Berner Fachhochschule
Haute école spécialisée bernoise
Bern University of Applied Sciences

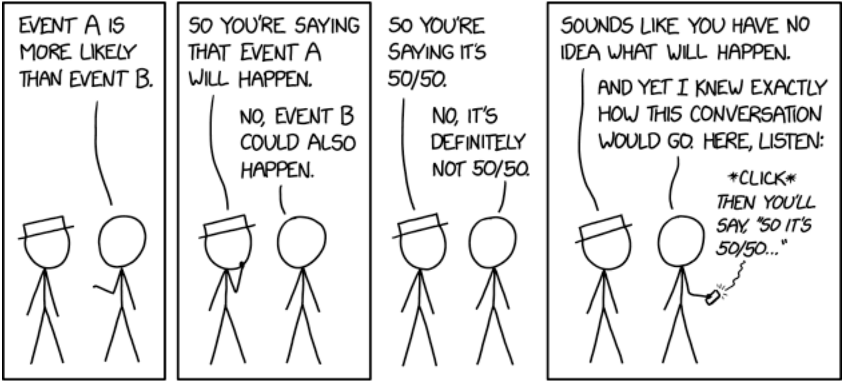
Applied Statistics

Introduction to Probability Theory

Release FS24

Why probability theory?

"The most important questions of life are, for the most part, really only problems of probability." – Pierre Simon de Laplace



(xkcd.com)

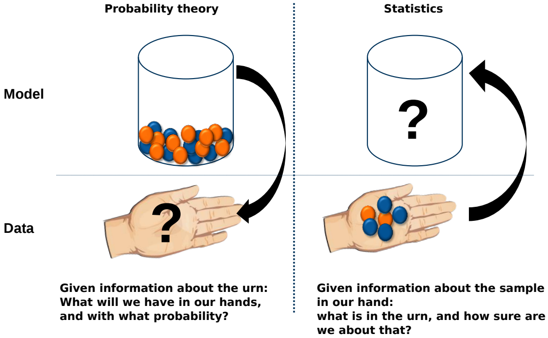
Why probability theory?

- Statistics is closely linked to probability theory
- Aim of probability theory: **modeling phenomena with uncertainty**
- Aim of statistics: performing **inference for probabilistic models**

Statistics and probability theory

- Probability theory:
  - modeling systems with uncertainty
  - predicting data generated by such a system
- Statistics:
  - collecting data
  - describing a model
  - inferring model parameters

Statistics and probability theory



(Source: Meier (2014))

Learning objectives

- Know the basic concepts of probability: event, sample space, probability measure
- Draw and read Venn diagrams
- Calculate a binomial coefficient and know its application

Probability theory

- Experiments are always "random": under "same conditions", we get different results
- Reasons for randomness:
  - Inherent randomness: e.g. the noise of a signal is by definition not deterministic
  - Incomplete control of experimental conditions: e.g. due to natural variation between individuals (even within individuals) $\Longrightarrow$ not fully controllable experimental environment
- Aim of probability theory: describe variability of the results

Important concepts

Definition (Elementary event, event, sample space, σ-algebra)

An **elementary event** ω is a possible outcome of an experiment.
The **sample space** Ω is the set of all elementary events.
An **event** A is a subset of the sample space ($A \subset \Omega$).
The σ-**algebra** $\mathcal{F}$ is the collection of all events considered.

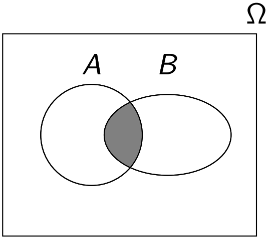
Examples:

1. Rolling a die. Sample space: all possible numbers we can get, hence $\Omega = \{1, 2, \ldots, 6\}$. Possible event "die shows an even number": $A = \{2, 4, 6\}$.
2. Sampling one person from this class. Elementary event: sampling one particular person, e.g. the lecturer. Example of an event: sampling a man.

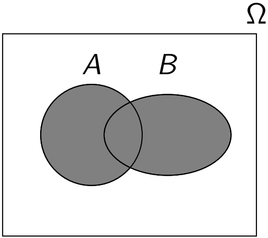
Venn diagram: visualization of events

Events and set operations visualized with **Venn diagrams**:

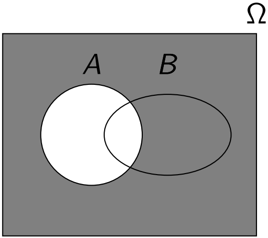
$A \cap B$:



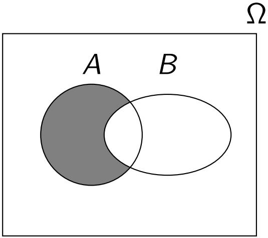
$A \cup B$:



A^c:



$A \setminus B$:



De Morgan's laws

Proposition (De Morgan's laws)

Let A and B be events. Then, $(A \cap B)^c = A^c \cup B^c$ *and* $(A \cup B)^c = A^c \cap B^c$.

Probability

Definition (Probability)

Let Ω be a sample space and $\mathcal{F}$ a σ-algebra. A **probability measure** is a function $\mathrm{P} : \mathcal{F} \to [0,1]$ that assigns a value between 0 and 1 to an event $A \subset \Omega$: $\mathrm{P}(A) \in [0,1]$.

It obeys the following properties (axioms of Kolmogorov):

i) $0 \leq \mathrm{P}(A) \leq 1$ for every event $A \subset \Omega$

ii) $\mathrm{P}(\Omega) = 1$

iii) $\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B)$ for *disjoint* (mutually exclusive) events A and B.

Probability

Examples (continued):

1. Rolling a die: for each number $\omega \in \{1, \ldots, 6\}$, the probability is $\mathrm{P}(\{\omega\}) = \frac{1}{6}$. The probability to get an even number is $\mathrm{P}(A) = \mathrm{P}(\{2\} \cup \{4\} \cup \{6\}) = \mathrm{P}(\{2\}) + \mathrm{P}(\{4\}) + \mathrm{P}(\{6\}) = \frac{1}{2}$.
2. Sampling one person from this class room (n people): the probability of sampling the lecturer is $\mathrm{P}(\{\omega\}) = \mathrm{P}(\{\mathsf{JW}\}) = \frac{1}{n}$. The probability of sampling a man depends on the number of men in this room.

Probability of unions

As a natural consequence of the axioms of Kolmogorov, we can calculate the probability of unions in general.

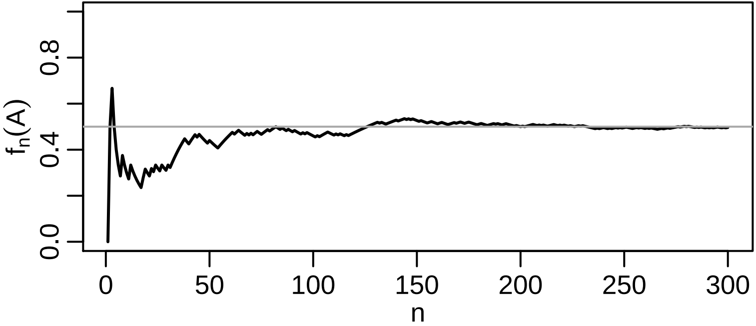
Proposition (Probability of unions)

Let A and B be events. Then, $\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cap B)$.
More general: let $A_1, A_2, \ldots, A_n$ be events. Then,

$$\mathrm{P}(A_1 \cup A_2 \cup \ldots \cup A_n) = \sum_{i_1=1}^{n} \mathrm{P}(A_{i_1}) - \sum_{i_1=1}^{n-1} \sum_{i_2=i_1+1}^{n} \mathrm{P}(A_{i_1} \cap A_{i_2}) + \sum_{i_1=1}^{n-2} \sum_{i_2=i_1+1}^{n-1} \sum_{i_3=i_2+1}^{n} \mathrm{P}(A_{i_1} \cap A_{i_2} \cap A_{i_3}) - \ldots$$

Interpretation of probabilities

Frequentist interpretation: $\mathrm{P}(A)$ is the relative frequency of event A in "infinitely many" experiments



Relative frequency of event $A =$ "head" after tossing a coin n times

Discrete probability spaces

- Assume finite (or countable) sample space: $\Omega = \{\omega_1, \omega_2, \dots\}$
- Probability of an event $A \subset \Omega : \mathrm{P}(A) = \sum_{i:\omega_i \in A} \mathrm{P}\left(\{\omega_i\}\right)$
- Normalization: $\mathrm{P}(\Omega) = \sum_{i \geq 1} \mathrm{P}\left(\{\omega_i\}\right) = 1$
- If Ω is *finite*, we often have $\mathrm{P}\left(\{\omega_i\}\right) = 1/|\Omega|$, i.e. every elementary event is equally probable. (Examples: rolling a die, sampling a person)

$$\Longrightarrow \mathrm{P}(A) = \frac{\#\text{favorable cases}}{\#\text{possible cases}} = \frac{|A|}{|\Omega|}$$

This is called a Laplace experiment.

Combinatorics

- Combinatorics is the *Art of Counting*
- **Binomial coefficient** counts number of possibilities to choose k out of n elements without replacement:

$$\binom{n}{k} := \frac{n!}{k!(n-k)!} = \frac{n \cdot (n-1) \cdot \ldots \cdot (n-k+1)}{k \cdot (k-1) \cdot \ldots \cdot 2 \cdot 1}$$

  - $\binom{n}{k} = \binom{n}{n-k} \quad (0 \leq k \leq n)$
  - $\binom{n}{0} = \binom{n}{n} = 1$
  - $\binom{n}{1} = \binom{n}{n-1} = n$

Example

- Probability to get 17 times head when flipping 50 times a coin:

$$\frac{|\#17\text{head}|}{|\Omega|} = \frac{\binom{50}{17}}{2^{50}} \approx 0.0087$$

- Probability that at least 2 out of 23 people have the same birthday: *What do you guess?*
  $\hookrightarrow$ see Problem 2.13?

References

Lukas Meier. Statistik und Wahrscheinlichkeitsrechnung. *Lecture notes*, 2014.