Lecture 3 - Conditional Probability

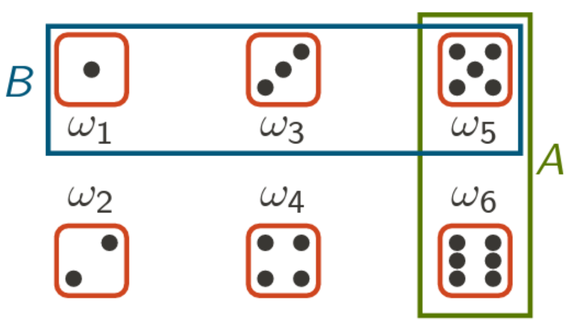
Introduction

- We often ask for a probability of the occurrence of some event A under condition that another event B occurred.
  - What is the chance of rolling at least 5 on a single die (A) if we know the outcome is odd (B)?
  - We choose a random word from an English book. What is the chance that the second letter of the word is 'h' (A) if the first letter is 't' (B)?
  - What is the chance that a randomly select student got a grade 5 from mathematics (A) if he or she received grade 2 from physics (B)?
- Such probability is called the conditional probability and is denoted by:

$$P(A|B)$$

Example

- What is the probability of getting at least 5 in a single roll of a die (A) if the outcome is odd (B)?



$$A = \{\omega_5, \omega_6\}$$
$$B = \{\omega_1, \omega_3, \omega_5\}$$
$$A \cap B = \{\omega_5\}$$

- If B occurred, the outcomes which could occur are only $\{\omega_1, \omega_2, \omega_3\}$ (the sample space Ω reduced to B) Among those left events, only $A \cap B = \{\omega_5\}$

support event A

$$P(A|B) = \frac{|A \cap B|}{|B|} = \frac{1}{3} = \frac{P(A \cap B)}{P(B)}$$

- The conditional probability of event A given event B, for which $P(B) > 0$, is a number defined as:

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

- Remark: The measure $P(A|B)$ as a function of A with B fixed satisfies the Kolmogorov axioms, therefore has all the properties of the probability:
- $P(A|B) = 1 - P(A'|B)$
- $P(A_1 \cup A_2|B) = P(A_1|B) + P(A_2|B) - P(A_1 \cap A_2|B)$

The chain rule

$$P(A_1 \cap A_2) = P(A_1)P(A_2|A_1)$$

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

- If events $A_1, A_2, \ldots, A_n$ satisfy:

$$P(A_1 \cap A_2 \cap \ldots \cap A_{n-1}) > 0$$

- then:

$$P(A_1 \cap A_2 \cap \ldots \cap A_n) = P(A_1)P(A_2|A_1) \cdot \ldots \cdot P(A_n|A_1 \cap A_2 \cap \ldots \cap A_{n-1})$$

Partition of the sample space

- We say that events $A_1, A_2, \ldots, A_n \subseteq \Omega$ form a partition of Ω, if

1. They are disjoint: $A_i \cap A_j = \emptyset$ for $i \neq j$
2. Their sum cover the sample space: $A_1 \cup A_2 \cup \ldots \cup A_n = \Omega$

The law of total probability

- Suppose that the events $A_1, \ldots, A_n$ form a partition. Then, for any event B:

$$P(B) = \sum_{i=1}^{n} P(A_i)P(B|A_i)$$

Simpson's paradox

- In two cities A and B, data concerning burglaries to houses and flats were collected:

City	% of burglaries in flats	% of burglaries in houses	# of flats	# of houses
A	2%	6%	1000	1000
B	3%	7%	1800	200

- In which city the chances of having burglary in a random apartment is higher?
- By looking at only the percentages of burglaries, most people would say that the city B is more dangerous, but actually, when computing the probability of burglary in both cities, given the number of flats and houses, the city A has slightly higher chance of getting robbed.

Bayes' rule

- If $A_1, \ldots, A_n$ form a partition with $P(A_i) > 0$ for all $i = 1, \ldots, n$ and B is an event with $P(B) > 0$, then:

$$P(A_1|B) = \frac{P(B|A_i)P(A_i)}{P(B)} = \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^{n} P(B|A_j)P(A_j)}$$