Lecture 5: §1.4. Conditional probability: it's the probability that an event A happens, given that you know that another has happened.

- Example: experiment: we toss a coin 3 times.

\[ \Omega = \{hhh, hht, hth, htt, thh, tht, tth, ttt\} , \quad \#(\Omega) = 8. \]

We bet on: \( A = \text{"we get at least 2 heads"} = \{hhh, hht, hth, thh\} \),

so \( \#(A) = 4 \) and \( P(A) = \frac{\#(A)}{\#(\Omega)} = \frac{4}{8} = \frac{1}{2} \).

(we win if A happens, which is ..)

Consider now the event

\( H = \text{"the first toss lands heads"} \).

Intuitively, if we know \( H \) happens, we are more likely to win.

\( H = \{hhh, hht, hth, htt\} \)

If \( H \) happens, we win in 3 out of 4 cases! We say that "the conditional probability of \( A \) given \( H \) is \( \frac{3}{4} \)" and write:

\[ P(A|H) = \frac{3}{4}, \quad \text{which is the proportion of favorable outcomes within } H. \]

Now, if \( H \) happens, the 3 favorable outcomes are

\( \{hhh, hht, hth\} = A \cap H \) or \( AH \) (both \( A \) and \( H \) occur)

out of the 4 outcomes in \( H \). So:

\[ P(A|H) = \frac{\#(AH)}{\#(H)} . \]

Therefore:

\[ P(A|H) = \frac{\#(AH)}{\#(H)} = \frac{P(AH)}{P(H) - B}. \]
For a generic outcome space \( \Omega \) and probability distribution

**Definition:** for an event \( B \in \Omega \) such that \( P(B) > 0 \), the conditional prob. of \( A \) given \( B \) is

\[
P(A | B) = \frac{P(AB)}{P(B)}
\]

(*)

**Geometric interpretation.**

\[ P(A | B) \text{ = "proportion of } B \text{ which is occupied by } A." \]

**Remark:** we may rewrite (*) as follows:

\[
P(AB) = P(A | B)P(B)
\]

"multiplication rule"

(the probability that \( A \) and \( B \) occur at the same time
\[ = \text{ prob. that } B \text{ occurs, times the prob. that } A \text{ occurs, once I know that } B \text{ has occurred}. \]

**Remark:** Look at the VENN DIAGRAM:

\[ AB \text{ and } AB^c \text{ are a partition of } A \]
\[ (\text{because } A = AB \cup AB^c \text{ and } AB \cap AB^c = \emptyset) : \]
\[ \text{so } P(A) = P(AB) + P(AB^c). \]

(**)

If \( P(B) > 0 \) and \( P(B^c) > 0 \) then, by the multiplication rule:
\[ P(AB) = P(A | B)P(B) \text{ and } P(AB^c) = P(A | B^c)P(B^c), \]

so (** becomes
\[ P(A) = P(A | B)P(B) + P(A | B^c)P(B^c) \]

"Rule of AVERAGE CONDITIONAL PROBABILITIES"
*Example: we have 2 urns:

URN #1: contains 2 white balls & 1 red ball
URN #2: contains 3 white balls & 2 red balls.

Experiment: i) We first toss a fair coin:
- if heads, choose urn #1
- if tails, choose urn #2

ii) Then we pick a ball from the chosen urn.

W = "we pick a white ball"; \( P(W) = ? \)

Call: H = "the coin shows heads"; \( R = \) "we pick a red ball".

\[ H^c = \text{"tails"}. \]

\[ \begin{array}{c}
\text{H} \\
\frac{1}{2} \\
\text{W}
\end{array} \quad \begin{array}{c}
\text{ urn #1} \\
\text{ urn #2}
\end{array} \]

\[ \begin{array}{c}
\wedge \quad \text{P} \left( W \mid H \right) = \frac{2}{3} \\
\wedge \quad \text{P} \left( R \mid H \right) = \frac{1}{3}
\end{array} \]

\[ \begin{array}{c}
\text{T=H}^c \quad \begin{array}{c}
\text{W} \\
\frac{1}{2} \\
\text{R}
\end{array} \quad \begin{array}{c}
\text{ urn #1} \\
\text{ urn #2}
\end{array} \]

\[ \begin{array}{c}
\wedge \quad \text{P} \left( W \mid H^c \right) = \frac{3}{5} \\
\wedge \quad \text{P} \left( R \mid H^c \right) = \frac{2}{5}
\end{array} \]

\[ \text{So} \quad P(W) = P(W \mid H) P(H) + P(W \mid H^c) P(H^c) \]

\[ = \frac{8}{3} \cdot \frac{1}{2} + \frac{3}{5} \cdot \frac{1}{2} = \frac{1}{3} + \frac{3}{10} = \frac{19}{30} \]

The situation above is normally represented by the following "TREE DIAGRAM".

\[ \begin{array}{c}
\frac{2}{3} \\
\frac{1}{3} = \text{P(WH)} \frac{19}{30} \\
\frac{1}{6} = \text{P(RH)} \frac{5}{30} \\
\frac{3}{10} = \text{P(WT)} \frac{9}{30} \\
\frac{2}{10} = \text{P(RT)} \frac{6}{30}
\end{array} \]

They sum to 1 because WH, BH, WT, BT is a partition of the sample space \( \Omega \).
• Generalization of the Rule of Average Conditional Probabilities.

Let $B_1, B_2, \ldots, B_n$ be a partition of $\Omega$. i.e.: $\Omega = B_1 \cup B_2 \cup \ldots \cup B_n$

- $B_i \cap B_j = \emptyset$ for $i \neq j$

We have that $AB_1, AB_2, \ldots, AB_n$ is a partition of $A$, therefore:

$$P(A) = P(AB_1) + P(AB_2) + \ldots + P(AB_n)$$

and

$$P(AB_i) = P(A|B_i)P(B_i)$$

therefore

$$P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \ldots + P(A|B_n)P(B_n)$$

(also called "TOTAL PROBABILITY THEOREM"

Example: Experiment: 3 urns: \[
\begin{align*}
\#1 &: 5W, 1R \text{ balls} \\
\#2 &: 0W, 7R \text{ balls} \\
\#3 &: 3W, 3R \text{ balls}
\end{align*}
\]

- Roll a die: $i$ if $B_i = \{1, 2, 3\}$ → urn #1
- $i$ if $B_2 = \{4\}$ → urn #2
- $i$ if $B_3 = \{5, 6\}$ → urn #3

Tree diagram

$$P(W) = \frac{5}{6} \cdot \frac{1}{2} + 0 \cdot \frac{1}{6} + \frac{1}{2} \cdot \frac{1}{3}$$

$$= \frac{5}{12} + \frac{1}{6} = \frac{7}{12}$$