भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Mathematical Foundations for Data Science (Probability)

Lecture 01: Sample Space, Events, Probability Measure and its Properties, Examples of Probability Assignment

Karthik P. N.
Assistant Professor, Department of AI
Email: pnkarthik@ai.iith.ac.in

10 August 2024

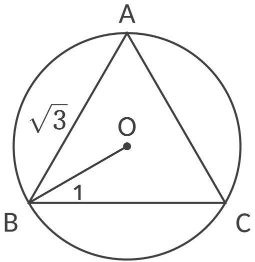
Probability Theory – Humble Beginnings

- Bernoulli (1713) and de Moivre (1718) gave the first definition of probability:

$$\text{probability of an event} = \frac{\text{\# favourable outcomes}}{\text{total number of outcomes}}.$$

- Cournot (1843):
  "An event with very small probability is morally impossible; an event with very high probability is morally certain."
- French mathematicians of the day were satisfied with the "frequentist" approach to probability, but not the German and English mathematicians of the day
- Frequentist approach could not satisfactorily explain certain paradoxes

Bertrand's Paradox

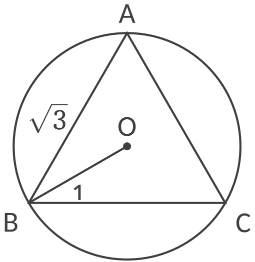


Take a circle with unit radius and inscribe an equilateral triangle in it. Draw a random chord. What is the probability that the length of the "random chord" is greater than $\sqrt{3}$?

Bertrand's perfectly valid arguments:

- Mid-point of chord should lie inside incircle of radius $1/2$

Bertrand's Paradox

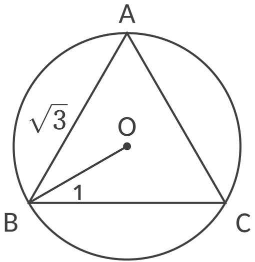


Take a circle with unit radius and inscribe an equilateral triangle in it. Draw a random chord. What is the probability that the length of the "random chord" is greater than $\sqrt{3}$?

Bertrand's perfectly valid arguments:

- Mid-point of chord should lie inside incircle of radius $1/2$

Bertrand's Paradox

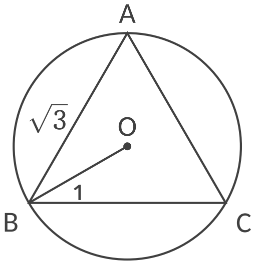


Take a circle with unit radius and inscribe an equilateral triangle in it. Draw a random chord. What is the probability that the length of the "random chord" is greater than $\sqrt{3}$?

Bertrand's perfectly valid arguments:

- Mid-point of chord should lie inside incircle of radius $1/2$
  Answer: $1/4$
- Angle between chord and tangent at A should be between $\pi/3$ and $2\pi/3$

Bertrand's Paradox

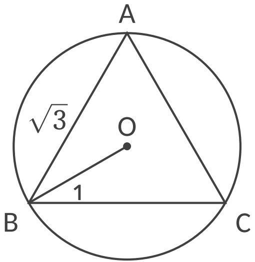


Take a circle with unit radius and inscribe an equilateral triangle in it. Draw a random chord. What is the probability that the length of the "random chord" is greater than $\sqrt{3}$?

Bertrand's perfectly valid arguments:

- Mid-point of chord should lie inside incircle of radius $1/2$
  Answer: $1/4$
- Angle between chord and tangent at A should be between $\pi/3$ and $2\pi/3$

Bertrand's Paradox

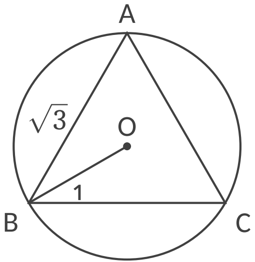


Take a circle with unit radius and inscribe an equilateral triangle in it. Draw a random chord. What is the probability that the length of the "random chord" is greater than $\sqrt{3}$?

Bertrand's perfectly valid arguments:

- Mid-point of chord should lie inside incircle of radius $1/2$
  Answer: $1/4$
- Angle between chord and tangent at A should be between $\pi/3$ and $2\pi/3$
  Answer: $1/3$
- Mid-point of chord should be between O and projection of O onto side BC

Bertrand's Paradox

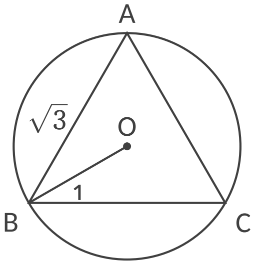


Take a circle with unit radius and inscribe an equilateral triangle in it. Draw a random chord. What is the probability that the length of the "random chord" is greater than $\sqrt{3}$?

Bertrand's perfectly valid arguments:

- Mid-point of chord should lie inside incircle of radius $1/2$
  Answer: $1/4$
- Angle between chord and tangent at A should be between $\pi/3$ and $2\pi/3$
  Answer: $1/3$
- Mid-point of chord should be between O and projection of O onto side BC

Bertrand's Paradox



Take a circle with unit radius and inscribe an equilateral triangle in it. Draw a random chord. What is the probability that the length of the "random chord" is greater than $\sqrt{3}$?

Bertrand's perfectly valid arguments:

- Mid-point of chord should lie inside incircle of radius $1/2$
  Answer: $1/4$
- Angle between chord and tangent at A should be between $\pi/3$ and $2\pi/3$
  Answer: $1/3$
- Mid-point of chord should be between O and projection of O onto side BC
  Answer: $1/2$

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Borel to the Rescue

- Contributions to Measure Theory by Borel (1894) provided a shift in perspective
- Countable unions played a key role in Borel's theory
- Kolmogorov's genius was in applying Borel's theory to formalise the axioms of probability, laying the foundation stone for modern probability theory
- For more details on the history of probability, see [Shafer and Vovk, 2018] and [Kolmogorov, 2004]

Sample Space

We begin with two universally accepted entities:

- Random experiment
- Outcome (denoted by ω) – source of randomness

Definition (Sample Space)

The sample space (denoted by Ω) of a random experiment is the set of all possible outcomes of the random experiment.

Example: Tossing a coin once

- If our interest is in the face that shows, then $\Omega = \{H, T\}$

Sample Space

We begin with two universally accepted entities:

- Random experiment
- Outcome (denoted by ω) – source of randomness

Definition (Sample Space)

The sample space (denoted by Ω) of a random experiment is the set of all possible outcomes of the random experiment.

Example: Tossing a coin once

- If our interest is in the face that shows, then $\Omega = \{H, T\}$
- If our interest is in the velocity with which the coin lands on ground, then $\Omega = [0, \infty) = \mathbb{R}_+$

Sample Space

We begin with two universally accepted entities:

- Random experiment
- Outcome (denoted by ω) – source of randomness

Definition (Sample Space)

The sample space (denoted by Ω) of a random experiment is the set of all possible outcomes of the random experiment.

Example: Tossing a coin once

- If our interest is in the face that shows, then $\Omega = \{H, T\}$
- If our interest is in the velocity with which the coin lands on ground, then $\Omega = [0, \infty) = \mathbb{R}_+$
- If our interest is in the number of times coin flips in air, then $\Omega = \mathbb{N}$

Example: Toss a coin n times, for some $n < \infty$.
Interest: faces that show up
$\Omega = \{H, T\}^n$

Example: Toss a coin infinitely many times.
Interest: faces that show up
$\Omega = \{H, T\}^\infty$

Event

Informal Definition (Event)

Informally,[a] an event is a subset of outcomes "of interest" to us.

[a]We shall give a more formal definition of an event later.

Example: Toss a coin 3 times; interest is in the faces that show up
$\Omega = \{H, T\}^3 = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$

Event A of interest: at least 2 heads show up

Event

Informal Definition (Event)

Informally,[a] an event is a subset of outcomes "of interest" to us.

[a] We shall give a more formal definition of an event later.

Example: Toss a coin 3 times; interest is in the faces that show up
$\Omega = \{H, T\}^3 = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$

Event A of interest: at least 2 heads show up
$A = \{HHH, THH, HTH, HHT\}$

Note

If an outcome $\omega \in A$ occurs, we say that the event A occurs.

Algebra

Definition (Algebra)

Let Ω be a sample space.
A collection $\mathscr{A}$ of subsets of Ω is called an algebra if it satisfies the following properties:

1. $\Omega \in \mathscr{A}$.
2. $A \in \mathscr{A} \implies A^c \in \mathscr{A}$ (closure under complements).
3. $A, B \in \mathscr{A} \implies A \cup B \in \mathscr{A}$.

Property 3 above implies, by mathematical induction, that

$$A_1, A_2, \ldots, A_n \in \mathscr{A} \implies \bigcup_{i=1}^{n} A_i \in \mathscr{A} \quad \text{for all } n \in \mathbb{N} \quad \text{(closure under finite unions).}$$

Exercise

Show that an algebra is closed under finite intersections.

Algebra – Examples

$\Omega = \{1, 2, \ldots, 6\}$ – outcomes of single throw of dice

- $\mathscr{A} = \{\emptyset, \Omega\}$
- $\mathscr{A} = 2^{\Omega}$ = collection of all subsets of Ω
- $\mathscr{A} = \Big\{ \emptyset, \Omega, \{1\}, \{2, 3\},$

Algebra – Examples

$\Omega = \{1, 2, \ldots, 6\}$ – outcomes of single throw of dice

- $\mathscr{A} = \{\emptyset, \Omega\}$
- $\mathscr{A} = 2^{\Omega}$ = collection of all subsets of Ω
- $\mathscr{A} = \left\{\emptyset, \Omega, \{1\}, \{2, 3\}, \{1, 2, 3\}, \{4, 5, 6\}, \{1, 4, 5, 6\}, \{2, 3, 4, 5, 6\}\right\}$

σ-Algebra – Motivation

Toss a coin until first head shows up

$\Omega = \{H, TH, TTH, TTTH, \ldots\}$

$$\mathscr{A} = \Big\{\emptyset, \Omega, \{H\}, \{TH\}, \{TTH\}, \{TTTH\}, \ldots \\ \{H, TH\}, \{H, TTH\}, \{H, TTTH\}, \ldots \\ \{H, TH, TTH\}, \{H, TH, TTTH\}, \ldots \\ \{H, TH, TTH, TTTH\}, \ldots \Big\}$$

Event of interest A = # of tosses is even

$A = \{TH, TTTH, \ldots\}$

σ-Algebra – Motivation

Toss a coin until first head shows up

$\Omega = \{H, TH, TTH, TTTH, \ldots\}$

$$\mathscr{A} = \Big\{ \emptyset, \Omega, \{H\}, \{TH\}, \{TTH\}, \{TTTH\}, \ldots \\ \{H, TH\}, \{H, TTH\}, \{H, TTTH\}, \ldots \\ \{H, TH, TTH\}, \{H, TH, TTTH\}, \ldots \\ \{H, TH, TTH, TTTH\}, \ldots \Big\}$$

Event of interest A = # of tosses is even

$A = \{TH, TTTH, \ldots\} \notin \mathscr{A}$

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

σ-Algebra

Definition (σ-Algebra)

Let Ω be a sample space.
A collection $\mathscr{F}$ of subsets of Ω is called a σ-algebra if it satisfies the following properties:

- $\Omega \in \mathscr{F}$.
- $A \in \mathscr{F} \implies A^c \in \mathscr{F}$ (closed under complements).
- $A_1, A_2, \ldots \in \mathscr{F} \implies \bigcup_{i=1}^{\infty} A_i \in \mathscr{F}$ (closure under countably infinite unions).

Remarks:

- Elements of a σ-algebra are called events

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

σ-Algebra

Definition (σ-Algebra)

Let Ω be a sample space.
A collection $\mathscr{F}$ of subsets of Ω is called a σ-algebra if it satisfies the following properties:

- $\Omega \in \mathscr{F}$.
- $A \in \mathscr{F} \implies A^c \in \mathscr{F}$ (closed under complements).
- $A_1, A_2, \ldots \in \mathscr{F} \implies \bigcup_{i=1}^{\infty} A_i \in \mathscr{F}$ (closure under countably infinite unions).

Remarks:

- Elements of a σ-algebra are called events
- An event $A \in \mathscr{F}$ is also referred to as an $\mathscr{F}$-measurable set

σ-Algebra

Definition (σ-Algebra)

Let Ω be a sample space.
A collection $\mathscr{F}$ of subsets of Ω is called a σ-algebra if it satisfies the following properties:

- $\Omega \in \mathscr{F}$.
- $A \in \mathscr{F} \implies A^c \in \mathscr{F}$ (closed under complements).
- $A_1, A_2, \ldots \in \mathscr{F} \implies \bigcup_{i=1}^{\infty} A_i \in \mathscr{F}$ (closure under countably infinite unions).

Remarks:

- Elements of a σ-algebra are called events
- An event $A \in \mathscr{F}$ is also referred to as an $\mathscr{F}$-measurable set
- Every σ-algebra is also an algebra, but the converse is not true

σ-Algebra

Definition (σ-Algebra)

Let Ω be a sample space.

A collection $\mathscr{F}$ of subsets of Ω is called a σ-algebra if it satisfies the following properties:

- $\Omega \in \mathscr{F}$.
- $A \in \mathscr{F} \implies A^c \in \mathscr{F}$ (closed under complements).
- $A_1, A_2, \ldots \in \mathscr{F} \implies \bigcup_{i=1}^{\infty} A_i \in \mathscr{F}$ (closure under countably infinite unions).

Remarks:

- Elements of a σ-algebra are called events
- An event $A \in \mathscr{F}$ is also referred to as an $\mathscr{F}$-measurable set
- Every σ-algebra is also an algebra, but the converse is not true
- The pair $(\Omega, \mathscr{F})$ is called a measurable space

Probability Measure

Fix a measurable space $(\Omega, \mathscr{F})$.

Definition (Probability Measure)

A function $\mathbb{P} : \mathscr{F} \to [0, 1]$ is called a probability measure if the following properties are satisfied:

1. $\mathbb{P}(\emptyset) = 0$.
2. $\mathbb{P}(\Omega) = 1$.
3. If $A_1, A_2, \ldots$ is a countable collection of mutually disjoint sets, with $A_i \in \mathscr{F}$ for each $i \in \mathbb{N}$ and $A_i \cap A_j = \emptyset$ for all $i \neq j$, then

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

The triplet $(\Omega, \mathscr{F}, \mathbb{P})$ is called a probability space

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Properties of Probability Measures

Properties of Probability Measures

Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$.

- For any two disjoint sets $A, B \in \mathscr{F}$,

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B).$$

Properties of Probability Measures

Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$.

- For any $n \in \mathbb{N}$ and a collection of mutually disjoint sets $B_1, \ldots, B_n \in \mathscr{F}$,

$$\mathbb{P}\left(\bigcup_{i=1}^{n} B_i\right) = \sum_{i=1}^{n} \mathbb{P}(B_i).$$

Properties of Probability Measures

Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$.

- For any set $A \in \mathscr{F}$,

$$\mathbb{P}(A^c) = 1 - \mathbb{P}(A).$$

Properties of Probability Measures

Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$.

- (Monotonicity property)
  For any two sets $A, B \in \mathscr{F}$ such that $A \subseteq B$,
  $$\mathbb{P}(A) \leq \mathbb{P}(B).$$

Corollary

Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. For all $A, B \in \mathscr{F}$ such that $A \subseteq B$,

$$\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A).$$

Properties of Probability Measures

Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$.

- For any two sets $A, B \in \mathscr{F}$,

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).$$

Properties of Probability Measures

Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$.

- (Inclusion-Exclusion Principle)
  For any $n \in \mathbb{N}$ and sets $A_1, \ldots, A_n \in \mathscr{F}$,

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mathbb{P}(A_i) - \sum_{i<j} \mathbb{P}(A_i \cap A_j) + \sum_{i<j<k} \mathbb{P}(A_i \cap A_j \cap A_k) - \cdots + (-1)^{n+1}\, \mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right).$$

Continuity of Probability Measure – 1

Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$.

- If $A_1 \subseteq A_2 \subseteq A_3 \subseteq \cdots$, where $A_i \in \mathscr{F}$ for each $i \in \mathbb{N}$, then

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \mathbb{P}\left(\lim_{n\to\infty} A_n\right) = \lim_{n\to\infty} \mathbb{P}(A_n).$$

Continuity of Probability Measure – 2

Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$.

- If $A_1 \supseteq A_2 \supseteq A_3 \supseteq \cdots$, where $A_i \in \mathscr{F}$ for each $i \in \mathbb{N}$, then

$$\mathbb{P}\left(\bigcap_{i=1}^{\infty} A_i\right) = \mathbb{P}\left(\lim_{n\to\infty} A_n\right) = \lim_{n\to\infty} \mathbb{P}(A_n).$$

Union Bound

Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$.

- For any $A_1, A_2, \ldots \in \mathscr{F}$,

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_i\right) \leq \sum_{n=1}^{\infty} \mathbb{P}(A_n)$$

Union Bound

Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$.

- For any $A_1, A_2, \ldots \in \mathscr{F}$,

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_i\right) \leq \sum_{n=1}^{\infty} \mathbb{P}(A_n)$$

Corollary

Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. For any two sets $A, B \in \mathscr{F}$,

$$\mathbb{P}(A \cup B) \leq \mathbb{P}(A) + \mathbb{P}(B).$$

More generally, for any $n \in \mathbb{N}$ and sets $A_1, \ldots, A_n \in \mathscr{F}$,

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} \mathbb{P}(A_i).$$

Probability Assignment – Examples

- $\Omega = \{H, T\}, \quad \mathscr{F} = 2^{\Omega} = \left\{\emptyset, \Omega, \{H\}, \{T\}\right\}$

- $\Omega = \{1, 2, \ldots, 6\},$
  $\mathscr{F} = \left\{\emptyset, \Omega, \{1\}, \{2, 3\},\right.$

Probability Assignment – Examples

- $\Omega = \{H, T\}, \quad \mathscr{F} = 2^{\Omega} = \Big\{\emptyset, \Omega, \{H\}, \{T\}\Big\}$

- $\Omega = \{1, 2, \ldots, 6\},$
  $\mathscr{F} = \Big\{\emptyset, \Omega, \{1\}, \{2, 3\}, \{1, 2, 3\}, \{4, 5, 6\}, \{1, 4, 5, 6\}, \{2, 3, 4, 5, 6\}\Big\}$

Some remarks:

- $\mathbb{P}(A) = 0 \;\not\Longrightarrow\; A = \emptyset$
- $\mathbb{P}(A) = 1 \;\not\Longrightarrow\; A = \Omega$

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

References

- Kolmogorov, A. N. (2004).
  From the heritage of A. N. Kolmogorov: The theory of probability.
  Theory of Probability & Its Applications, 48(2):191–220.
- Shafer, G. and Vovk, V. (2018).
  The origins and legacy of Kolmogorov's Grundbegriffe.
  arXiv preprint arXiv:1802.06071.