Introduction to probability

2. Axiomatic definition of probability

Wojciech Kotłowski

Institute of Computing Science, PUT
http://www.cs.put.poznan.pl/wkotlowski/

09.03.2021

Classical/geometric definition of probability

- Sample space Ω
- Event $A \subseteq \Omega$ is a subset of the sample space
- Probability of an event A: $P(A) = \frac{|A|}{|\Omega|}$
  - Classical definition: $|A|$ is the cardinality of A
  - Geometric definition: $|A|$ is the measure of A

Classical/geometric definition of probability

- Sample space Ω
- Event $A \subseteq \Omega$ is a subset of the sample space
- Probability of an event A: $P(A) = \frac{|A|}{|\Omega|}$
  - Classical definition: $|A|$ is the cardinality of A
  - Geometric definition: $|A|$ is the measure of A

Limits of the classical definition:

- The principle of indifference is ambiguous (examples in a moment)

Classical/geometric definition of probability

- Sample space Ω
- Event $A \subseteq \Omega$ is a subset of the sample space
- Probability of an event A: $P(A) = \frac{|A|}{|\Omega|}$
  - Classical definition: $|A|$ is the cardinality of A
  - Geometric definition: $|A|$ is the measure of A

Limits of the classical definition:

- The principle of indifference is ambiguous (examples in a moment)
- Each outcomes assigned the same probability
  - e.g., biased coin

Classical/geometric definition of probability

- Sample space Ω
- Event $A \subseteq \Omega$ is a subset of the sample space
- Probability of an event A: $P(A) = \frac{|A|}{|\Omega|}$
  - Classical definition: $|A|$ is the cardinality of A
  - Geometric definition: $|A|$ is the measure of A

Limits of the classical definition:

- The principle of indifference is ambiguous (examples in a moment)
- Each outcomes assigned the same probability
  - e.g., biased coin
- Limited to two specific spaces (finite sets or subsets of $\mathbb{R}^n$)
  - e.g., tossing a coin until first head is observed

Classical/geometric definition of probability

- Sample space Ω
- Event $A \subseteq \Omega$ is a subset of the sample space
- Probability of an event A: $P(A) = \frac{|A|}{|\Omega|}$
  - Classical definition: $|A|$ is the cardinality of A
  - Geometric definition: $|A|$ is the measure of A

Limits of the classical definition:

- The principle of indifference is ambiguous (examples in a moment)
- Each outcomes assigned the same probability
  - e.g., biased coin
- Limited to two specific spaces (finite sets or subsets of $\mathbb{R}^n$)
  - e.g., tossing a coin until first head is observed
- Mathematical inconsistency (example in a moment)

The problem of d'Alembert

Consider the following game: we toss two coins and win if at least one head is observed, otherwise we lose. What is the probability of winning?

The problem of d'Alembert

Consider the following game: we toss two coins and win if at least one head is observed, otherwise we lose. What is the probability of winning?

Answer: Let A be the event that the game is won

Since $\Omega = \{HH, HT, TH, TT\}$, $A = \{HH, HT, TH\}$, and of course $P(A) = \frac{3}{4}$

The problem of d'Alembert

Consider the following game: we toss two coins and win if at least one head is observed, otherwise we lose. What is the probability of winning?

Answer: Let A be the event that the game is won

Since $\Omega = \{HH, HT, TH, TT\}$, $A = \{HH, HT, TH\}$, and of course $P(A) = \frac{3}{4}$

D'Alembert's answer: If the first coin comes up head, the second toss will not happen as the game is already settled.

Therefore, $\Omega = \{H, TH, TT\}$, $A = \{H, TH\}$, so $P(A) = \frac{2}{3}$

Ambiguity in assigning equal probabilities to all outcomes!

Geometric probability: Bertrand's paradox

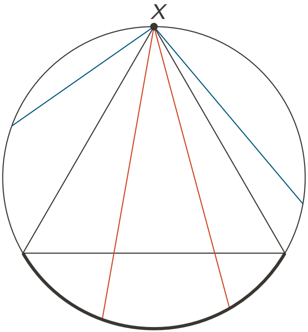
Consider an equilateral triangle inscribed in a unit circle. Suppose a chord of the circle is chosen at random. What is the probability that the chord is longer than a side of the triangle?

Geometric probability: Bertrand's paradox

Consider an equilateral triangle inscribed in a unit circle. Suppose a chord of the circle is chosen at random. What is the probability that the chord is longer than a side of the triangle?

A – event „the chord longer than a side of the triangle"

The chords from A marked in red, outside A in blue



Consider chords starting at X.

Outcomes: the other endpoint of a chord described by an angle in $[0, 2\pi]$.

$\Omega = [0, 2\pi)$

The chords from A correspond to an arc marked in bold: $A = (\frac{2}{3}\pi, \frac{4}{3}\pi)$

$$P(A) = \frac{\frac{2}{3}\pi}{2\pi} = \frac{1}{3}$$

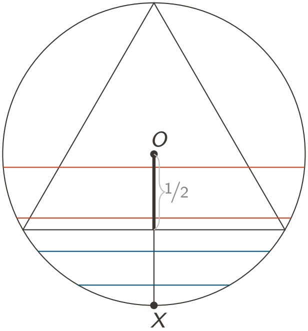
This probability does not depend on the initial point X.

Geometric probability: Bertrand's paradox

Consider an equilateral triangle inscribed in a unit circle. Suppose a chord of the circle is chosen at random. What is the probability that the chord is longer than a side of the triangle?

A – event „the chord longer than a side of the triangle"

The chords from A marked in red, outside A in blue



Consider the chords perpendicular to the radius OX.

Outcomes: chord position determined by the distance from the center of the circle in $[0, 1]$.

$\Omega = [0, 1]$

Chords from A correspond to the interval marked in bold: $A = [0, \frac{1}{2}]$

$$P(A) = \frac{\frac{1}{2}}{1} = \frac{1}{2}$$

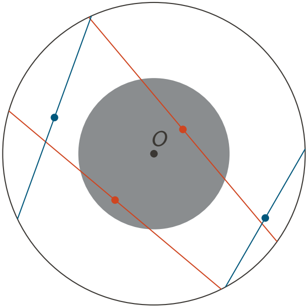
This probability does not depend on the radius OX.

Geometric probability: Bertrand's paradox

Consider an equilateral triangle inscribed in a unit circle. Suppose a chord of the circle is chosen at random. What is the probability that the chord is longer than a side of the triangle?

A – event „the chord longer than a side of the triangle"

The chords from A marked in red, outside A in blue



Every chord uniquely determined by the position of its center.

Samples: Points within the circle

$\Omega = K(O, 1)$ (ball of radius 1 and center O)

Chords from A correspond to points in the gray disk: $A = K(O, \frac{1}{2})$

$$P(A) = \frac{\frac{1}{4}\pi}{\pi} = \frac{1}{4}$$

Geometric probability: Bertrand's paradox

Consider an equilateral triangle inscribed in a unit circle. Suppose a chord of the circle is chosen at random. What is the probability that the chord is longer than a side of the triangle?

What is the source of this paradox?

In each case we used a different sample space

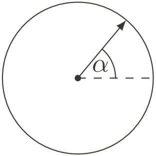
1. $\Omega = [0, 2\pi)$ (angle)
2. $\Omega = [0, 1]$ (distance)
3. $\Omega = K(O, 1)$ (point)

The fact that all outcomes are equally likely in one space does not necessarily mean that they are equally likely in another space!

Thus, all these cases concern different random experiments!

Non-measurable sets*

[Jakubowski, Sztencel: Rachunek prawdopodobieństwa dla prawie każdego, dodatek A.3]



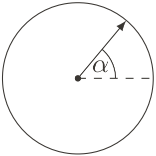
We draw an angle from $[0, 2\pi)$

One can show that one can divide $\Omega = [0, 2\pi)$ into countably infinite sets $A_1, A_2, A_3, \ldots$, such that:

1. All sets are disjoint: $A_i \cap A_j = \emptyset$ dla $i \neq j$
2. $A_1 \cup A_2 \cup \ldots = [0, 2\pi)$ (they cover the space Ω)
3. They are congruent, i.e., any A_i can be obtained from A_1 by a rotation by some angle

Non-measurable sets*

[Jakubowski, Sztencel: Rachunek prawdopodobieństwa dla prawie każdego, dodatek A.3]



We draw an angle from $[0, 2\pi)$

One can show that one can divide $\Omega = [0, 2\pi)$ into countably infinite sets $A_1, A_2, A_3, \ldots,$ such that:

1. All sets are disjoint: $A_i \cap A_j = \emptyset$ dla $i \neq j$
2. $A_1 \cup A_2 \cup \ldots = [0, 2\pi)$ (they cover the space Ω)
3. They are congruent, i.e., any A_i can be obtained from A_1 by a rotation by some angle

$$|\Omega| \;=\; |A_1 \cup A_2 \cup \ldots| \;=\; \sum_{i=1}^{\infty} |A_i|$$

$|\Omega| = 2\pi$

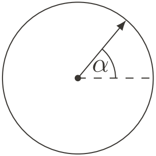
follows from 2.

follows from 1.

$|A_i| = |A_1|$ from 3.

Non-measurable sets*

[Jakubowski, Sztencel: Rachunek prawdopodobieństwa dla prawie każdego, dodatek A.3]



We draw an angle from $[0, 2\pi)$

One can show that one can divide $\Omega = [0, 2\pi)$ into countably infinite sets $A_1, A_2, A_3, \ldots,$ such that:

1. All sets are disjoint: $A_i \cap A_j = \emptyset$ dla $i \neq j$
2. $A_1 \cup A_2 \cup \ldots = [0, 2\pi)$ (they cover the space Ω)
3. They are congruent, i.e., any A_i can be obtained from A_1 by a rotation by some angle

$|\Omega| = 2\pi$

$$|\Omega| = |A_1 \cup A_2 \cup \ldots| = \sum_{i=1}^{\infty} |A_i|$$

follows from 2.

follows from 1.

$|A_i| = |A_1|$ from 3.

We thus have $2\pi = \infty \cdot |A_1|$ which is a contradiction!

We thus cannot assign a measure to A_i, and so neither the probability!
non-measurable sets

Probabilistic space

$$(\Omega, \mathcal{F}, P)$$

Sample space

Collection of events (σ-algebra)

probability measure

Sample space

- Outcome ω: a single result of a random experiment
- Sample space Ω: set of all possible outcomes
- Ω can be infinite or even uncountable

Sample space

- Outcome ω: a single result of a random experiment
- Sample space Ω: set of all possible outcomes
- Ω can be infinite or even uncountable

Examples:

- Rolling a die:

Sample space

- Outcome ω: a single result of a random experiment
- Sample space Ω: set of all possible outcomes
- Ω can be infinite or even uncountable

Examples:

- Rolling a die: $\Omega = \{1, 2, 3, 4, 5, 6\}$

Sample space

- Outcome ω: a single result of a random experiment
- Sample space Ω: set of all possible outcomes
- Ω can be infinite or even uncountable

Examples:

- Rolling a die: $\Omega = \{1, 2, 3, 4, 5, 6\}$
- Number of coin tosses until the first head:

Sample space

- Outcome ω: a single result of a random experiment
- Sample space Ω: set of all possible outcomes
- Ω can be infinite or even uncountable

Examples:

- Rolling a die: $\Omega = \{1, 2, 3, 4, 5, 6\}$
- Number of coin tosses until the first head: $\Omega = \{1, 2, 3, \ldots\}$

Sample space

- Outcome ω: a single result of a random experiment
- Sample space Ω: set of all possible outcomes
- Ω can be infinite or even uncountable

Examples:

- Rolling a die: $\Omega = \{1, 2, 3, 4, 5, 6\}$
- Number of coin tosses until the first head: $\Omega = \{1, 2, 3, \ldots\}$
- Value of voltage in an electrical network:

Sample space

- Outcome ω: a single result of a random experiment
- Sample space Ω: set of all possible outcomes
- Ω can be infinite or even uncountable

Examples:

- Rolling a die: $\Omega = \{1, 2, 3, 4, 5, 6\}$
- Number of coin tosses until the first head: $\Omega = \{1, 2, 3, \ldots\}$
- Value of voltage in an electrical network: $\Omega = \mathbb{R}$

Sample space

- Outcome ω: a single result of a random experiment
- Sample space Ω: set of all possible outcomes
- Ω can be infinite or even uncountable

Examples:

- Rolling a die: $\Omega = \{1, 2, 3, 4, 5, 6\}$
- Number of coin tosses until the first head: $\Omega = \{1, 2, 3, \ldots\}$
- Value of voltage in an electrical network: $\Omega = \mathbb{R}$
- Hard drive lifespan:

Sample space

- Outcome ω: a single result of a random experiment
- Sample space Ω: set of all possible outcomes
- Ω can be infinite or even uncountable

Examples:

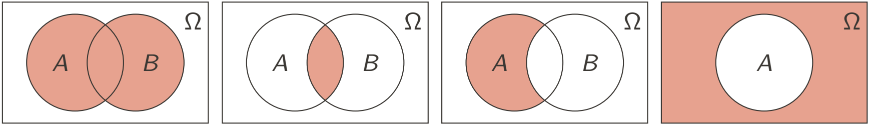
- Rolling a die: $\Omega = \{1, 2, 3, 4, 5, 6\}$
- Number of coin tosses until the first head: $\Omega = \{1, 2, 3, \ldots\}$
- Value of voltage in an electrical network: $\Omega = \mathbb{R}$
- Hard drive lifespan: $\Omega = [0, \infty)$

Events

Events are subsets of the sample space Ω

Events

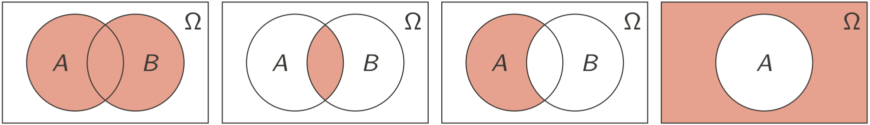
Events are subsets of the sample space Ω



sum $A \cup B$ intersection $A \cap B$ difference $A \setminus B$ complement A'

Events

Events are subsets of the sample space Ω

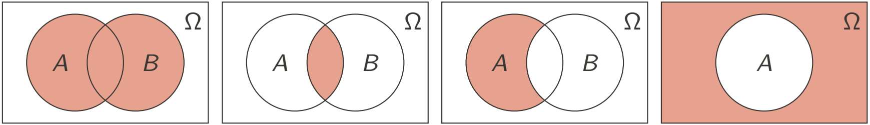


sum $A \cup B$ intersection $A \cap B$ difference $A \setminus B$ complement A'

- We say that an event A occurred, if the outcome $\omega \in A$

Events

Events are subsets of the sample space Ω

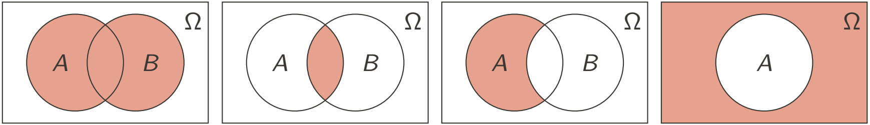


sum $A \cup B$ intersection $A \cap B$ difference $A \setminus B$ complement A'

- We say that an event A occurred, if the outcome $\omega \in A$
- The event $A = \Omega$ is called certain

Events

Events are subsets of the sample space Ω

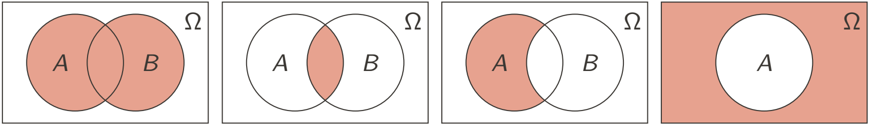


sum $A \cup B$ intersection $A \cap B$ difference $A \setminus B$ complement A'

- We say that an event A occurred, if the outcome $\omega \in A$
- The event $A = \Omega$ is called certain
- The event $A = \emptyset$ is called impossible

Events

Events are subsets of the sample space Ω

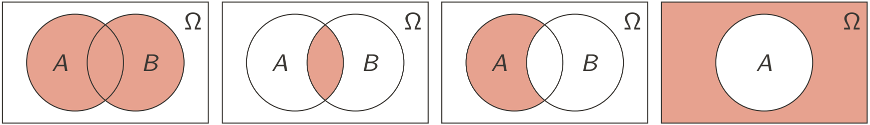


sum $A \cup B$ | intersection $A \cap B$ | difference $A \setminus B$ | complement A'

- We say that an event A occurred, if the outcome $\omega \in A$
- The event $A = \Omega$ is called certain
- The event $A = \emptyset$ is called impossible
- The event complementary to A is A'

Events

Events are subsets of the sample space Ω

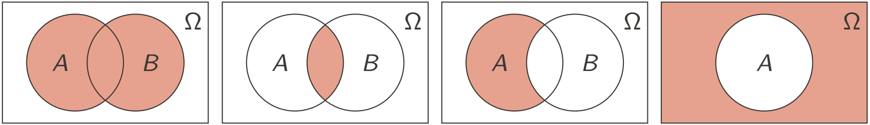


sum $A \cup B$ | intersection $A \cap B$ | difference $A \setminus B$ | complement A'

- We say that an event A occurred, if the outcome $\omega \in A$
- The event $A = \Omega$ is called certain
- The event $A = \emptyset$ is called impossible
- The event complementary to A is A'
- Events A and B are disjoint (mutually exclusive) if $A \cap B = \emptyset$

Events

Events are subsets of the sample space Ω



sum $A \cup B$ intersection $A \cap B$ difference $A \setminus B$ complement A'

- We say that an event A occurred, if the outcome $\omega \in A$
- The event $A = \Omega$ is called certain
- The event $A = \emptyset$ is called impossible
- The event complementary to A is A'
- Events A and B are disjoint (mutually exclusive) if $A \cap B = \emptyset$
- More generally: events $A_1, A_2, \ldots$ are disjoint if $A_i \cap A_j = \emptyset$ for each $i \neq j$

Collection of events

A collection of events $\mathcal{F}$ is a collection of subsets of Ω, which contains all possible events; so $\mathcal{F} \subseteq 2^{\Omega}$
(2^{Ω} is the power set, i.e., the set of all subsets of Ω)

Collection of events

A collection of events $\mathcal{F}$ is a collection of subsets of Ω, which contains all possible events; so $\mathcal{F} \subseteq 2^{\Omega}$
(2^{Ω} is the power set, i.e., the set of all subsets of Ω)

Why can't we simply assume that every subset of Ω is an event, i.e. to set $\mathcal{F} = 2^{\Omega}$???

Collection of events

A collection of events $\mathcal{F}$ is a collection of subsets of Ω, which contains all possible events; so $\mathcal{F} \subseteq 2^{\Omega}$
(2^{Ω} is the power set, i.e., the set of all subsets of Ω)

Why can't we simply assume that every subset of Ω is an event, i.e. to set $\mathcal{F} = 2^{\Omega}$???

- For countable Ω we indeed can and most often do.
- For uncountable Ω (e.g., $\Omega = \mathbb{R}$) we cannot, as there will be non-measurable sets in $\mathcal{F}$, to which no probability measure can be assigned!

Collection of events

A collection of events $\mathcal{F}$ is a collection of subsets of Ω, which contains all possible events; so $\mathcal{F} \subseteq 2^{\Omega}$
(2^{Ω} is the power set, i.e., the set of all subsets of Ω)

Why can't we simply assume that every subset of Ω is an event, i.e. to set $\mathcal{F} = 2^{\Omega}$???

- For countable Ω we indeed can and most often do.
- For uncountable Ω (e.g., $\Omega = \mathbb{R}$) we cannot, as there will be non-measurable sets in $\mathcal{F}$, to which no probability measure can be assigned!

No matter what exactly $\mathcal{F}$ is, we always want to be able to apply all set-theoretic operations to events, such as sum, intersection, complement, set difference, etc. In other words, we want the outcomes of these operations to be events as well (i.e. to belong to $\mathcal{F}$).

This is guaranteed if we assume that $\mathcal{F}$ is a σ-algebra.

σ-algebra

A collection of $\mathcal{F} \subseteq 2^{\Omega}$ is called a σ-algebra (σ-field), if:

1. $\Omega \in \mathcal{F}$
2. If $A \in \mathcal{F}$ then $A' \in \mathcal{F}$
3. If $A_1, A_2, \ldots \in \mathcal{F}$ then $A_1 \cup A_2 \cup \ldots \in \mathcal{F}$

Remark: property 3 for any countable sum of events.

Properties of σ-agebras

$\boxed{\text{Fac: The empty set belongs to } \mathcal{F}\text{: } \emptyset \in \mathcal{F}}$

Properties of σ-agebras

$\boxed{\text{Fac: The empty set belongs to } \mathcal{F}\text{: } \emptyset \in \mathcal{F}}$

Proof: Since $\Omega \in \mathcal{F}$, a $\Omega' = \emptyset$, then from property 2 it holds $\emptyset \in \mathcal{F}$.

Properties of σ-algebras

$\boxed{\text{Fact: If } A, B \in \mathcal{F} \text{ then also } A \cap B \in \mathcal{F}}$

Properties of σ-algebras

Fact: If $A, B \in \mathcal{F}$ then also $A \cap B \in \mathcal{F}$

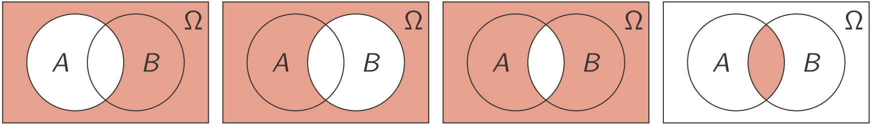
Proof:

(a) From property 2 we have: $A' \in \mathcal{F}$ and $B' \in \mathcal{F}$

(b) From property 3 we have: $C = A' \cup B' \in \mathcal{F}$

(c) From property 2 we have: $C' \in \mathcal{F}$

(d) But from the de Morgan's law* we have $C' = A \cap B$



A' $\quad$ B' $\quad$ $C = A' \cup B'$ $\quad$ $C' = A \cap B$

*De Morgan's law: $(E \cap F)' = E' \cup F'$

Properties of σ-algebras

During the exercise classes, we will prove that if $A, B \in \mathcal{F}$ then so does $A \setminus B \in \mathcal{F}$

Conclusion: A σ-algebra is closed under any set-theoretic operations such as sum, intersection, difference, complement, etc.

Example of a σ-algebra: the power set

If Ω is countable, we may simply take $\mathcal{F} = 2^{\Omega}$, i.e., all subsets of the sample space are events

Example of a σ-algebra: the Borel algebra

If $\Omega = \mathbb{R}$ (uncountable), we cannot take $\mathcal{F} = 2^{\Omega}$, as it is not possible to define a probability measure over such collection (non-measurable sets).

Example of a σ-algebra: the Borel algebra

If $\Omega = \mathbb{R}$ (uncountable), we cannot take $\mathcal{F} = 2^{\Omega}$, as it is not possible to define a probability measure over such collection (non-measurable sets).

Assume that $\mathcal{F}$ at least contains all events of the form: „the outcome less than a", „the outcome between a and b" ($a, b \in \mathbb{R}$), etc.

Then, $\mathcal{F}$ must contain all possible intervals, open or closed, finite or infinite, e.g., $[a, b)$, (a, b), $(-\infty, a]$, (b, ∞), etc.

Example of a σ-algebra: the Borel algebra

If $\Omega = \mathbb{R}$ (uncountable), we cannot take $\mathcal{F} = 2^\Omega$, as it is not possible to define a probability measure over such collection (non-measurable sets).

Assume that $\mathcal{F}$ at least contains all events of the form: „the outcome less than a", „the outcome between a and b" ($a, b \in \mathbb{R}$), etc.

Then, $\mathcal{F}$ must contain all possible intervals, open or closed, finite or infinite, e.g., $[a, b)$, (a, b), $(-\infty, a]$, (b, ∞), etc.

From the properties of σ-algebra, $\mathcal{F}$ also contains countable sums and intersections of intervals (including single points) .

Such a collection is called the Borel σ-algebra.
It contains all "practical" subsets of $\mathbb{R}$ (even such sets as the Cantor set or the set of rational numbers).

Taking the Cartesian products of subsets, one can generalize the Borel σ-algebra to $\mathbb{R}^2$ (place), $\mathbb{R}^3$ (3D space), etc.

The probability measure

Kolmogorov axioms (1933):

A probability measure is a real-valued function P defined on a σ-algebra $\mathcal{F} \subseteq 2^{\Omega}$, which satisfies:

1. Nonnegativity: $P(A) \geqslant 0$ for all $A \in \mathcal{F}$
2. Normalization: $P(\Omega) = 1$
3. Additivity: For any sequence of disjoint* events $A_1, A_2, \ldots \in \mathcal{F}$:

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

Andrey Kolmogorov (1903-1987)

Remark: symbol $\bigcup_{i=1}^{\infty} A_i$ means $A_1 \cup A_2 \cup \ldots$

*Recall: $A_i \cap A_j = \emptyset$ for each $i \neq j$

The properties of probability

Fact: The probability of the empty set is zero: $P(\emptyset) = 0$

The properties of probability

Fact: The probability of the empty set is zero: $P(\emptyset) = 0$

Proof: Take $A_1 = A_2 = \ldots = \emptyset$. Then $\bigcup_{i=1}^{\infty} A_i = \emptyset$.

From axiom 3 we have $P(\emptyset) = \sum_{i=1}^{\infty} P(\emptyset)$, which is only possible if $P(\emptyset) = 0$.

The properties of probability

Fact (finite additivity): For any disjoint events $A_1, \ldots, A_n$ we have $P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)$

The properties of probability

Fact (finite additivity): For any disjoint events $A_1, \ldots, A_n$ we have $P\Big(\bigcup_{i=1}^n A_i\Big) = \sum_{i=1}^n P(A_i)$

Proof: Take an infinite sequence of events $A_1, A_2, \ldots$, in which $A_{n+1} = A_{n+2} = \ldots = \emptyset$.

All events are disjoint $A_i \cap \emptyset = \emptyset$ for $i = 1, \ldots, n$.

Moreover, $\bigcup_{i=1}^\infty A_i = \bigcup_{i=1}^n A_i$.

Therefore:

$$P\Big(\bigcup_{i=1}^{n} A_i\Big) = P\Big(\bigcup_{i=1}^{\infty} A_i\Big) \overset{(3)}{=} \sum_{i=1}^{\infty} P(A_i) \overset{(*)}{=} \sum_{i=1}^{n} P(A_i),$$

where in $(*)$ we used $P(\emptyset) = 0$.

The properties of probability

Fact: for any event A it holds $P(A') = 1 - P(A)$

Conclusion: for any event A it holds $P(A) \leqslant 1$

The properties of probability

Fact: for any event A it holds $P(A') = 1 - P(A)$

Conclusion: for any event A it holds $P(A) \leqslant 1$

Proof: Since $A \cup A' = \Omega$, and A, A' are disjoint:

$$P(\Omega) = P(A) + P(A') = 1.$$

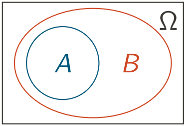
Therefore $P(A) = 1 - P(A')$.

Conclusion follows from the proven fact and from $P(A') \geqslant 0$ (axiom 1)

The properties of probability

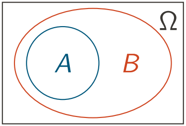
Fact: If $A \subseteq B$ then $P(B \setminus A) = P(B) - P(A)$

Conclusion: If $A \subset B$ then $P(B) \geqslant P(A)$.



The properties of probability

Fact: If $A \subseteq B$ then $P(B \setminus A) = P(B) - P(A)$
Conclusion: If $A \subset B$ then $P(B) \geqslant P(A)$.



Proof: We write down B as a disjoint sum $B = A \cup (B \setminus A)$.

Therefore $P(B) = P(A) + P(B \setminus A)$, from which both statements follow.

The properties of probability

Fact: For any events A and B it holds:

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

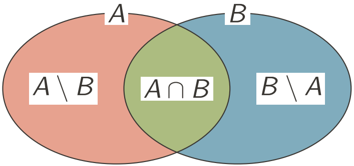
The properties of probability

Fact: For any events A and B it holds:

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

Proof: We divide subsets A, B and $A \cup B$ into disjoint parts:

$$\begin{aligned} A &= (A \setminus B) \cup (A \cap B) \\ B &= (B \setminus A) \cup (A \cap B) \\ A \cup B &= (A \setminus B) \cup (A \cap B) \cup (B \setminus A) \end{aligned}$$



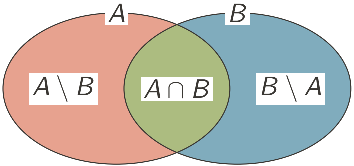
The properties of probability

Fact: For any events A and B it holds:

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

Proof: We divide subsets A, B and $A \cup B$ into disjoint parts:

$$\begin{aligned} A &= (A \setminus B) \cup (A \cap B) \\ B &= (B \setminus A) \cup (A \cap B) \\ A \cup B &= (A \setminus B) \cup (A \cap B) \cup (B \setminus A) \end{aligned}$$



From finite additivity:

$$\begin{aligned} P(A) &= P(A \setminus B) + P(A \cap B) \\ P(B) &= P(B \setminus A) + P(A \cap B) \\ P(A \cup B) &= P(A \setminus B) + P(A \cap B) \cup P(B \setminus A) \end{aligned}$$

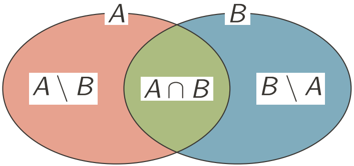
The properties of probability

Fact: For any events A and B it holds:

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

Proof: We divide subsets A, B and $A \cup B$ into disjoint parts:

$$\begin{aligned} A &= (A \setminus B) \cup (A \cap B) \\ B &= (B \setminus A) \cup (A \cap B) \\ A \cup B &= (A \setminus B) \cup (A \cap B) \cup (B \setminus A) \end{aligned}$$



From finite additivity:

$$\begin{aligned} P(A) &= P(A \setminus B) + P(A \cap B) \\ P(B) &= P(B \setminus A) + P(A \cap B) \\ P(A \cup B) &= P(A \setminus B) + P(A \cap B) \cup P(B \setminus A) \end{aligned}$$

Substituting the first and the second equality into the third one finishes the proof.

The properties of probability

Fact: For any events A and B it holds:

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

Conclusion: For any events A and B it holds:

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

The properties of probability

Fact: For any events A and B it holds:

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

Conclusion: For any events A and B it holds:

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

Union bound: For any events $A_1, \ldots, A_n$ it holds:

$$P(A_1 \cup \ldots \cup A_n) \leqslant P(A_1) + \ldots + P(A_n)$$

A simple proof by induction using the conclusion above.

Inclusion-exclusion principle

For any three events $A_1, A_2, A_3 \in \Omega$:

$$\begin{aligned} P(A_1 \cup A_2 \cup A_3) \ &= \ P(A_1) + P(A_2) + P(A_3) \\ &\quad - P(A_1 \cap A_2) - P(A_1 \cap A_3) - P(A_2 \cap A_3) \\ &\quad + P(A_1 \cap A_2 \cap A_3) \end{aligned}$$

Proved during the exercise classes

Inclusion-exclusion principle

For any three events $A_1, A_2, A_3 \in \Omega$:

$$\begin{aligned} P(A_1 \cup A_2 \cup A_3) &= P(A_1) + P(A_2) + P(A_3) \\ &\quad - P(A_1 \cap A_2) - P(A_1 \cap A_3) - P(A_2 \cap A_3) \\ &\quad + P(A_1 \cap A_2 \cap A_3) \end{aligned}$$

Proved during the exercise classes

More generally: for any events $A_1, \ldots, A_n$:

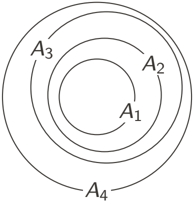
$$\begin{aligned} P(A_1 \cup \ldots \cup A_n) = \sum_{i=1}^{n} P(A_i) - \sum_{1 \leqslant i < j \leqslant n} P(A_i \cap A_j) + \\ + \sum_{1 \leqslant i < j < k \leqslant n} P(A_i \cap A_j \cap A_k) - \ldots + (-1)^{n-1} P(A_1 \cap \ldots \cap A_n) \end{aligned}$$

We skip a tedious proof by induction

The properties of probability*

We call sequence of events $A_1, A_2, A_3, \dots$ increasing if:

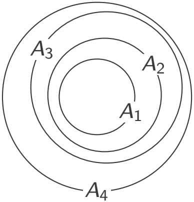
$$A_1 \subset A_2 \subset A_3 \subset \dots$$



The properties of probability*

We call sequence of events $A_1, A_2, A_3, \dots$ increasing if:

$$A_1 \subset A_2 \subset A_3 \subset \dots$$

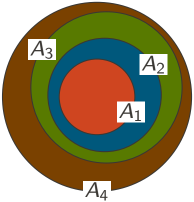


Fact (on continuity): If $A_1, A_2, \dots$ is increasing and $A = \bigcup_{n=1}^{\infty} A_n$ then:

$$P(A) = \lim_{n\to\infty} P(A_n)$$

The properties of probability*

We call sequence of events $A_1, A_2, A_3, \ldots$ increasing if:



$$A_1 \subset A_2 \subset A_3 \subset \ldots$$

Fact (on continuity): If $A_1, A_2, \ldots$ is increasing and $A = \bigcup_{n=1}^{\infty} A_n$ then:

$$P(A) = \lim_{n \to \infty} P(A_n)$$

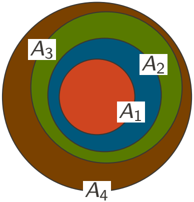
Proof: Define disjoint events:

$B_1 = A_1$, $B_2 = A_2 \setminus A_1$, $B_3 = A_3 \setminus A_2$, $B_4 = A_4 \setminus A_3, \ldots$

The properties of probability*

We call sequence of events $A_1, A_2, A_3, \ldots$ increasing if:

$$A_1 \subset A_2 \subset A_3 \subset \ldots$$



Fact (on continuity): If $A_1, A_2, \ldots$ is increasing and $A = \bigcup_{n=1}^{\infty} A_n$ then:

$$P(A) = \lim_{n \to \infty} P(A_n)$$

Proof: Define disjoint events:

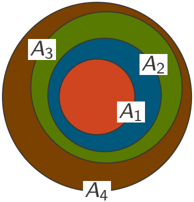
$B_1 = A_1$, $B_2 = A_2 \setminus A_1$, $B_3 = A_3 \setminus A_2$, $B_4 = A_4 \setminus A_3, \ldots$

We write down A_n as a disjoint sum: $A_n = B_1 \cup B_2 \cup \ldots \cup B_n$

Analogously, $A = B_1 \cup B_2 \cup \ldots$

The properties of probability*

We call sequence of events $A_1, A_2, A_3, \ldots$ increasing if:



$$A_1 \subset A_2 \subset A_3 \subset \ldots$$

Fact (on continuity): If $A_1, A_2, \ldots$ is increasing and $A = \bigcup_{n=1}^{\infty} A_n$ then:

$$P(A) = \lim_{n\to\infty} P(A_n)$$

Proof: Define disjoint events:

$B_1 = A_1$, $B_2 = A_2 \setminus A_1$, $B_3 = A_3 \setminus A_2$, $B_4 = A_4 \setminus A_3, \ldots$

We write down A_n as a disjoint sum: $A_n = B_1 \cup B_2 \cup \ldots \cup B_n$

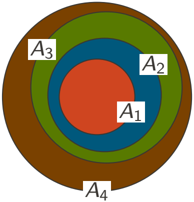
Analogously, $A = B_1 \cup B_2 \cup \ldots$

$$P(A) \stackrel{(3)}{=} \sum_{i=1}^{\infty} P(B_i) = \lim_{n\to\infty} \sum_{i=1}^{n} P(B_i)$$

The properties of probability*

We call sequence of events $A_1, A_2, A_3, \ldots$ increasing if:

$$A_1 \subset A_2 \subset A_3 \subset \ldots$$



Fact (on continuity): If $A_1, A_2, \ldots$ is increasing and $A = \bigcup_{n=1}^{\infty} A_n$ then:

$$P(A) = \lim_{n\to\infty} P(A_n)$$

Proof: Define disjoint events:

$B_1 = A_1$, $B_2 = A_2 \setminus A_1$, $B_3 = A_3 \setminus A_2$, $B_4 = A_4 \setminus A_3, \ldots$

We write down A_n as a disjoint sum: $A_n = B_1 \cup B_2 \cup \ldots \cup B_n$

Analogously, $A = B_1 \cup B_2 \cup \ldots$

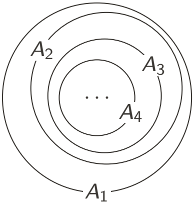
$$P(A) \overset{(3)}{=} \sum_{i=1}^{\infty} P(B_i) = \lim_{n\to\infty} \sum_{i=1}^{n} P(B_i)$$

$$\overset{(3)}{=} \lim_{n\to\infty} P(B_1 \cup \ldots \cup B_n) = \lim_{n\to\infty} P(A_n)$$

The properties of probability*

We call a sequence of events $A_1, A_2, A_3, \dots$ decreasing if:

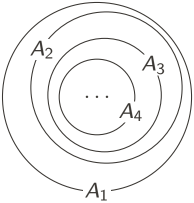
$$A_1 \supset A_2 \supset A_3 \supset \dots$$



The properties of probability*

We call a sequence of events $A_1, A_2, A_3, \ldots$ decreasing if:

$$A_1 \supset A_2 \supset A_3 \supset \ldots$$



Fact: If $A_1, A_2, \ldots$ is decreasing and $A = \bigcap_{n=1}^{\infty} A_n$ then:

$$P(A) = \lim_{n\to\infty} P(A_n)$$

Proof: Note that $A_1', A_2', \ldots$ is increasing and $A' = \bigcup_{n=1}^{\infty} A_n'$ (from the de Morgan's law).

Using the previous fact, $P(A') = \lim_{n\to\infty} P(A_n')$.

Therefore:

$$P(A) = 1 - P(A') = \lim_{n\to\infty} (1 - P(A_n')) = \lim_{n\to\infty} P(A_n)$$

Classical probability satisfies the axioms of Kolmogorov

$$P(A) = \frac{|A|}{|\Omega|}$$

Classical probability satisfies the axioms of Kolmogorov

$$P(A) = \frac{|A|}{|\Omega|}$$

Axiom 1. Obvious

Classical probability satisfies the axioms of Kolmogorov

$$P(A) = \frac{|A|}{|\Omega|}$$

Axiom 1. Obvious

Axiom 2. $P(\Omega) = \frac{|\Omega|}{|\Omega|} = 1$

Classical probability satisfies the axioms of Kolmogorov

$$P(A) = \frac{|A|}{|\Omega|}$$

Axiom 1. Obvious

Axiom 2. $P(\Omega) = \frac{|\Omega|}{|\Omega|} = 1$

Axiom 3. Since Ω is finite, we only consider finite sequences of events. If $A_1, \ldots, A_n$ are disjoint, then:

$$\left|A_1 \cup A_2 \cup \ldots \cup A_n\right| \;=\; |A_1| + |A_2| + \ldots + |A_n|.$$

So:

$$P\Big(\bigcup_{i=1}^{n} A_i\Big) \;=\; \frac{\left|\bigcup_{i=1}^{n} A_i\right|}{|\Omega|} \;=\; \sum_{i=1}^{n} \frac{|A_i|}{|\Omega|} \;=\; \sum_{i=1}^{n} P(A_i).$$

Classical probability satisfies the axioms of Kolmogorov

$$P(A) = \frac{|A|}{|\Omega|}$$

Axiom 1. Obvious

Axiom 2. $P(\Omega) = \frac{|\Omega|}{|\Omega|} = 1$

Axiom 3. Since Ω is finite, we only consider finite sequences of events. If $A_1, \ldots, A_n$ are disjoint, then:

$$\left|A_1 \cup A_2 \cup \ldots \cup A_n\right| \; = \; |A_1| + |A_2| + \ldots + |A_n|.$$

So:

$$P\Big(\bigcup_{i=1}^{n} A_i\Big) \; = \; \frac{\left|\bigcup_{i=1}^{n} A_i\right|}{|\Omega|} \; = \; \sum_{i=1}^{n} \frac{|A_i|}{|\Omega|} \; = \; \sum_{i=1}^{n} P(A_i).$$

(similar proof for geometric probability)

Probability over the countable sample space

Informally: If Ω is countable, it suffices to assign a "probability value" to every outcome, and then the probability of any event A is just a sum of the probability values of all outcomes which belong to A.

Probability over the countable sample space

Formally: Let $\Omega = \{\omega_1, \omega_2, \ldots\}$ be a countable set and let $\mathcal{F} = 2^\Omega$.

To every ω_n assign a real number $p_n \geqslant 0$, such that

$$\sum_{n=1}^{\infty} p_n = 1.$$

The probability of any event $A \subseteq \Omega$ is defined as a sum of p_n over all $\omega_n \in A$:

$$P(A) \;=\; \sum_{n:\,\omega_n \in A} p_n,$$

Therefore, $p_n = P(\{\omega_n\})$ is the probability of outcome ω_n.

Of course, all of this also holds for a finite sample space Ω.

We will verify at exercise classes that that this probability satisfies the Kolmogorov axioms

Example: rolling two dice

We are only interested in the sum of the numbers on both dice

Example: rolling two dice

We are only interested in the sum of the numbers on both dice

- Sample space $\Omega = \{\omega_2, \omega_3, \ldots, \omega_{12}\}$:

Example: rolling two dice

We are only interested in the sum of the numbers on both dice

- Sample space $\Omega = \{\omega_2, \omega_3, \ldots, \omega_{12}\}$:
- Probabilities of outcomes:

ω_n	p_n	ω_n	p_n	ω_n	p_n
ω_2	$^1/_{36}$	ω_6	$^5/_{36}$	ω_{10}	$^3/_{36}$
ω_3	$^2/_{36}$	ω_7	$^6/_{36}$	ω_{11}	$^2/_{36}$
ω_4	$^3/_{36}$	ω_8	$^5/_{36}$	ω_{12}	$^1/_{36}$
ω_5	$^4/_{36}$	ω_9	$^4/_{36}$		

Example: rolling two dice

We are only interested in the sum of the numbers on both dice

- Sample space $\Omega = \{\omega_2, \omega_3, \ldots, \omega_{12}\}$:
- Probabilities of outcomes:

ω_n	p_n	ω_n	p_n	ω_n	p_n
ω_2	$^1/_{36}$	ω_6	$^5/_{36}$	ω_{10}	$^3/_{36}$
ω_3	$^2/_{36}$	ω_7	$^6/_{36}$	ω_{11}	$^2/_{36}$
ω_4	$^3/_{36}$	ω_8	$^5/_{36}$	ω_{12}	$^1/_{36}$
ω_5	$^4/_{36}$	ω_9	$^4/_{36}$		

- Probabilities of events:
  - „We rolled 7":

    $A = \{\omega_7\}$, $P(A) = \frac{6}{36}$
  - „We rolled at least 10":

    $A = \{\omega_{10}, \omega_{11}, \omega_{12}\}$, $P(A) = \frac{3}{36} + \frac{2}{36} + \frac{1}{36} = \frac{1}{6}$

Example: tossing a coin till the first head

We are only interested in the number of tosses

Example: tossing a coin till the first head

We are only interested in the number of tosses

- Sample space $\Omega = \{\omega_1, \omega_2, \omega_3, \ldots\}$
  (where ω_n means „the first head in the n-th toss")

Example: tossing a coin till the first head

We are only interested in the number of tosses

- Sample space $\Omega = \{\omega_1, \omega_2, \omega_3, \ldots\}$
  (where ω_n means „the first head in the n-th toss")
- Probability of outcome ω_n:

Example: tossing a coin till the first head

We are only interested in the number of tosses

- Sample space $\Omega = \{\omega_1, \omega_2, \omega_3, \dots\}$
  (where ω_n means „the first head in the n-th toss")
- Probability of outcome ω_n: $p_n = \frac{1}{2^n}$

Example: tossing a coin till the first head

We are only interested in the number of tosses

- Sample space $\Omega = \{\omega_1, \omega_2, \omega_3, \dots\}$
  (where ω_n means „the first head in the n-th toss")
- Probability of outcome ω_n: $p_n = \frac{1}{2^n}$
- Probability of event A „more than 5 tosses":

Example: tossing a coin till the first head

We are only interested in the number of tosses

- Sample space $\Omega = \{\omega_1, \omega_2, \omega_3, \ldots\}$
  (where ω_n means „the first head in the n-th toss")
- Probability of outcome ω_n: $p_n = \frac{1}{2^n}$
- Probability of event A „more than 5 tosses":

$$\begin{aligned} A &= \{\omega_6, \omega_7, \ldots\} = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5\}' \\ P(A) &= 1 - \sum_{n=1}^{5} p_n = 1 - \frac{1}{2} - \frac{1}{4} - \frac{1}{8} - \frac{1}{16} - \frac{1}{32} = \frac{1}{32} \end{aligned}$$

Example: tossing a coin till the first head

We are only interested in the number of tosses

- Sample space $\Omega = \{\omega_1, \omega_2, \omega_3, \dots\}$
  (where ω_n means „the first head in the n-th toss")
- Probability of outcome ω_n: $p_n = \frac{1}{2^n}$
- Probability of event A „more than 5 tosses":

$$
\begin{aligned}
A &= \{\omega_6, \omega_7, \dots\} = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5\}' \\
P(A) &= 1 - \sum_{n=1}^{5} p_n = 1 - \frac{1}{2} - \frac{1}{4} - \frac{1}{8} - \frac{1}{16} - \frac{1}{32} = \frac{1}{32}
\end{aligned}
$$

- Probability of event B „even number of tosses":

Example: tossing a coin till the first head

We are only interested in the number of tosses

- Sample space $\Omega = \{\omega_1, \omega_2, \omega_3, \dots\}$
  (where ω_n means „the first head in the n-th toss")
- Probability of outcome ω_n: $p_n = \frac{1}{2^n}$
- Probability of event A „more than 5 tosses":

$$\begin{aligned} A &= \{\omega_6, \omega_7, \dots\} = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5\}' \\ P(A) &= 1 - \sum_{n=1}^{5} p_n = 1 - \frac{1}{2} - \frac{1}{4} - \frac{1}{8} - \frac{1}{16} - \frac{1}{32} = \frac{1}{32} \end{aligned}$$

- Probability of event B „even number of tosses":

$$\begin{aligned} B &= \{\omega_2, \omega_4, \omega_6, \dots\} \\ P(B) &= \sum_{n=1}^{\infty} p_{2n} = \sum_{n=1}^{\infty} \frac{1}{4^n} = \frac{1}{4} \cdot \frac{1}{1 - \frac{1}{4}} = \frac{1}{3} \end{aligned}$$

Example: natural numbers

Let $\Omega = \mathbb{N} = \{1, 2, 3, \ldots\}$. Can we anyhow assign equal probabilities to all natural numbers?

Example: natural numbers

Let $\Omega = \mathbb{N} = \{1, 2, 3, \ldots\}$. Can we anyhow assign equal probabilities to all natural numbers?

We cannot! To every number $n \in \mathbb{N}$ assign probability p_n
From the normalization condition:

$$\sum_{n=1}^{\infty} p_n = 1$$

If $p_n = p$ for all n, then the sum on the left gives zero (for $p = 0$) or infinity (dla $p > 0$)!

Example: natural numbers

Let $\Omega = \mathbb{N} = \{1, 2, 3, \ldots\}$. Can we anyhow assign equal probabilities to all natural numbers?

We cannot! To every number $n \in \mathbb{N}$ assign probability p_n

From the normalization condition:

$$\sum_{n=1}^{\infty} p_n \; = \; 1$$

If $p_n = p$ for all n, then the sum on the left gives zero (for $p = 0$) or infinity (dla $p > 0$)!

A necessary condition for the convergence of the series $\sum_{n=1}^{\infty} p_n$ is the convergence of the summand to zero:

$$\lim_{n\to\infty} p_n = 0$$

This condition is not sufficient.

Example: natural numbers

Can we assign probabilities to natural numbers in such a way that for every $n \in \mathbb{N}$: (a) $p_n \propto \frac{1}{n}$; (b) $p_n \propto \frac{1}{n^2}$? ($\propto$ means „proportional to")

Example: natural numbers

Can we assign probabilities to natural numbers in such a way that for every $n \in \mathbb{N}$: (a) $p_n \propto \frac{1}{n}$; (b) $p_n \propto \frac{1}{n^2}$? ($\propto$ means „proportional to")

(a) We cannot. The series:

$$\sum_{n=1}^{\infty} \frac{1}{n}$$

is divergent, so there exists no constant c, such that $p_n = \frac{c}{n}$ and $\sum_{n=1}^{\infty} p_n = 1$

Example: natural numbers

Can we assign probabilities to natural numbers in such a way that for every $n \in \mathbb{N}$: (a) $p_n \propto \frac{1}{n}$; (b) $p_n \propto \frac{1}{n^2}$? ($\propto$ means „proportional to")

(a) We cannot. The series:

$$\sum_{n=1}^{\infty} \frac{1}{n}$$

is divergent, so there exists no constant c, such that $p_n = \frac{c}{n}$ and $\sum_{n=1}^{\infty} p_n = 1$

(b) We can. We have:

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6} \qquad \text{(the Basel problem)},$$

so that:

$$p_n = \frac{6}{\pi^2 n^2}$$

Example: natural numbers

Can we assign probabilities to natural numbers in such a way that for every $n \in \mathbb{N}$: (a) $p_n \propto \frac{1}{n}$; (b) $p_n \propto \frac{1}{n^2}$? ($\propto$ means „proportional to")

(a) We cannot. The series:

$$\sum_{n=1}^{\infty} \frac{1}{n}$$

is divergent, so there exists no constant c, such that $p_n = \frac{c}{n}$ and $\sum_{n=1}^{\infty} p_n = 1$

(b) We can. We have:

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6} \qquad \text{(the Basel problem)},$$

so that:

$$p_n = \frac{6}{\pi^2 n^2}$$

Generally: the sum normalizes when $p_n \propto \frac{1}{n^\alpha}$ with $\alpha > 1$.

Uncountable sample space

Can we still assign a probability value to each outcome and, based on that, compute the probability of every event?

Uncountable sample space

Can we still assign a probability value to each outcome and, based on that, compute the probability of every event?

Unfortunately not. Take $\Omega = [0, 1]$ and the points drawn according to a geometric probability model.

- Every point $x \in [0, 1]$ is an outcome.
- Every point $x \in [0, 1]$ has zero length, and so zero probability, $P(\{x\}) = 0$.
- But every interval $[a, b] \subseteq \Omega$ with $a \neq b$ has non-zero length and consequently non-zero probability.

More about this at lecture on continuous random variables.

The interpretation of probability

Kolmogorov axioms describe the mathematical properties of the probability measure, but say nothing about how to specify its actual values.

What is the interpretation of the value of probability?

The interpretation of probability

Kolmogorov axioms describe the mathematical properties of the probability measure, but say nothing about how to specify its actual values.

What is the interpretation of the value of probability?

- Classical (Laplace): all outcomes are equally likely ✗
- Frequentist: probability as the limit of a frequency ✓
- Subjective: probability as a measure of belief ✓

Frequentist interpretation

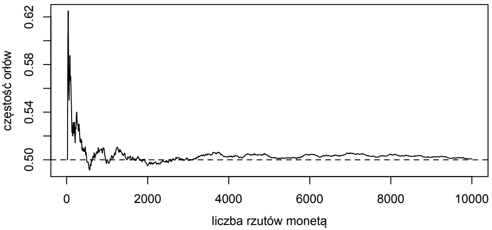
Concerns repeatable random experiments.

Repeat the random experiment N times.

For any event A, let N_A denote the number of experiments in which A occurred.

The probability of event A is the limit of the frequency of A:

$$P(A) \;=\; \lim_{N\to\infty} \frac{N_A}{N}.$$



Subjective (Bayesian) interpretation

The probability does not exists in an objective way, and only measures the subjective degree of belief

Subjective (Bayesian) interpretation

The probability does not exists in an objective way, and only measures the subjective degree of belief

Allows assigning probabilities to any, not necessarily repeatable, events:

- What is the chance that the sun will rise tomorrow?
- What is the chance that Poland will the next World Cup?
- What is the chance that X will win the coming presidential election?

Subjective (Bayesian) interpretation

The probability does not exists in an objective way, and only measures the subjective degree of belief

Allows assigning probabilities to any, not necessarily repeatable, events:

- What is the chance that the sun will rise tomorrow?
- What is the chance that Poland will the next World Cup?
- What is the chance that X will win the coming presidential election?

Can be assessed by means of a personal gambling strategy:

> *The probability of event A is such a number p, that for any price S (positive or negative) I am willing to pay $p \cdot S$ for entering the game, in which I win S when A occurs, or win nothing if A does not occur*

Subjective (Bayesian) interpretation

The probability does not exists in an objective way, and only measures the subjective degree of belief

Allows assigning probabilities to any, not necessarily repeatable, events:

- What is the chance that the sun will rise tomorrow?
- What is the chance that Poland will the next World Cup?
- What is the chance that X will win the coming presidential election?

Can be assessed by means of a personal gambling strategy:

> *The probability of event A is such a number p, that for any price S (positive or negative) I am willing to pay $p \cdot S$ for entering the game, in which I win S when A occurs, or win nothing if A does not occur*

Subjective probability is updated based on the observations using the *Bayes rule*