Overview of Probability

Probability and Statistics for Data Science

Carlos Fernandez-Granda

These slides are based on the book Probability and Statistics for Data Science by Carlos Fernandez-Granda, available for purchase here. A free preprint, videos, code, slides and solutions to exercises are available at https://www.ps4ds.net

Motivation

Describing uncertain phenomena

Who will win the next presidential election?

How much will a certain stock cost tomorrow?

Will the New York Knicks win the next NBA championship?

Strategy

Interpret uncertain phenomenon as an experiment that can be repeated over and over

$$\mathrm{P}\left(\text{Knicks win}\right) = \frac{\text{times Knicks win}}{\text{total repetitions}}$$

Plan

- Probability spaces
- Conditional probability
- Estimating probabilities from data
- Independence
- Conditional independence
- The Monte Carlo method

Probability space

1. Model phenomenon of interest as experiment with mutually exclusive outcomes

2. Group outcomes in sets called events

3. Assign a probability to each event

Six-sided die

Set of outcomes (sample space) $\Omega = \{1, 2, 3, 4, 5, 6\}$

Examples of events:

$A := \{1, 3, 5\}$

$B := \{4\}$

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

Probability measure

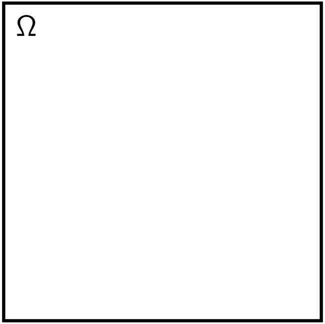
Assigns probability to each event

Intuitive definition: If we repeat the experiment many times

$$\mathrm{P}(\text{event}) = \frac{\text{number of times event occurs}}{\text{total repetitions}}$$

Rule 1

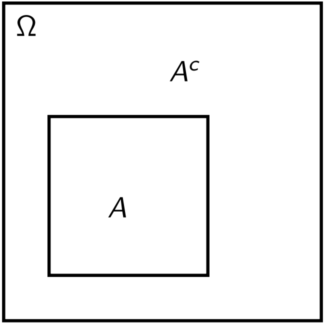
We must assign a probability to the whole sample space Ω



$$\mathrm{P}(\text{sample space}) = 1$$

Rule 2

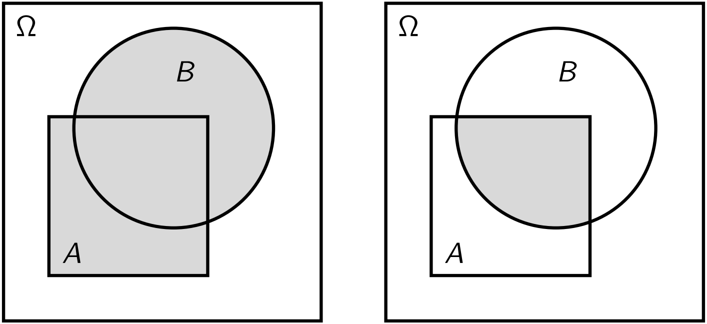
If we assign a probability to A then we need to assign a probability to A^c



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

Rule 3

If we assign a probability to A and B we need to assign a probability to $A \cup B$ and $A \cap B$



If A and B are disjoint $\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B)$

Important

Probability spaces might look very abstract, but they just implement our intuitive definition of probability

Conditional probability

Imagine that Knicks are first seed in playoffs

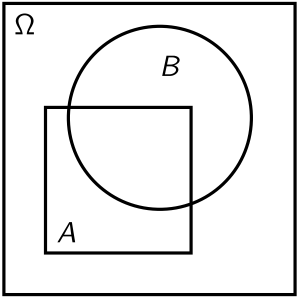
We need to update the probability that they win

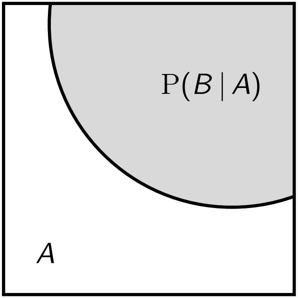
$$\mathrm{P}\left(\text{Knicks win} \,|\, \text{Knicks are 1st seed}\right) = \frac{\text{times Knicks are 1st seed and win}}{\text{times Knicks are 1st seed}}$$

Conditional probability

The conditional probability of an event B given A is

$$\mathrm{P}(B \mid A) := \frac{\mathrm{P}(B \cap A)}{\mathrm{P}(A)}$$



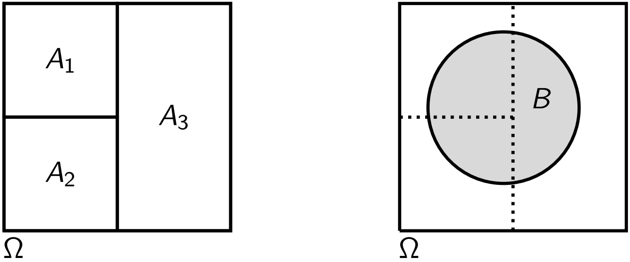


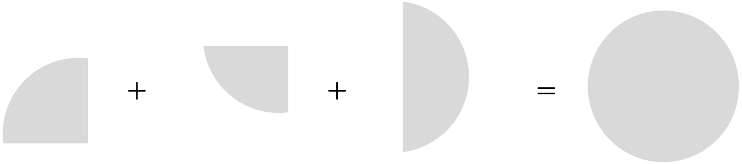
Chain rule

From definition of conditional probability $\mathrm{P}(B \mid A) := \frac{\mathrm{P}(B \cap A)}{\mathrm{P}(A)}$

$$\mathrm{P}(A \cap B) = \mathrm{P}(A)\,\mathrm{P}(B|A)$$

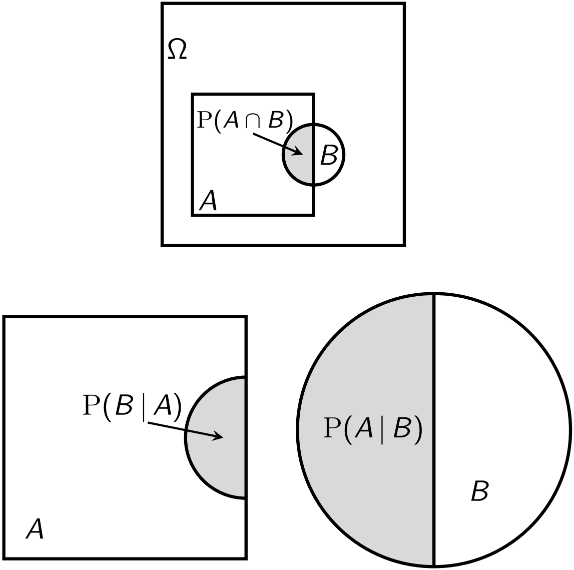
Law of Total Probability





$$\mathrm{P}(A_1 \cap B) \quad + \quad \mathrm{P}(A_2 \cap B) \quad + \quad \mathrm{P}(A_3 \cap B) \quad = \quad \mathrm{P}(B)$$

$\mathrm{P}(A|B) \neq \mathrm{P}(B|A)$



Bayes' Rule

For any events A and B

$$\mathrm{P}(A \mid B) = \frac{\mathrm{P}(A)\,\mathrm{P}(B \mid A)}{\mathrm{P}(B)}$$

as long as $\mathrm{P}(B) > 0$

Great, but in practice

How do we estimate estimate probabilities from data?

Intuitive definition of probability

$$\mathrm{P}(\text{event}) = \frac{\text{number of times event occurs}}{\text{total repetitions}}$$

Six-sided die

Data collection: We roll the die 60 times and observe 8 twos

Probability of event *Rolling a two*?

Empirical probability

Let A be an event and $X := \{x_1, x_2, \ldots, x_n\}$ a dataset

The empirical probability of A is

$$\mathrm{P}_X(A) := \frac{\sum_{i=1}^{n} 1_{x_i \in A}}{n}$$

where $1_{x_i \in A}$ is one if $x_i \in A$ and zero otherwise

Conditional probability

Data collection: We roll the die 60 times and observe 8 twos, 6 fours and 10 sixes

Conditional probability of event *Rolling a two* given *Roll is even*?

Empirical conditional probability

Let A and B be events, and $X := \{x_1, x_2, \ldots, x_n\}$ a dataset

The empirical conditional probability of B given A is

$$\mathrm{P}_X(B \,|\, A) := \frac{\sum_{i=1}^n 1_{x_i \in A \cap B}}{\sum_{i=1}^n 1_{x_i \in A}}$$

where $1_{x_i \in S}$ is one if $x_i \in S$ and zero otherwise

Independence of two events

Two events A, B are independent if

$$\mathrm{P}(B \mid A) = \mathrm{P}(B)$$

or equivalently

$$\mathrm{P}(A \cap B) = \mathrm{P}(A)\,\mathrm{P}(B \mid A) = \mathrm{P}(A)\,\mathrm{P}(B)$$

Multiple events

If A, B and C are pairwise independent, then

$$\mathrm{P}\left(C \mid A \cap B\right) = \mathrm{P}\left(C\right)?$$

Independence of multiple events

The events $A_1, A_2, \ldots, A_n \in \mathcal{F}$ are mutually independent
if and only if for any $\{i_1, i_2, \ldots, i_m\} \subseteq \{1, 2, \ldots, n\}$

$$\mathrm{P}\left(\cap_{j=1}^{m} A_{i_j}\right) = \prod_{j=1}^{m} \mathrm{P}\left(A_{i_j}\right)$$

Conditional independence

A, B are conditionally independent given C if

$$\mathrm{P}(A \mid B, C) = \mathrm{P}(A \mid C)$$

or equivalently

$$\mathrm{P}(A \cap B \mid C) = \mathrm{P}(A \mid C)\,\mathrm{P}(B \mid C)$$

Independence and conditional independence

Does independence imply conditional independence? No!

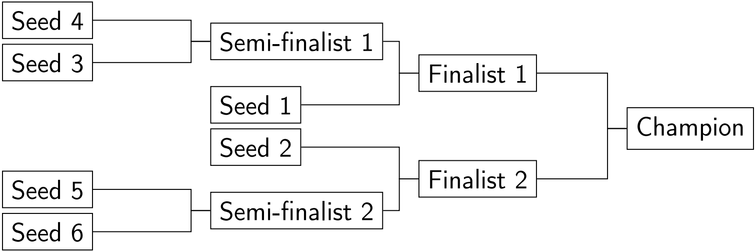
Does conditional independence imply independence? No!

Life is not a homework problem

In practice, we often cannot compute probabilities exactly, even when we have all the necessary information!

Tournament

Group stage followed by bracket



Intuitive definition of probability

$$\mathrm{P}(\text{event}) = \frac{\text{number of times event occurs}}{\text{total repetitions}}$$

Monte Carlo method

To approximate the probability of an event A, we

1. Generate n simulated outcomes: $s_1, s_2, \ldots, s_n$

2. Compute the fraction of the outcomes in A,

$$\mathrm{P}_{\mathsf{MC}}(A) := \frac{\sum_{i=1}^{n} 1_{s_i \in A}}{n}$$

   where $1_{x_i \in A}$ is one if $s_i \in A$ and zero otherwise

Summary

- Probability spaces
- Conditional probability
- Estimating probabilities from data
- Independence
- Conditional independence
- The Monte Carlo method