Brush-Up Maths for Data Science (2025)

Lecture Slides, Aug. 30th

Nicklas S. Andersen

University of Southern Denmark (SDU)

Department of Mathematics & Computer Science (IMADA)

Probability

In the most general sense:

- Probability provides us a way to measure the likelihood that something will happen

In this context, we are interested in:

1. A **random experiment**: An activity with a result that cannot be predicted ahead of time
2. An **outcome** $s \in S$: Any result from conducting an experiment
3. A **sample space** S: The set of all possible outcomes of a random experiment
4. An **event** $E \subseteq S$: A subset of the sample space and describes a collection of outcomes

We will now:

- Take a look at some examples
- Formally define what "probability" actually is

Probability

- Example 1

Consider the following:

1. The **random experiment**: We roll a (six-sided) die
2. An **outcome**: Only one outcome will occur (we can only roll a single number)
3. The **sample space**: There are six possible outcomes:

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

4. Possible **events** are:

   - "Rolling a 2" contains only one outcome:

$$E = \{2\} \subseteq S$$

   - "Rolling a number greater than 2" contains multiple outcomes:

$$E = \{3, 4, 5, 6\} \subseteq S$$

Probability

- Example 2

Consider the following:

1. The **random experiment**: Two coins are tossed at the same time
2. An **outcome**: We have the following possibilities
   - Both coins may land heads up (we write HH)
   - The first coin might land heads up and the second one tails up (we write HT)
   - … and so on
3. The **sample space**: The sample space for this experiment is

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

4. Let us consider the **event**: "Getting at least one heads". We can write the outcomes in this event as:

$$E = \{HH, HT, TH\} \subseteq S$$

Probability

- Definition

What is probability?

- Probability is a numerical measure of the likelihood that an event will occur
- It is expressed using the function notation P
- An **event** is typically denoted by a capital letter (e.g., E)
- The probability that event E occurs is written as:

$$P(E)$$

A theoretical probability is based on a mathematical model:

$$P(E) = \frac{\text{number of outcomes in event } E}{\text{number of outcomes in the sample space } S} = \frac{|E|}{|S|}.$$

In this model, each single outcome in the sample space has the same probability:

$$P(\{s\}) = \frac{1}{|S|} \quad \text{for each } s \in S.$$

Theoretical Probability

- Example

Consider the example we looked at earlier:

1. The **random experiment**: We roll a (six-sided) die
2. An **outcome**: Only one outcome will occur (we can only roll a single number)
3. The **sample space**: There are six possible outcomes:

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

4. Possible **events** and their probability are:

"Rolling a 2" contains only one outcome:

$$E_1 = \{2\} \subseteq S \;\Rightarrow\; P(E_1) = \frac{|E_1|}{|S|} = \frac{1}{6}$$

"Rolling an even number" contains outcomes:

$$E_2 = \{2, 4, 6\} \subseteq S \;\Rightarrow\; \frac{|E_2|}{|S|} = \frac{3}{6}$$

"Rolling a number greater than 4" contains outcomes:

$$E_3 = \{5, 6\} \subseteq S \;\Rightarrow\; P(E_3) = \frac{|E_3|}{|S|} = \frac{2}{6}$$

"Rolling a 7" contains no outcomes:

$$E_4 = \{\} = \emptyset \subseteq S \;\Rightarrow\; \frac{|E_2|}{|S|} = \frac{0}{6} = 0$$

"Rolling a number less than 7" contains outcomes:

$$E_4 = \{1, 2, 3, 4, 5, 6\} = \emptyset \subseteq S \;\Rightarrow\; \frac{|E_4|}{|S|} = \frac{6}{6} = 1$$

Probability Properties

These earlier examples illustrate some important properties:

- The number of outcomes in an event can never be less than 0:
  - The smallest probability is 0
  - If $P(E) = 0$, the event is impossible
- The number of outcomes in an event can never exceed the total outcomes in the sample space:
  - The largest probability is 1
  - If $P(E) = 1$, the event is certain

Key takeaway:

- The probability of any event E always falls between 0 and 1, inclusive:

$$0 \leq P(E) \leq 1$$

Complement

- Definition

Consider the random experiment of rolling a six-sided die. In this case we let:

- $S = \{1, 2, 3, 4, 5, 6\}$ be the sample space
- E be the event of rolling an even number:

$$E = \{2, 4, 6\}$$

- $\bar{E}$ be the event of rolling an odd number:

$$\bar{E} = \{1, 3, 5\}$$

The complement of an event E is defined as:

$$\bar{E} = S \setminus E$$

Note: E and $\bar{E}$ are mutually exclusive:

- They cannot happen at the same time
- Together cover the entire sample space

The sum of their probabilities is always:

$$P(E) + P(\bar{E}) = 1$$

The probability of the complement is:

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

Since $P(E) = \frac{|E|}{|S|} = \frac{3}{6}$:

$$P(\bar{E}) = 1 - \frac{3}{6} = \frac{3}{6} = \frac{1}{2}$$

Thus, the probability of rolling an odd number is:

$$P(\bar{E}) = \frac{1}{2}$$

Empirical Probability

- Definition

We learned earlier that:

- A theoretical probability is based on a mathematical model
- It is assumed that all outcomes are equally likely to occur

What is an empirical probability then?

- An empirical probability is based on collected data
- It is the relative frequency of the event occurring

Empirical Probability

- Example

Consider the example from earlier:

1. The **random experiment**: Two coins are tossed at the same time
2. An **outcome**: We have the following possibilities
   - Both coins may land heads up (we write HH)
   - The first coin might land heads up and the second tails up (we write HT)
   - ... and so on
3. The **sample space**: The sample space for this experiment is:

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

4. Suppose that we want to find the probability of getting 1 heads and 1 tails:

$$E = \{HT, TH\} \subseteq S$$

The theoretical probability is:

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

If we perform the experiment 10 times and observe HT or TH 7 times, the empirical probability is:

$$P_{\text{empirical}}(E) = \frac{7}{10} \approx 70\%$$

Here, we notice a discrepancy between the theoretical probability (50%) and the empirical probability (70%).

Law of Large Numbers

- Intuition

This leads us to the Law of Large Numbers.

To investigate the discrepancy, consider the results of tossing two coins multiple times:

Number of Trials	Times HT or TH Observed	Empirical Probability of E
10	7	$\frac{7}{10} \approx 70\%$
20	13	$\frac{13}{20} \approx 65\%$
30	17	$\frac{17}{30} \approx 56\%$
40	22	$\frac{22}{40} \approx 55\%$
50	26	$\frac{26}{50} \approx 52\%$

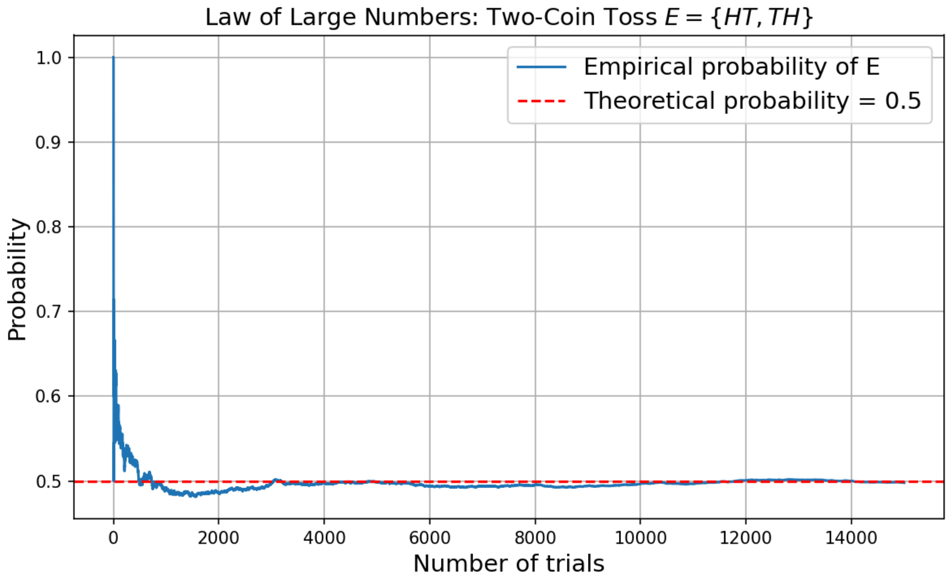
Notice that as the number of trials increases, the empirical probability begins to approach the theoretical probability of: $P(E) = \frac{1}{2}$.

Law of Large Numbers

- Definition

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

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



"As the number of trials increases, the empirical probability of an event approaches its theoretical probability."

Exercise Set

A coin is flipped, and a six-sided die is rolled:

1. List the sample space
2. Write the event: “roll an odd number”
3. Write the event: “get tails and roll an even number”

One card is drawn from a standard card deck. Find:

4. $P(\text{ace})$
5. $P(\text{red card})$

Two dice are rolled. Find:

6. $P(\text{sum} = 5)$
7. $P(\text{sum} > 10)$

One card is drawn from a standard 52-card deck:

8. Find the probability that the card is NOT a heart