e^{-\frac{x}{2}}$ for $0<x$ . □

**9.20. Definition:** The above $\eta_n$ is called $\chi^2$ i.e.**chi-squared** ("khí-négyzet") r.v.of parameter $n$. □

**9.21. Definition:** For any $n \in N$ and independent r.v. $\zeta, \xi_1,\ldots,\xi_n \sim N(0,1)$ the r.v. $\theta_n = \frac{\zeta}{\sqrt{\frac{\sum_{i=1}^{n} \xi_i^2}{n}}}$ is called **Student-** or **t- distribution** of parameter $n$. □

**9.22. Theorem:** The density function is

$$f_\theta(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\cdot\Gamma\left(\frac{n}{2}\right)\cdot\left(1+\frac{x^2}{n}\right)^{\frac{n+1}{2}}}$$

where $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$ is the **gamma** function (eg. $\Gamma(n)=(n-1)!$ ).

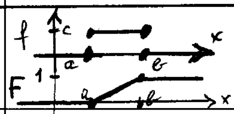
**9.23. Theorem:** If $\xi \sim N(0,1)$ then the distribution and density functions of $\eta = e^\xi$ , the so called **lognormal r.v.** ("lognormális eloszlás") are:

$$F_\eta(x) = \begin{cases} 0 & \text{if} \quad x \le 0 \\ \Phi(\ln x) & \text{if} \quad x > 0 \end{cases}, \quad f_\eta(x) = \begin{cases} \frac{1}{\sqrt{2\pi}} \frac{1}{x} e^{-\frac{(\ln x)^2}{2}} & \text{if} \ x > 0 \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad M(\eta)=\sqrt{e} .$$

**Proof:** $F_\eta = P(\eta<x) = P(e^\xi<x) = P(\xi<\ln(x)) = \Phi(\ln(x))$, $f_\eta = F'_\eta = \ldots$ . □

The following table can be found on my webpage in higher resolution:
http://math.uni-pannon.hu/~szalkai/Eloszlasok(pdf)+kezjav+.gif

## Nevezetes eloszlások

| | Név | Paraméter | Definíció | $M=\mu$ | $\sigma^2$ | Alkalmazás |
|---|---|---|---|---|---|---|
| *diszkrét eloszlások* | Indikátorváltozó | $p$ | $p_1 = p,\ p_0 = 1-p$ | $p$ | $p(1-p)$ | Egy $p$ valószínűségű esemény bekövetkezik-e? |
| | Binomiális eloszlás | $n, p$ | $p_k = \binom{n}{k} p^k (1-p)^{n-k}\ (k=0,1,\dots,n)$ | $np$ | $np(1-p)$ | Visszatevéses mintavétel. Egy $p$ valószínűségű esemény $n$ független ismétlés során hányszor következik be? |
| | Geometriai eloszlás | $p$ | $p_k = (1-p)^{k-1} \cdot p$ | $\frac{1}{p}$ | $\frac{(1-p)}{p^2}$ | |
| | Hipergeometrikus eloszlás | $N, K, n$ | $p_k = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}\ (k=0,1,\dots,n)$ | $\frac{Kn}{N}$ $= n \cdot p$ | $\frac{K}{N}\left(1-\frac{K}{N}\right)\frac{N-n}{N-1}$ $= n \cdot p \cdot q \left(\frac{N-n}{N-1}\right)$ | Visszatevés nélküli mintavétel. *Megjegyzés:* A $p = \frac{K}{N}$ jelöléssel $\mu$ és $\sigma$ egyszerűbb alakra írható. |
| | Poisson-eloszlás | $\lambda (>0)$ | $p_k = \frac{\lambda^k}{k!} e^{-\lambda}\ (k=0,1,\dots)$ | $\lambda$ | $\lambda$ | Egységnyi idő alatt megfigyelt események száma. Egységnyi területre eső (hibák) száma. |
| *folytonos eloszlások* | Standard normális eloszlás | — | $f(x) = \varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$ | 0 | 1 | Normális eloszlás standardizáltja. |
| | Normális eloszlás | $\mu, \sigma (>0)$ | $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$ | $\mu$ | $\sigma^2$ | Populáció egyedeinek méretei, tömegei. Gyártási folyamatban fellépő méretingadozások. Fizikai mennyiségek. |
| | Egyenletes eloszlás | $a, b$ |  $c = \frac{1}{b-a}$ $F(x) = \frac{x-a}{b-a}$ | $\frac{a+b}{2}$ | $\left(\frac{b-a}{\sqrt{12}}\right)^2$ | |
| | Exponenciális eloszlás | $\lambda (>0)$ | $f(x) = \lambda e^{-\lambda x}$ ha $x > 0$; $F(x) = 1 - e^{-\lambda x}$ ha $x > 0$ | $\frac{1}{\lambda}$ | $\frac{1}{\lambda^2}$ | Élettartam, várakozási idő. Időintervallumok a Poisson-eloszláshoz kapcsolódóan. |

# 10. Laws of large numbers

("Nagy számok törvényei")

The forthcoming formulas and inequalities give us *general approximations* for any type of r.v. The error of these approximations decreases when the number of measurements is high. Of course for specific r.v. these inequalities could be highly improved.

**10.1. Theorem: Markov-inequality** ("Markov egyenlőtlenség")

*If* M(ξ) *of the r.v.* ξ *does exist, then for every positive number* $a \in R^+$ *we have*

$$\boxed{P(\xi \geq a) \;\leq\; \frac{M(\xi)}{a}} \;. \;\square$$

**Explanation:** The above inequality says that the outcome of the measuring (ξ) may be large, but the *probability* of this is small. The higher of the measure is ($\xi \geq a$) the lower is its probability. When $a$ goes to infinity ($a \to \infty$) the probability of measuring higher than $a$ is smaller than M/$a$ , which tends to 0 .

**10.2. Theorem: Chebyshev-inequality** ("Csebisev egyenlőtlenség")

*If* M(ξ) *and D(ξ) of the r.v.* ξ *do exist, then for every positive numbers* $k, \varepsilon \in \mathrm{R}^+$ *we have*

$$\boxed{P\big(\, |\xi - M(\xi)| \geq kD(\xi) \,\big) \;\leq\; \frac{1}{k^2}}\,, \qquad \text{i.e.} \qquad \boxed{P\big(\, |\xi - M(\xi)| \geq \varepsilon \,\big) \;\leq\; \frac{D^2(\xi)}{\varepsilon^2}}\;,$$

*or, the negation of the event:*

$$\boxed{P\big(\, |\xi - M(\xi)| \leq \varepsilon \,\big) \;\geq 1 - \frac{D^2(\xi)}{\varepsilon^2}} \;. \;\square$$

**Explanation:** We think that the results of a measuring (ξ) are about, not far from the mean **(**M(ξ)**)**, the subformula |ξ-M(ξ)| calculates the distance of these quantities. Chebyshev proved, that the *probability* of large differences is small. More precisely (see the first two formulas): the difference |ξ-M| can be sometimes larger than $k \cdot$M(ξ) or ε , but with probability less than $1/k^2$ and $D(\xi)^2/\varepsilon^2$ , both these upper bounds go to 0 when $k$ and ε go to ∞ . On other hand, if D(ξ) is smaller, the upper bound $D(\xi)^2/\varepsilon^2$ is smaller, too. In other words (see the third formula): the probability of small distances is close to 1 , think on the case when $\varepsilon \to 0$ .

**10.3. Theorem: Law of large numbers by Bernoulli**

("Bernoulli-féle nagy számok törvénye")

*Let* $A \subseteq \Omega$ *be any fixed event with* P($A$)=$p$ *and* $q$:=1-$p$. *Denote* $\xi_n$ *the frequency of the event* (i.e. how many times occured) *A during we executed n many independent experiments. Then for any* $\varepsilon \in \mathrm{R}^+$ *and* $n \in \mathrm{N}$ *we have*

$$\boxed{P\left(\left|\frac{\xi_n}{n} - p\right| \geq \varepsilon\right) \leq \frac{pq}{\varepsilon^2 n}}\,, \qquad \text{i.e.} \qquad \boxed{P\left(\left|\frac{\xi_n}{n} - p\right| < \varepsilon\right) \geq 1 - \frac{pq}{\varepsilon^2 n}}\,.$$

In other words:

*For any* ε>0 *and* δ>0 *there is an integer number* $n_0$ ("treshold") such that for every $n > n_0$ :

$$\boxed{P\left(\left|\frac{\xi_n}{n} - p\right| \geq \varepsilon\right) \leq \delta}\,, \qquad \text{i.e.} \qquad \boxed{P\left(\left|\frac{\xi_n}{n} - p\right| < \varepsilon\right) \geq 1 - \delta}\,.$$

*where* $\delta = \dfrac{pq}{\varepsilon^2 n}$ , *so* $\delta \to 0$ *whenever* $n \to \infty$ . □

**Explanation:** Observe first, that $\xi_n/n$ is the *relative frequency* of the event *A*.
The theorem calculates the difference of the theoretical *probability* ($p$) and the empirical (practical, greek, "tapasztalati") *relative frequency* ($\xi_n/n$). To be more precise, the theorem calculates "only" the *probability* of large differences (see the left hand side formulas), and results that this probability is *small*, moreover this probability tends to 0 if $n$ (the number of experiments) goes to infinity, i.e. very large. This means (see also the right hand side formulas), that the statement "the difference is small" is valild with almost 100% probability, i.e. the negation "the difference is not small" can have probability less than any small number $\varepsilon$ . Though, these formulas do not say, that the relative frequency $\xi_n/n$ would converge to the (theoretical) probability $p$.

If we do *not* know $p$, then we can use the general inequality $p(1-p)\leq 1/4$ (valid for all $p\in\mathrm{R}$) and state $\delta\leq 1/4\varepsilon^2 n$ which also tends to 0 when $n\to\infty$ .

**10.4. Theorem: The weak law of large numbers by Chebyshev**
("A nagy számok gyenge törvénye, Csebisev-alak)"

*Let* $\xi_1 , \xi_2 , \ldots , \xi_n$ *be independent r.v. with the same distribution, which do have (the same) mean* $m:=\mathrm{M}(\xi_i)$ *and dispersion* $\sigma:=\mathrm{D}(\xi_i)$ *(i=1,2,…,n,... ). Let further* $S_n:= \xi_1+\xi_2+\ldots+\xi_n$ . *Then for every* $\varepsilon\in\mathrm{R}^+$ we have:

$$\boxed{P\left(\left|\frac{S_n}{n}-m\right|\leq\varepsilon\right)\geq 1-\frac{\sigma^2}{n\varepsilon^2}}, \quad \textit{equivalently} \quad \boxed{P\left(\left|\frac{S_n}{n}-m\right|>\varepsilon\right)\leq\frac{\sigma^2}{n\varepsilon^2}} \ . \ \square$$

**Explanation:** Since the r.v. $\xi_1 , \xi_2 , \ldots , \xi_n$ have the same distribution, the represent the repeated, independent measurements of the same quantity. So $S_n/n$ calculates the empirical (practical) arithmetical mean, i.e. the average. This means, that the formulas above calculate the difference of the *practical* average ($S_n/n$) and the *theoretical* mean ($m$). More precisely, the formulas count the *probability* of this difference ($|S_n/n - m|$) to be small ($\leq\varepsilon$) or not small ($>\varepsilon$). Observe, that for any fixed $\varepsilon>0$ the upper bound $\sigma^2/n\varepsilon^2 \to 0$ as $n \to \infty$ , i.e. when the number of experiments is large. So, by the left hand side formula, the statement "the difference is small" can be valid with almost 100% probability.

Bernoulli's Theorem 10.3. is a special case of the present formulas since in theorem 10.3. we have $m=\mathrm{M}(\xi_n/n)=p$ and $\sigma^2=pq$ .

**10.5. Theorem: The strong law of large numbers,** or **Central limit theorem**
("Nagy számok erős törvénye, vagy Központi (=centrális) határeloszlás tétel")

*Let* $\xi_1 , \xi_2 , \ldots , \xi_n$ *be independent r.v. with the same distribution, which do have (the same) mean* $m:=\mathrm{M}(\xi_i)$ *and dispersion* $\sigma:=\mathrm{D}(\xi_i)$ *(i=1,2,…,n,... ). If*

$$\varsigma_n=\frac{\xi_1+\ldots+\xi_n-n\cdot m}{\sigma\sqrt{n}} \ , \quad \textit{then} \quad \lim_{n\to\infty}P(\varsigma_n<y)=\Phi(y) \ . \qquad \square$$

**Explanation:** The r.v. $\zeta_n$ calculates the *standardized version* of the sum $\Sigma_n :=\xi_1+\xi_2+\ldots+\xi_n$ , i.e. $\zeta_n = (\Sigma_n-\mathrm{M}(\Sigma_n))/\mathrm{D}(\Sigma_n)$ because the original sum $\Sigma_n$ would go to $+\infty$ ($\lim_{n\to\infty} \Sigma_n = +\infty$) since all the terms $\xi_i$ have the same mean $m$ . However, the standardized $\zeta_n$ has $\mathrm{M}(\zeta_n)=0$ and $\mathrm{D}(\zeta_n)=1$ for each $n$. Now, the final conclusion of the theorem says that the limit of the distribution function of $\zeta_n$ is $\Phi$, i.e. the limit of $\zeta_n$ is the standard normal distribution! This limit was explained in Section 9 as "*the sum of many small +/- quantities always results a normal distribution*" .

**10.6. Moivre-Laplace Theorem:** *For any numbers* $0 \le p \le 1$ *and* $u, v \in \mathbb{R}$

$$\lim_{n\to\infty}\left(\sum_{u\le k\le v}\binom{n}{k}p^k q^{n-k}\right) = \Phi(v^*) - \Phi(u^*) \ ,$$

*or in simpler form*

$$\sum_{u\le k\le v}\binom{n}{k}p^k q^{n-k} \approx \Phi(v^*) - \Phi(u^*)$$

*where*

$$u^* = \frac{u-np}{\sqrt{npq}} = \frac{u-m}{\sigma} \quad \text{and} \quad v^* = \frac{v-np}{\sqrt{npq}} = \frac{v-m}{\sigma} \ . \quad \square$$

**Explanation:** Inside the summation (Σ) we see the formula of the (discrete) binomial distribution. This means, that repeating the same experiment, the binomial distribution can be approximated with a normal distribution. In fact, this theorem is a speial case of the central limit theorem. But how to understand and memorize the above formulas. The second, approximating formula is, in fact the well known one

$$P(u<\xi<v) = F(v)-F(u) = \Phi(v^*)-\Phi(u^*) = \Phi\left(\frac{v-m}{\sigma}\right) - \Phi\left(\frac{u-m}{\sigma}\right) .$$

**Summary:** We learned, that for large $n$ (larger than 30) hypergeometric distributions can be approximated by binomial ones, which, further can be approximated by Poisson r.v., and finally, all these can be approximated by normal distributions (r.v.).

The below example gives an important explanation of this summary!

**10.7. Main example:** We have $N$=1000 many animals in a farm, and a non epidemic illness kills each animal, independently with probability 0.3 . What is the probability that the number of surviving animals is between $\frac{N}{3}$ and $\frac{2N}{3}$ ?

**Solution:**

**Part 1.:** The problem is, in fact a binomial (Bernoulli) r.v.:

$N$=1000 , p=P(recovers)=0,7 , q=1-p=0,3 , ξ:=how many animals survives.

*Binomial r.v.:* $P(\xi=k)=\binom{n}{k}p^k(1-p)^{n-k}$ (k=0;1;…;$n$=1000)

$$P(334 \le \xi \le 666) = \sum_{k=334}^{k=666}\binom{1000}{k}0{,}7^k 0{,}3^{1000-k} \ .$$

Since there are many terms in the above calculations, too large and too small, we need an approximating method.

**Part 2.:** We use the Poisson approximation:

$P(\xi=k)=\frac{\lambda^k}{k!}e^{-\lambda}$ ($k$=0;1;…) , λ = np = 1000·0,7 = 700 , so

$$P(334 \le \xi \le 666) = \sum_{k=334}^{k=666}\frac{700^k}{k!}e^{-700} \ .$$

**Part 3.:** The above formula is still hard to calculate, we use the Moivre-Laplace theorem.

$\sum_{u\le k\le v}\binom{n}{k}p^k q^{n-k} \approx \Phi(v^*) - \Phi(u^*)$ , where: $u^*=\frac{u-np}{\sqrt{npq}} = \frac{u-m}{\sigma}$ and $v^*=\frac{v-np}{\sqrt{npq}} = \frac{v-m}{\sigma}$ ,

M=m=np= 1000·0.7= 700 , D=σ=$\sqrt{npq} = \sqrt{1000x0{,}7x0{,}3} = \sqrt{210} \approx 14.49$ ,

$$P(334 \le \xi \le 666) = \sum_{k=334}^{k=666}\binom{1000}{k}0{,}7^k 0{,}3^{1000-k} \approx \Phi\left(\frac{666-700}{\sqrt{210}}\right) - \Phi\left(\frac{334-700}{\sqrt{210}}\right) =$$

$= \Phi(-2{,}35) - \Phi(-25{,}26) = 1 - \Phi(2{,}35) - 0 = 1 - 0{,}99065 =$ **0.00935** . □

Appendix

# Probability theory - Mathematical dictionary

| **Probabiliy Theory** | **Mathematics** |
|---|---|
| Sample space (=all outcomes of an experiment) | $H \neq \emptyset$ (arbitrary) base **set** (or $\Omega$) |
| **outcome** of an experiment | $x \in H$ any **element** of H |
| **event** (actual outcome of an experiment) | $A \subseteq H$ (any) **subset** |
| elementary event | $\{x\} \subseteq H$ a subset consisting of a single element (singleton) |
| "A event happened/realized" | $x \in A$ |
| "A event did not happen/realized" | $x \notin A$ |
| sure event | $H \subseteq H$ (base set itself), or: if P(A)=1 |
| impossible event | $\emptyset \subseteq H$ (empty set), or: if P(A)=0 |
| negation of an event | $A^-$ (complement set) |
| sum of the events A+B ("or") | $A \cup B$ (union) |
| product of events $A \cdot B$ ("and") | $A \cap B$ (intersection) |
| difference of events A-B | A\B (difference) |
| excluding events | $A \cap B = \emptyset$ (disjont sets), or: if $P(A \cap B)=0$ |
| B follows from A ( A implies B) | $A \subseteq B$ (A is a subset of B) |
| | |
| **Probability** P(A) | **P**: $P(H) \to R$ any function ($P$(H) is the power set of H) with the axioms of Kolmogorov (like area) |
| independent events A,B | $P(A \cap B) = P(A) \cdot P(B)$ |
| complete system of events | a partition of H |
| random variable (numerical outcome of a measuring) | $\xi : H \to R$ any function |
| discrete random variable | $Im(\xi) = \{x_1, x_2, \ldots, x_n, \ldots\}$ any enumerable set |
| the distribution of the discrete r.v. | $\{p(x_1), p(x_2), \ldots, p(x_n), \ldots\}$ |
| continuous random variables | there is an $[a,b] \subseteq Im(\xi)$ interval |
| | |
| **distribution function** | $F : R \to R$ any function with the axioms<br>or: $F(t) := P(\xi < t)$<br>or: primitive function of $f$: $F(t) = \int_{-\infty}^{t} f(x)dx$ |
| **density function** | $f : R \to R$ any function with axioms<br>or: derivative function of F : $f(x)=F'(x)$ |
| $P(a \leq \xi < b) = \int_a^b f(x)dx = F(b)-F(a)$ | Newton-Leibniz 's Rule |
| mean or expected value $M(\xi) = E(\xi)$ | average, arithmetical mean |
| dispersion $D(\xi)$ | sprinkling/scattering of the measurings values |

## Table of the standard normal distribution function (Φ)

| X | Φ(x) | | z | Φ(x) | | x | Φ(x) | | x | Φ(x) |
|---|---|---|---|---|---|---|---|---|---|---|
| 0,00 | 0,5000 | | 0,34 | 0,6331 | | 0,68 | 0,7517 | | 1,02 | 0,8461 |
| 0,01 | 0,5040 | | 0,35 | 0,6368 | | 0,69 | 0,7549 | | 1,03 | 0,8485 |
| 0,02 | 0,5080 | | 0,36 | 0,6406 | | 0,70 | 0,7580 | | 1,04 | 0,8508 |
| 0,03 | 0,5120 | | 0,37 | 0,6443 | | 0,71 | 0,7611 | | 1,05 | 0,8531 |
| 0,04 | 0,5160 | | 0,38 | 0,6480 | | 0,72 | 0,7642 | | 1,06 | 0,8554 |
| 0,05 | 0,5199 | | 0,39 | 0,6517 | | 0,73 | 0,7673 | | 1,07 | 0,8577 |
| 0,06 | 0,5239 | | 0,40 | 0,6554 | | 0,74 | 0,7704 | | 1,08 | 0,8599 |
| 0,07 | 0,5279 | | 0,41 | 0,6591 | | 0,75 | 0,7734 | | 1,09 | 0,8621 |
| 0,08 | 0,5319 | | 0,42 | 0,6628 | | 0,76 | 0,7764 | | 1,10 | 0,8643 |
| 0,09 | 0,5359 | | 0,43 | 0,6664 | | 0,77 | 0,7794 | | 1,11 | 0,8665 |
| 0,10 | 0,5398 | | 0,44 | 0,6700 | | 0,78 | 0,7823 | | 1,12 | 0,8686 |
| 0,11 | 0,5438 | | 0,45 | 0,6736 | | 0,79 | 0,7852 | | 1,13 | 0,8708 |
| 0,12 | 0,5478 | | 0,46 | 0,6772 | | 0,80 | 0,7881 | | 1,14 | 0,8729 |
| 0,13 | 0,5517 | | 0,47 | 0,6808 | | 0,81 | 0,7910 | | 1,15 | 0,8749 |
| 0,14 | 0,5557 | | 0,48 | 0,6844 | | 0,82 | 0,7939 | | 1,16 | 0,8770 |
| 0,15 | 0,5596 | | 0,49 | 0,6879 | | 0,83 | 0,7967 | | 1,17 | 0,8790 |
| 0,16 | 0,5636 | | 0,50 | 0,6915 | | 0,84 | 0,7995 | | 1,18 | 0,8810 |
| 0,17 | 0,5675 | | 0,51 | 0,6950 | | 0,85 | 0,8023 | | 1,19 | 0,8830 |
| 0,18 | 0,5714 | | 0,52 | 0,6985 | | 0,86 | 0,8051 | | 1,20 | 0,8849 |
| 0,19 | 0,5753 | | 0,53 | 0,7019 | | 0,87 | 0,8078 | | 1,21 | 0,8869 |
| 0,20 | 0,5793 | | 0,54 | 0,7054 | | 0,88 | 0,8106 | | 1,22 | 0,8888 |
| 0,21 | 0,5832 | | 0,55 | 0,7088 | | 0,89 | 0,8133 | | 1,23 | 0,8907 |
| 0,22 | 0,5871 | | 0,56 | 0,7123 | | 0,90 | 0,8159 | | 1,24 | 0,8925 |
| 0,23 | 0,5910 | | 0,57 | 0,7157 | | 0,91 | 0,8186 | | 1,25 | 0,8944 |
| 0,24 | 0,5948 | | 0,58 | 0,7190 | | 0,92 | 0,8212 | | 1,26 | 0,8962 |
| 0,25 | 0,5987 | | 0,59 | 0,7224 | | 0,93 | 0,8238 | | 1,27 | 0,8980 |
| 0,26 | 0,6026 | | 0,60 | 0,7257 | | 0,94 | 0,8264 | | 1,28 | 0,8997 |
| 0,27 | 0,6064 | | 0,61 | 0,7291 | | 0,95 | 0,8289 | | 1,29 | 0,9015 |
| 0,28 | 0,6103 | | 0,62 | 0,7324 | | 0,96 | 0,8315 | | 1,30 | 0,9032 |
| 0,29 | 0,6141 | | 0,63 | 0,7357 | | 0,97 | 0,8340 | | 1,31 | 0,9049 |
| 0,30 | 0,6179 | | 0,64 | 0,7389 | | 0,98 | 0,8365 | | 1,32 | 0,9066 |
| 0,31 | 0,6217 | | 0,65 | 0,7422 | | 0,99 | 0,8389 | | 1,33 | 0,9082 |
| 0,32 | 0,6255 | | 0,66 | 0,7454 | | 1,00 | 0,8413 | | 1,34 | 0,9099 |
| 0,33 | 0,6293 | | 0,67 | 0,7486 | | 1,01 | 0,8438 | | 1,35 | 0,9115 |

| x | Φ(x) | x | Φ(x) | x | Φ(x) | x | Φ(x) |
|---|---|---|---|---|---|---|---|
| 1,36 | 0,9131 | 1,70 | 0,9554 | 2,08 | 0,9812 | 2,76 | 0,9971 |
| 1,37 | 0,9147 | 1,71 | 0,9564 | 2,10 | 0,9821 | 2,78 | 0,9973 |
| 1,38 | 0,9162 | 1,72 | 0,9573 | 2,12 | 0,9830 | 2,80 | 0,9974 |
| 1,39 | 0,9177 | 1,73 | 0,9582 | 2,14 | 0,9838 | 2,82 | 0,9976 |
| 1,40 | 0,9192 | 1,74 | 0,9591 | 2,16 | 0,9846 | 2,84 | 0,9977 |
| 1,41 | 0,9207 | 1,75 | 0,9599 | 2,18 | 0,9854 | 2,86 | 0,9979 |
| 1,42 | 0,9222 | 1,76 | 0,9608 | 2,20 | 0,9861 | 2,88 | 0,9980 |
| 1,43 | 0,9236 | 1,77 | 0,9616 | 2,22 | 0,9868 | 2,90 | 0,9981 |
| 1,44 | 0,9251 | 1,78 | 0,9625 | 2,24 | 0,9875 | 2,92 | 0,9982 |
| 1,45 | 0,9265 | 1,79 | 0,9633 | 2,26 | 0,9881 | 2,94 | 0,9984 |
| 1,46 | 0,9279 | 1,80 | 0,9641 | 2,28 | 0,9887 | 2,96 | 0,9985 |
| 1,47 | 0,9292 | 1,81 | 0,9649 | 2,30 | 0,9893 | 2,98 | 0,9986 |
| 1,48 | 0,9306 | 1,82 | 0,9656 | 2,32 | 0,9898 | 3,00 | 0,9987 |
| 1,49 | 0,9319 | 1,83 | 0,9664 | 2,34 | 0,9904 | 3,05 | 0,9989 |
| 1,50 | 0,9332 | 1,84 | 0,9671 | 2,36 | 0,9909 | 3,10 | 0,9990 |
| 1,51 | 0,9345 | 1,85 | 0,9678 | 2,38 | 0,9913 | 3,15 | 0,9992 |
| 1,52 | 0,9357 | 1,86 | 0,9686 | 2,40 | 0,9918 | 3,20 | 0,9993 |
| 1,53 | 0,9370 | 1,87 | 0,9693 | 2,42 | 0,9922 | 3,25 | 0,9994 |
| 1,54 | 0,9382 | 1,88 | 0,9699 | 2,44 | 0,9927 | 3,30 | 0,9995 |
| 1,55 | 0,9394 | 1,89 | 0,9706 | 2,46 | 0,9931 | 3,35 | 0,9996 |
| 1,56 | 0,9406 | 1,90 | 0,9713 | 2,48 | 0,9934 | 3,40 | 0,9997 |
| 1,57 | 0,9418 | 1,91 | 0,9719 | 2,50 | 0,9938 | 3,45 | 0,9997 |
| 1,58 | 0,9429 | 1,92 | 0,9726 | 2,52 | 0,9941 | 3,50 | 0,9998 |
| 1,59 | 0,9441 | 1,93 | 0,9732 | 2,54 | 0,9945 | 3,55 | 0,9998 |
| 1,60 | 0,9452 | 1,94 | 0,9738 | 2,56 | 0,9948 | 3,60 | 0,9998 |
| 1,61 | 0,9463 | 1,95 | 0,9744 | 2,58 | 0,9951 | 3,65 | 0,9999 |
| 1,62 | 0,9474 | 1,96 | 0,9750 | 2,60 | 0,9953 | 3,70 | 0,9999 |
| 1,63 | 0,9484 | 1,97 | 0,9756 | 2,62 | 0,9956 | 3,75 | 0,9999 |
| 1,64 | 0,9495 | 1,98 | 0,9761 | 2,64 | 0,9959 | 3,80 | 0,9999 |
| 1,65 | 0,9505 | 1,99 | 0,9767 | 2,66 | 0,9961 | | |
| 1,66 | 0,9515 | 2,00 | 0,9772 | 2,68 | 0,9963 | | |
| 1,67 | 0,9525 | 2,02 | 0,9783 | 2,70 | 0,9965 | | |
| 1,68 | 0,9535 | 2,04 | 0,9793 | 2,72 | 0,9967 | | |
| 1,69 | 0,9545 | 2,06 | 0,9803 | 2,74 | 0,9969 | | |

## Bibliography:

("Irodalom")

[1] **Szalkai, István:** *Homepage* http://math.uni-pannon.hu/~szalkai
[2] **Dr. Mihálykóné dr. Orbán, Éva:** *Probability theory and mathematical statistics for IT students*, http://math.uni-pannon.hu/~szalkai/ME-ProbTh-JAV-2014.pdf
[3] **Solt György:** *Valószínűségszámítás*, példatár, (Probability theory problem book, in Hungarian), Műszaki Kiadó, Budapest,

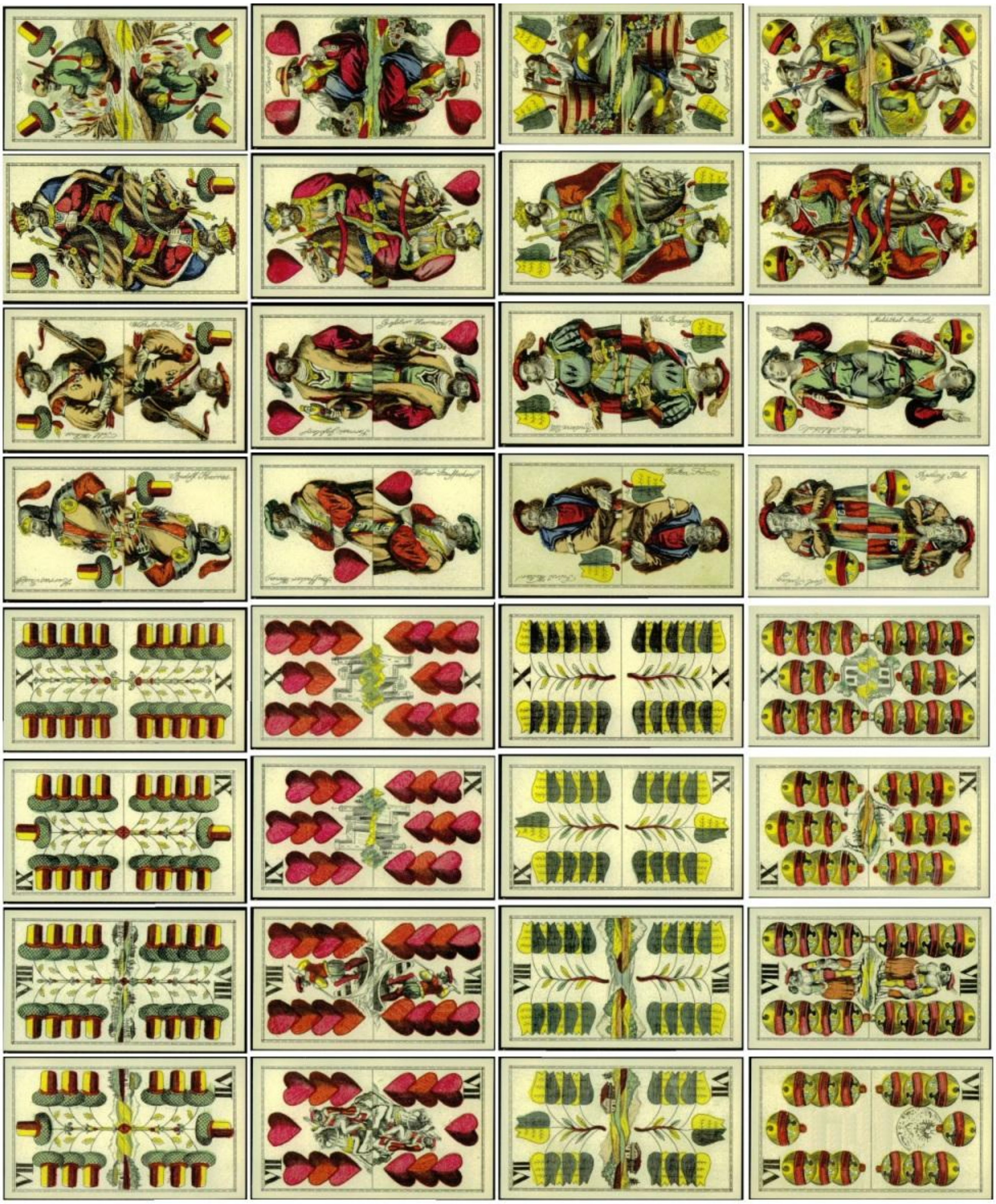
## Biographies:

("Életrajzok")

The below data are only for information, not for learning.

| | | |
|---|---|---|
| **Bayes**, Thomas | (1702-1761), | English mathematician |
| **Bernoulli**, Jacob | (1654-1705), | Swiss mathematician |
| **Boole**, George | (1815-1864), | English mathematician |
| **Cantor,** Georg | (1845-1918), | Russian born German mathematician |
| **Chebyshev,** Pafnuty Lvovich (Csebisev, Pafnutyij Lvovics) | (1821-1894), | Russian mathematician |
| **Euler,** Leonhard | (1707-1783), | Swiss mathematician and physician |
| **Galton,** Francis | (1822-1911), | English mathematician |
| **Gauss**, Karl Friedrich | (1777-1855), | German mathematician |
| **Kolmogorov**, Andrey Nikolaevich | (1903-1987), | Russian mathematician |
| **Laplace**, Pierre Simon | (1749-1827), | French mathematician |
| **Leibniz**, Gottfried Wilhelm | (1646-1716), | German mathematician |
| **Liouville**, Joseph | (1809-1882), | French mathematician |
| **Markov**, Andrei Andreyevich | (1856-1922), | Russian mathematician |
| **Moivre**, Abraham | (1667-1754), | French mathematician |
| **De Morgan**, Augustus | (1806-1871), | English mathematician, born in India |
| **Newton**, Isaac | (1643-1727), | English mathematician and physician |
| **Poisson**, Siméon Denis | (1781-1840), | French mathematician |
| **Schiller,** Friedrich | (1759-1805), | German writer and poet |

**Tell**, William (14th century), Swiss legendary folk hero
in the four languages of Switzerland: German: **Wilhelm Tell**; French: **Guillaume Tell**; Italian: **Guglielmo Tell**; Romansh: **Guglielm Tell**; in Hungarian: **Tell Vilmos**



**The "Hungarian" deck**