Mathematics Brush-Up for Data Science

Lecture Slides 15: Probability

Nicklas S. Andersen

University of Southern Denmark (SDU)
Department of Mathematics & Computer Science (IMADA)

Updated: 2026-08-31 18:49 CEST

Probability

Probability measures how likely an event is to occur.

We use two complementary viewpoints:

- Theoretical probability comes from a mathematical model
- Empirical probability is estimated from observed data

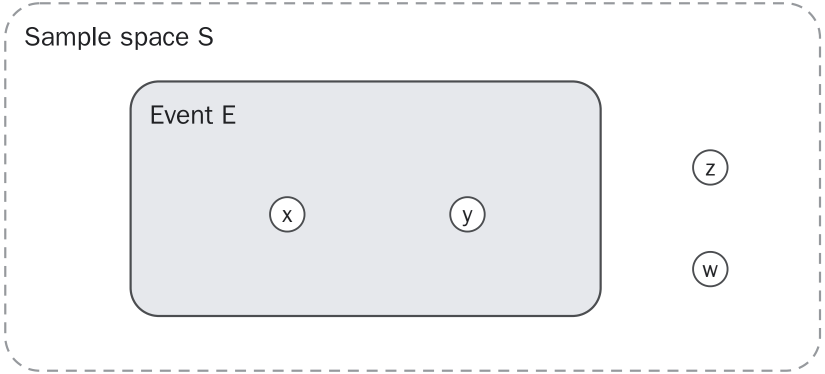
Both use the same language of experiments, outcomes, sample spaces, and events.

Probability Terminology

- Definitions

Before assigning probabilities, we define a common language:

Term	Meaning
Random experiment	is an activity whose result cannot be predicted in advance.
Outcome	is one possible result of the experiment.
Sample space S	is the set of all possible outcomes.
Event $E \subseteq S$	is a collection of outcomes we want to study.



The event E occurs when the experiment produces one of the outcomes, in this illustration, x or y.

Sample Spaces & Events

- Example: Rolling a Die

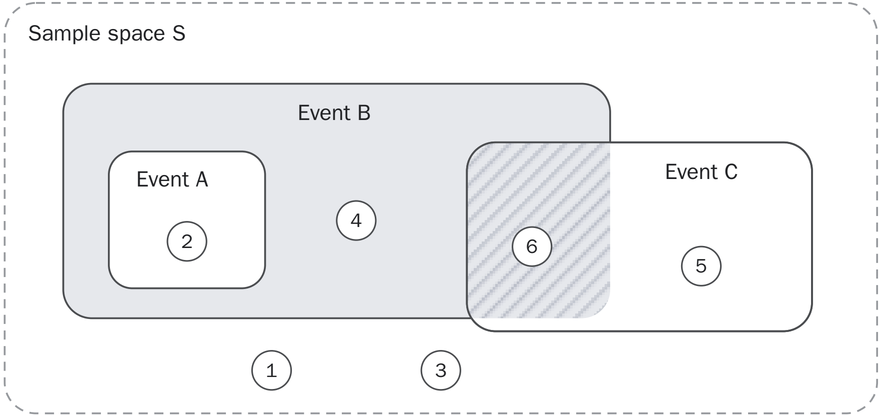
We roll a fair six-sided die. The possible outcomes are the numbers 1 through 6, so the sample space is

$$S = \{1, 2, 3, 4, 5, 6\}.$$

We are interested in the following events:

- $A = \{2\}$ — roll a 2
- $B = \{2, 4, 6\}$ — roll an even number
- $C = \{5, 6\}$ — roll a number greater than 4

Thus, an event may contain one, several, all, or none of the outcomes in S.



Sample Spaces & Events

- Example: Combining Two Experiments

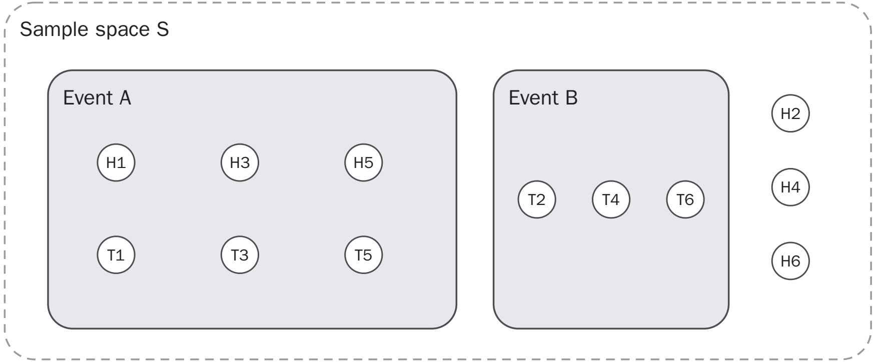
We flip a fair coin and then roll a fair six-sided die. Each outcome records both results, so the sample space is

$$S = \{H1, H2, \ldots, H6, T1, T2, \ldots, T6\}.$$

We are interested in the following events:

- $A = \{H1, H3, H5, T1, T3, T5\}$ — roll an odd number
- $B = \{T2, T4, T6\}$ — get tails and roll an even number

Since $A \cap B = \emptyset$, no outcome satisfies both conditions.
$\rightsquigarrow$ the events are mutually exclusive, or disjoint.



Exercise Round 1

- Sample Spaces & Events

Suppose we:

- draw one card numbered from 1 to 4
- then flip a fair coin

An outcome such as $3H$ records both results. We are interested in the following events:

- A — draw a number strictly greater than 2
- B — draw an odd number and get tails

1. List the sample space S.

2. Write the events A and B as sets.

3. Find $A \cap B$.

4. Are A and B mutually exclusive?

Theoretical Probability

- Basic Properties

A probability model assigns a number between 0 and 1 to each event. Every valid probability model satisfies the following properties.

Impossible and certain events

$$P(\emptyset) = 0, \qquad P(S) = 1.$$

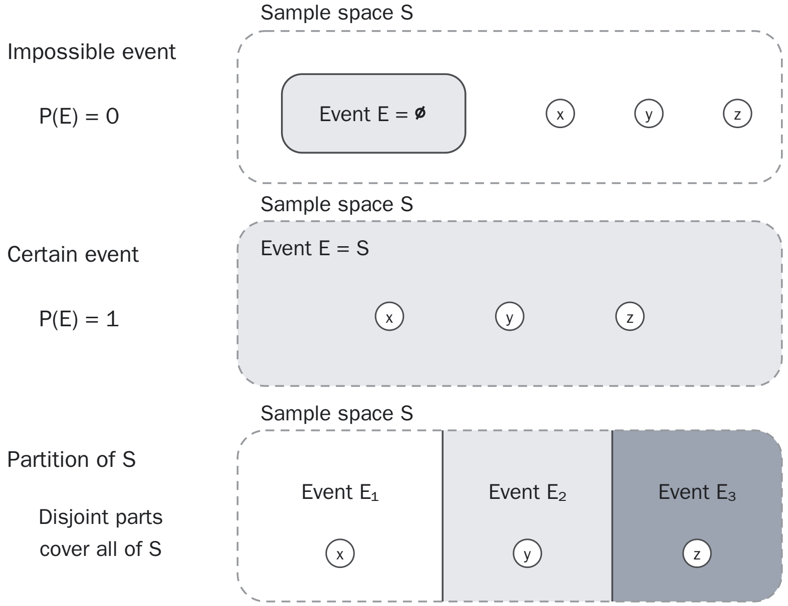
Every event satisfies $0 \le P(E) \le 1$.

Partition of the sample space

If mutually exclusive events $E_1, \ldots, E_k$ cover S, then

$$P(E_1) + \cdots + P(E_k) = 1.$$

Each outcome is counted exactly once.



Theoretical Probability

- Probability as a Function

A probability model assigns a number between 0 and 1 to each event. Every valid probability model satisfies the following properties.

Impossible and certain events

$$P(\emptyset) = 0, \qquad P(S) = 1.$$

Every event satisfies $0 \le P(E) \le 1$.

Partition of the sample space

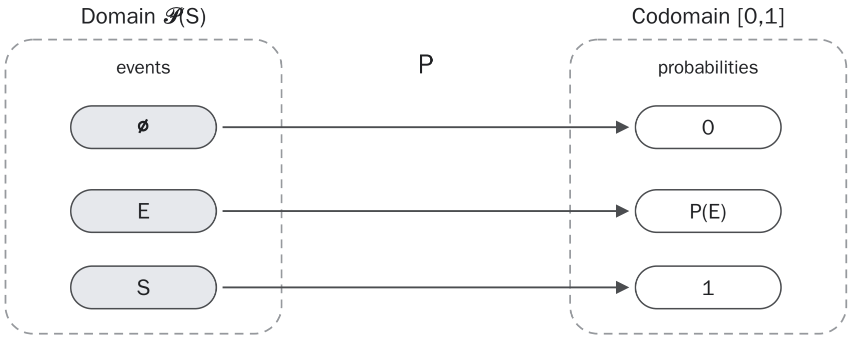
If mutually exclusive events $E_1, \ldots, E_k$ cover S, then

$$P(E_1) + \cdots + P(E_k) = 1.$$

Each outcome is counted exactly once.

The possible events form the power set $\mathcal{P}(S)$, the set of all subsets of S. A probability model is a function satisfying these properties:

$$P : \mathcal{P}(S) \to [0,1], \qquad E \longmapsto P(E).$$



Theoretical Probability

- Equally Likely Outcomes

A theoretical probability is based on a mathematical model. One such model assumes equally likely outcomes and can be described at two levels.

Individual outcomes

For a finite sample space S, one valid model treats the singleton events as an equally likely partition of S:

$$P(\{s\}) = \frac{1}{|S|}.$$

To extend this from individual outcomes to events, note that any event $E \subseteq S$ is the union of $|E|$ of these singleton events.

Events: direct-counting formula

Since the singleton events are mutually exclusive, their probabilities add:

$$P(E) = \sum_{s \in E} P(\{s\}) = \frac{|E|}{|S|}.$$

Thus, the formula divides:

- The number of outcomes in E (favorable outcomes)
- The total number of possible outcomes

The result $P(E)$ is the probability that event E occurs.

Theoretical Probability

- Model 1: A Fair Die

We roll a fair six-sided die. The possible outcomes are the numbers 1 through 6, so the sample space is

$$S = \{1, 2, 3, 4, 5, 6\}.$$

We are interested in the following events:

- $A = \{2\}$ — roll a 2
- $B = \{2, 4, 6\}$ — roll an even number
- $C = \{5, 6\}$ — roll a number greater than 4
- $\emptyset$ — roll a 7 (impossible)
- S — roll a number less than 7 (certain)

Applying the direct-counting formula, we get:

$$P(A) = \frac{|A|}{|S|} = \frac{1}{6},$$

$$P(B) = \frac{|B|}{|S|} = \frac{3}{6} = \frac{1}{2},$$

$$P(C) = \frac{|C|}{|S|} = \frac{2}{6} = \frac{1}{3}.$$

The impossible and certain events give the extreme probabilities:

$$P(\emptyset) = \frac{|\emptyset|}{|S|} = \frac{0}{6} = 0,$$

$$P(S) = \frac{|S|}{|S|} = \frac{6}{6} = 1.$$

Theoretical Probability

- Model 2: A Loaded Die

We roll a loaded six-sided die. The sample space is the same as before:

$$S = \{1, 2, 3, 4, 5, 6\}.$$

Unlike the fair-die model, this model assigns unequal probabilities:

$$\begin{aligned} P(\{1\}) &= 0.15 \\ &\vdots \\ P(\{5\}) &= 0.15, \\ P(\{6\}) &= 0.25. \end{aligned}$$

The values are nonnegative and the probabilities sum to 1, so they define a valid probability model.

Now, suppose we are interested in the event:

- $A = \{2, 4, 6\}$ — roll an even number

Its outcomes are not equally likely, so we add their assigned probabilities:

$$\begin{aligned} P(A) &= P(\{2\}) + P(\{4\}) + P(\{6\}) \\ &= 0.15 + 0.15 + 0.25 \\ &= 0.55. \end{aligned}$$

Applying the direct counting formula we get:

$$\frac{|A|}{|S|} = \frac{3}{6} = \frac{1}{2} \neq P(A) = 0.55.$$

Complements

- Definition

For an event $E \subseteq S$, the complement E' contains every outcome in S that is not in E:

$$E' = S \setminus E.$$

Every outcome belongs to exactly one of E and E'. Therefore, we must have that:

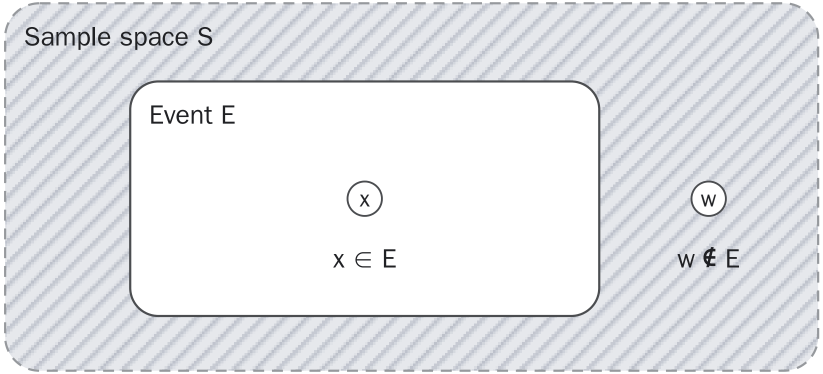
$$E \cap E' = \emptyset \qquad \text{and} \qquad E \cup E' = S.$$

That is, E and E' partition S, so:

$$P(E) + P(E') = 1$$

Rearranging gives the equivalent complement rule:

$$P(E) + P(E') = 1 \quad \Longleftrightarrow \quad P(E') = 1 - P(E).$$



Complements

- Example: Even and Odd Die Outcomes

We roll a fair six-sided die. The possible outcomes are the numbers 1 through 6, so the sample space is

$$S = \{1, 2, 3, 4, 5, 6\}.$$

We are interested in the following events:

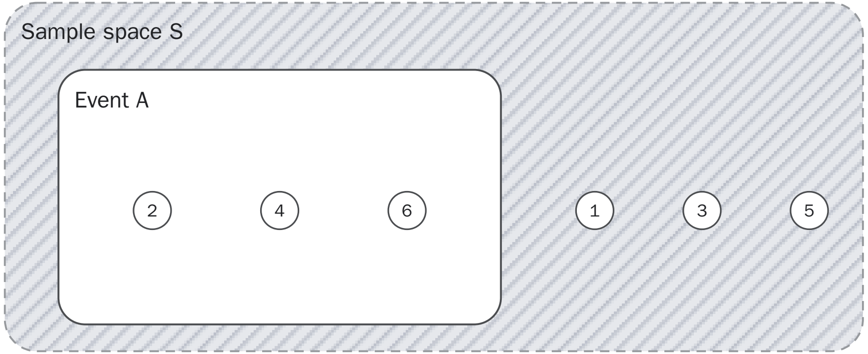
- $A = \{2, 4, 6\}$ — roll an even number
- $A' = S \setminus A = \{1, 3, 5\}$ — roll an odd number

Direct counting formula gives us the probability:

$$P(A') = \frac{|A'|}{|S|} = \frac{3}{6} = \frac{1}{2},$$

while the complement rule gives us the same result:

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



Exercise Round 2

- Theoretical Probability & Complements

We roll two fair six-sided dice and record the result as an ordered pair, such as $(2, 6)$.

We are interested in the following events:

- A — the sum is 8
- B — at least one die shows a 6
- C — the two dice show the same number

1. Describe the sample space S and state $|S|$.

2. Complete both parts:
(a) List the outcomes in A and C.
(b) Find $P(A)$ and $P(C)$.

3. Use the complement rule to find $P(B)$.

4. Find and interpret $P(A')$.

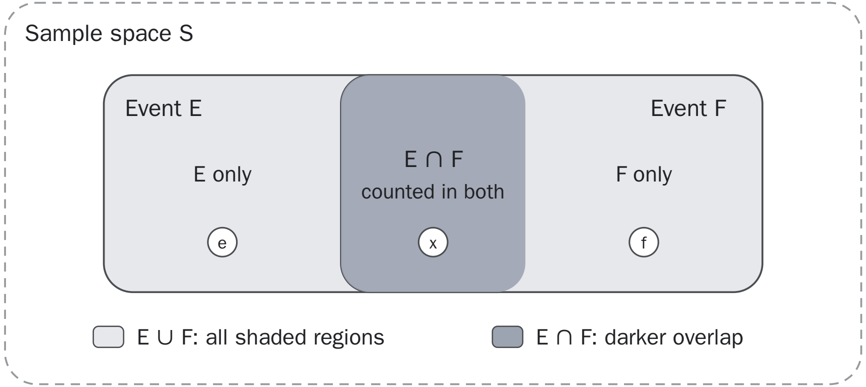
Combining Events

- Unions, Intersections & the Addition Rule

Probability questions often involve two conditions, represented by events E and F. An outcome may then belong to E only, F only, both events, or neither.

This leads to three related concepts:

- $E \cup F$ occurs when E, F, or both occur
- $E \cap F$ occurs when both occur
- E and F are mutually exclusive when $E \cap F = \emptyset$



Suppose we want the probability that E or F occurs:

$$P(E \cup F)$$

If we were to just add $P(E)$ and $P(F)$ then the event $E \cap F$ would be counted twice!

We can solve this by subtracting the overlap once:

$$P(E \cup F) = P(E) + P(F) - P(E \cap F).$$

For mutually exclusive events, the overlap is empty, so

$$P(E \cup F) = P(E) + P(F).$$

Combining Events

- Example: Union & Intersection

We roll a fair six-sided die. The possible outcomes are the numbers 1 through 6, so the sample space is

$$S = \{1, 2, 3, 4, 5, 6\}.$$

We are interested in the following events:

- $A = \{2, 4, 6\}$ — roll an even number
- $B = \{4, 5, 6\}$ — roll a number greater than or equal to 4

Their shared outcomes form the intersection:

$$A \cap B = \{4, 6\}.$$

All outcomes without repetition form the union:

$$A \cup B = \{2, 4, 5, 6\}.$$

Apply the addition rule: We subtract the intersection to remove the outcomes that would be counted twice:

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\ &= \frac{3}{6} + \frac{3}{6} - \frac{2}{6} = \frac{2}{3}. \end{aligned}$$

As a check, we can count the four outcomes in the union directly:

$$P(A \cup B) = \frac{|A \cup B|}{|S|} = \frac{4}{6} = \frac{2}{3}.$$

Conditional Probability

- Definition

Suppose we learn that an event F has occurred.
Outcomes outside F are no longer relevant.

If $P(F) > 0$, the probability that E also occurs is

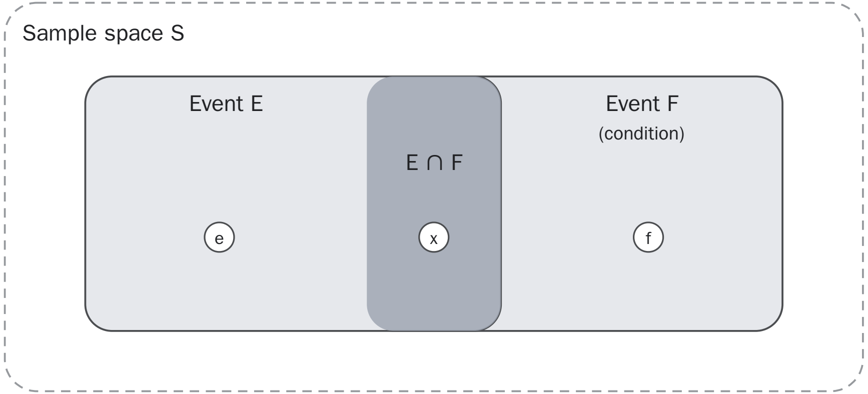
$$P(E \mid F) = \frac{P(E \cap F)}{P(F)}.$$

The denominator restricts attention to F, while the numerator retains the part that also belongs to E.

Rearranging the definition gives the multiplication rule

$$P(E \cap F) = P(E \mid F)P(F).$$

Interpretation: first reach F, then reach E within F.

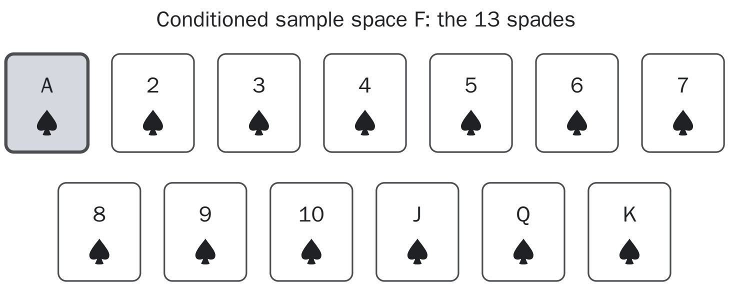


Conditional Probability

- Example: Drawing a Card

We draw one card from a well-shuffled standard 52-card deck. Suppose we know that the card is a spade and want to find the probability that it is an ace. Let

- E be "the card is an ace"
- F be "the card is a spade"



Suppose F occurred. Only the 13 spades remain, and exactly one is also an ace:

$$|F| = 13, \qquad E \cap F = \{A\spadesuit\}, \qquad |E \cap F| = 1.$$

Applying the direct counting formula within the restricted sample space F, we get:

$$P(E \mid F) = \frac{|E \cap F|}{|F|} = \frac{1}{13}.$$

Applying the conditional-probability formula in the full sample space gives the same result:

$$P(E \mid F) = \frac{P(E \cap F)}{P(F)} = \frac{1/52}{13/52} = \frac{1}{13}.$$

Independent Events

- Definition

Events E and F are independent if learning that one occurred does not change the probability of the other.

If $P(F) > 0$, this means

$$P(E \mid F) = P(E).$$

Similarly, if $P(E) > 0$, then $P(F \mid E) = P(F)$.

Using the multiplication rule,

$$\begin{aligned} P(E \cap F) &= P(E \mid F)P(F) \\ &= P(E)P(F). \end{aligned}$$

Thus, independence is equivalently tested by

$$P(E \cap F) = P(E)P(F).$$

Finally, do not confuse the terms:

Mutually exclusive events cannot occur together:

$$P(E \cap F) = 0.$$

Independent events do not affect one another's probabilities:

$$P(E \cap F) = P(E)P(F).$$

Independent Events

- Example: A Coin and a Die

We flip a fair coin and roll a fair six-sided die.

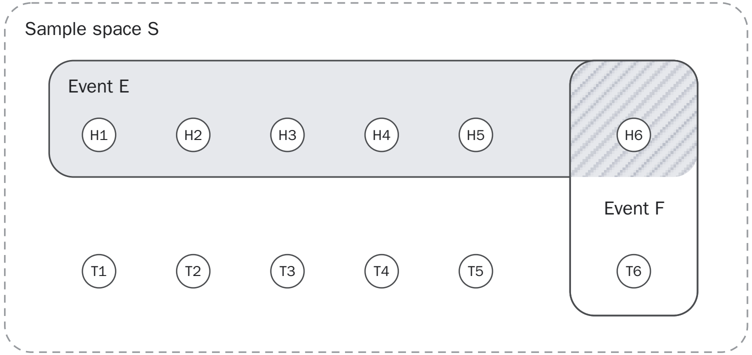
We are interested in the following events:

- $E = \{H1, H2, \ldots, H6\}$ — get heads
- $F = \{H6, T6\}$ — roll a 6

Together, the two stages give 12 equally likely outcomes. We test whether learning that one event occurred changes the probability of the other.

From the full sample space,

$$P(E) = \frac{6}{12} = \frac{1}{2},$$
$$P(F) = \frac{2}{12} = \frac{1}{6}.$$



The highlighted intersection contains only $H6$, so

$$P(E \cap F) = \frac{1}{12} = \frac{1}{2} \cdot \frac{1}{6} = P(E)P(F).$$

Because $P(E \cap F) = P(E)P(F)$, the events are independent: knowing one result does not change the probability of the other.

Exercise Round 3

- Combining, Conditioning & Independence

Suppose we:

- draw one card from a standard 52-card deck
- record its rank and suit

An outcome such as $Q\heartsuit$ records both features. We are interested in the following events:

- A — the card is red
- B — the card is a face card (a jack, queen, or king)

1. Find $P(A)$, $P(B)$ and $P(A \cap B)$.

2. Use the addition rule to find $P(A \cup B)$.

3. Find $P(A \mid B)$.

4. Decide whether A and B are:

(a) mutually exclusive
(b) independent

Empirical Probability

- From Models to Data

We learned earlier that:

- A theoretical probability is based on a mathematical model
- In the equally likely model, every individual outcome has the same probability

What is an empirical probability, then?

- An empirical probability is based on collected data
- It is the relative frequency with which the event occurs

Empirical Probability

- A Data-Based Estimate

The empirical probability of an event E is its observed relative frequency.

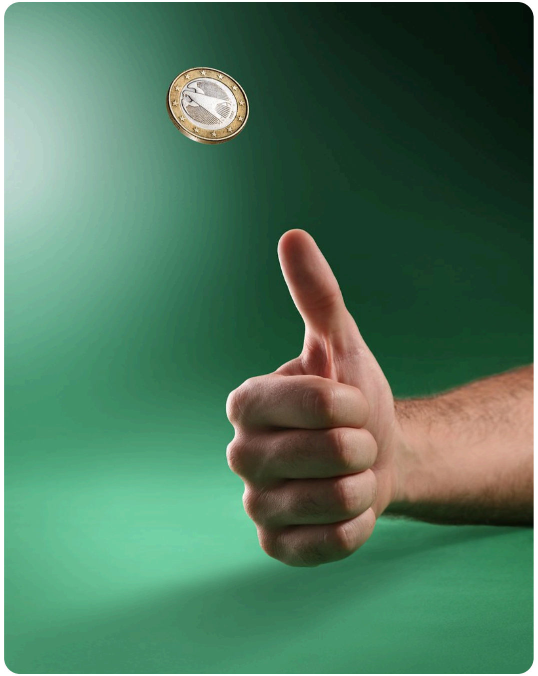
In particular, after n trials, if E occurs k times, then:

$$\widehat{P}(E) = \frac{k}{n}$$

The notation distinguishes two quantities:

- $P(E)$ is the probability specified by a model
- $\widehat{P}(E)$ is an estimate obtained from data

An empirical estimate can differ from the theoretical probability, especially when the number of trials is small.



Empirical Probability

- Example: Two Coin Tosses

We toss two fair coins and record the ordered pair of results. An outcome such as HT means the first coin shows heads and the second tails. The sample space is

$$S = \{HH, HT, TH, TT\}.$$

We are interested in the following event:

- $E = \{HT, TH\}$ — get exactly one head and one tail

In the equally likely model, the theoretical probability is

$$P(E) = \frac{|E|}{|S|} = \frac{2}{4} = \frac{1}{2} = 50\%.$$

Now suppose we repeat the two-coin experiment 10 times and observe E in 7 trials. The empirical probability is

$$\widehat{P}(E) = \frac{7}{10} = 70\%.$$

Here, the empirical estimate (70%) differs from the theoretical probability (50%).

The Law of Large Numbers

- Intuition

This leads us to the Law of Large Numbers.

To investigate the discrepancy from the previous slide, let us repeat the two-coin experiment.

Number of trials	Times E observed	Empirical probability $\widehat{P}(E)$
10	7	$\frac{7}{10} = 70\%$
20	13	$\frac{13}{20} = 65\%$
30	17	$\frac{17}{30} \approx 56.7\%$
40	22	$\frac{22}{40} = 55\%$
50	26	$\frac{26}{50} = 52\%$

As the number of trials increases, the empirical probability begins to move toward the theoretical probability $P(E) = \frac{1}{2}$.

The Law of Large Numbers

- Stabilizing Empirical Probabilities

The Law of Large Numbers explains the relationship between empirical and theoretical probability:

- The probability of an event is meaningful over many repeated trials, not just a few
- It is normal for an empirical probability based on a small number of trials to differ from the theoretical
- Over the long run, the empirical probability tends to converge to the theoretical probability