



Trainer: Bondor Cosmina-Ioana, PhD

Introduction to Medical Probability



ALWAYS



SEEK



KNOWLEDGE

Objectives

- **Foundations of probability**
 - definitions, events, complements, addition and multiplication rules
- **Conditional probability and Bayes' theorem**
 - dependence, independence
- **Random variables**
 - expected value, variance
- **Probability distributions**
 - normal, binomial and Poisson distributions

Learning objectives

- By the end of the lecture, students should be able to:
 - Define probability and distinguish theoretical from empirical probability.
 - Calculate probabilities of events using addition and multiplication rules.
 - Explain conditional probability, independence, and Bayes' theorem.
 - Recognize binomial and normal probability distributions.

Why probability matters in medicine

- uncertain diagnoses, prognoses, and treatment outcomes

Examples:

- a treatment with statins  reduce cholesterol?
- positive quick test for alcohol  the driver consume alcohol?

Why probability matters in medicine

- uncertain diagnoses, prognoses, and treatment outcomes

Examples:

- a treatment with statins  reduce cholesterol?
- positive quick test for alcohol  the driver consume alcohol?

Patient have high cholesterol. Medical doctor prescribe statins as a treatment

- What is the probability of effective treatment? (lowering the cholesterol by 20 units)
 - it depend on
 - the performance of the drug

The police stopped the driver for a quick alcohol test.

- The result was !!! positive
- What is the probability of being wrong? (false positive)
 - it depend on
 - the performance of the test
 - driver pretest probability
- the quick test should be confirmed by the blood test

Definitions



experiment = an action / process

random experiment = with unpredictable outcome

trial = a repetition of the experiment

events = possible outcomes noted with A, B, C etc.

$$\text{Probability (A)} = \frac{\text{the number of times that A appears}}{\text{the total number of trials}}$$

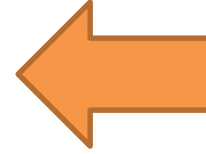
This is the classical definition of the probability.

Key Requirements

- **Finite outcomes:** The total number of possible outcomes must be countable and limited.
- **Equally likely:** Every individual outcome must have the exact same chance of occurring as any other.

Patient have high cholesterol. Medical doctor prescribe statine as a treatment

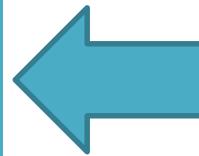
- What is the probability of effective treatment? (lower the cholesterol by 20 units)
 - it depend on
 - the performance of the drug



experiment – a patient taken statine
outcome unknown/uncertain
trial – repeating the experiment on many patients
events – effective/not effective

The police stopped the driver for a quick alcohol test.

- The result was !!! positive
- What is the probability of being wrong? (false positive)
 - it depend on
 - the performance of the test
 - driver pretest probability
- the quick test should be confirmed by the blood test



experiment – a driver making a quick test for alcohol
outcome unknown/uncertain
trial – repeating the experiment on many drivers
events – positive/negative

Patient have high cholesterol. Medical doctor prescribe statins as a treatment

possible outcome: effective/not effective

- What is the probability of effective treatment? (lower the cholesterol by 20 units)
 - it depend on
 - the performance of the drug

The police stopped the driver for a quick alcohol test.

- The result was !!! positive
- What is the probability of being wrong? (false positive)
 - it depend on
 - the performance of the test
 - driver pretest probability
- the quick test should be confirmed by the blood test

Problem

the occurrence of the possible events are not equally likely!!!



use the Empirical Probability

Example 1

P (death by Ebola virus) = ?

The event = death by ebola virus in 2014 Outbreack

Ebola virus disease (EVD) – acute infectious disease with high mortality risk.

CDC. 2014 Ebola Outbreak in West Africa - Case Counts. Nov.2015. Available at: <http://www.cdc.gov/vhf/ebola/outbreaks/2014-west-africa/case-counts.html>

Country	Cases	Deaths	Survival	Death probability	Survival probability
Siera Leone	14,122	3,955		?	

$$P(\text{death Ebola illness}) = \frac{\text{the number of deaths}}{\text{the total number of cases}}$$

- $P(\text{death Ebola illness}) = \frac{3955}{14122} = 0.28$
- The probability that a random person infected with Ebola will die is 0.28 (28% fatality rate)

Country	Cases	Deaths	Survival	Death probability	Survival probability
Siera Leone	14122	3955		=3955/14122 =0.28	

Axiomatic definition of probability

The event space

- The *event space* contains all possible outcomes of an trial
- *Event* = a specified subset
- *Elementary event* = an event with just one outcome from the event space
- *Compound event* = Any event whose realization depends on at least two elementary events

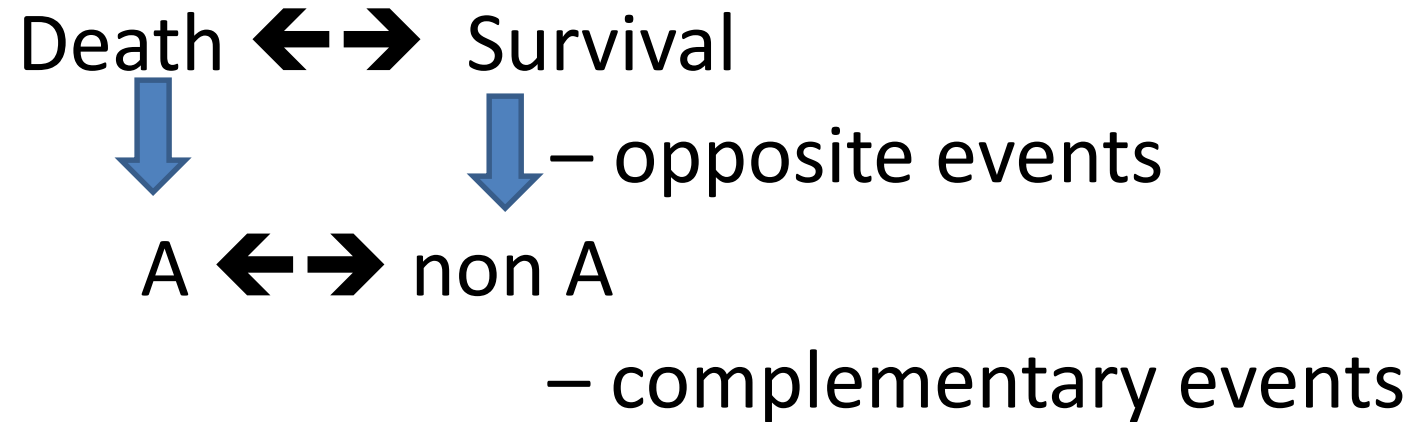
Certain event

- an event that is guaranteed to happen
- it occurs in every possible outcome of an experiment.
- The event space = the certain event

Complementary events

- also called opposite events
- two events that together include all possible outcomes, but cannot happen at the same time

Example 2



No. of deaths + No. of survival = Total no. of people

$$\text{Deaths (\%)} + \text{Survival (\%)} = 100 \%$$

$$P(\text{survival Ebola virus}) = P(\text{non A}) = ?$$

$$P(\text{survival Ebola illness}) = \frac{10167}{14122} = 0.72$$

The probability that a random person infected with Ebola will survive is 0.72 (72% survival rate)

$$P(A) + P(\text{non } A) = 1$$

$$\Rightarrow P(\text{non } A) = 1 - P(A) = 1 - 0.28 = 0.72$$

Country	Cases	Deaths	Survivals	Death probability	Survival probability
Siera Leone	14122	3955	10167	0.28	0.72 = 1 - 0.28 or =10167/14122

Events A and B:

- **intersection** $D=A \cap B$ - the event that occurs only when A and B occur simultaneously

Impossible event

- Is an event that cannot occur at any time
- The null set ϕ = the impossible event

Mutually exclusive events

- events that **cannot happen at the same time.**

Example 3

$$P(\text{A and B}) = P(\text{Type 0 and Type A}) = ?$$

Blood type	Frequency	P(Blood type)
O	400	0.40
A	450	0.45
B	142	0.142
AB	8	0.008
Total	1000	1.0

$P(A \text{ and } B) = 0$ for mutually exclusive events

Blood type	Frequency	P(Blood type)
O	400	0.40
A	450	0.45
B	142	0.142
AB	8	0.008
Total	1000	1.0

$P(\text{Type 0 and Type A}) = ?$

Type 0 and Type A are mutually exclusive:

$P(\text{Type 0 and Type A}) = 0$

Two events are mutually exclusive = they can't happen at the same time

Events A and B:

- **reunion** $C=A\cup B$ - an event that occurs if at least one of events A **or** B takes place

The Axiomatic Definition of Probability

The probability function $P(A)$ is called probability of the event A , if the following axioms are satisfied:

- $0 \leq P(A) \leq 1$, for any event A
- $P(E) = 1$, where E is the certain event
- If A și B are incompatible (cannot be in the same time), then

$$P(A \cup B) = P(A) + P(B)$$

where $A \cup B$ – adding event

Example 4

$$P(\text{A or B}) = P(\text{Type O or Type A}) = ?$$

Blood type	Frequency	P(Blood type)
O	400	0.40
A	450	0.45
B	142	0.142
AB	8	0.008
Total	1000	1.0

$P(A \text{ or } B) = P(A) + P(B)$ for mutually exclusive event

Blood type	Frequency	P(Blood type)
O	400	0.40
A	450	0.45
B	142	0.142
AB	8	0.008
Total	1000	1.0

$$\begin{aligned} P(\text{A person have type O or A}) &= \\ &= P(\text{Type O or Type A}) = P(\text{Type O}) + P(\text{Type A}) = \\ &= \frac{400}{1000} + \frac{450}{1000} = \frac{850}{1000} = 0.85 \end{aligned}$$

Two events are mutually exclusive/incompatible if they can't happen at the same time

Independent events

- events where the occurrence (or non-occurrence) of one **does not affect** the probability of the other

Example 5

Birth	Frequency					
Girl	500	$P(A=\text{girl}) = 0.5$ $P(B=\text{boy}) = 0.5$				
Boy	500					
Total	1000					

A, B two independent events
 $P(A \text{ and } B) = P(\text{Boy and Boy}) = ?$

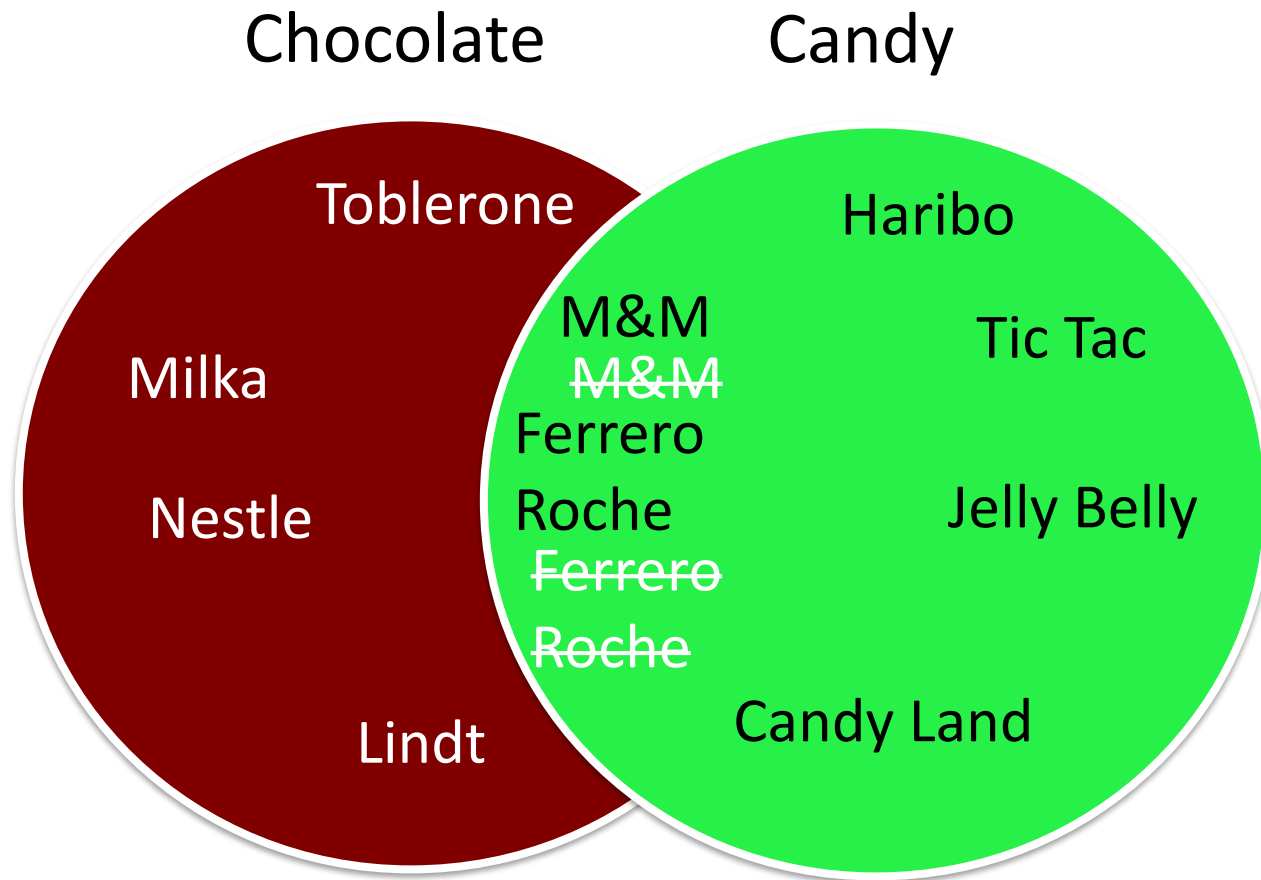
Families with 2 children

Birth	Frequency		Family with two children	P("and")	Frequency	Probability
Girl	500	$P(A=\text{girl}) = 0.5$ $P(B=\text{boy}) = 0.5$	girl, girl	$P(A \text{ and } A)$	250	$0.5 * 0.5 = 0.25$
Boy	500		boy, boy	$P(B \text{ and } B)$	250	$0.5 * 0.5 = 0.25$
			girl, boy	$P(A \text{ and } B)$	250	$0.5 * 0.5 = 0.25$
			boy, girl	$P(B \text{ and } A)$	250	$0.5 * 0.5 = 0.25$
Total	1000				1000	

Example 6

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

for any events A, B



$$\begin{aligned}
 P(A \text{ or } B) &= \\
 &= P(\text{chocolate or candy}) = \\
 &= P(A) + P(B) - P(A \text{ and } B)
 \end{aligned}$$

Probability

$$0 \leq \Pr(A) \leq 1$$



Summary

- A,B **mutually exclusive** events

$$P(A \text{ and } B)=0$$

- B the **complementary** event of A

$$P(\text{complementary event of } A) = P(\text{non } A) = 1 - P(A)$$

- A the **certain** event

$$P(\text{certain event}) = 1$$

- A the **impossible** event

$$P(\text{impossible event}) = 0$$



Summary

A, B two events:

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

A, B two independent events:

$$P(A \text{ and } B) = P(A) * P(B)$$

A, B two independent events:

$$P(A \text{ or } B) = P(A) + P(B) - P(A) * P(B)$$

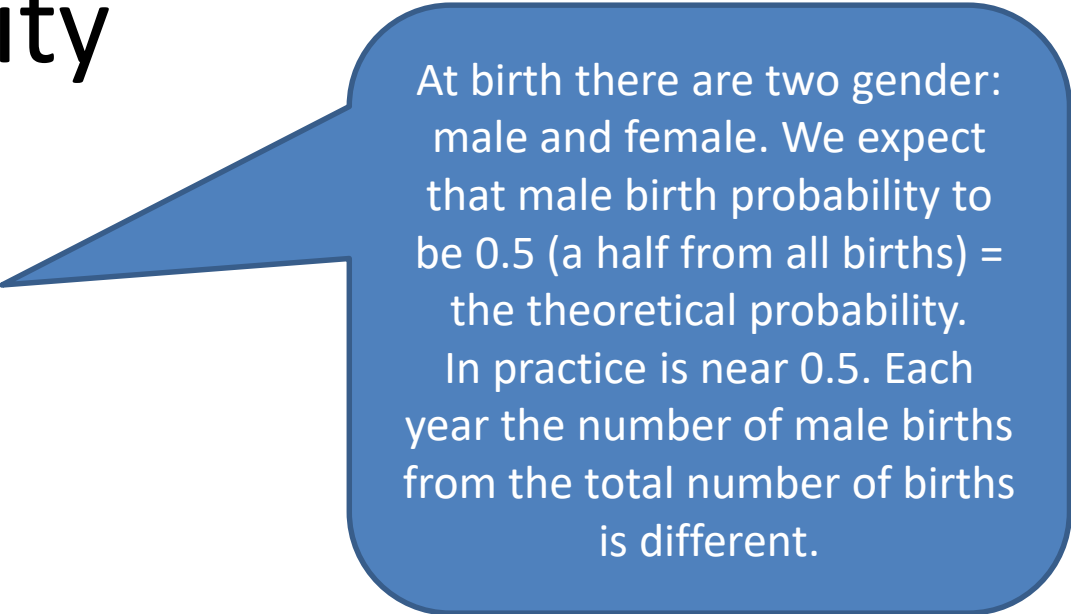
A, B two mutually exclusive events:

$$P(A \text{ or } B) = P(A) + P(B)$$



Probability

- Theoretical
 - how should be
- Empirical
 - how it is if we experiment
- Subjective
 - the opinion of a person about how it should be
- Example: gender
 - 0.5 is the theoretical probability of a male birth (it means 50%)
 - 0.494 is the empirical probability of a male birth in 2023 in Romania



At birth there are two gender: male and female. We expect that male birth probability to be 0.5 (a half from all births) = the theoretical probability. In practice is near 0.5. Each year the number of male births from the total number of births is different.

Empirical probability

- Events with unknown probability and unequal
- if we cannot estimate the theoretical probability
- then we should make experimental trials (in very big number) to estimate the relative frequency of the event = the empirical probability
- The theoretical probability can be estimate with the empirical probability i.e. **relative frequency**

Dependent events



event B dependent on event A

– occurrence of B influenced by the occurrence of A

$$P(\text{B **dependent** on A}) = P(B|A)$$

$P(A | B)$ = conditional probability

- The probability of occurrence of event A given that event B is already happen

$$P(A | B) = P(A \text{ and } B) / P(B)$$



Bayes' theorem

Dependent events

Conditional probability—the probability of an outcome depending on an earlier outcome.

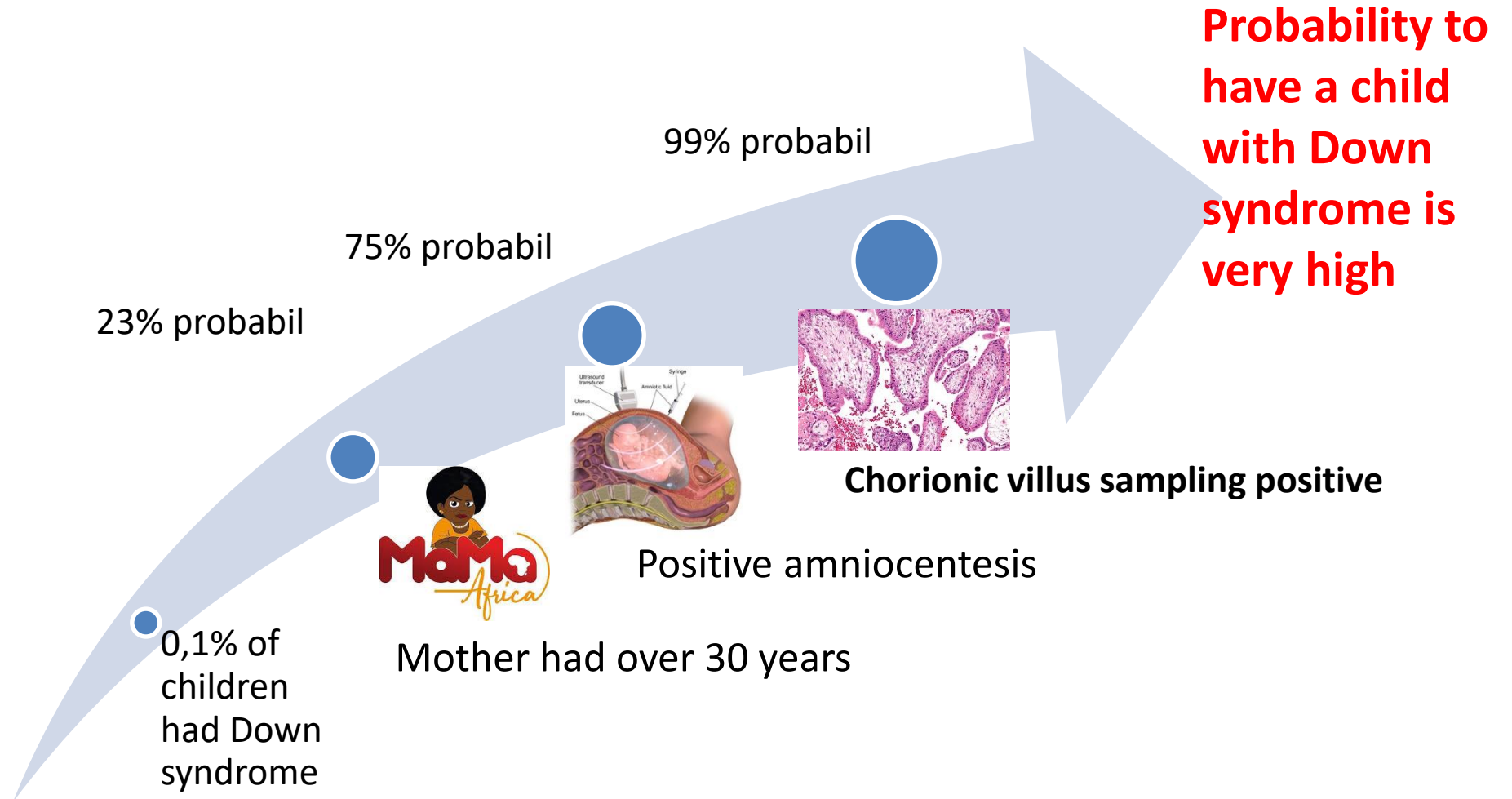
$$P(A|B) = \frac{P(B|A)*P(A)}{P(B)}.$$

Involved in making medical decision – reasoning process - interpreting diagnostic procedures.



Down syndrome during pregnancy

Will the child have the disease? Should we interrupt the pregnancy?



Bayes' theorem - to improve what we know after something happen

- That's how computers learn
- That's how spam filter (email) works
- That's how medical doctors put a diagnostic
- That's how we know what our friends like

Population – representative sample

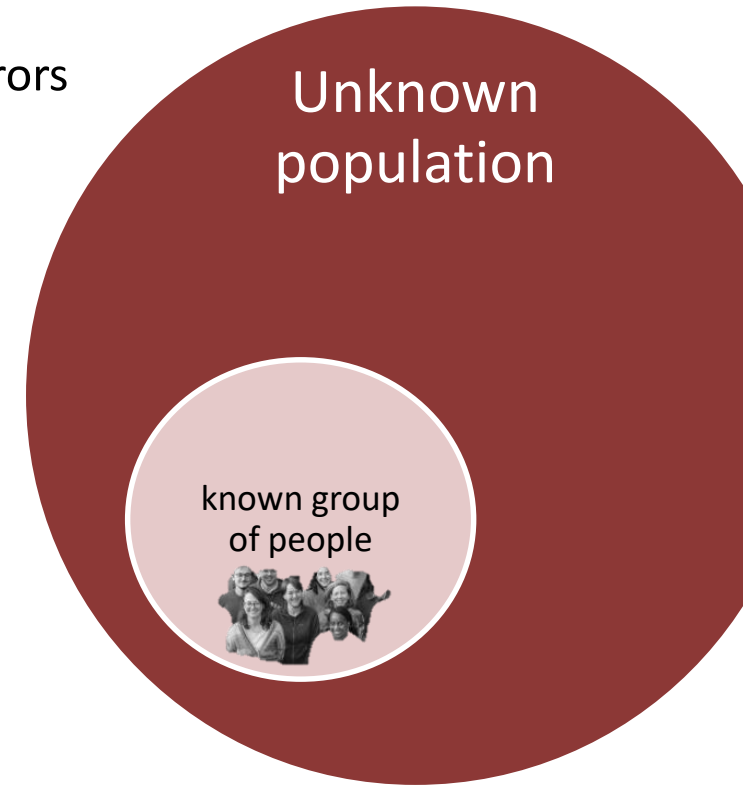
Population



- from statistical point of view
 - a set of elements with the same characteristic
- medical dentistry
 - patients
 - dental office
- medicine
 - patients
 - hospital

we have a question

- we should follow **whole population** during their life time to find it?
 - we cannot
 - too much money, too many employee, too many data, a lot of errors
- we should follow **2000 people** during life time?
 - we can
 - cost relative small, few staff, few errors
- we should follow **100 people** during life time?
 - we can, but the results will be not reliable

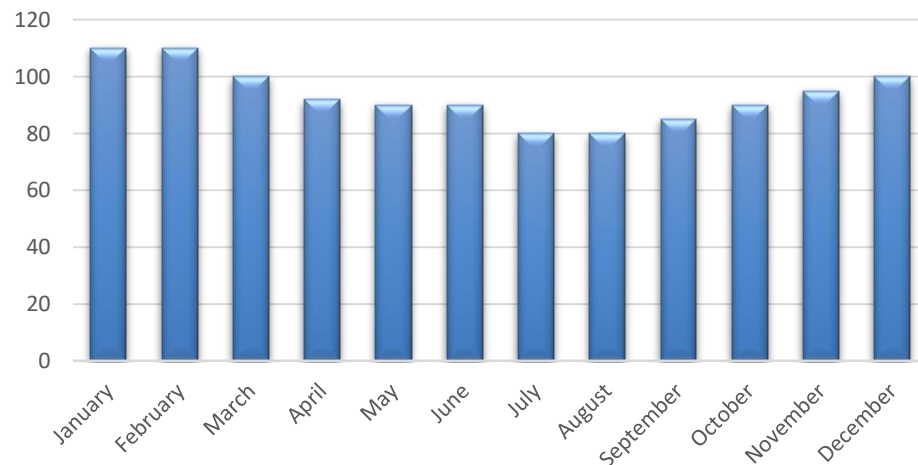


Conclusion: we need a sample

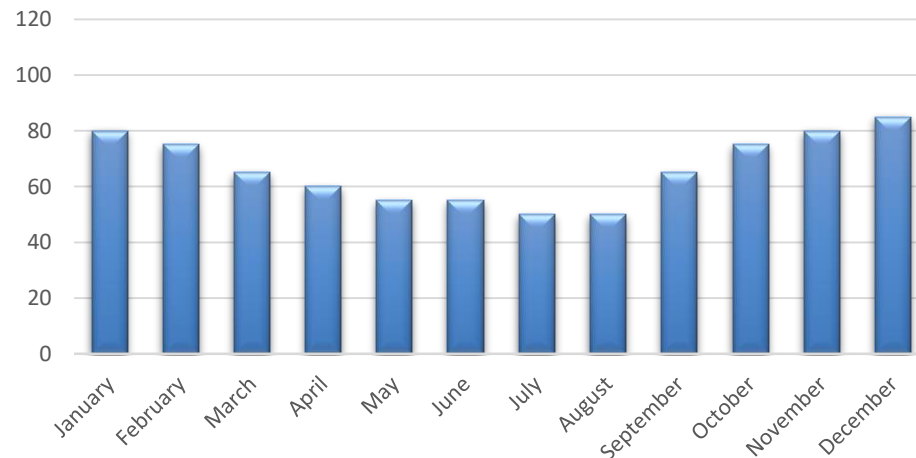
Sample

- a subset of a population
 - give as a glimpse of whole population
 - give us an idea of the shape of the data
-
- why shape is important? if we want to compare two groups

Last year consumption of electricity



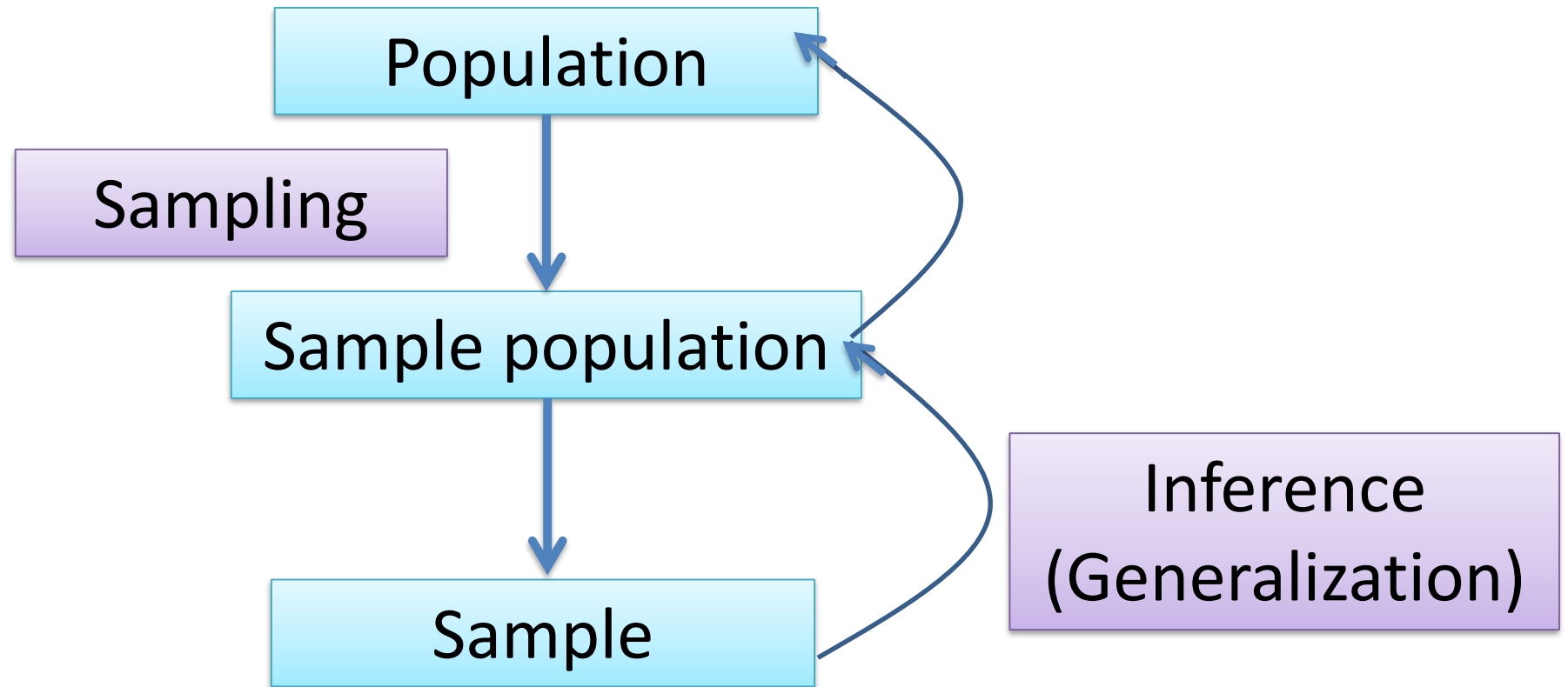
This year consumption of electricity



Why to study samples instead of whole population?

- More quickly
- Less expansive
- Less dangerous
- More accurate conclusions

- **Less dangerous** – if we should test a new drug and this drug is dangerous, then only few people will be affected, not whole population
- **More accurate conclusions** – if we investigate a phenomenon or we put a diagnostic or interpret a radiography and we have several investigators (whole population – we need more than one investigator), each one will have a different subjective approach, so we will have a lot of errors, not all of them understand the process of investigation.



Representative sample for the population

- have the same distribution of important characteristics as the population

where:

- important characteristics = in connection with the studied characteristics

select random



high probability to select a
representative sample



Why random selection?



- Reducing / eliminating experimental errors = decrease / eliminate selection **bias (error)**
- The probability to select a representative sample is high

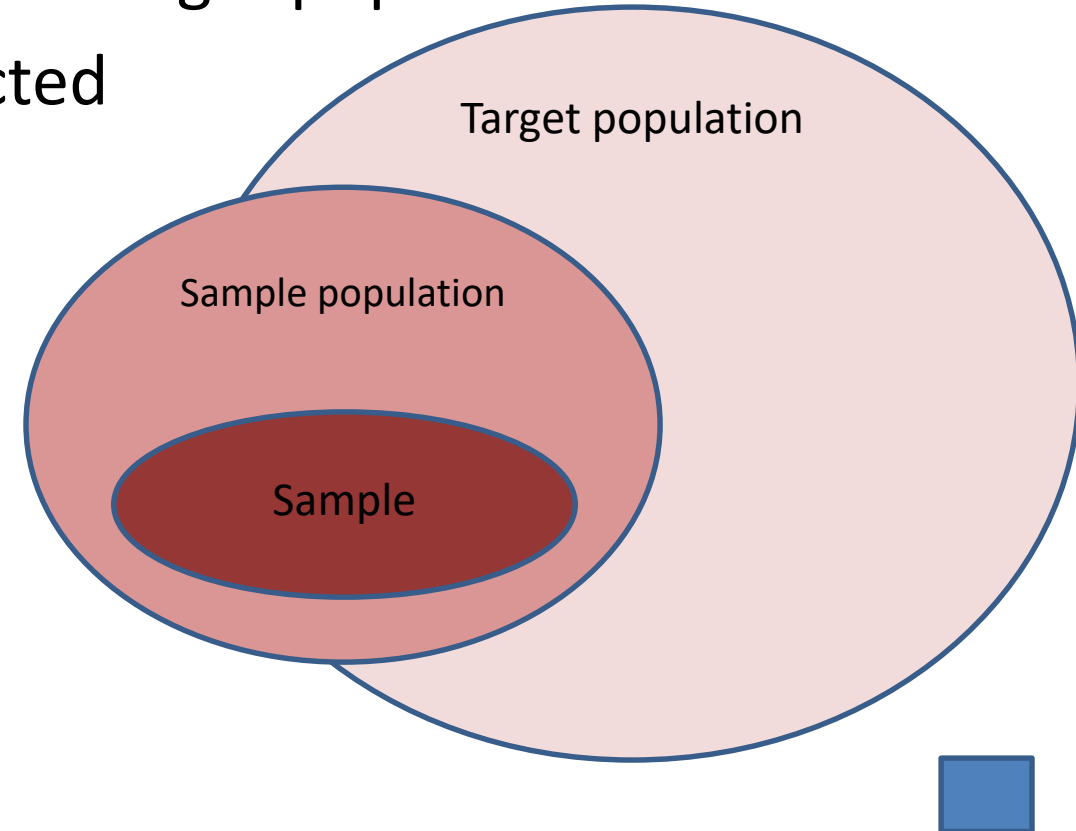
Random selection

- **Random selection** – each individual from the population have the same probability to be selected in the sample
- Ex. Go to town hall take all ID numbers of the people who live in the city. Extract randomly ID's



Errorr (bias) of selection

- to select wrong
 - sample population is not a subset of the target population
 - a part of the population was not selected



Even if we select random there is still probability of errors

- there is only a small probability to select with errors if we select random
- ex. prevalence of smoking – there are chances that even if we randomly select, all the people selected are playing a sport
- What to do? we **replicate** the study ! Only one study cannot demonstrate a hypothesis

Probabilistic - Sampling methods



Probabilistics – the probability of each subject to be selected is known

- **Random** sampling
 - each subject has the same probability to be selected
- **Systematic** sampling
 - each k^{th} subject is selected
- **Stratified** sampling
 - population is divided into subgroups and a random sample is selected from each subgroups
- **Cluster** sampling
 - population is divided into clusters and a random sample is selected from each cluster (cluster = geographic zones)

Non-probabilistic - Sampling methods



Non-probabilistic – the probability of each subject to be selected is unknown

- **Snowball** sample – respondent are asked to help recruit people who fit the criteria of inclusion
- **Convenient sample** – subjects are selected because they are available to the researcher

Census

- **Census** – all population are selected
 - in this case we do not need inferential statistics

Sample with random selection

```
graph TD; A[Sample with random selection] --> B[Measurements]; B --> C[random result]; C --> D[Result = random variable];
```

The diagram is a vertical flowchart with four rectangular boxes. The first box is blue and contains the text 'Sample with random selection'. A light blue arrow points down from its bottom right corner to the top right corner of the second box. The second box is green and contains the text 'Measurements'. A light green arrow points down from its bottom right corner to the top right corner of the third box. The third box is light green and contains the text 'random result'. A light orange arrow points down from its bottom right corner to the top right corner of the fourth box. The fourth box is orange and contains the text 'Result = random variable'. The boxes are staggered to the left, creating a descending staircase effect.

Measurements

random result

Result = random variable

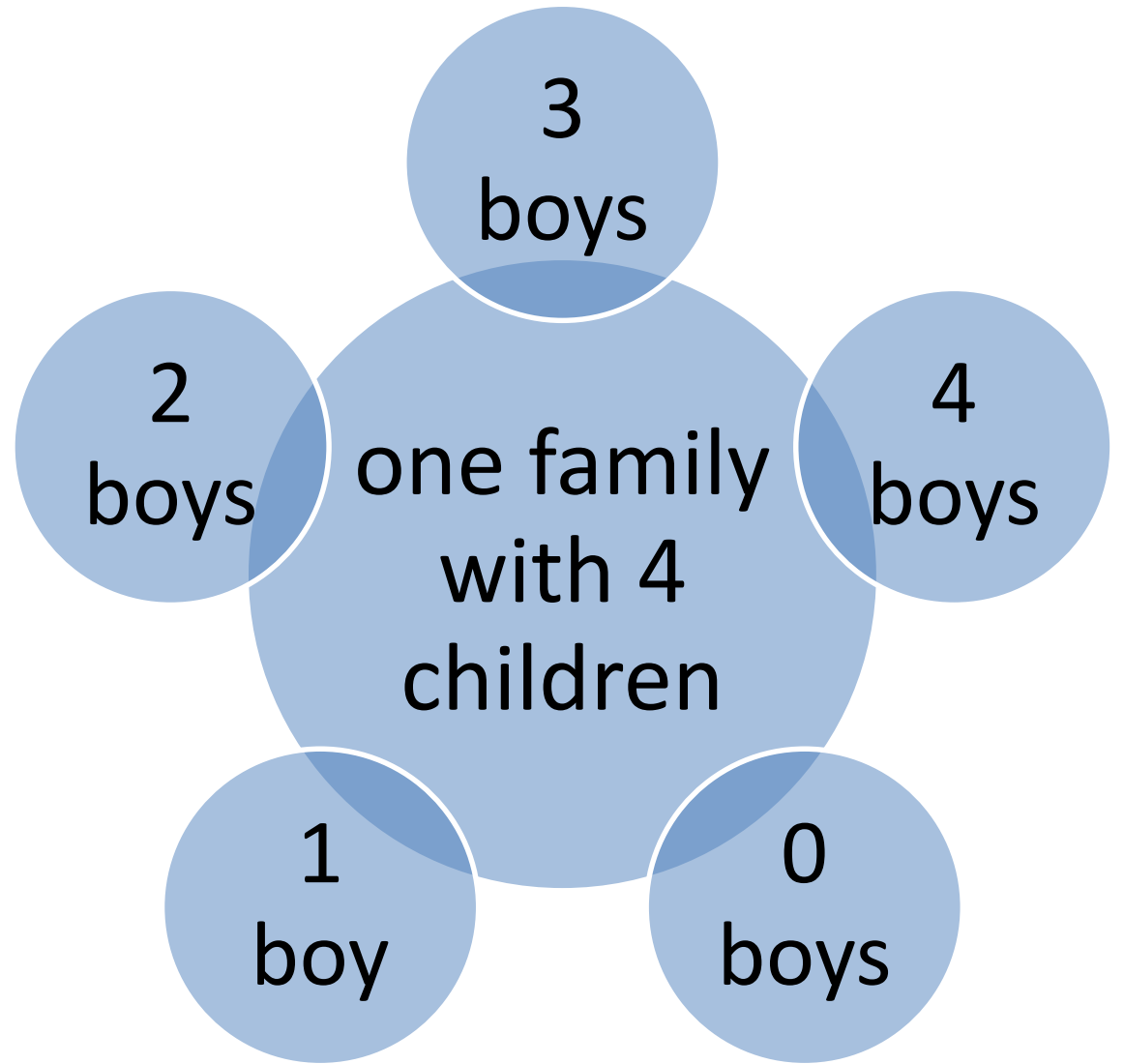


Probability distribution

- The values of a random variable can be summarized in a frequency distribution which we call **probability distribution**

Example: Experiment

Probability of a
child born being
male ≈ 0.5 (50%
of the cases)



In **100 family with 4 children selected randomly?**

In **randomly selected 100 families with 4 children?**

Frequency distribution of variable X the number of times occur the possible values of the variable X :

Number of boys	0	1	2	3	4	Total
No. of family	4	29	40	24	9	100

- Number of boys in a family – random **variable**

Probability distribution

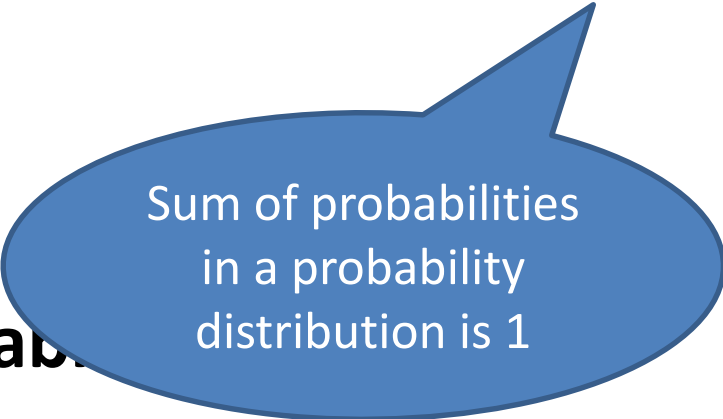
Number of boys	0	1	2	3	4	Total
No. of family	4	29	40	24	9	100
Probability	0.04	0.29	0.40	0.24	0.09	1

Probability distribution

We call **probability distribution** of variable X the number of times occur the possible values of the variable X divide each by the total

Number of boys	0	1	2	3	4	Total
Probability	0.04	0.29	0.40	0.24	0.09	1

- Number of boys in a family – random variable



Sum of probabilities
in a probability
distribution is 1

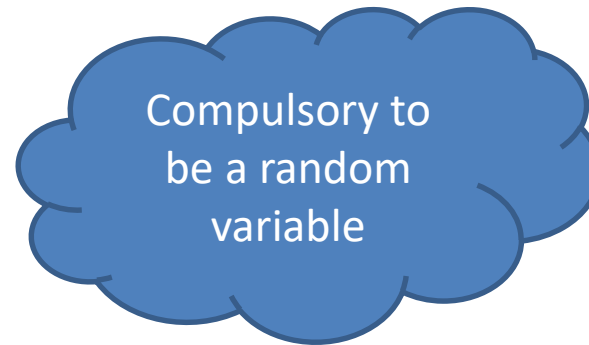
Probability distribution

- Distribution probability in case of a finite variable X :

$$X: \begin{pmatrix} x_1 & x_2 & \dots & x_n \\ P(x_1) & P(x_2) & \dots & P(x_n) \end{pmatrix}$$

- Properties

$$P(x_1) + P(x_2) + \dots + P(x_n) = 1$$



- The **mean** of a discrete probability distribution (called also **expected value**) is given by the formula:

$$M(x) = \sum_{i=1}^n P(x_i)$$

- Represents the weighted average of possible values, each value being weighted by its probability of occurrence

- **Variance:** is a weighted average of the squared deviations in X

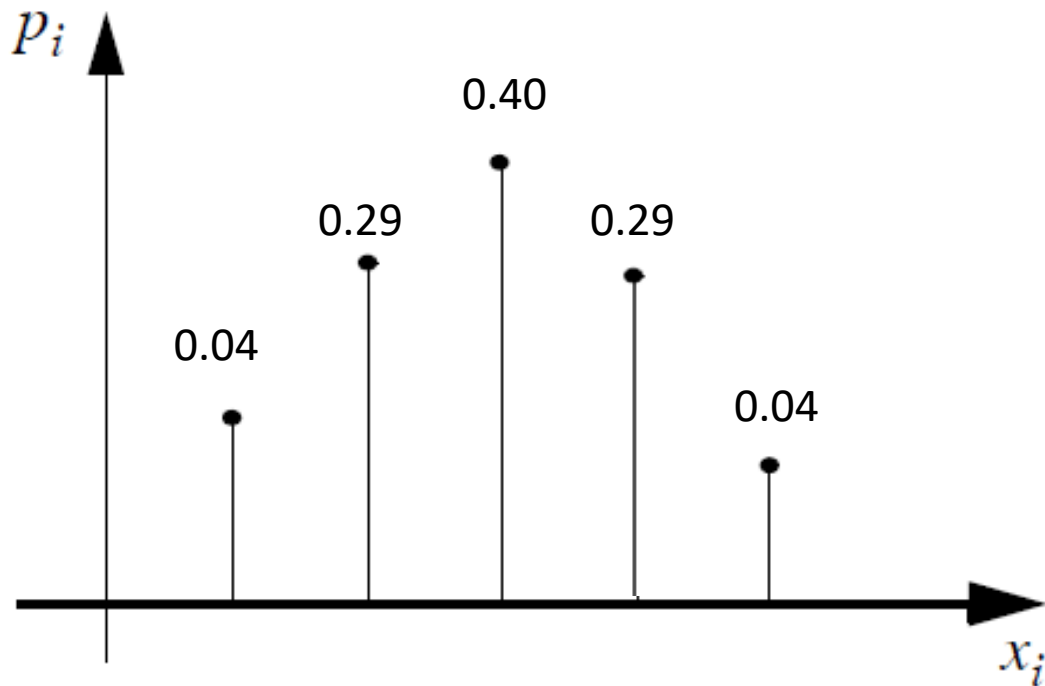
$$V(x) = \sum_{i=1}^n (x_i - M(x))^2 \cdot P(x_i)$$

- **Standard deviation:**

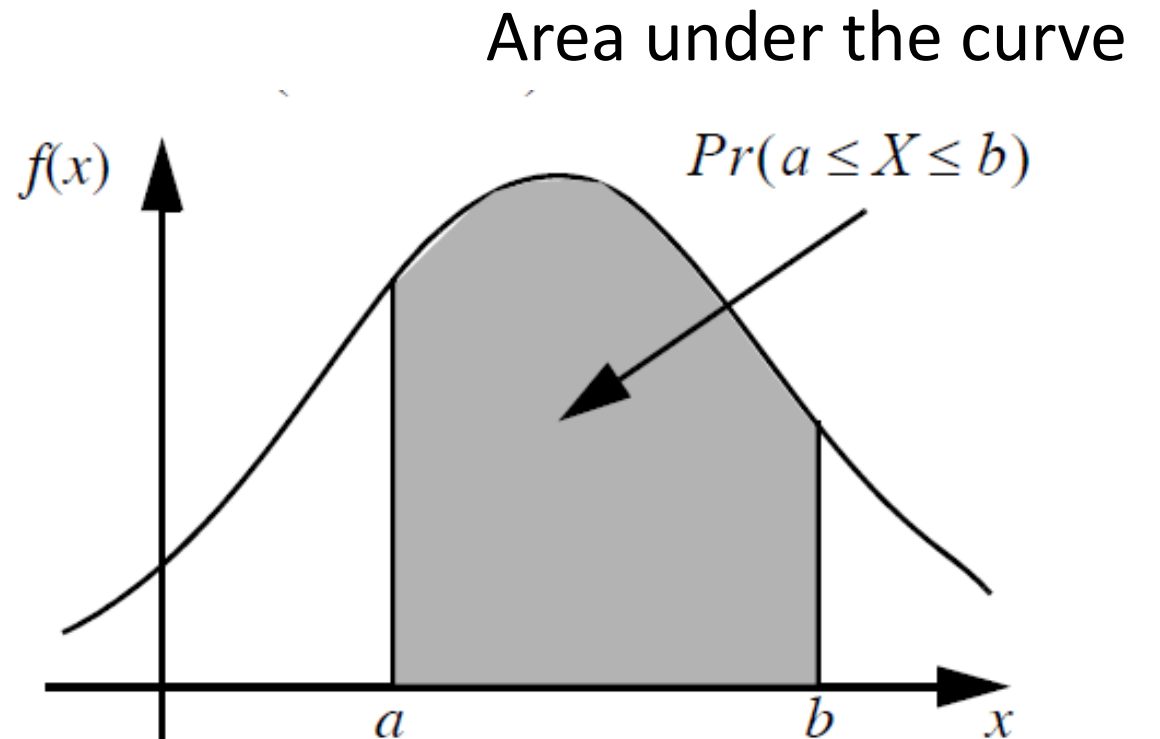
$$\sigma(x) = \sqrt{V(x)}$$

Random variable

Finite



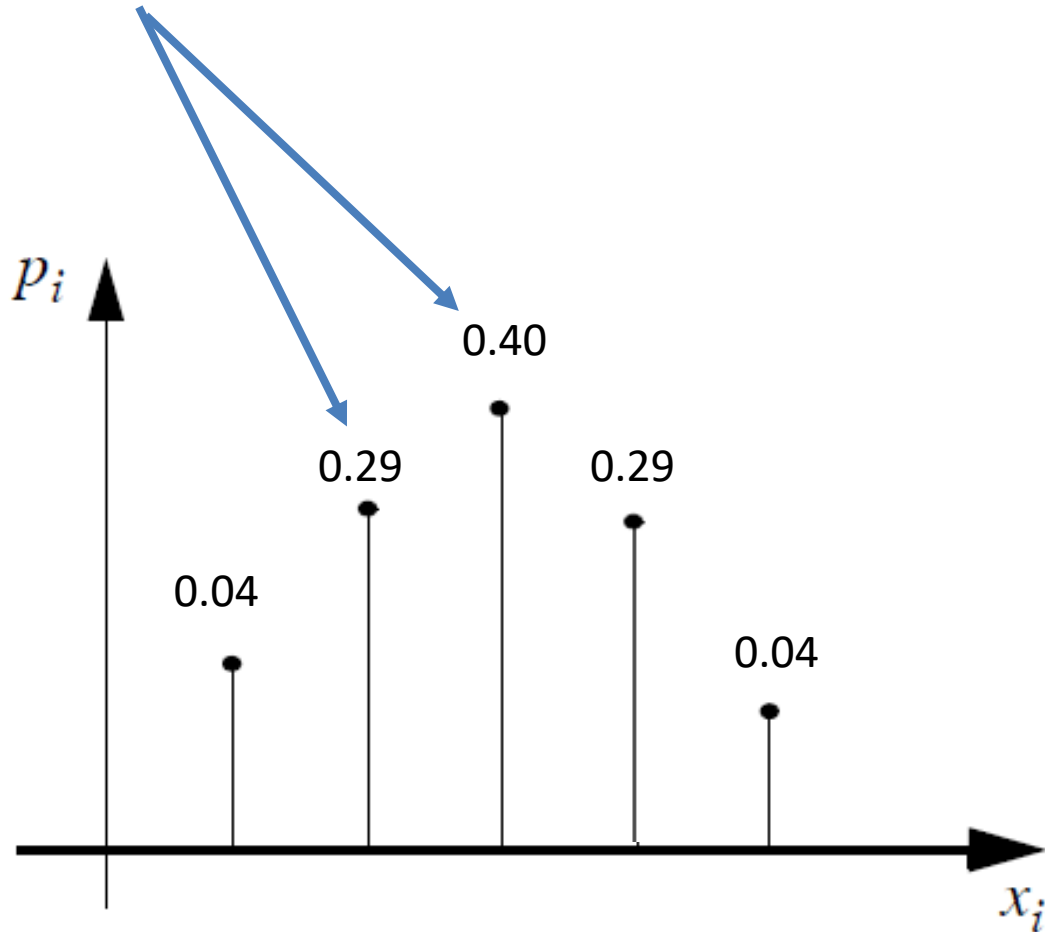
Infinite



Random variable

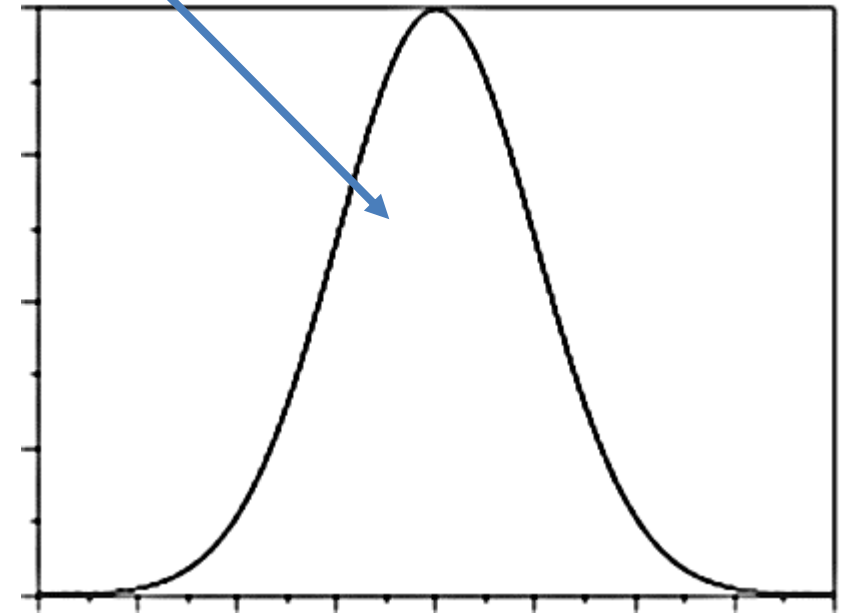
Finite

Sum of probabilities = 1



Infinite

Area under the curve = 1



How we compute probability distribution?



If the variable is finite

Empirical – experiment – probability distribution

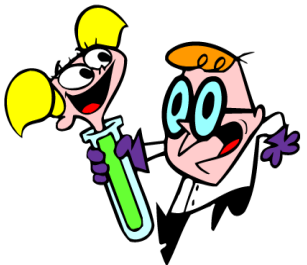


If is ∞ ?

! we are lucky – we found

Formula

Law

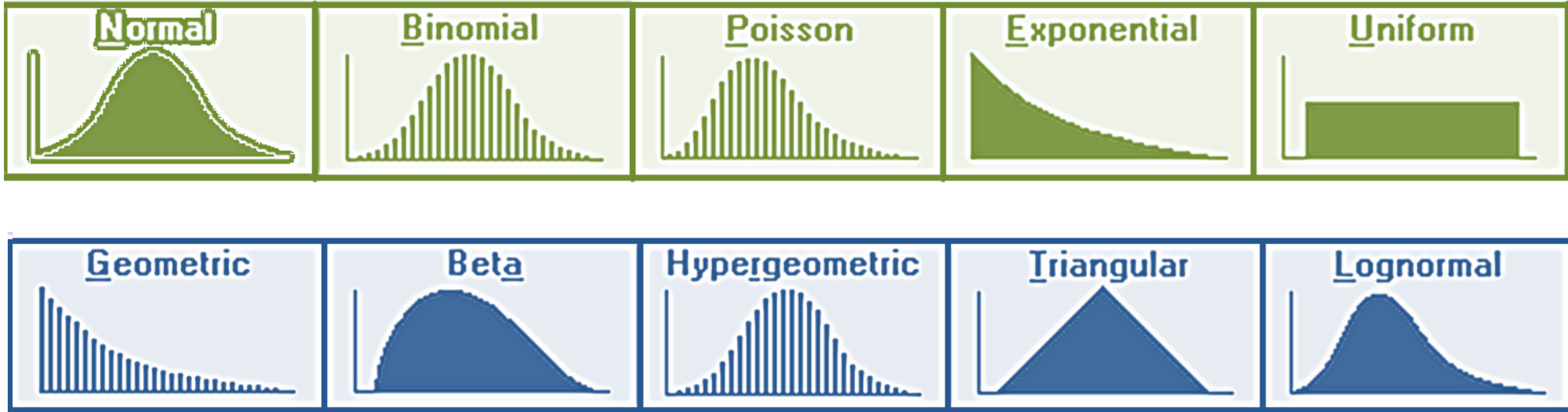


If we are not lucky:

We model (approximate) after a theoretical probability distribution (a known one – one we were lucky with)



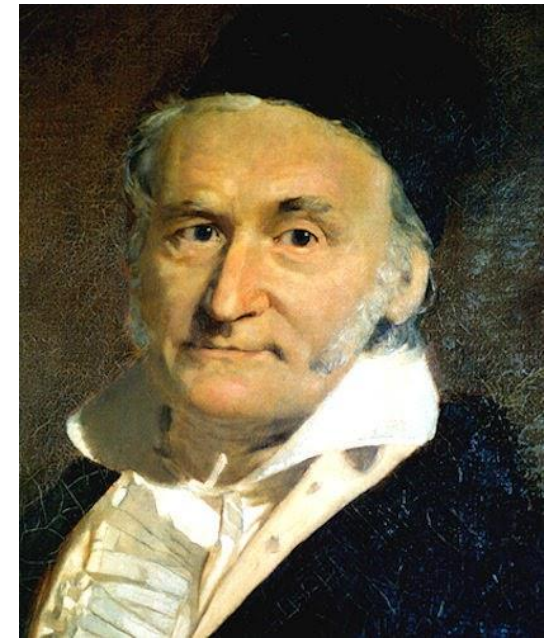
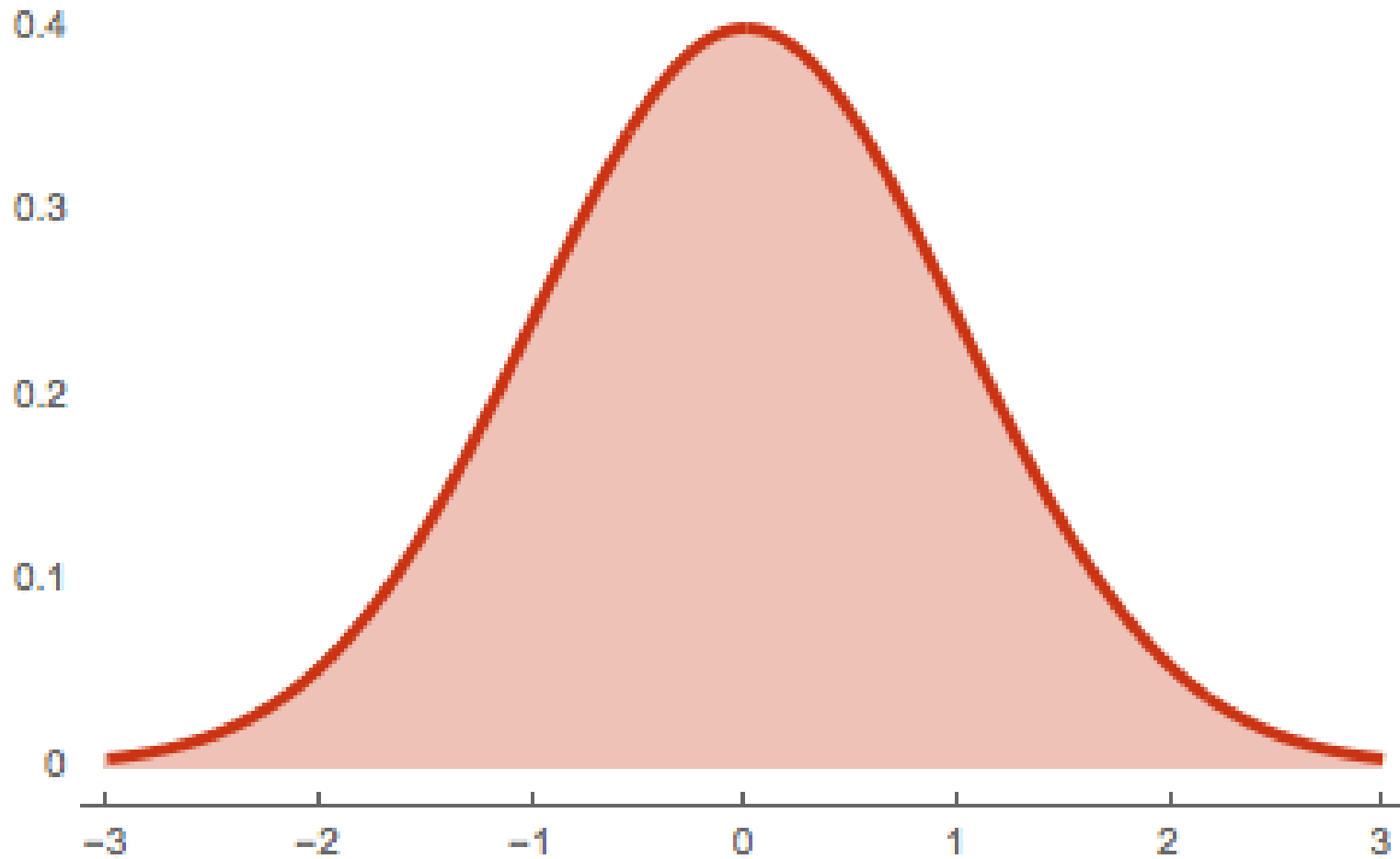
Known distribution law of probability



Normal distribution

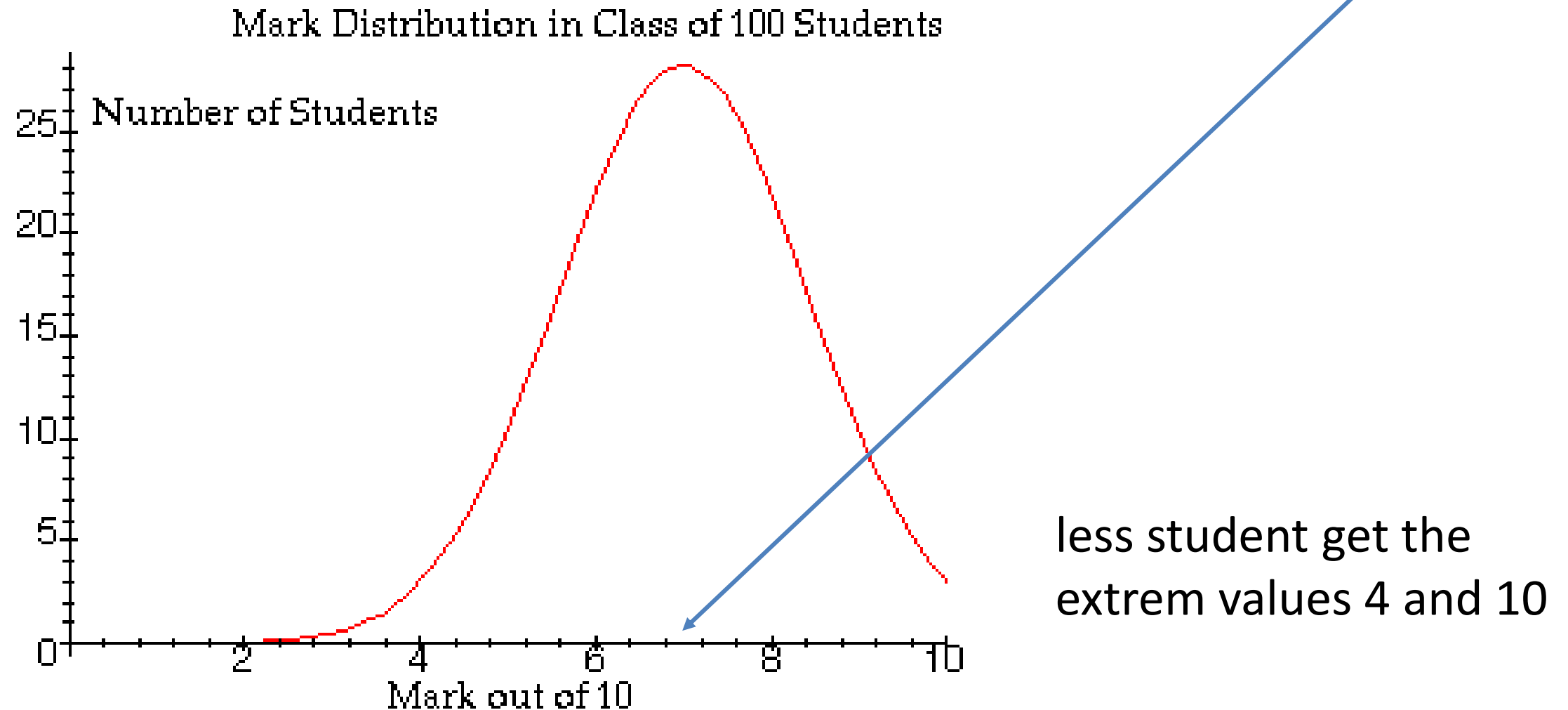
Normal distribution (bell shape, Gauss)

It is “normal” to happen like this



Carl Friedrich Gauss
(30 April 1777 – 23 February 1855)

The normality – most of the student get 7 (the middle of the mark between 4 to 10)



Normal distribution

Random variable **X** is **normal** **N**(μ , σ) if the distribution depend on two parameters: mean μ and standard deviation σ

Formula:

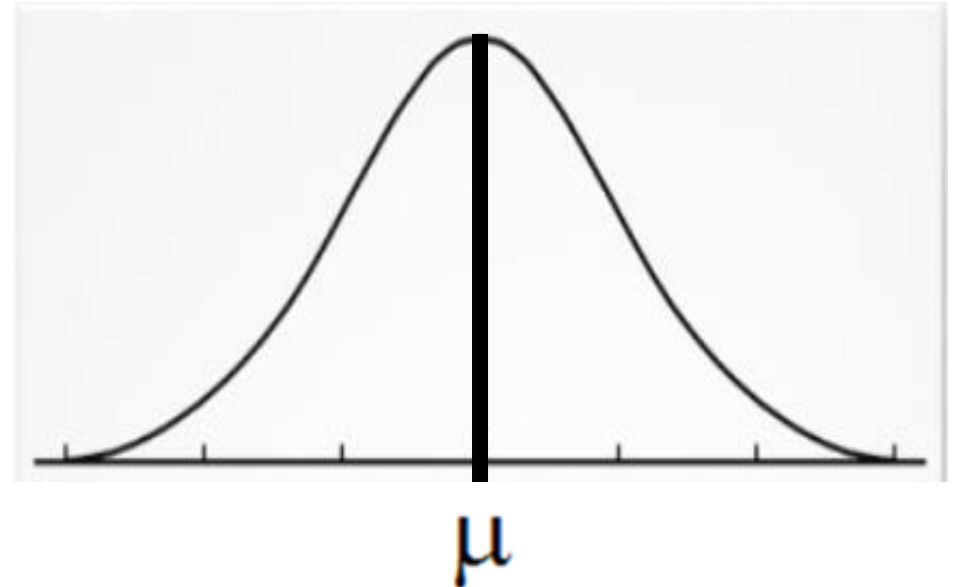
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} .$$

we can check if a variable is normally distributed



Rule: A numerical variables is normal distributed if

- ✓ Arithmetic mean = Median = Mode (or near equal)
- ✓ Quartile 1, Quartile 3 are simetrical with the mean (or near simetrical)
- ✓ Skewness ≈ 0 (between -1 to 1)
- ✓ Kurtosis ≈ 0 (between -1 to 1)

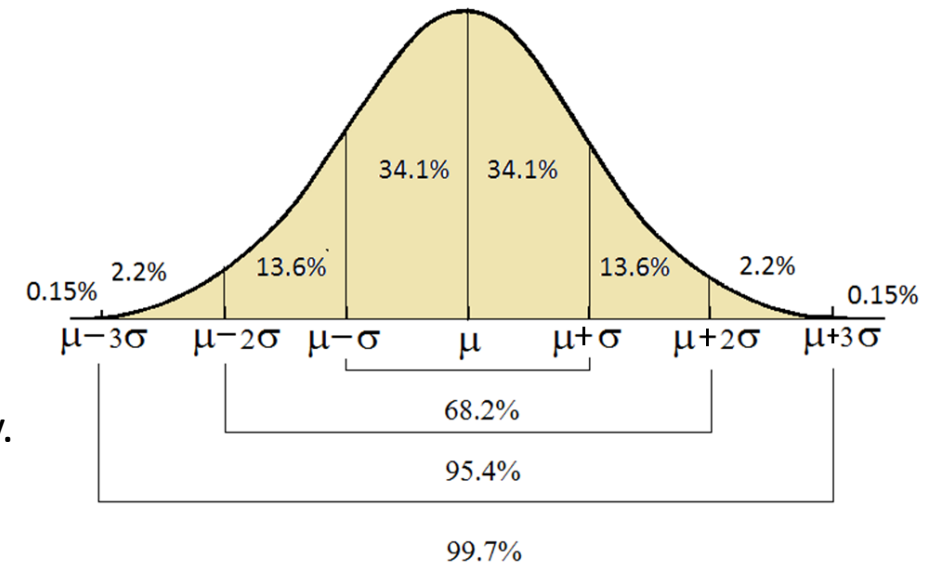


Other properties of normal distribution

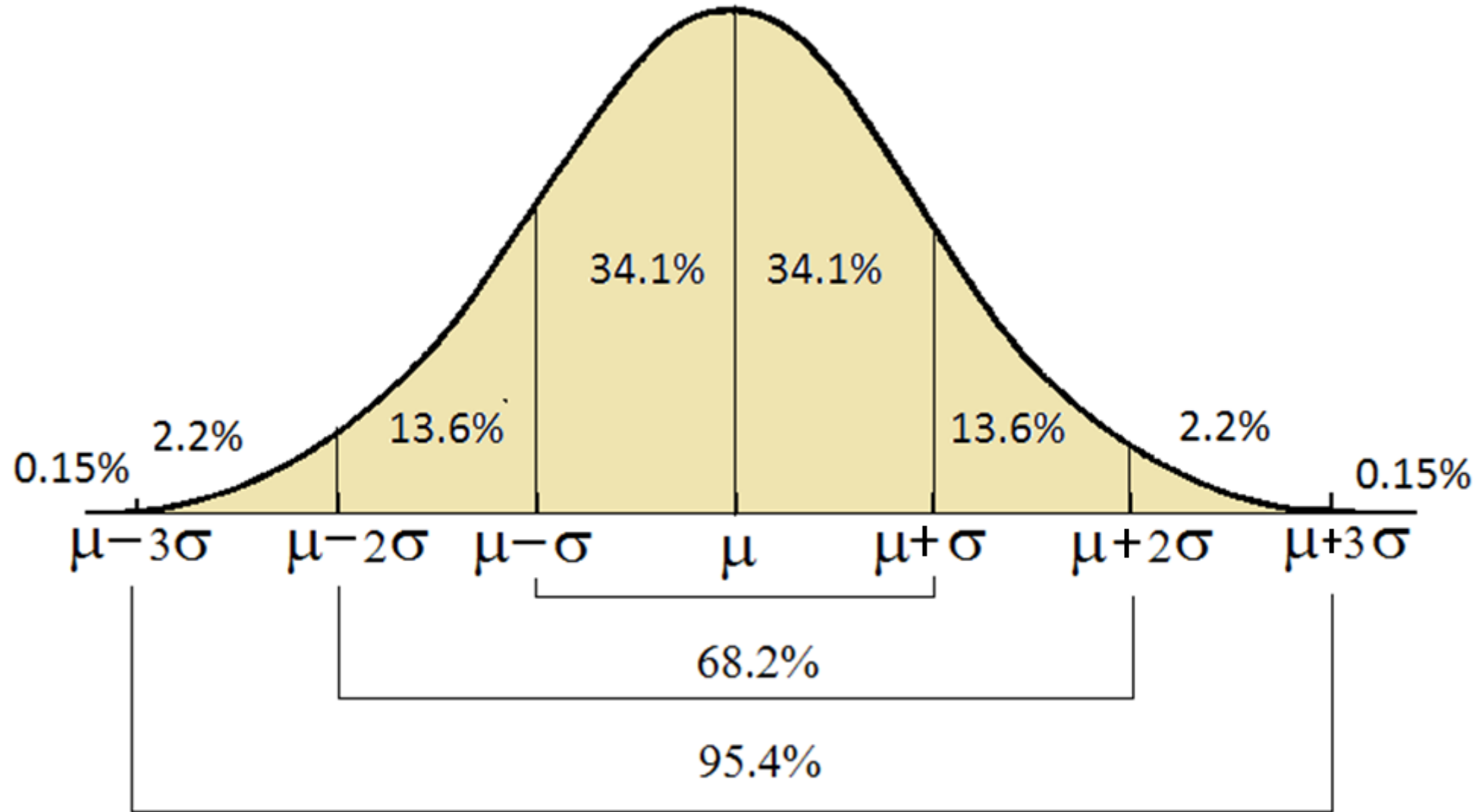
- ✓ In the interval: $\text{mean} \pm \text{st.dev.}$ there are minimum 68.2% of data
- ✓ In the interval: $\text{mean} \pm 2^* \text{ st.dev.}$ there are minimum 95.4% of data
- ✓ In the interval: $\text{mean} \pm 3^* \text{ st.dev.}$ there are minimum 99.7% of data



Where
st.dev. – standard deviation
mean – arithmetic mean
 $\text{mean} \pm \text{st.dev.}$ – the interval between mean-st.dev and mean+st.dev.
 μ – the population arithmetic mean
 σ – the standard deviation



Normal distribution



Where

μ - the population arithmetic mean

σ - the standard deviation

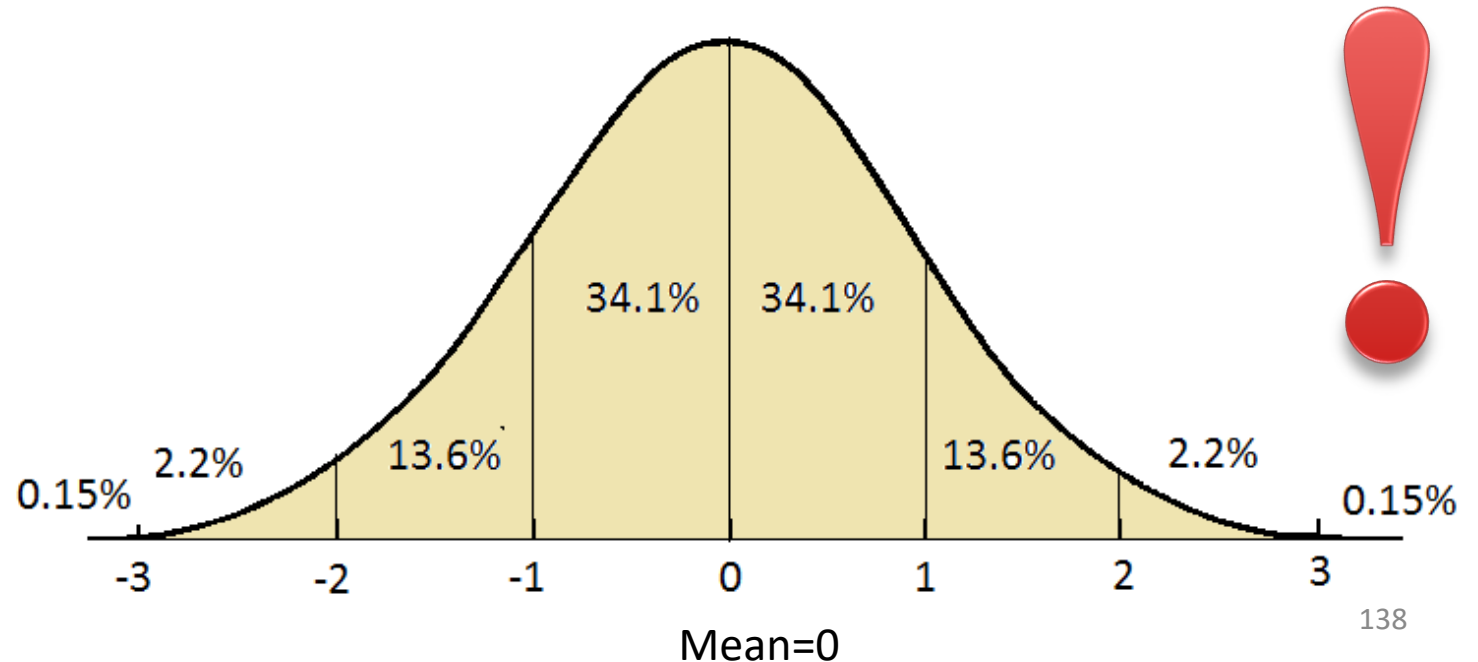
The **Standard** Normal Distribution

A normal distribution with $\mu=0$ and $\sigma=1$. We change the variable with:

$$Z = \frac{X - \mu}{\sigma}$$

Formula:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$



THE BINOMIAL LAW

- n repeated trials
- p – probability of success
- q – probability of failure
- Probability of k successes:

$$\Pr(X = k) = C_n^k p^k q^{n-k}$$

where k-combination of an n-set

$$C_n^k = \frac{n!}{k! * (n-k)!}$$

k factorial

$$k! = k \times (k-1) \times (k-2) \times \dots \times 2 \times 1$$



- 10 Covid-19 patients in the ICU
- the probability of survival is $p = 0.7$
- the probability of death is $q = 0.3$
- What is the probability that exactly one will survive?
- $$P(1 \text{ survive from } 10) = \frac{10!}{1!(10-1)!} 0.7^1 0.3^9 = \frac{3,628,800}{362,880} * 0.7 * 0.000019683 = 0.000138$$



POISSON LAW



1781–1842

- infinite discrete variable
- Poisson distribution model: The number of occurrences of an event within a given interval of time or space
- Example:
 - number of hospital admissions per year,
 - number of ambulances required per month,



POISSON LAW

- characterized by a parameter θ – the theoretical (expected) average number of occurrences of the event within the interval under consideration
- The probability that the event occurs k times within the interval under consideration:

$$\Pr(X = k) = \frac{e^{-\theta} \theta^k}{k!}$$

where $e \approx 2.71828$



Example: Crohn disease cases in 1 year

- The frequency of Crohn's disease (intestinal inflammation) = 3 in 100 people in one year = 0.03
- What is the probability of having 6 people with Crohn's disease in a year?
- $\theta = 0.03$,
- $\Pr(k=6/\text{year}) = \frac{e^{-0.03} \cdot 0.03^6}{6!} = 0.83 \cdot 10^{-13}$



Conclusion

- A medical question
- Select a sample randomly from the target population
- Make inference to the population by modeling the phenomenon
 - with known law distribution



Thank you!