Conditional 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

Plan

Define conditional probability

Describe three fundamental results:

- The chain rule
- The law of total probability
- Bayes' rule

Motivation: Flights and rain

How likely is a flight delay if it rains?

Events of interest: L (flight is late), R (it rains)

From past data:

$$\mathrm{P}(L \cap R^c) = \frac{2}{20} \quad \mathrm{P}(L^c \cap R^c) = \frac{14}{20}$$

$$\mathrm{P}(L \cap R) = \frac{3}{20} \quad \mathrm{P}(L^c \cap R) = \frac{1}{20}$$

Probability of flight being late

Intuitively,

$$\mathrm{P}(L) = \frac{\text{times airplane is late}}{\text{total repetitions}}$$

$L = (L \cap R^c) \cup (L \cap R)$, so

$$\begin{aligned} \mathrm{P}(L) &= \mathrm{P}(L \cap R^c) + \mathrm{P}(L \cap R) \\ &= \frac{1}{4} \end{aligned}$$

but we want probability of flight being late if it rains

$$\mathrm{P}(L \cap R^c) = \frac{2}{20} \quad \mathrm{P}(L^c \cap R^c) = \frac{14}{20} \quad \mathrm{P}(L \cap R) = \frac{3}{20} \quad \mathrm{P}(L^c \cap R) = \frac{1}{20}$$

Intuitive definition

$$\begin{aligned}
\mathrm{P}(L \mid R) &= \frac{\text{times airplane is late and it rains}}{\text{times it rains}} \\
&= \frac{\text{times airplane is late and it rains}}{\text{total repetitions}} \cdot \frac{\text{total repetitions}}{\text{times it rains}} \\
&= \frac{\mathrm{P}(L \cap R)}{\mathrm{P}(R)}
\end{aligned}$$

Flights and rain

$$\mathrm{P}(L \mid R) = \frac{\mathrm{P}(L \cap R)}{\mathrm{P}(R)}$$
$$= \frac{\mathrm{P}(L \cap R)}{\mathrm{P}(L \cap R) + \mathrm{P}(L^c \cap R)}$$
$$= \frac{3}{4}$$

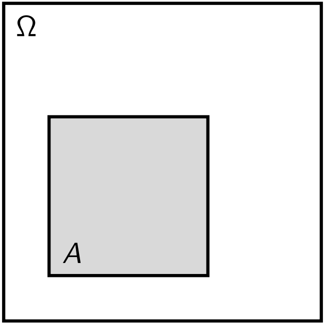
Three times larger than $\mathrm{P}(L)$!

$$\mathrm{P}\left(L \cap R^c\right) = \frac{2}{20} \quad \mathrm{P}\left(L^c \cap R^c\right) = \frac{14}{20} \quad \mathrm{P}\left(L \cap R\right) = \frac{3}{20} \quad \mathrm{P}\left(L^c \cap R\right) = \frac{1}{20}$$

Updating the probability space

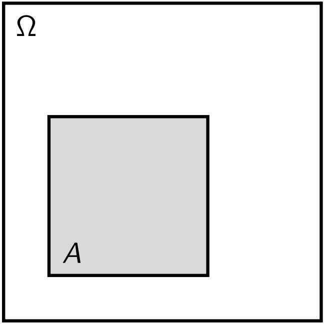
Let $(\Omega, \mathcal{C}, \mathrm{P})$ be a probability space

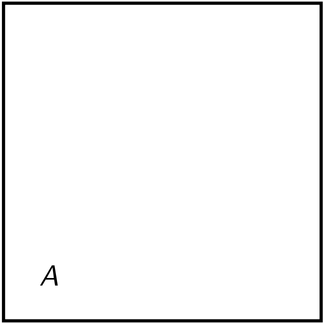
We find out that the outcome is in A



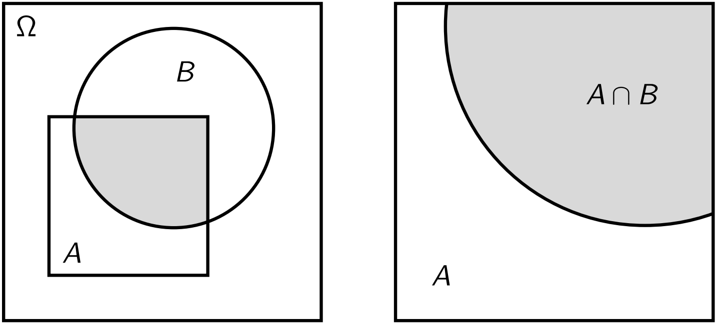
How do we update the probability space?

New sample space?





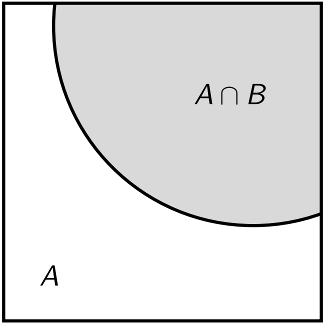
New collection of events?



Replace each event B by $A \cap B$

New probability measure?

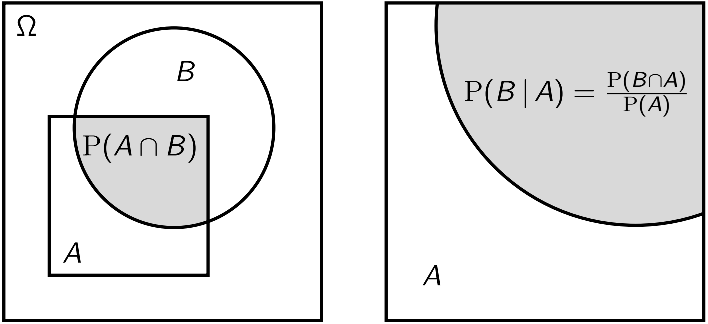
Replace $P(B)$ by $\mathrm{P}(A \cap B)$?



Probability of the whole sample space? $\mathrm{P}(A \cap A) = \mathrm{P}(A) \neq 1$

Normalize by $1/\mathrm{P}(A)$!

Conditional probability given A



Conditional probability

The conditional probability of an event $B \in \mathcal{C}$ given A is

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

$\mathrm{P}(B)$ is the prior probability

$\mathrm{P}(B \mid A)$ is the posterior probability

Notation

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

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

Chain rule

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

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

Chain rule

For any three events A, B, C

$$\begin{aligned} \mathrm{P}(A, B, C) &= \mathrm{P}(A)\, P(B, C \mid A) \\ &= \mathrm{P}(A)\, P(B \mid A)\, P(C \mid A, B) \end{aligned}$$

$$\begin{aligned} \mathrm{P}(C \mid A, B) &= \frac{\mathrm{P}(A, B, C)}{\mathrm{P}(A, B)} \\ &= \frac{\mathrm{P}(B, C \mid A)}{\mathrm{P}(B \mid A)} \end{aligned}$$

Order is completely arbitrary

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

Chain rule

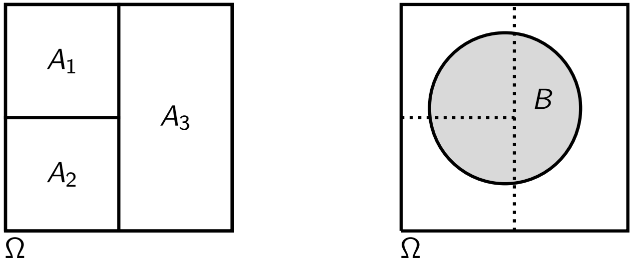
For any sequence of events $S_1, S_2, S_3, \ldots$

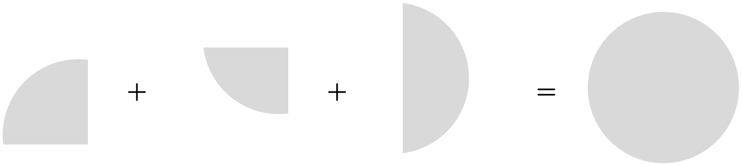
$$\begin{aligned} \mathrm{P}\left(\cap_i S_i\right) &= \mathrm{P}\left(S_1\right)\mathrm{P}\left(S_2|S_1\right)\mathrm{P}\left(S_3|S_1,S_2\right)\ldots \\ &= \prod_i \mathrm{P}\left(S_i|\cap_{j=1}^{i-1} S_j\right) \end{aligned}$$

Order is completely arbitrary

Important to choose order wisely!

Law of Total Probability





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

Law of Total Probability

$A_1, A_2, \ldots \in \mathcal{C}$ is a partition of Ω if

- A_i and A_j are disjoint if $i \neq j$
- $\Omega = \cup_i A_i$

For any event $S \in \mathcal{C}$

$$\begin{aligned}
\mathrm{P}(S) &= \sum_i \mathrm{P}(\cup_i (A_i \cap S)) \\
&= \sum_i \mathrm{P}(A_i, S) \\
&= \sum_i \mathrm{P}(A_i)\,\mathrm{P}(S|A_i)
\end{aligned}$$

Flights and rain

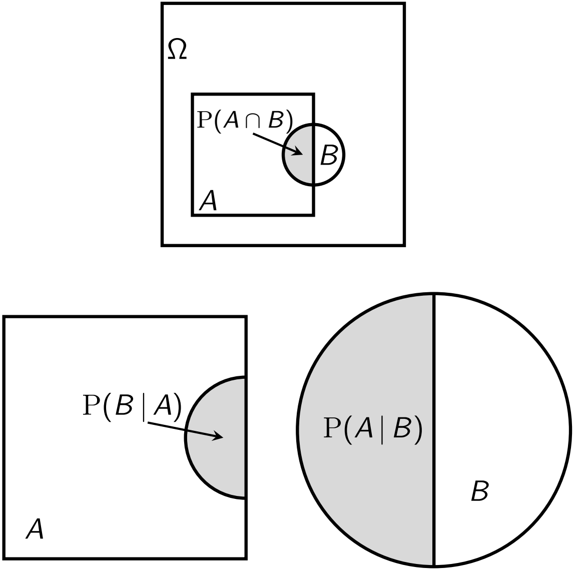
$$\mathrm{P}(R) = 0.2 \quad \mathrm{P}(L|R) = 0.75 \quad \mathrm{P}(L|R^c) = 0.125$$

$$\begin{aligned}
\mathrm{P}(L) &= \mathrm{P}(R, L) + \mathrm{P}(R^c, L) \\
&= \mathrm{P}(R)\,\mathrm{P}(L|R) + \mathrm{P}(R^c)\,\mathrm{P}(L|R^c) \\
&= 0.2 \cdot 0.75 + 0.8 \cdot 0.125 = 0.25
\end{aligned}$$

Important!

$$\mathrm{P}\left(A|B\right) = \frac{\mathrm{P}(A, B)}{\mathrm{P}(B)} \neq \frac{\mathrm{P}(A, B)}{\mathrm{P}(A)} = \mathrm{P}\left(B|A\right)$$

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



Bayes' Rule

For any events A and B in a probability space $(\Omega, \mathcal{C}, \mathrm{P})$

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

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

Flights and rain

$$\mathrm{P}(R) = 0.2 \quad \mathrm{P}(L|R) = 0.75 \quad \mathrm{P}(L|R^c) = 0.125$$

$$\begin{aligned}
\mathrm{P}(R|L) &= \frac{\mathrm{P}(R,L)}{\mathrm{P}(L)} \\
&= \frac{\mathrm{P}(L|R)\,\mathrm{P}(R)}{\mathrm{P}(L,R) + \mathrm{P}(L,R^c)} \\
&= \frac{\mathrm{P}(L|R)\,\mathrm{P}(R)}{\mathrm{P}(L|R)\,\mathrm{P}(R) + \mathrm{P}(L|R^c)\,\mathrm{P}(R^c)} \\
&= \frac{0.75 \cdot 0.2}{0.75 \cdot 0.2 + 0.125 \cdot 0.8} = 0.6
\end{aligned}$$

What have we learned?

Intuitive and formal definition of conditional probability

Three fundamental results:

- The chain rule
- The law of total probability
- Bayes' rule