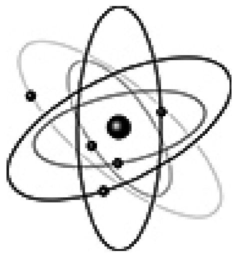
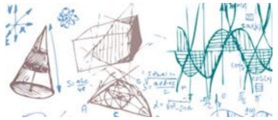
Introduction to probability, statistics and data handling

Agnieszka Obłąkowska-Mucha
Tomasz Szumlak

Faculty of Physics and Applied Computer Science
AGH University of Krakow





Intro

- Lectures 30 hours
- Tutorials 20 hours (**compulsory!**)
- Computer Labs 10 hours (**compulsory!**)
- Final grades – 0.2 lecture activity 0.4 tutorial + 0.4 labs

  Note, you may not fail neither tutorial nor labs
- Please make sure in advance that you have computer account with the Uni!
- Our contact details
  - ❑ Agnieszka Obłąkowska-Mucha (lectures): amucha@agh.edu.pl
    Saliha Bashir (tutorials and labs): bashir@agh.edu.pl
  - ❑ Building D11/110

Intro

- Course web page: https://agnieszkamucha.github.io/Statistics/
- Textbook: Introductory Statistics, OpenStax Access online
- Lectures: slides with discussions
- Computer Laboratory: follow the instructions, complete tasks
- Tutorials:
    1. find assignments BEFORE the day of tutorial
    2. find and read matching lecture or chapter in textbook
    3. solve or think over the assignments

Data and Experiments

- Our primary concern in Statistics is **analysing data** collected from an **experiment** – this is an essential notion and we should define it well
- Just as in the lab, we define an **experiment as a procedure** to be followed – the outcome of this procedure **constitutes the result** which can be represented as a **single quantity** or a **set of quantities** or a **distribution**
- These measured quantities can be **discrete** or **continuous**
- Note: no matter how accurately all conditions of the experiment are maintained, its results will in general differ – the measurement has an **intrinsic random component**
- This can be attributed to the very **nature** of the observed phenomenon or **limited accuracy** of the measurement
- This is why we should always use **statistics** to **process** the results of an experiment and understand them

$$\boldsymbol{Result = Value \pm Uncertainty\ [UNIT]}$$

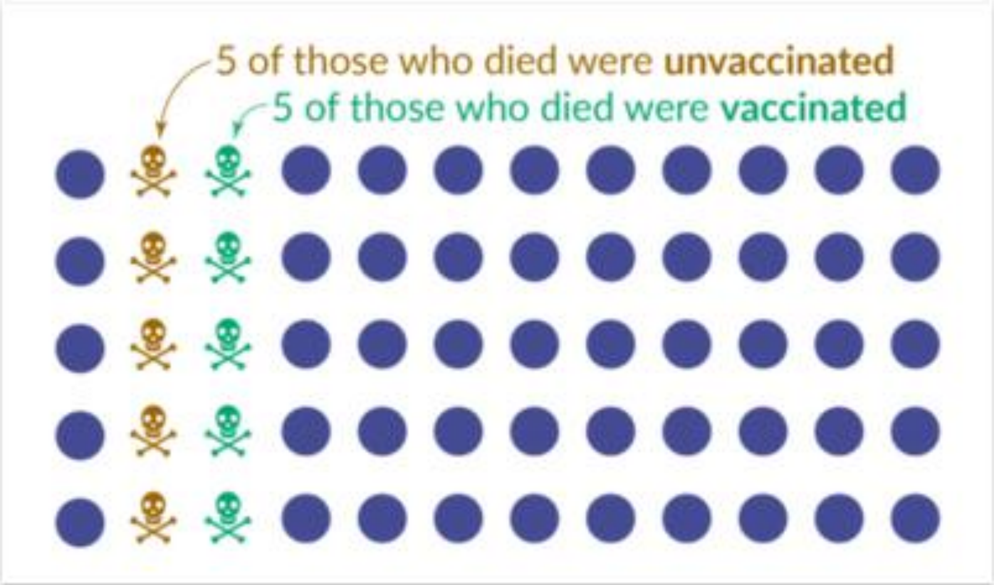
Where can you find statistics?

Everywhere…

Data

- Imagine we live in a place with a population of 60 people.
- 10 people died because of Covid-19. And we learn that 50% of them were vaccinated.

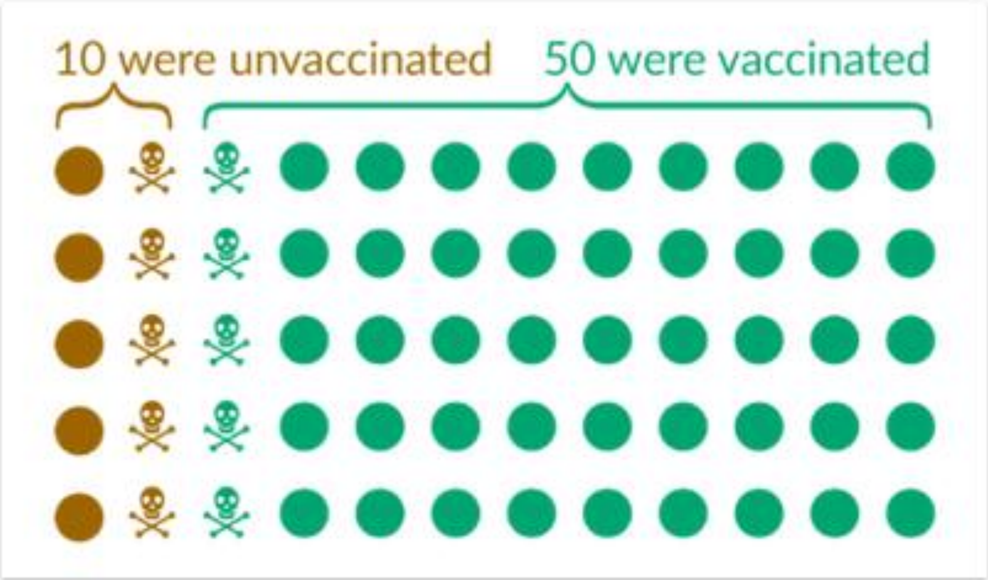
The newspaper may read, “Half of those who died from the virus were vaccinated.”



We need to know about those who did *not* die: how many people in this population were vaccinated? And how many were not vaccinated?

https://ourworldindata.org/covid-deaths-by-vaccination

Data



- Now we have all the information we need and can calculate the death *rates*:
- Of 10 unvaccinated people, 5 died → **the death rate among the unvaccinated is 50%**.
- Of 50 vaccinated people, 5 died → **the death rate among the vaccinated is 10%**.
- We, therefore, see that the death rate among the vaccinated is **5 times lower** than among the unvaccinated.

Probability vs statistics

Probability serves as the foundational framework for statistical analysis, providing the theoretical basis for understanding uncertainty and randomness in data.

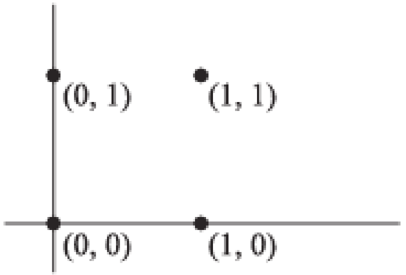
Statistics utilizes probability theory to make inferences, draw conclusions, and quantify the likelihood of different outcomes within a given dataset.

Sample space (I)

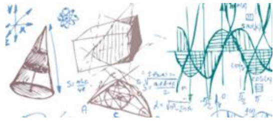
- **Def.1** A set Ω that consists of all possible outcomes of an experiment is called a **sample space**, then each outcome is called a **sample point**. Often, we can define more than one sample space.

- **Ex.1** Imagine we toss a die once – a sample space of all possible results we can get is given by $\Omega = \{1, 2, \ldots, 5, 6\}$

- **Ex. 2** Let's toss a coin twice. We can use the following: 0 == tails and 1== heads. The sample space can be then represented on a graph like this:



- The above corresponds to a space: $\Omega = \{HH, HT, TH, TT\}$



Sample space (II)

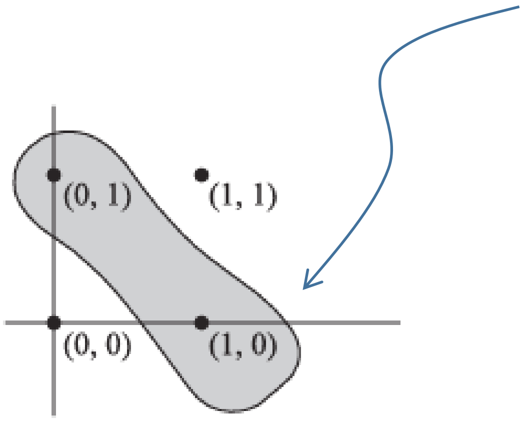
- **Def. 2** If a sample space has a finite number of points, it is called a **finite sample** space.
- **Def. 3** If a sample space has as many points as there are natural numbers (1, 2, 3, …, N, …), it is called a **countably infinite space**.
- **Def. 4** If a sample space has as many points as there are in an any interval on the x-axis ($a \leq x \leq b$), it is called a non-countably infinite space.
- **Def. 5** A sample space that is finite or countably infinite is called a **discrete sample space**.
- **Def. 6** A sample space that is non-countably infinite is called a **continuous sample space**.

Events (I)

- **Def. 7** An **event** is a subset $\mathbb{A}$ of the sample space Ω, i.e., it is a set of possible outcomes that we are interested in.

- **Def. 8** If the outcome of an experiment is an element of $\mathbb{A}$ we say that the event $\mathbb{A}$ has **occurred**. An event consisting of a single point, belonging to sample space, is called an **elementary event**.

- **Ex. 3** We can use the sample space from ex. 2 to define an event $\mathbb{A}$: ‚only one head comes up'.



Events (II)

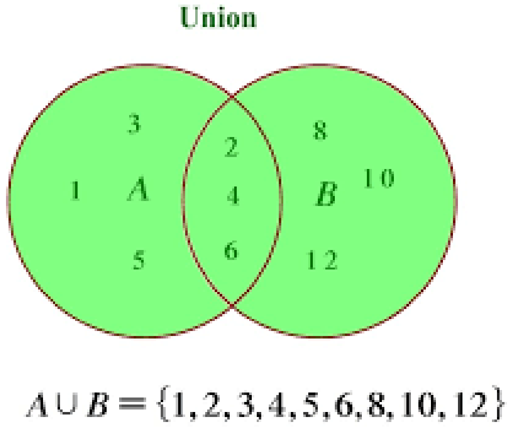
- **Def. 9** We call the sample space the **certain event**, since an element that belongs to Ω must occur in our experiment.

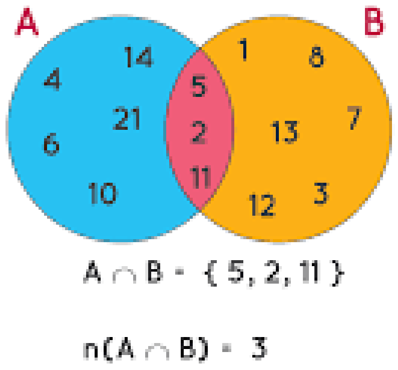
- **Def. 10** By analogy, an empty set ∅, is called the **impossible event**.

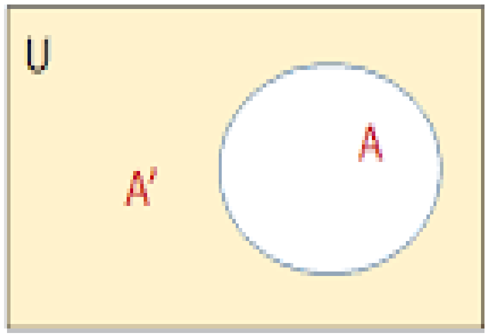
  So, using set operations on events we can obtains other events!

  - The **union**, $\mathbb{A} \cup \mathbb{B}$, of $\mathbb{A}$ and $\mathbb{B}$ means „either $\mathbb{A}$ or $\mathbb{B}$ or both"
  - The **intersection**, $\mathbb{A} \cap \mathbb{B}$, of $\mathbb{A}$ and $\mathbb{B}$ means „both $\mathbb{A}$ and $\mathbb{B}$"
  - The **complement,** $\mathbb{A}'$, means „not $\mathbb{A}$"

  Event $\mathbb{A} - \mathbb{B} = \mathbb{A} \cap \mathbb{B}'$, means „$\mathbb{A}$ but not $\mathbb{B}$". We have in particular, $\mathbb{A}' = \Omega - \mathbb{A}$







Sets and examples

Prove the following distributive law:
Let A, B, and C be subsets of a Universal set U.

$$1)\ A \cap (B \cup C) = (A \cap B) \cup (A \cap C).$$
$$2)\ (A \cap B) - C = (A - C) \cap (B - C)$$

U = {1,2,3,4,5}
A = {1,2}
B = {2,3}
C = {1,2,5}

Events (III)

- **Def. 11** If the sets corresponding to events $\mathbb{A}$ and $\mathbb{B}$ are **disjoint**, $\mathbb{A} \cap \mathbb{B} = \emptyset$, we say that the events are **mutually exclusive**. They cannot both occur at the same time
- **Def. 12** We say that a collection of events $\mathbb{A}_1, \mathbb{A}_2, \dots, \mathbb{A}_n$ is mutually exclusive if and only if every pair in the collection is mut. excl.

Events (IV)

Ex. 4 Using further ex. 2, let's use the set operations on the following events: „at least one head occurs" and „the second toss result is a tail". Determine the outcome of all the operations listed.

1. Are these sets mutually exclusive?
2. What is $\mathbb{A} \cup \mathbb{B}$?
3. What is $\mathbb{A} \cap \mathbb{B}$?
4. What is $\mathbb{A}'$?

Probability (I)

- The main premise here is that we can assign to **events numbers** that can **measure** the probability
- **Def. 13** If an event can occur in n different ways out of a total number of N possible ways, all of which are **equally likely**, then the probability of the event equals to n/N
- **Ex. 5** In the case of a **fair coin** toss we have **two equally** likely events, so it seems reasonable to assign them probability $p(H) = p(T) = 1/2$. If in an experiment **we measure a bias** in the number of heads or tails we will call the **coin loaded**
- **Def. 14** If after N repetitions of an experiment, where N **should be large**, a particular event is observed to occur n times, then the probability of the event is n/N

Probability (II)

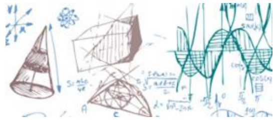
- There are some serious troubles with both definitions given in the previous slide…

  How can we tell if events are **equally likely**?

  What does it mean that a sample should be **large**?

- These issues are „cured" by the **axiomatic** approach to the probability. The core element in the axiomatic definition is a notion of a probability function $p(\mathbb{A})$, which gives a number related with each event.

  - **Axiom 1** For every event: $p(\mathbb{A}) \geq 0$
  - **Axiom 2** For the certain event $p(\Omega) = 1$
  - **Axiom 3** For any number of mutually exclusive events $\mathbb{A}_1, \mathbb{A}_2, \ldots, \mathbb{A}_n$ $p(\mathbb{A}_1 \cup \mathbb{A}_2 \cup \cdots \cup \mathbb{A}_n) = p(\mathbb{A}_1) + p(\mathbb{A}_2) + \cdots + p(\mathbb{A}_n)$



Probability (III)

Theorem 1

If $\mathbb{A}_1 \subset \mathbb{A}_2 \rightarrow p(\mathbb{A}_1) \leq p(\mathbb{A}_2)$

Theorem 2

For every event $\mathbb{A} \rightarrow 0 \leq p(\mathbb{A}) \leq 1$

Theorem 3

The impossible event has probability zero $p(\emptyset) = 0$

Theorem 4

If $\mathbb{A}'$ is the complement of $\mathbb{A}$, then $p(\mathbb{A}') = 1 - p(\mathbb{A})$

Theorem 5

If $\mathbb{A}_1$ and $\mathbb{A}_2$ are any two events, then

$p(\mathbb{A}_1 \cup \mathbb{A}_2) = p(\mathbb{A}_1) + p(\mathbb{A}_2) - p(\mathbb{A}_1 \cap \mathbb{A}_2)$

Theorem 6

For any events $\mathbb{A}_1$ and $\mathbb{A}_2$ $p(\mathbb{A}_1) = p(\mathbb{A}_1 \cap \mathbb{A}_2) + p(\mathbb{A}_1 \cap \mathbb{A}'_2)$

Examples

- **Ex. 5** Lottery

A container holds 49 balls, each with a number 1 through 49. During the drawing six of them are taken out without replacement. What is the probability that a player has chosen exactly the same numbers?

Say, $p(1)$ is the probability to chose the first number and is equal: $p(1) = \frac{1}{49}$, then the probability to chose by the player the second number is $p(2) = \frac{1}{48}$… and so on. At the end we have:

$$p(1,..,6) = \frac{1}{49 \cdot 48 \cdot 47 \cdot 46 \cdot 45 \cdot 44} = \frac{43!}{49!}$$

Now, the order is not important!

$$p(win) = \frac{6!\,43!}{49!} = \frac{1}{\binom{49}{6}} = \frac{1}{C_6^{49}}$$

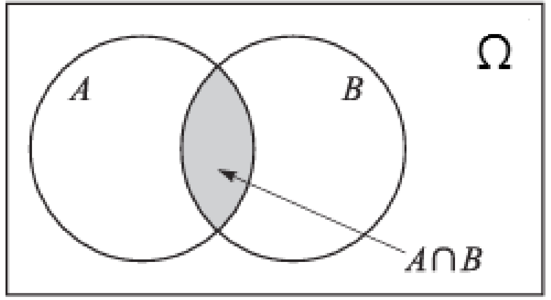
Where, C_6^{49} is the number of combinations of 6 elements out of 49.

Conditional probability

- **Conditional probability**, i.e., probability is not absolute (subjective point of view may have a significant impact on a value you assign to an event!)

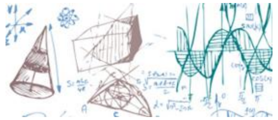
- **Def.1** Let $\mathbb{A}$ and $\mathbb{B}$ be two events and assume that $p(\mathbb{A}) > 0$. We denote by $p(\mathbb{B}|\mathbb{A})$ the probability of event $\mathbb{B}$ given that event $\mathbb{A}$. Since we know that $\mathbb{A}$ has occurred it becomes the new sample space instead of the original one Ω. Thus, we define the conditional probability as:

$$p(\mathbb{B}|\mathbb{A}) = \frac{p(\mathbb{A} \cap \mathbb{B})}{p(\mathbb{A})}$$



$$p(\mathbb{A} \cap \mathbb{B}) = p(\mathbb{B}|\mathbb{A})p(\mathbb{A})$$

- By **requiring** the occurrence of event $\mathbb{A}$ we make the **event space** to **collapse** – original probabilities **are redefined** – we could say that values of probability that we assign **depend on our knowledge**!



Conditional probability

- **Ex. 1** We have a fair die which is tossed once. Calculate the probability that a single toss of a die will result in a number less than 4. Repeat the math if it is given that the toss resulted in an odd number.
  - If no additional information is available we estimate the probability of turning 4 as a union of fundamental events 1, 2, 3:

$$p(\mathbb{B}) = p(1) + p(2) + p(3) = 1/2$$

  - Now, we know that an event „an odd number turned” has occurred, we have:

$$p(\mathbb{A}) = \frac{3}{6} = \frac{1}{2}$$

$$p(\mathbb{A} \cap \mathbb{B}) = \frac{2}{6} = \frac{1}{3} \rightarrow p(\mathbb{B}|\mathbb{A}) = \frac{1/3}{1/2} = \frac{2}{3}$$

  - In this case we „added knowledge” to our calculations! We knew that an odd number turned.
  - So, the values of probabilities we are going to assign to events depends on the extent of our knowledge about this event. In time it can change!

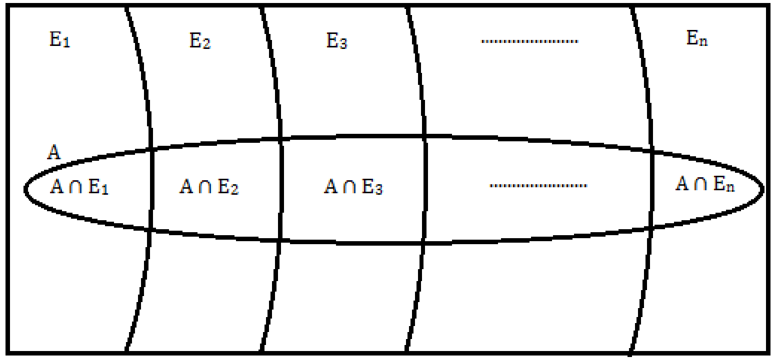
Some theorems on CP

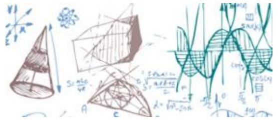
- **Theorem.1** Let say we have three events $\mathbb{A}_1, \mathbb{A}_2, \mathbb{A}_3$. The probability that they all occur can be calculated as follow (can easily be generalised to any number of events):

$$p(\mathbb{A}_1 \cap \mathbb{A}_2 \cap \mathbb{A}_3) = p(\mathbb{A}_1)p(\mathbb{A}_2|\mathbb{A}_1)p(\mathbb{A}_3|\mathbb{A}_1 \cap \mathbb{A}_2)$$

- **Theorem.2** Imagine that we have an event $\mathbb{A}$ that can be represented by n mutually exclusive events $\mathbb{E}_i$ ($\Omega = \mathbb{E}_1 \cup \mathbb{E}_2 \cup \ldots \cup \mathbb{E}_n$), then

$$p(\mathbb{A}) = \sum_{i/0}^{i/n} p(\mathbb{A} \cap \mathbb{E}_i) = p(\mathbb{E}_1)p(\mathbb{A}|\mathbb{E}_1) + \cdots + p(\mathbb{E}_n)p(\mathbb{A}|\mathbb{E}_n)$$





Independent events

- It is kind of easy to notice, that if there is no influence of event $\mathbb{A}$ on event $\mathbb{B}$ (so it does not matter if the former event occur or not), then

$$p(\mathbb{B}|\mathbb{A}) = p(\mathbb{B})$$

- In other words we say that $\mathbb{A}$ and $\mathbb{B}$ are **independent events**

$$\boxed{p(\mathbb{A} \cap \mathbb{B}) = p(\mathbb{B}|\mathbb{A})p(\mathbb{A}) = p(\mathbb{B})p(\mathbb{A})}$$

- This definition can easily be extended to any number of events. Say, we have three events $\mathbb{A}_1, \mathbb{A}_2, \mathbb{A}_3$ and they are pair-wise independent

$$p\left(\mathbb{A}_i \cap \mathbb{A}_j\right) = p(\mathbb{A}_i)p\left(\mathbb{A}_j\right), i \neq j, i, j = 1,2,3$$

$$p(\mathbb{A}_1 \cap \mathbb{A}_2 \cap \mathbb{A}_3) = p(\mathbb{A}_1)p(\mathbb{A}_2)p(\mathbb{A}_3)$$

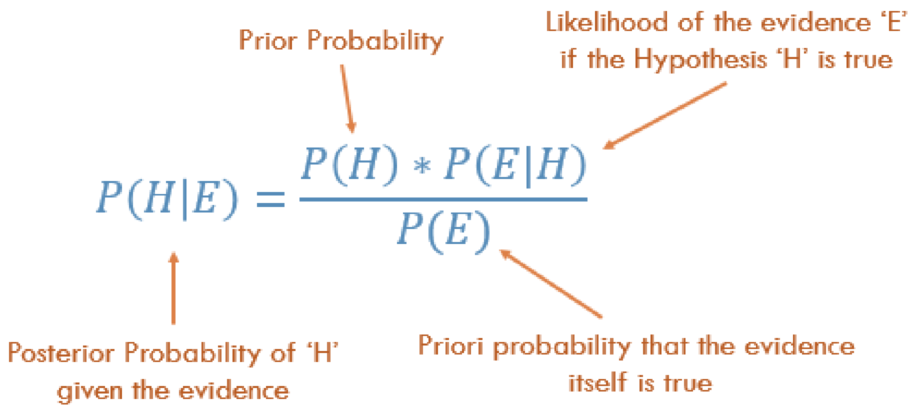
- Now, putting this all together, we come to a very interesting conclusion that is called…

Bayes' Theorem (I)

- Let's assume that the sample space can be divided into mutually exclusive events: $\Omega = \mathbb{E}_1 \cup \mathbb{E}_2 \cup \ldots \cup \mathbb{E}_n$ (see also plot on slide 16), by this we enforce that any of $\mathbb{E}_i$ **must occur**. Now, if $\mathbb{A}$ is any event from this sample space then we have the following theorem:

$$p(\mathbb{E}_i|\mathbb{A}) = \frac{p(\mathbb{E}_i)p(\mathbb{A}|\mathbb{E}_i)}{\sum_{i/0}^{i/n} p(\mathbb{E}_i)p(\mathbb{A}|\mathbb{E}_i)} = \frac{p(\mathbb{E}_i)p(\mathbb{A}|\mathbb{E}_i)}{p(\mathbb{A})}$$

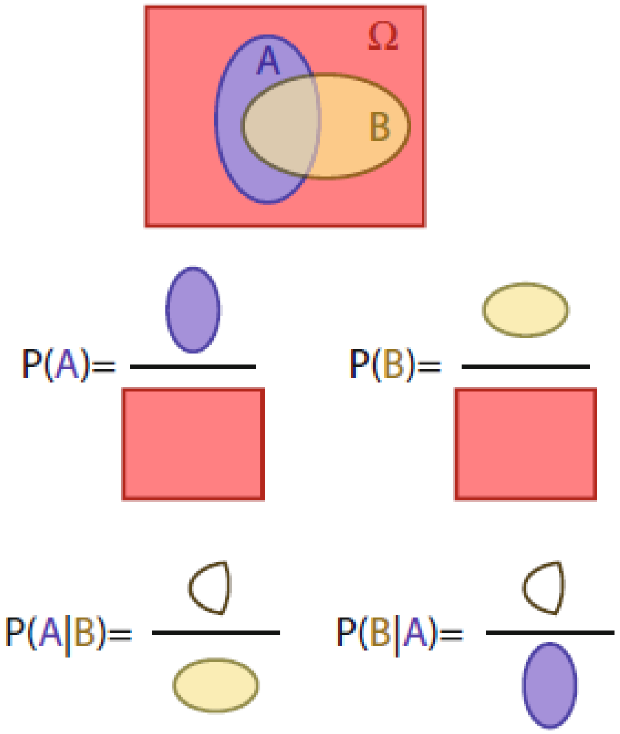
- By means of Bayes theorem we can estimate probability of various events that can cause $\mathbb{A}$ to occur. Therefore, it is also called a theorem on prob. of causes.



agh.edu.pl

Bayes' Theorem (II)

Pictures are nice, so…

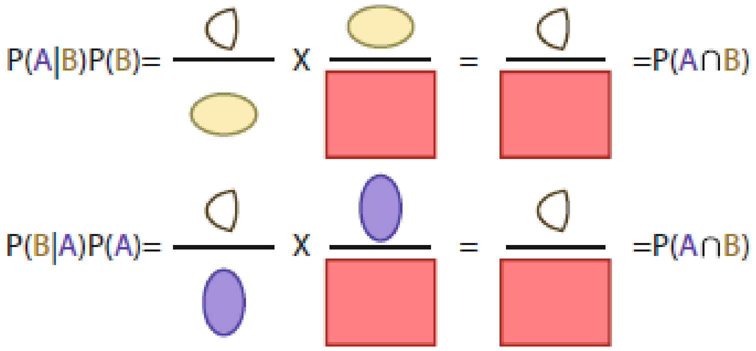


$\boldsymbol{p}(\mathbb{A})$ - **prob. of A, we have no knowledge of B**

$$p(\mathbb{A}|\mathbb{B}) = \frac{p(\mathbb{A})p(\mathbb{B}|\mathbb{A})}{p(\mathbb{B})}$$

$\boldsymbol{p}(\mathbb{A}|\mathbb{B})$ - **here, we are smarter, we know B has occurred**

Bayes' Theorem (III)



- One amazing thing about Bayes approach is capability of working out probabilities of events **that cannot be even considered by frequentists** (which deal with repeatable cases)
- A critical difference (that is very useful for experimentalists) between frequentist and Bayesian methods is that the latter can give a probability that **a value of unknown parameter** (something that came from theory for instance) **lies within a certain interval**

  Note, since a parameter is just a number and cannot be considered a random variable, **this procedure would not be possible in a classical probability**

Bayes' Theorem (IV)

- For a long time Bayes theorem was treated as not so interesting rule that is a simple consequence of conditional probability definition. However, recently the interpretation of the rule was revisited what sparked a new branch of statistics called Bayesian approach.
- Often treated as alternative/complementary way of assigning probabilities
- Let's rewrite the theorem as follow:

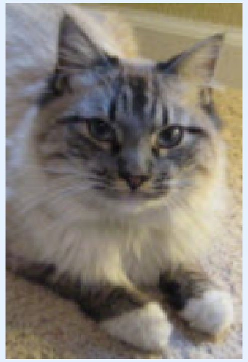
$$p(theory|data) \propto p(data|theory)p(theory)$$

- Here we use **subjective probabilities**, that express a **degree of believe** that something is true – that is the core of the Bayesian approach
- „Theory" represents a hypothesis, „data" represent the outcome of an experiment
- $p(theory)$ - represent a **prior** probability, or a degree of believe before measurement, $p(data|theory)$ - represent **likelihood** of getting the data given the theory is true
- The $p(theory|data)$, **posterior**, tells us, how the **prior** probability should be changed given the observed data

Bayes’ Theorem Example

Example: Allergy or Not?

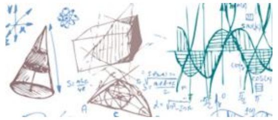
Hunter says she is itchy. There is a test for Allergy to Cats, but this test is not always right:



- For people that **really do** have the allergy, the test says "Yes" **80%** of the time
- For people that **do not** have the allergy, the test says "Yes" **10%** of the time ("false positive")

If 1% of the population have the allergy, and **Hunter's test says "Yes"**, what are the chances that Hunter really has the allergy?

https://www.mathsisfun.com/data/bayes-theorem.html



Bayes' Theorem Example

We want to know the chance of having the allergy when test says "Yes", written **P(Allergy|Yes)**

Let's get our formula:

$$P(\text{Allergy}|\text{Yes}) = \frac{P(\text{Allergy})\ P(\text{Yes}|\text{Allergy})}{P(\text{Yes})}$$

- P(Allergy) is Probability of Allergy = 1%
- P(Yes|Allergy) is Probability of test saying "Yes" for people with allergy = 80%
- P(Yes) is Probability of test saying "Yes" (to anyone) = ??%

Oh no! We **don't know** what the **general** chance of the test saying "Yes" is ...

... but we can calculate it by adding up those **with**, and those **without** the allergy:

- 1% have the allergy, and the test says "Yes" to 80% of them
- 99% do **not** have the allergy and the test says "Yes" to 10% of them

	Test says "Yes"	Test says "No"
Have allergy	**80%**	**20% "False Negative"**
Don't have it	**10% "False Positive"**	**90%**

Bayes' Theorem Example

$$P(A|B) = \frac{P(A)P(B|A)}{P(A)P(B|A) + P(\text{not } A)P(B|\text{not } A)}$$

- A in this case is "actually has the allergy"
- B in this case is "test says Yes"

P(A|B) means "The probability that Hunter actually has the allergy given that the test says Yes"

P(B|A) means "The probability that the test says Yes given that Hunter actually has the allergy"

To be clearer, let's change A to **has** (actually has allergy) and B to **Yes** (test says yes):

$$P(\text{has}|\text{Yes}) = \frac{P(\text{has})P(\text{Yes}|\text{has})}{P(\text{has})P(\text{Yes}|\text{has}) + P(\text{not has})P(\text{Yes}|\text{not has})}$$

And put in the numbers:

$$P(\text{has}|\text{yes}) = \frac{0{,}01 \times 0{,}8}{0{,}01 \times 0{,}8 + 0{,}99 \times 0{,}1} = 0{,}0748\ldots$$

Which is about **7%**

Another example

Example: Picnic Day

You are planning a picnic today, but the morning is cloudy

- Oh no! 50% of all rainy days start off cloudy!
- But cloudy mornings are common (about 40% of days start cloudy)
- And this is usually a dry month (only 3 of 30 days tend to be rainy, or 10%)

What is the chance of rain during the day?

Answer is 0.125…