Statistical analysis of experimental data
Concept of probability

Aleksander Filip Żarnecki

FACULTY OF PHYSICS
UNIVERSITY OF WARSAW

Lecture 01

October 3, 2024

Concept of probability

Concept of probability

Statistical analysis of experimental data

Quantitative research, not only in physics, involve different stages:

- design of the experiment
  It can require building a dedicated experiment (physics, astronomy), preparing survey forms (social sciences) or defining test and control population (medical sciences).

Statistical analysis of experimental data

Quantitative research, not only in physics, involve different stages:

- design of the experiment
  It can require building a dedicated experiment (physics, astronomy), preparing survey forms (social sciences) or defining test and control population (medical sciences).
- running the experiment: actual measurement
  We run the experiment and collect data. We usually try to express the result of the experiment in numerical form...

Introduction

Statistical analysis of experimental data

Quantitative research, not only in physics, involve different stages:

- design of the experiment
  It can require building a dedicated experiment (physics, astronomy), preparing survey forms (social sciences) or defining test and control population (medical sciences).
- running the experiment: actual measurement
  We run the experiment and collect data. We usually try to express the result of the experiment in numerical form...

  Experiment can result in single measurement (single number). But we usually prefer to have a large number of equivalent measurements...

Statistical analysis of experimental data

Quantitative research, not only in physics, involve different stages:

- design of the experiment
  It can require building a dedicated experiment (physics, astronomy), preparing survey forms (social sciences) or defining test and control population (medical sciences).
- running the experiment: actual measurement
  We run the experiment and collect data. We usually try to express the result of the experiment in numerical form...

  Experiment can result in single measurement (single number). But we usually prefer to have a large number of equivalent measurements...
- Statistical analysis and interpretation of the results
  This is what we will focus on in this course...

Statistical analysis of experimental data

Quantitative research, not only in physics, involve different stages:

- design of the experiment
  It can require building a dedicated experiment (physics, astronomy), preparing survey forms (social sciences) or defining test and control population (medical sciences).
- running the experiment: actual measurement
  We run the experiment and collect data. We usually try to express the result of the experiment in numerical form...

  Experiment can result in single measurement (single number). But we usually prefer to have a large number of equivalent measurements...
- Statistical analysis and interpretation of the results
  This is what we will focus on in this course...

Statistical analysis is also required for proper experiment design.

It is similarly important for interpretation of "single measurements"...

Statistical analysis of experimental data

The goal of this lecture is to present basic concepts and methods of statistical analysis of experimental data...

Statistical analysis of experimental data

The goal of this lecture is to present basic concepts and methods of statistical analysis of experimental data...

We will quite often refer to problems encountered in the elementary particle physics, as this is the perfect test ground for statistical analysis

Introduction

Statistical analysis of experimental data

The goal of this lecture is to present basic concepts and methods of statistical analysis of experimental data...

We will quite often refer to problems encountered in the elementary particle physics, as this is the perfect test ground for statistical analysis

This is because all particles (of the same type) are identical.
We can easily assure same initial conditions for the repeated measurement...
This is often a fundamental problem in social or medical sciences...

Statistical analysis of experimental data

The goal of this lecture is to present basic concepts and methods of statistical analysis of experimental data...

We will quite often refer to problems encountered in the elementary particle physics, as this is the perfect test ground for statistical analysis

This is because all particles (of the same type) are identical.
We can easily assure same initial conditions for the repeated measurement...
This is often a fundamental problem in social or medical sciences...

The knowledge of particle physics is not required (but for some fundamental concepts like the invariant mass or mean lifetime) and the presented analysis methods can clearly be used also in other fields of science...

Typical problems adapted from book by S.Brandt

- We measure properties of the selected individuals from a large population (could be elementary particles of a given type).
  What is the precision of the measurement?
  How large the test sample need to be to obtain given precision?

Introduction

Typical problems adapted from book by S.Brandt

- We measure properties of the selected individuals from a large population (could be elementary particles of a given type).
  What is the precision of the measurement?
  How large the test sample need to be to obtain given precision?

- A certain experimental result has been obtained. It can be compared with other experiments or with different theoretical predictions.
  How to decide, if the results is in agreement with the predicted theoretical value or with previous experiments?
  When can we claim that given theoretical model is excluded?

Typical problems adapted from book by S.Brandt

- We measure properties of the selected individuals from a large population (could be elementary particles of a given type).
  What is the precision of the measurement?
  How large the test sample need to be to obtain given precision?

- A certain experimental result has been obtained. It can be compared with other experiments or with different theoretical predictions.
  How to decide, if the results is in agreement with the predicted theoretical value or with previous experiments?
  When can we claim that given theoretical model is excluded?

- A general model describing the process studied is known, but parameters of this model must be obtained from experiment (very common case in particle physics).
  What is the optimum procedure for extracting model parameters from the data? How can the experiment be optimised to give strongest constraints on the model parameters?

Course plan

14 lectures, Thursdays, 9:15 – 12:00, room B2.38
From October 3th, 2024 to January 23th, 2025 (without January 9th!)

Course plan

14 lectures, Thursdays, 9:15 – 12:00, room B2.38
From October 3th, 2024 to January 23th, 2025 (without January 9th!)

Homework exercises
Solutions have to be uploaded to Kampus within two weeks!
Until midnight before the lecture! Proposed solutions will be presented at the lecture.

Course plan

14 lectures, Thursdays, 9:15 – 12:00, room B2.38
From October 3th, 2024 to January 23th, 2025 (without January 9th!)

Homework exercises

Solutions have to be uploaded to Kampus within two weeks!

Until midnight before the lecture! Proposed solutions will be presented at the lecture.

Written exam, five problems to be solved

Solutions to be uploaded to Kampus within one week!

By uploading the solutions to Kampus you declare that they resulted from your own work and that you have not shared nor discussed them with anyone.

Course plan

14 lectures, Thursdays, 9:15 – 12:00, room B2.38
From October 3th, 2024 to January 23th, 2025 (without January 9th!)

Homework exercises
Solutions have to be uploaded to Kampus within two weeks!
Until midnight before the lecture! Proposed solutions will be presented at the lecture.

Written exam, five problems to be solved
Solutions to be uploaded to Kampus within one week!
By uploading the solutions to Kampus you declare that they resulted from your own work and that you have not shared nor discussed them with anyone.

Assessment criteria

Minimum of 50% of points collected from exercises and exam (with the same weight).
Assessment in September: 50% of points from written examination only.

Solutions

Solutions (dedicated files or printouts of the notebook with the final output) should be uploaded to Kampus in readable format (PDF preferred).

Upload the Python notebook or add the link to your notebook in Google Colab as a comment!

Please share your Google Colab space with a.zarneckiuw.edu.pl, so I can access your notebook.

More details to be given next week

Solutions

Solutions (dedicated files or printouts of the notebook with the final output) should be uploaded to Kampus in readable format (PDF preferred).

Upload the Python notebook or add the link to your notebook in Google Colab as a comment!

Please share your Google Colab space with a.zarneckiuw.edu.pl, so I can access your notebook.

More details to be given next week

Web resources

Kampus platform will be used for home exercises and final exam:

https://kampus-kursy.ckc.uw.edu.pl/course/view.php?id=802

All information, including lecture slides will also be available from the dedicated web page:

http://www.fuw.edu.pl/~zarnecki/SAED/

accessible without USOS account...

Some references

- G. Bohm and G. Zech, Introduction to Statistics and Data Analysis for Physicsts, Verlag Deutsches Elektronen-Synchrotron, 3rd edition;
- S. Brandt, Data Analysis: Statistical and Computational Methods for Scientists and Engineers, Springer 2014;
- **M. Bonamente, Statistics and Analysis of Scientific Data**, Springer 2017;

- R.J. Barlow, Practical Statistics for Particle Physics, PDF from arXiv;
- Max Bramer, Principles of Data Mining, Springer 2016;
- David Forsyth, Probability and Statistics for Computer Science, Springer 2018;
- Particle Physics Reference Library Volume 2 (chapter 15), Springer 2020.

Concept of probability

Experiments

We perform an experiment to collect data.
To allow for statistical analysis we make many measurements...

However, there are always some random factors and results of subsequent measurements are usually different. It is important to understand where these variations come from!

Experiments

We perform an experiment to collect data.
To allow for statistical analysis we make many measurements...

However, there are always some random factors and results of subsequent measurements are usually different. It is important to understand where these variations come from!

1. Fluctuations in the measurement results can be due to the actual nature of the physics process studied. Examples: coin toss, roll of a die,

Experiments

We perform an experiment to collect data.
To allow for statistical analysis we make many measurements...

However, there are always some random factors and results of subsequent measurements are usually different. It is important to understand where these variations come from!

1. Fluctuations in the measurement results can be due to the actual nature of the physics process studied. Examples: coin toss, roll of a die, but also particle decay time measurement or measurement of source radioactivity...

Experiments

We perform an experiment to collect data.
To allow for statistical analysis we make many measurements...

However, there are always some random factors and results of subsequent measurements are usually different. It is important to understand where these variations come from!

1. Fluctuations in the measurement results can be due to the actual nature of the physics process studied. Examples: coin toss, roll of a die, but also particle decay time measurement or measurement of source radioactivity...

   These fluctuations are unavoidable, we can not reduce them.
   But usually this is also the most interesting case for us...

Experiments

We perform an experiment to collect data.
To allow for statistical analysis we make many measurements...

However, there are always some random factors and results of subsequent measurements are usually different. It is important to understand where these variations come from!

2. Fluctuations can result from the measurement method: finite precision of the instruments leading to different numerical results, also varying measurement conditions etc.
   Examples: measurement of the coin diameter or the die mass,

Experiments

We perform an experiment to collect data.
To allow for statistical analysis we make many measurements...

However, there are always some random factors and results of subsequent measurements are usually different. It is important to understand where these variations come from!

2. Fluctuations can result from the measurement method: finite precision of the instruments leading to different numerical results, also varying measurement conditions etc.
   Examples: measurement of the coin diameter or the die mass, measurements of the proton mass, electron charge, speed of light...

Experiments

We perform an experiment to collect data.
To allow for statistical analysis we make many measurements...

However, there are always some random factors and results of subsequent measurements are usually different. It is important to understand where these variations come from!

2. Fluctuations can result from the measurement method: finite precision of the instruments leading to different numerical results, also varying measurement conditions etc.
   Examples: measurement of the coin diameter or the die mass, measurements of the proton mass, electron charge, speed of light...

   We can try to reduce the fluctuations (and thus improve precision) by adjusting the measurement procedure...

Experiments

We perform an experiment to collect data.
To allow for statistical analysis we make many measurements...

However, there are always some random factors and results of subsequent measurements are usually different. It is important to understand where these variations come from!

3. Fluctuations can reflect inhomogeneity of the population studied.
   Examples: measurement of the drug effectiveness, answers to questionnaires in various types of surveys (eg. asking for age),

Experiments

We perform an experiment to collect data.
To allow for statistical analysis we make many measurements...

However, there are always some random factors and results of subsequent measurements are usually different. It is important to understand where these variations come from!

3. Fluctuations can reflect inhomogeneity of the population studied.
   Examples: measurement of the drug effectiveness, answers to questionnaires in various types of surveys (eg. asking for age), mass spectrometry, composition and energy spectra of cosmic rays...

Experiments

We perform an experiment to collect data.
To allow for statistical analysis we make many measurements...

However, there are always some random factors and results of subsequent measurements are usually different. It is important to understand where these variations come from!

3. Fluctuations can reflect inhomogeneity of the population studied.
   Examples: measurement of the drug effectiveness, answers to questionnaires in various types of surveys (eg. asking for age), mass spectrometry, composition and energy spectra of cosmic rays...

   Results will usually depend on the way the tested population sample is selected.
   This selection has to be well defined!

Experiments

It is crucial to correctly identify the source of fluctuations!

Example: coin diameter measurement.

Fluctuations in numerical results can be due to:

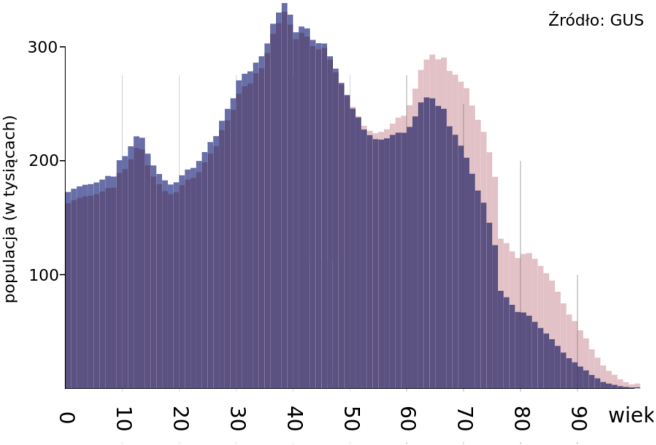
- finite precision of the instrument there is always some measurement error
- the measurement method, how we define the diameter
  in particular when the coin is not exactly round
- fluctuations in the actual coin size,
  if we measure a set of coins, not a single one

$\Rightarrow$ we need to define the problem properly!

In addition to measurement fluctuations, we also need to consider possible systematic effects.
We will come back to this later...

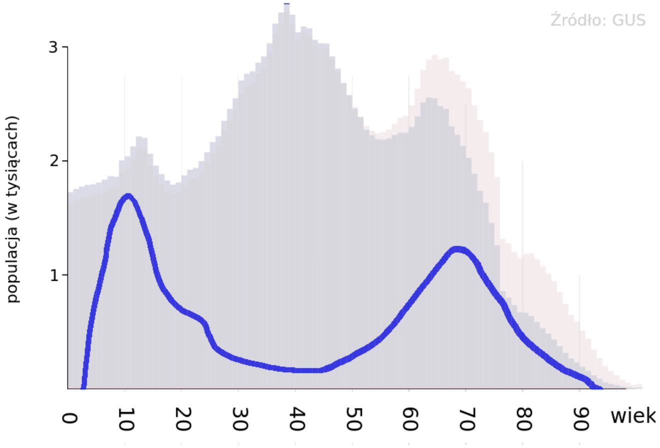
Experiments Example: influence of the test sample choice.

Age structure of the population of Poland (GUS estimate for 2022).



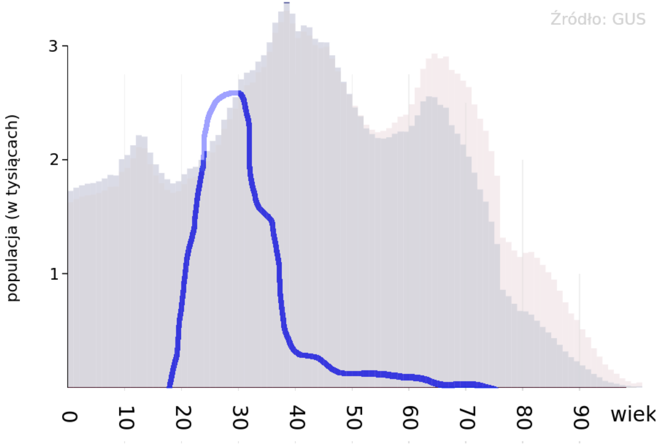
Experiments Example: influence of the test sample choice.

Possible result for survey in public park (on working day).



Experiments Example: influence of the test sample choice.

Possible result for survey at the university campus.



Elementary event

Outcome of a single experiment (measurement):

- result of the roll of a die
- observation of N charged particles in the particle collision
- observation of nuclear decay after given time
- observation of given process, eg. decay $K^+ \to \mu^+ + \nu_\mu$

Basic terms

Elementary event

Outcome of a single experiment (measurement):

- result of the roll of a die
- observation of N charged particles in the particle collision
- observation of nuclear decay after given time
- observation of given process, eg. decay $K^+ \rightarrow \mu^+ + \nu_\mu$

Sample space is the set of all possible outcomes of the experiment

Basic terms

Elementary event

Outcome of a single experiment (measurement):

- result of the roll of a die
- observation of N charged particles in the particle collision
- observation of nuclear decay after given time
- observation of given process, eg. decay $K^+ \to \mu^+ + \nu_\mu$

Sample space is the set of all possible outcomes of the experiment

Roll of a die

Six possible outcomes in the sample space: $\Omega = \{1, 2, 3, 4, 5, 6\}$

Basic terms

Elementary event

Outcome of a single experiment (measurement):

- result of the roll of a die
- observation of N charged particles in the particle collision
- observation of nuclear decay after given time
- observation of given process, eg. decay $K^+ \rightarrow \mu^+ + \nu_\mu$

Sample space is the set of all possible outcomes of the experiment

Charged particle multiplicity

Particle counting gives non-negative integer number: $\Omega = \{0, 1, 2, \ldots\}$

Basic terms

Elementary event

Outcome of a single experiment (measurement):

- result of the roll of a die
- observation of N charged particles in the particle collision
- observation of nuclear decay after given time
- observation of given process, eg. decay $K^+ \to \mu^+ + \nu_\mu$

Sample space is the set of all possible outcomes of the experiment

Decay time measurement

Time interval measured is a real number: $\Omega = \mathbb{R}_+$

Basic terms

Elementary event

Outcome of a single experiment (measurement):

- result of the roll of a die
- observation of N charged particles in the particle collision
- observation of nuclear decay after given time
- observation of given process, eg. decay $K^+ \to \mu^+ + \nu_\mu$

Sample space is the set of all possible outcomes of the experiment

Observation of K^+ decay:

Sample space should include all possible (observable) decay channels:

$$\Omega = \{K^+ \to \pi^+\pi^\circ, K^+ \to \pi^+\pi^\circ\pi^\circ, K^+ \to \pi^+\pi^+\pi^-, \quad K^+ \to e^+\nu_e, K^+ \to \mu^+\nu_\mu,$$

$$K^+ \to \pi^\circ e^+\nu_e, K^+ \to \pi^\circ\mu^+\nu_\mu, K^+ \to \pi^\circ\pi^\circ e^+\nu_e, K^+ \to \pi^+\pi^- e^+\nu_e, \quad K^+ \to \pi^+\mu^+e^-\}$$

Basic terms

Sample space

Sample space Ω is the set of all possible outcomes of the experiment

Event

An event A is a given subset of Ω, $A \subset \Omega$,
it can represent a number of possible outcomes for the experiment (!)

Basic terms

Sample space

Sample space Ω is the set of all possible outcomes of the experiment

Event

An event A is a given subset of Ω, $A \subset \Omega$,
it can represent a number of possible outcomes for the experiment (!)

Examples of events

Roll of a die

- six: $A_1 = \{6\}$
- odd number: $A_2 = \{1, 3, 5\}$
- even number: $A_3 = \{2, 4, 6\}$
- any number: $A_4 = \Omega = \{1, 2, 3, 4, 5, 6\}$

Sample space

Sample space Ω is the set of all possible outcomes of the experiment

Event

An event A is a given subset of Ω, $A \subset \Omega$,
it can represent a number of possible outcomes for the experiment (!)

Examples of events

Charged particle multiplicity

- pair production: $A_1 = \{2\}$
- odd number: $A_2 = \{1, 3, 5, \ldots\}$
- even number: $A_3 = \{2, 4, 6, \ldots\}$

Sample space

Sample space Ω is the set of all possible outcomes of the experiment

Event

An event A is a given subset of Ω, $A \subset \Omega$,
it can represent a number of possible outcomes for the experiment (!)

Examples of events

Observation of K^+ decay:

- Hadronic decays
$$A_1 = \{K^+ \to \pi^+\pi^\circ, K^+ \to \pi^+\pi^\circ\pi^\circ, K^+ \to \pi^+\pi^+\pi^-\}$$
- Leptonic decays
$$A_2 = \{K^+ \to e^+\nu_e, K^+ \to \mu^+\nu_\mu\}$$
- LFV decay (forbidden in SM)
$$A_3 = \{K^+ \to \pi^+\mu^+e^-\}$$

Events

From the formal point of view it is useful to introduce two events which exist for each experiment:

- impossible (empty) event: $A = \emptyset$
- sure event: $A = \Omega$

Events

From the formal point of view it is useful to introduce two events which exist for each experiment:

- impossible (empty) event: $A = \emptyset$
- sure event: $A = \Omega$

We also define:

- union of events $C = A \cup B$: all outcomes belonging to A **or** B
- intersection of events $D = A \cap B$: all outcomes belonging to A **and** B

Events

From the formal point of view it is useful to introduce two events which exist for each experiment:

- impossible (empty) event: $A = \emptyset$
- sure event: $A = \Omega$

We also define:

- union of events $C = A \cup B$: all outcomes belonging to A **or** B
- intersection of events $D = A \cap B$: all outcomes belonging to A **and** B
- complementary event $E = \bar{A}$: all outcomes not belonging to A

Events

From the formal point of view it is useful to introduce two events which exist for each experiment:

- impossible (empty) event: $A = \emptyset$
- sure event: $A = \Omega$

We also define:

- union of events $C = A \cup B$: all outcomes belonging to A **or** B
- intersection of events $D = A \cap B$: all outcomes belonging to A **and** B
- complementary event $E = \bar{A}$: all outcomes not belonging to A
- mutually exclusive events: two events with no common outcome
  $\Rightarrow$ their intersection is an empty event: $A \cap B = \emptyset$

Concept of probability

Definition of Probability

How to define probability?

We do have an intuitive understanding of the probability concept...

Probability $P(A)$ of an event A should describes the odds of an outcome of single measurement (experiment) to belong to A

Definition of Probability

How to define probability?

We do have an intuitive understanding of the probability concept...

Probability $P(A)$ of an event A should describes the odds of an outcome of single measurement (experiment) to belong to A

Probability is a number between 0 and 1

- probability of an empty event: $P(\emptyset) = 0$
- probability of a sure event: $P(\Omega) = 1$

How to define probability?

We do have an intuitive understanding of the probability concept...

Probability $P(A)$ of an event A should describes the odds
of an outcome of single measurement (experiment) to belong to A

Probability is a number between 0 and 1

- probability of an empty event: $P(\emptyset) = 0$
- probability of a sure event: $P(\Omega) = 1$

Classical definition as developed in the 18th–19th centuries

If the sample space contains N_Ω elementary events (possible outcomes of the experiment) and the considered event A contains N_A elementary events, then, **assuming all elementary events are equally probable**

$$P(A) = \frac{N_A}{N_\Omega}$$

Classical definition - example

see also in Wikipedia

In a straight poker game, consider player dealing five cards from a deck of 52.
Sampling space contains all possible "hands", all possible sets of 5 cards selected out of 52 (order is not relevant):

$$N_{\Omega} = \binom{52}{5} = 2\,598\,960$$

Classical definition - example

see also in Wikipedia

In a straight poker game, consider player dealing five cards from a deck of 52.
Sampling space contains all possible "hands", all possible sets of 5 cards selected out of 52 (order is not relevant):

$$N_\Omega = \binom{52}{5} = 2\,598\,960$$

What is the probability of a "straight flush" (five cards in a sequence, of the same suit)?

Classical definition - example

see also in Wikipedia

In a straight poker game, consider player dealing five cards from a deck of 52.
Sampling space contains all possible “hands”, all possible sets of 5 cards selected out of 52 (order is not relevant):

$$N_\Omega = \binom{52}{5} = 2\,598\,960$$

What is the probability of a “straight flush” (five cards in a sequence, of the same suit)?
We have four suits and the sequence can start from 9 numbers (from 1 to 9):

$$N_{flush} = 4 \times 9 = 36 \Rightarrow p_{flush} = \frac{36}{2\,598\,960} \approx 1.35 \cdot 10^{-5}$$

Classical definition - example

see also in Wikipedia

In a straight poker game, consider player dealing five cards from a deck of 52.
Sampling space contains all possible "hands", all possible sets of 5 cards selected out of 52 (order is not relevant):

$$N_\Omega = \binom{52}{5} = 2\,598\,960$$

What is the probability of a "straight flush" (five cards in a sequence, of the same suit)?
We have four suits and the sequence can start from 9 numbers (from 1 to 9):

$$N_{flush} = 4 \times 9 = 36 \Rightarrow p_{flush} = \frac{36}{2\,598\,960} \approx 1.35 \cdot 10^{-5}$$

What is the probability to get four cards of the same rank?

$$N_{four} = 13 \times 48 = 624 \Rightarrow p_{four} = \frac{624}{2\,598\,960} \approx 2.4 \cdot 10^{-4}$$

where we had to take into account the number of options (48) for the fifth card !

Classical definition

The classical definition of the probability works fine in many simple problems, the gambling games in particular.

However, the sampling space and elementary events have to be uniquely defined!

This is usually not a problem for experimental results given by discrete numbers (eg. roll of a die, card games, etc.).

But problems arise when this approach is to be applied to experiments with continuous spectra of possible measurement results...

This is well illustrated by the "Bertrand paradox"

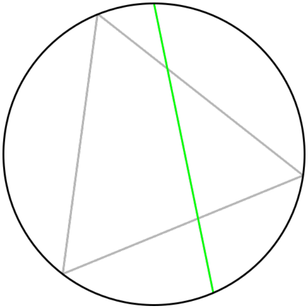
Bertrand paradox

description adapted from R. J. Barlow
pictures from Wikipedia

Definition of the problem:

In a circle of radius R an equilateral triangle is drawn.

What is the probability that the length of a random chord is greater than the triangle side?



Bertrand paradox

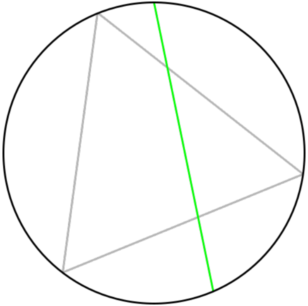
description adapted from R. J. Barlow
pictures from Wikipedia

Definition of the problem:

In a circle of radius R an equilateral triangle is drawn.

What is the probability that the length of a random chord is greater than the triangle side?

The problem seems to be simple, but we realize that the main question is: how the random chord should be defined?



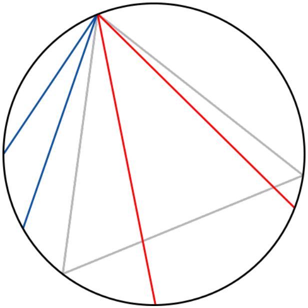
Bertrand paradox

description adapted from R. J. Barlow
pictures from Wikipedia

Definition of the problem:
In a circle of radius R an equilateral triangle is drawn.

What is the probability that the length of a random chord is greater than the triangle side?

Solution 1:
A random chord can be defined as a line connecting two random points on the circle.



Without loss of generality, we can move one of its ends to the vertex of the triangle.

The chord will be longer than the side of the triangle, if its other end is between the two other vertices $\Rightarrow$ probability is $1/3$

Bertrand paradox

description adapted from R. J. Barlow
pictures from Wikipedia

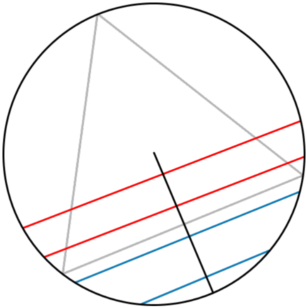
Definition of the problem:

In a circle of radius R an equilateral triangle is drawn.

What is the probability that the length of a random chord is greater than the triangle side?

Solution 2:

Without loss of generality, we can rotate a random chord in such a way that its centre is on the indicated radius of the circle.



The chord will be longer than the side of the triangle, if its centre is inside the triangle.

The side of the triangle cuts the radius in the middle

$\Rightarrow$ probability is $1/2$

Bertrand paradox

description adapted from R. J. Barlow
pictures from Wikipedia

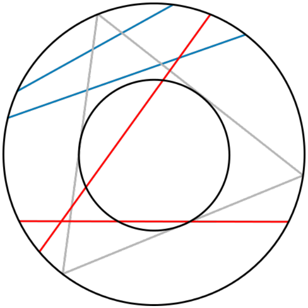
Definition of the problem:

In a circle of radius R an equilateral triangle is drawn.

What is the probability that the length of a random chord is greater than the triangle side?

Solution 3:

We can draw a chord by choosing a random point in the circle as its centre (unique definition of the chord).



The chord will be longer than the side of the triangle, if its centre is inside the circle of radius $R/2$.

The surface of the circle is proportional to radius squared

$\Rightarrow$ probability is $1/4$.

Bertrand paradox

Base on the classical probability definition, we can get three contradictory answers! Why?

Bertrand paradox

Base on the classical probability definition, we can get three contradictory answers! Why?

In the considered continuous sample space

"equally probable" elementary events are not uniquely defined!

We need to define how the actual experiment is performed...

Bertrand paradox

Base on the classical probability definition, we can get three contradictory answers! Why?

In the considered continuous sample space
"equally probable" elementary events are not uniquely defined!
We need to define how the actual experiment is performed...

Frequentist definition of probability

When repeating the same experiment a large number of times, $N \gg 1$,
the probability of A

$$P(A) \quad = \quad \lim_{N \to \infty} \frac{N(A)}{N}$$

where $N(A)$ is the number of occurrences of the event A
While this is not visible in the formula, the probability still depends on the considered sample space Ω, which reflects the way the experiment is done.

Definition of Probability

The frequentist (sometimes also called "classical") definition of probability gives direct recipe for the analysis of experimental data...

However, it is not always possible to perform the experiment a very large (infinite) number of times... We need some additional guidance to know how to define the probability.

Definition of Probability

The frequentist (sometimes also called "classical") definition of probability gives direct recipe for the analysis of experimental data...

However, it is not always possible to perform the experiment a very large (infinite) number of times... We need some additional guidance to know how to define the probability.

Kolmogorov Axioms

Kolmogorov (1933) formulated the three conditions
which have to be fulfilled by probability $P(A)$ of an event $A \subset \Omega$:

1. probability is a non-negative number: $P(A) \geq 0$
2. probability of all possible outcomes (sample space): $P(\Omega) = 1$
3. if A and B are mutually exclusive events: $P(A \cup B) = P(A) + P(B)$

We can derive all properties of the probability from these three axioms...

Concept of probability

Fundamental properties

Following properties can be derived from the Kolmogorov axioms:

- probability of the empty event is zero:

$$P(\emptyset) = 0$$

$\emptyset$ and Ω are mutually exclusive and $\Omega = \Omega \cup \emptyset$

$\Rightarrow P(\Omega) = P(\Omega) + P(\emptyset) \quad \Rightarrow P(\emptyset) = 0$

Fundamental properties

Following properties can be derived from the Kolmogorov axioms:

- probability of the empty event is zero:

$$P(\emptyset) = 0$$

- probability of complementary event:

$$P(\bar{A}) = 1 - P(A)$$

A and $\bar{A}$ are mutually exclusive and by definition $A \cup \bar{A} = \Omega$

$\Rightarrow P(A) + P(\bar{A}) = P(\Omega) = 1 \quad \Rightarrow P(\bar{A}) = 1 - P(A)$

Fundamental properties

Following properties can be derived from the Kolmogorov axioms:

- probability of the empty event is zero:

$$P(\emptyset) = 0$$

- probability of complementary event:

$$P(\bar{A}) = 1 - P(A)$$

- probability of the union of two events:

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

Union of A and B can be decomposed as:

$$P(A \cup B) = P(A \cap \bar{B}) + P(B \cap \bar{A}) + P(A \cap B)$$

Fundamental properties

Following properties can be derived from the Kolmogorov axioms:

- probability of the empty event is zero:

$$P(\emptyset) = 0$$

- probability of complementary event:

$$P(\bar{A}) = 1 - P(A)$$

- probability of the union of two events:

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

and then one can note that

$$P(A \cap \bar{B}) = P(A) - P(A \cap B) \quad \text{and} \quad P(B \cap \bar{A}) = P(B) - P(B \cap A)$$

Statistical Independence

Two events A and B are said to be statistically independent if and only if

$$P(A \cap B) = P(A) \cdot P(B)$$

Two important properties follow:

- mutually exclusive (nonempty) events cannot be independent
- if A is subset of B, $A \subset B$, they cannot be independent, unless $B = \Omega$

Statistical Independence

Two events A and B are said to be statistically independent if and only if

$$P(A \cap B) = P(A) \cdot P(B)$$

Two important properties follow:

- mutually exclusive (nonempty) events cannot be independent
- if A is subset of B, $A \subset B$, they cannot be independent, unless $B = \Omega$

Conditional Probability

When two events are not independent, we can consider probability of event A given that another event B is observed:

$$P(A|B) \quad = \quad \frac{P(A \cap B)}{P(B)} \qquad \text{or 0 if } P(B) = 0$$

Statistical Independence

Two events A and B are said to be statistically independent if and only if

$$P(A \cap B) = P(A) \cdot P(B)$$

Two important properties follow:

- mutually exclusive (nonempty) events cannot be independent
- if A is subset of B, $A \subset B$, they cannot be independent, unless $B = \Omega$

Conditional Probability

When two events are not independent, we can consider probability of event A given that another event B is observed:

$$P(A|B) \quad = \quad \frac{P(A \cap B)}{P(B)} \qquad \text{or 0 if } P(B) = 0$$

For independent events: $P(A|B) = P(A)$

Example (1)

Rolling a single die: $N = 6$ possible outcomes (elementary events).
We assume each outcome has the same probability $\Rightarrow p_e = P(\Omega)/N = 1/6$

Consider following events:

- odd number: $A_1 = \{1,3,5\}$ $\Rightarrow$ $p_1 = 1/2$
- even number: $A_2 = \{2,4,6\}$ $\Rightarrow$ $p_2 = 1/2$
- 1 or 6: $A_3 = \{1,6\}$ $\Rightarrow$ $p_3 = 1/3$
- at least 4: $A_4 = \{4,5,6\}$ $\Rightarrow$ $p_4 = 1/2$

Example (1)

Rolling a single die: $N = 6$ possible outcomes (elementary events).
We assume each outcome has the same probability $\Rightarrow p_e = P(\Omega)/N = 1/6$

Consider following events:

- odd number: $A_1 = \{1, 3, 5\}$ $\Rightarrow$ $p_1 = 1/2$
- even number: $A_2 = \{2, 4, 6\}$ $\Rightarrow$ $p_2 = 1/2$
- 1 or 6: $A_3 = \{1, 6\}$ $\Rightarrow$ $p_3 = 1/3$
- at least 4: $A_4 = \{4, 5, 6\}$ $\Rightarrow$ $p_4 = 1/2$

A_1 and A_2 are not independent, they are mutually exclusive:

$$P(A_1 \cap A_2) = P(\emptyset) = 0$$

Example (1)

Rolling a single die: $N = 6$ possible outcomes (elementary events).
We assume each outcome has the same probability $\Rightarrow p_e = P(\Omega)/N = 1/6$

Consider following events:

- odd number: $A_1 = \{1, 3, 5\}$ $\Rightarrow$ $p_1 = 1/2$
- even number: $A_2 = \{2, 4, 6\}$ $\Rightarrow$ $p_2 = 1/2$
- 1 or 6: $A_3 = \{1, 6\}$ $\Rightarrow$ $p_3 = 1/3$
- at least 4: $A_4 = \{4, 5, 6\}$ $\Rightarrow$ $p_4 = 1/2$

A_1 and A_3 are independent:

$$A_1 \cap A_3 = \{1\}$$

$$P(A_1 \cap A_3) = p_e = 1/6 = P(A_1) \cdot P(A_3)$$

Example (1)

Rolling a single die: $N = 6$ possible outcomes (elementary events).
We assume each outcome has the same probability $\Rightarrow p_e = P(\Omega)/N = 1/6$

Consider following events:

- odd number: $A_1 = \{1, 3, 5\}$ $\Rightarrow$ $p_1 = 1/2$
- even number: $A_2 = \{2, 4, 6\}$ $\Rightarrow$ $p_2 = 1/2$
- 1 or 6: $A_3 = \{1, 6\}$ $\Rightarrow$ $p_3 = 1/3$
- at least 4: $A_4 = \{4, 5, 6\}$ $\Rightarrow$ $p_4 = 1/2$

A_1 and A_4 are NOT independent:

$$A_1 \cap A_4 = \{5\}$$

$$P(A_1 \cap A_4) = p_e = 1/6 \quad \neq \quad P(A_1) \cdot P(A_4) = 1/4$$

Example (1)

Rolling a single die: $N = 6$ possible outcomes (elementary events).
We assume each outcome has the same probability $\Rightarrow p_e = P(\Omega)/N = 1/6$

Consider following events:

- odd number: $A_1 = \{1,3,5\}$ $\Rightarrow$ $p_1 = 1/2$
- even number: $A_2 = \{2,4,6\}$ $\Rightarrow$ $p_2 = 1/2$
- 1 or 6: $A_3 = \{1,6\}$ $\Rightarrow$ $p_3 = 1/3$
- at least 4: $A_4 = \{4,5,6\}$ $\Rightarrow$ $p_4 = 1/2$

Probability of A_4 when A_1 is observed:

$$P(A_1 \cap A_4) = 1/6$$

$$P(A_4|A_1) = \frac{P(A_1 \cap A_4)}{P(A_1)} = \frac{1/6}{1/2} = 1/3 \neq P(A_4)$$

Example (2)

Rolling two dice (eg. red and blue): $N = 6 \cdot 6 = 36$ possible outcomes.
Sample space can be best presented as a table:

	1	2	3	4	5	6
1						
2						
3						
4						
5						
6						

We assume each outcome has the same probability $\Rightarrow p_e = P(\Omega)/N = \frac{1}{36}$

Example (2)

Rolling two dice (eg. red and blue): $N = 6 \cdot 6 = 36$ possible outcomes.
Sample space can be best presented as a table:

	1	2	3	4	5	6
1			A			B
2			A		B	
3			A	B		
4			$A \cap B$			
5		B	A			
6	B		A			

We assume each outcome has the same probability $\Rightarrow p_e = P(\Omega)/N = \frac{1}{36}$

Events: A - red die shows 3 and B - sum of two dice is 7

$$P(A \cap B) = 1/36 = P(A) \cdot P(B) \quad \Rightarrow \text{are independent}$$

Example (2)

Rolling two dice (eg. red and blue): $N = 6 \cdot 6 = 36$ possible outcomes.
Sample space can be best presented as a table:

	1	2	3	4	5	6
1			A	C		
2			$A \cap C$			
3		C	A			
4	C		A			
5			A			
6			A			

We assume each outcome has the same probability $\Rightarrow p_e = P(\Omega)/N = \frac{1}{36}$

Events: A - red die shows 3 and C - sum of two dice is 5

$$P(A \cap C) = 1/36 \neq P(A) \cdot P(C) = 1/6 \cdot 1/9 \quad \Rightarrow \text{NOT independent}$$

Example (2)

Rolling two dice (eg. red and blue): $N = 6 \cdot 6 = 36$ possible outcomes.
Sample space can be best presented as a table:

	1	2	3	4	5	6
1			A	C		
2			$A \cap C$			
3		C	A			
4	C		A			
5			A			
6			A			

We assume each outcome has the same probability $\Rightarrow p_e = P(\Omega)/N = \frac{1}{36}$

Events: A - red die shows 3 and C - sum of two dice is 5

$$P(A|C) = P(A \cap C)/P(C) = \frac{1}{36} \ / \ \frac{1}{9} = 1/4$$

Example (2)

Rolling two dice (eg. red and blue): $N = 6 \cdot 6 = 36$ possible outcomes.
Sample space can be best presented as a table:

	1	2	3	4	5	6
1			A	C		
2			$A \cap C$			
3		C	A			
4	C		A			
5			A			
6			A			

We assume each outcome has the same probability $\Rightarrow p_e = P(\Omega)/N = \frac{1}{36}$

Events: A - red die shows 3 and C - sum of two dice is 5

$$P(C|A) = P(A \cap C)/P(A) = \frac{1}{36} \; / \; \frac{1}{6} = 1/6$$

Partition of the sample space

It is a set of events A_i $(i = 1 \ldots n)$ with the following properties:

- all A_i are mutually exclusive

$$A_i \cap A_j = \emptyset, \qquad \forall i \neq j$$

- they cover the whole sampling space

$$\bigcup_{i=1}^{n} A_i \;=\; \Omega$$

Partition of the sample space

It is a set of events A_i $(i = 1 \ldots n)$ with the following properties:

- all A_i are mutually exclusive

$$A_i \cap A_j = \emptyset, \qquad \forall i \neq j$$

- they cover the whole sampling space

$$\bigcup_{i=1}^{n} A_i \; = \; \Omega$$

From the two conditions we realize that:

$$\sum_{i=1}^{n} P(A_i) \; = \; 1$$

We can be sure that one (and only one) A_i will always take place

Concept of probability

Total Probability Theorem

Given partition A_i of the sampling space, for any event B we can write

$$P(B) = \sum_{i=1}^{n} P(B \cap A_i) = \sum_{i=1}^{n} P(B|A_i) \cdot P(A_i)$$

Total probability of B can be calculated as a sum over probabilities calculated in separate sub-spaces. Very useful in many cases...

Total Probability Theorem

Given partition A_i of the sampling space, for any event B we can write

$$P(B) = \sum_{i=1}^{n} P(B \cap A_i) = \sum_{i=1}^{n} P(B|A_i) \cdot P(A_i)$$

Total probability of B can be calculated as a sum over probabilities calculated in separate sub-spaces. Very useful in many cases...

Example (1)

What is the probability of giving birth to twins in Europe?

We can addressing this problem by combining probabilities for different countries in Europe:

$$P(\text{Twins}) = \sum_{i=1}^{n} P(\text{Twins}|\text{Country}_i) \cdot \frac{N_i}{\sum_{i=1}^{n} N_i}$$

where N_i is the number off all births in country i

Total Probability Theorem

Given partition A_i of the sampling space, for any event B we can write

$$P(B) = \sum_{i=1}^{n} P(B \cap A_i) = \sum_{i=1}^{n} P(B|A_i) \cdot P(A_i)$$

Total probability of B can be calculated as a sum over probabilities calculated in separate sub-spaces. Very useful in many cases...

Example (2)

What is the probability of producing π^+ in e^+e^- annihilation into Z^0:

$$e^+e^- \rightarrow Z^0 \rightarrow \pi^+ + \ldots$$

We can divide the sampling space, all Z^0 decays, into separate decay channels, $Z^0 \rightarrow f\bar{f}$, where $f = e, \mu, \tau, \nu_e, \nu_\mu, \nu_\tau, u, d, s, c, b$. For some of these channels the answer is known.

eg. $P(\pi^+|Z \rightarrow \nu\bar{\nu}) = 0$

Bayes' Theorem

For events A and B the two conditional probabilities are related:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad \text{and} \quad P(B|A) = \frac{P(B \cap A)}{P(A)}$$

as $B \cap A = A \cap B$ we obtain: Bayes' Theorem

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

Bayes' Theorem

For events A and B the two conditional probabilities are related:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad \text{and} \quad P(B|A) = \frac{P(B \cap A)}{P(A)}$$

as $B \cap A = A \cap B$ we obtain: Bayes' Theorem

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

This can be also written in a more general form:

$$P(A_i|B) \quad = \quad \frac{P(B|A_i)\ P(A_i)}{\sum_{j=1}^{n} P(B|A_j)\ P(A_j)}$$

where A_i is the partition of the sampling space.

Bayes' Theorem

For events A and B the two conditional probabilities are related:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad \text{and} \quad P(B|A) = \frac{P(B \cap A)}{P(A)}$$

as $B \cap A = A \cap B$ we obtain: Bayes' Theorem

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

This can be also written in a more general form:

$$P(A_i|B) = \frac{P(B|A_i)\, P(A_i)}{\sum_{j=1}^{n} P(B|A_j)\, P(A_j)}$$

where A_i is the partition of the sampling space.

There is nothing new in this as long as A_i and B belong to the same sampling space...

Bayesian approach

Bayes theorem can be used to generalize the concept of probability.
In particular, one can consider "probability" of given hypothesis H
(theoretical model or model parameter, eg. Hubble constant)
when taking into known outcome D (data) of the experiment:

$$P(H|D) = \frac{P(D|H)}{P(D)} \cdot P(H)$$

Bayesian approach

Bayes theorem can be used to generalize the concept of probability.

In particular, one can consider "probability" of given hypothesis H

(theoretical model or model parameter, eg. Hubble constant)

when taking into known outcome D (data) of the experiment:

$$P(H|D) = \frac{P(D|H)}{P(D)} \cdot P(H)$$

There are two problems with this approach:

- H can not be considered an event, sampling space can not be defined (no experiment to repeat)
- we need to make a subjective assumption about the "prior" $P(H)$ describing our initial belief in hypothesis H

Bayesian approach

Bayes theorem can be used to generalize the concept of probability.
In particular, one can consider "probability" of given hypothesis H (theoretical model or model parameter, eg. Hubble constant) when taking into known outcome D (data) of the experiment:

$$P(H|D) = \frac{P(D|H)}{P(D)} \cdot P(H)$$

There are two problems with this approach:

- H can not be considered an event, sampling space can not be defined (no experiment to repeat)
- we need to make a subjective assumption about the "prior" $P(H)$ describing our initial belief in hypothesis H

For these reasons I rather use term "degree of belief" for the result of the Bayesian procedure applied "outside" the sampling space (not to random events)

Example of Bayesian approach

adapted from G. Cowan, 2011 CERN Summer Student Lectures on Statistics

How much should I worry, if I get a positive test result for virus X?

- Hypothesis: H - I am infected (I am a carrier of X)
- Data: D - test result is positive

How likely is it that I am a carrier of X, what is $P(H|D)$?

Example of Bayesian approach

adapted from G. Cowan, 2011 CERN Summer Student Lectures on Statistics

How much should I worry, if I get a positive test result for virus X?

- Hypothesis: H - I am infected (I am a carrier of X)
- Data: D - test result is positive

How likely is it that I am a carrier of X, what is $P(H|D)$?

We need to know efficiency and false rate for the test. Let as assume

$$\begin{aligned} P(D|H) &= 0.99 && \text{test efficiency} \\ P(D|\bar{H}) &= 0.01 && \text{false positive rate} \end{aligned}$$

We can also assume that 1 person per 1000 is an X carrier in our country

$$P(H) = 0.001$$

Example of Bayesian approach

We first use the total probability theorem to calculate the probability of the positive result:

$$\begin{aligned} P(D) &= P(D|H) \cdot P(H) + P(D|\bar{H}) \cdot P(\bar{H}) \\ &= 0.99 \cdot 0.001 + 0.01 \cdot 0.999 = 0.01098 \approx 0.011 \end{aligned}$$

Example of Bayesian approach

We first use the total probability theorem to calculate the probability of the positive result:

$$
\begin{aligned}
P(D) &= P(D|H) \cdot P(H) + P(D|\bar{H}) \cdot P(\bar{H}) \\
&= 0.99 \cdot 0.001 + 0.01 \cdot 0.999 = 0.01098 \approx 0.011
\end{aligned}
$$

and then apply Bayes' theorem:

$$
\begin{aligned}
P(H|D) &= \frac{P(D|H)}{P(D)} \cdot P(H) \\
&= \frac{0.99}{0.011} \cdot 0.001 \approx 0.09
\end{aligned}
$$

Example of Bayesian approach

We first use the total probability theorem to calculate the probability of the positive result:

$$P(D) = P(D|H) \cdot P(H) + P(D|\bar{H}) \cdot P(\bar{H})$$
$$= 0.99 \cdot 0.001 + 0.01 \cdot 0.999 = 0.01098 \approx 0.011$$

and then apply Bayes' theorem:

$$P(H|D) = \frac{P(D|H)}{P(D)} \cdot P(H)$$
$$= \frac{0.99}{0.011} \cdot 0.001 \approx 0.09$$

You can believe that your chances of being infected with X are 9%...

How useful is this information in your opinion?

Example of Bayesian approach

We first use the total probability theorem to calculate the probability of the positive result:

$$\begin{aligned} P(D) &= P(D|H) \cdot P(H) + P(D|\bar{H}) \cdot P(\bar{H}) \\ &= 0.99 \cdot 0.001 + 0.01 \cdot 0.999 = 0.01098 \approx 0.011 \end{aligned}$$

and then apply Bayes' theorem:

$$\begin{aligned} P(H|D) &= \frac{P(D|H)}{P(D)} \cdot P(H) \\ &= \frac{0.99}{0.011} \cdot 0.001 \approx 0.09 \end{aligned}$$

You can believe that your chances of being infected with X are 9%...

This number represents the probability of person with positive test result being infected.

Example of Bayesian approach

We first use the total probability theorem to calculate the probability of the positive result:

$$\begin{aligned} P(D) &= P(D|H) \cdot P(H) + P(D|\bar{H}) \cdot P(\bar{H}) \\ &= 0.99 \cdot 0.001 + 0.01 \cdot 0.999 = 0.01098 \approx 0.011 \end{aligned}$$

and then apply Bayes' theorem:

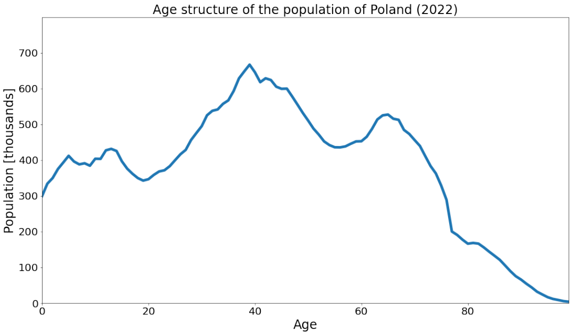
$$\begin{aligned} P(H|D) &= \frac{P(D|H)}{P(D)} \cdot P(H) \\ &= \frac{0.99}{0.011} \cdot 0.001 \approx 0.09 \end{aligned}$$

You can believe that your chances of being infected with X are 9%...

But this can hardly be interpreted as the probability or your infection (not a random variable)...

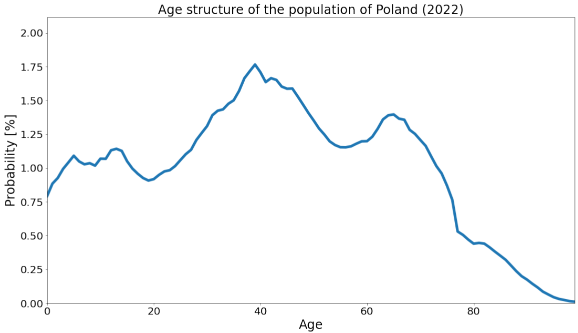
Prior problem

Let us look again at the statistical data on the population of Poland:



Prior problem

Let us look again at the statistical data on the population of Poland:



We can present it as the age probability (for randomly chosen person).

Can we draw any conclusions concerning the life expectation?

Prior problem

To simplify the problem, let us assume that the "survival probability" does not change in time:

$$p(y|b) \quad = \quad s(a)$$

where: b is the year of birth, y is the year of the population census, $a = y - b$ is age.

Prior problem

To simplify the problem, let us assume that the "survival probability" does not change in time:

$$p(y|b) \quad = \quad s(a)$$

where: b is the year of birth, y is the year of the population census, $a = y - b$ is age.

Unfortunately, what we measure is the age distribution:

$$p(a) \quad = \quad p(b|y) \qquad \text{for } b = y - a$$

Prior problem

To simplify the problem, let us assume that the "survival probability" does not change in time:

$$p(y|b) \quad = \quad s(a)$$

where: b is the year of birth, y is the year of the population census, $a = y - b$ is age.

Unfortunately, what we measure is the age distribution:

$$p(a) \quad = \quad p(b|y) \qquad \text{for } b = y - a$$

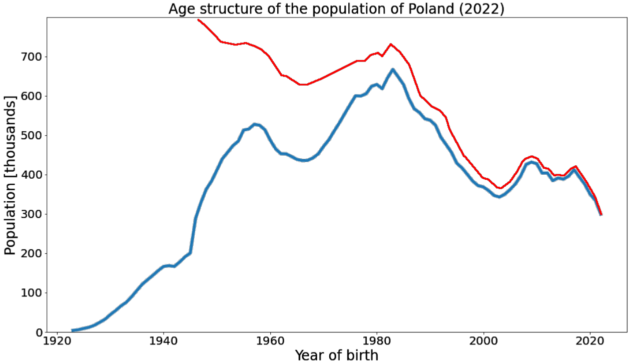
We can apply Bayes' Theorem to relate it to the "survival probability"

$$p(a) \quad = \quad \frac{s(a) \cdot u(y-a)}{\sum_b \, s(y-b) \cdot u(b)}$$

where $u(b)$ is the birth number distribution. **We need to know it!!!**

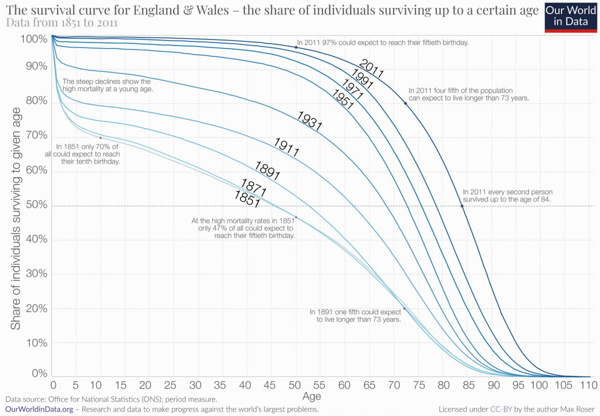
Prior problem

What is needed is the number of births in each year: shown is my guess



Without exact knowledge of this "prior" we can not draw any conclusions...

There are also other reasons, why extraction of survival curve from single population census is not possible. In particular, it depends significantly on the year of birth...

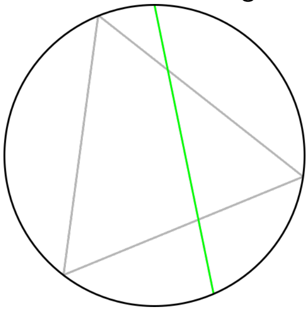


Source: ourworldindata.org

Concept of probability

Bertrand paradox

In a circle of radius R an equilateral triangle is drawn. What is the probability that the length of a random chord is greater than the side of the triangle?



What is, in your opinion, the correct answer to the problem?

Give arguments for the proper construction of random chord.

You can also propose your own definition/construction!

Solutions should be uploaded until October 16 (Wednesday)