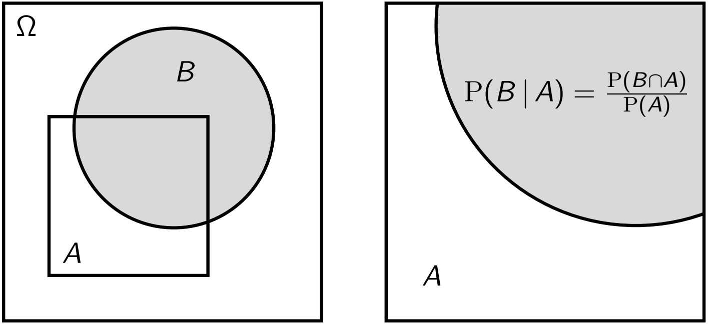
Independence

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

Conditional probability given an event A



Independence of two events

Two events A, B are independent if

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

or equivalently

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

House of Representatives 1984

		Duty-free exports	
		Yes	No
Budget	Yes	151	88
	No	21	140

$$\mathrm{P}(D) = \frac{172}{400} = 0.43$$

$$\mathrm{P}(D \mid B) = \frac{151}{239} = 0.632$$

House of Representatives 1984

		Immigration	
		Yes	No
Anti-satellite test ban	Yes	124	113
	No	89	93

$$\mathrm{P}(A) = \frac{237}{419} = 0.566 \quad \approx \quad \mathrm{P}(A \mid I) = \frac{124}{213} = 0.582$$

$$\mathrm{P}(I) = \frac{213}{419} = 0.508$$

$$\mathrm{P}(A, I) = \frac{124}{419} = 0.296 \quad \approx \quad 0.288 = \mathrm{P}(A)\mathrm{P}(I)$$

Brady and Hurricanes

Events of interest:

Tom Brady wins Super Bowl (T)

Category 5 hurricane in the North Atlantic Ocean (H)

Year	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21
Brady wins	✓	✗	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	✓	✗	✓	✗	✓
Hurricane	✗	✓	✓	✓	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✓	✓	✓	✓	✗	✗

$$\mathrm{P}(H) = \frac{8}{20} = 0.4$$

$$\mathrm{P}(H \mid T) = \frac{4}{7} = 0.571$$

Brady and Hurricanes

Events of interest:

Tom Brady wins Super Bowl (T)

Category 5 hurricane in the North Atlantic Ocean (H)

Year	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21
Brady wins	✓	✗	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	✓	✗	✓	✗	✓
Hurricane	✗	✓	✓	✓	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✓	✓	✓	✓	✗	✗

$$\mathrm{P}(H) = \frac{8}{20} = 0.4$$

$$\mathrm{P}(H \mid T) = \frac{4}{7} = 0.571$$

Brady and Hurricanes

What if Brady had won in 2011 and lost in 2016?

Year	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21
Brady wins	✓	✗	✓	✓	✗	✗	✗	✗	✗	✗	✓	✗	✗	✓	✗	✗	✗	✓	✗	✓
Hurricane	✗	✓	✓	✓	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✓	✓	✓	✓	✗	✗

$$\mathrm{P}(H) = \frac{8}{20} = 0.4$$

$$\mathrm{P}(H \mid T) = \frac{3}{7} = 0.43$$

Multiple events

If A, B and C are pairwise independent, then

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

Two coin flips

Probability space for two fair coin flips

Outcomes: *heads-heads*, *heads-tails*, *tails-heads*, *tails-tails*

Probability measure:

$$\mathrm{P}(\{\textsf{h-h}\}) = \mathrm{P}(\{\textsf{h-t}\}) = \mathrm{P}(\{\textsf{t-h}\}) = \mathrm{P}(\{\textsf{t-t}\}) = \frac{1}{4}$$

Events of interest:

$$\begin{aligned} A &:= \{\textsf{h-h}, \textsf{h-h}\} && \text{(first flip is heads)} \\ B &:= \{\textsf{h-h}, \textsf{t-h}\} && \text{(second flip is heads)} \\ C &:= \{\textsf{h-h}, \textsf{t-h}\} && \text{(flips are the same)} \end{aligned}$$

Two coin flips

$$\mathrm{P}(A) = \mathrm{P}(\{\text{h-h}\} \cup \{\text{h-t}\}) = \frac{1}{2}$$
$$\mathrm{P}(B) = \mathrm{P}(\{\text{h-h}\} \cup \{\text{t-h}\}) = \frac{1}{2}$$
$$\mathrm{P}(C) = \mathrm{P}(\{\text{h-h}\} \cup \{\text{t-t}\}) = \frac{1}{2}$$

$$\mathrm{P}(A, B) = \mathrm{P}(\{\text{h-h}\}) = \frac{1}{4} = \mathrm{P}(A)\mathrm{P}(B)$$
$$\mathrm{P}(A, C) = \mathrm{P}(\{\text{h-h}\}) = \frac{1}{4} = \mathrm{P}(A)\mathrm{P}(C)$$
$$\mathrm{P}(B, C) = \mathrm{P}(\{\text{h-h}\}) = \frac{1}{4} = \mathrm{P}(B)\mathrm{P}(C)$$

Two coin flips

A, B and C are pairwise independent

If the three events are truly independent we should have
$\mathrm{P}(C \mid A, B) = \mathrm{P}(C)$

$$
\begin{aligned}
\mathrm{P}(C \mid A, B) &= \frac{\mathrm{P}(A, B, C)}{\mathrm{P}(A, B)} \\
&= \frac{\mathrm{P}(\{\text{h-h}\})}{\mathrm{P}(\{\text{h-h}\})} \\
&= 1 \quad \neq \quad \frac{1}{2} = \mathrm{P}(C)
\end{aligned}
$$

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 $\{i_1, i_2, \ldots, i_m\} \subseteq \{1, 2, \ldots, n\}$

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

Then any conditional probability of A_i equals $\mathrm{P}(A_i)$

$$\begin{aligned}\mathrm{P}(A_3 \,|\, A_1, A_2) &= \frac{\mathrm{P}(A_1, A_2, A_3)}{\mathrm{P}(A_1, A_2)} \\ &= \frac{\mathrm{P}(A_1)\mathrm{P}(A_2)\mathrm{P}(A_3)}{\mathrm{P}(A_1)\mathrm{P}(A_2)} \\ &= \mathrm{P}(A_3)\end{aligned}$$

What have we learned?

Definition of independence

Pairwise independence does not imply mutual independence