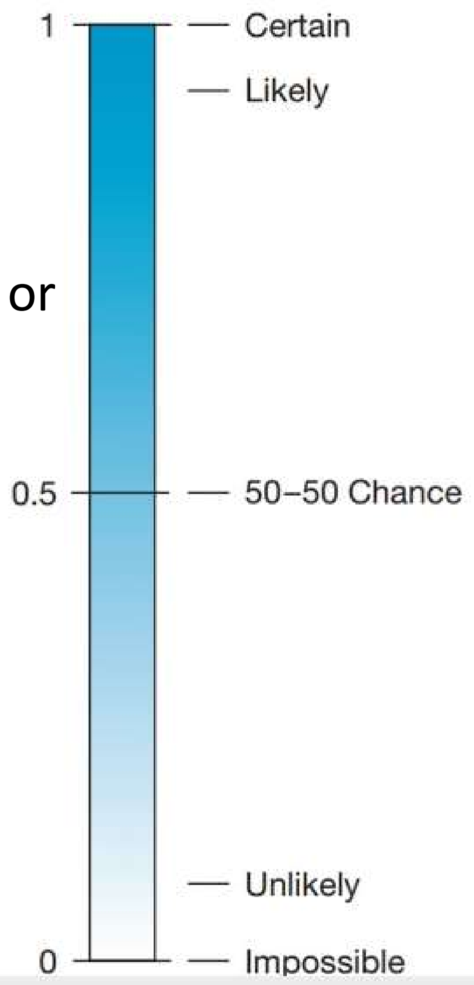
Basics concepts of Probability

Paola Rebora

Notation for Probabilities

- P denotes a probability.
- A, B, and C denote specific events (any collection of results or outcomes of a procedure).
- P(A) denotes the "probability of event A occurring."
- P($\bar{A}$) denotes the probability that event A does not occur

  P($\bar{A}$)=1-P(A)



Definition of Probability

1. **Relative Frequency Approximation of Probability** Conduct (or observe) a procedure and count the number of times that event A occurs. $P(A)$ is then *approximated* as follows:

$$P(A) = \frac{\text{number of times } A \text{ occurred}}{\text{number of times the procedure was repeated}}$$

When referring to relative frequency approximations of probabilities, this text will not distinguish between results that are exact probabilities and those that are approximations, so an instruction to "find the probability" could actually mean "*estimate* the probability."

2. **Classical Approach to Probability (Requires Equally Likely Outcomes)** If a procedure has n different simple events that are *equally likely,* and if event A can occur in s different ways, then

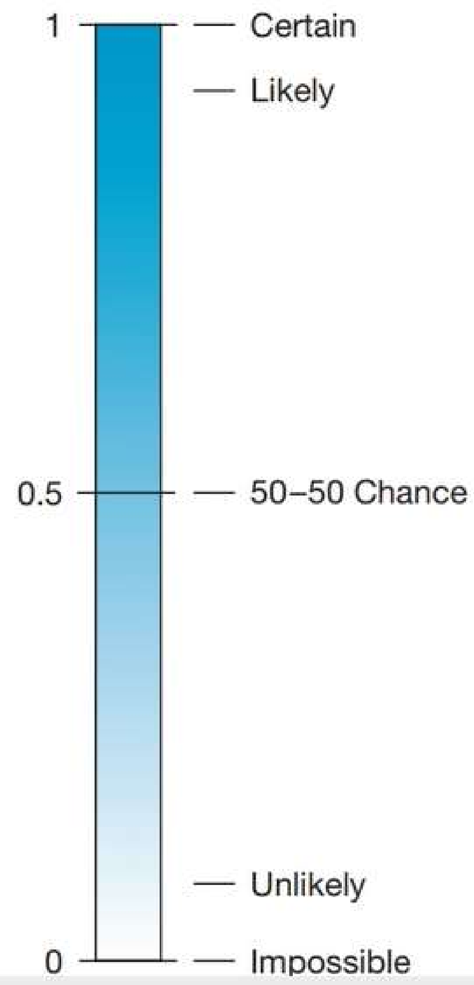
$$P(A) = \frac{\text{number of ways } A \text{ occurs}}{\text{number of different simple events}} = \frac{s}{n}$$

CAUTION When using the classical approach, always confirm that the outcomes are *equally likely*.

3. **Subjective Probabilities** $P(A)$, the probability of event A, is *estimated* by using knowledge of the relevant circumstances.

Important Principles for Probability

- The probability of an event is a fraction or decimal number between 0 and 1 inclusive.
- The probability of an impossible event is 0.
- The probability of an event that is certain to occur is 1.



Properties of probability: ADDITION RULE

- Probability is an additive function for mutually exclusive event (when one happens the other cannot happen.)

Ex. The probability that throwing a die will show a one or a four is the sum of the probabilities of the two events:

P(1 or 4)= P(1)+P(4)= 1/6 + 1/6 = 2/6.

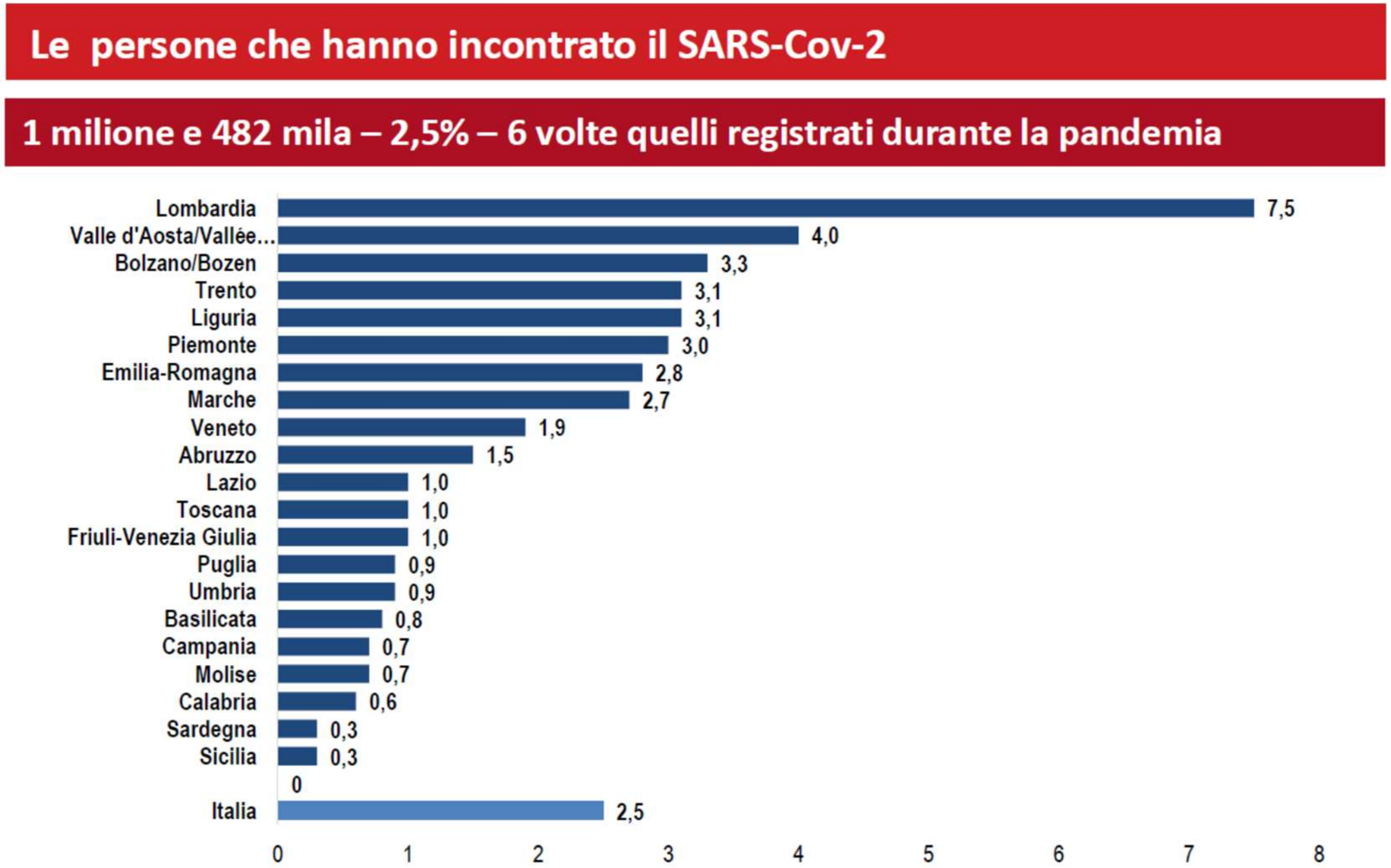
Properties of probability: ADDITION RULE

To find P(A or B), add the number of ways event A can occur and the number of ways event B can occur, but add in such a way that every outcome is counted only once. P(A or B) is equal to that sum, divided by the total number of outcomes in the sample space.

$$\mathbf{P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)}$$

where P(A and B) denotes the probability that A and B both occur at the same time as an outcome in a trial of a procedure.

Example: seroprevalence (IgG+) in Italy from ISTAT survey



Example: SARS–CoV2 seroprevalence (IgG+) in Italy from ISTAT survey

$$\textbf{P(SARS–CoV2 IgG+ } \boldsymbol{OR} \textbf{ Lombardy)} = \mathrm{P(SARS–CoV2\ IgG+\ )} + P(\mathrm{Lombardy}) - P(\mathrm{SARS–CoV2\ IgG+}\boldsymbol{AND}\ \mathrm{Lombardy})$$

SARS–CoV2	IgG+	IgG-	Tot
Lombardia	0.075*10 103 969= 757 798	9 346 171	10 103 969
Other regions	748 165	49 386 388	50 134 553
Italy	0.025*60 238 522= 1 505 963	58732559	60 238 522

$$\mathrm{P(SARS–CoV2\ IgG+\ } \boldsymbol{OR} \mathrm{\ Lombardy)} = \frac{1\,505\,963 + 10\,103\,969 - 757\,798}{60\,238\,522} =$$

$$=0.025+0.168\text{-}0.013=0.18$$

Example: SARS–CoV2 seroprevalence (IgG+) in Italy from ISTAT survey

$$\mathrm{P(SARS{-}CoV2\ IgG+}\ \boldsymbol{OR}\ \mathrm{Lombardy)} = \frac{1\,505\,963 + 10\,103\,969 - 757\,798}{60\,238\,522} = 0.18$$

SARS–CoV2	IgG+	IgG-	Tot
Lombardia	757 798	9 346 171	10 103 969
Other regions	748 165	49 386 388	50 134 553
Italy	1 505 963	58732559	60 238 522

$$\mathrm{P(SARS{-}CoV2\ IgG+}\ \boldsymbol{OR}\ \mathrm{Lombardy)} = \frac{757\,798 + 9\,346\,171 + 748\,165}{60\,238\,522} = 0.18$$

Properties of probability: MULTIPLICATION RULE

Suppose two events are independent, i.e. knowing one has happened tells us nothing about whether the other happens. Then the probability that both happen is the product of their probabilities.

Ex. suppose we toss two coins. One coin does not influence the other, so the results of the two tosses are independent, and the probability of two heads occurring is

$$\mathbf{P(H1\ and\ H2) = P(H1)P(H2)= 1/2 \times 1/2= 1/4}$$

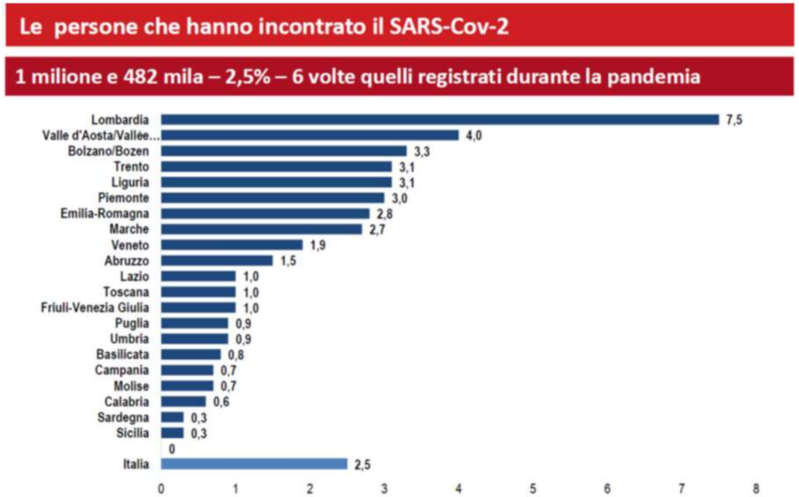
Independent events

Knowing one has happened tells us nothing about whether the other happens:

$$P(A|B) = P(A)$$

SARS–CoV2 $IgG +$ and Lombardy are NOT independent!

$$\text{P(SARS-CoV2 IgG+ | Lombardy)} \neq \text{P(SARS-CoV2 IgG+)}$$



Conditional Probability

$P(A|B)$=given B, the probability of A

It represents the probability of event A occurring after it is assumed that event B has already occurred. (Interpret as “event A occurs after event B has already occurred.”)

A conditional probability of an event is a probability obtained with the additional information that some other event has already occurred.

Conditional Probability

Ex. seroprevalence of SARS-CoV2 in Italy by ISTAT

P(SARS-CoV2 IgG+)=2.5%

P(SARS-CoV2 IgG+ | Lombardy)=7.5%

The probability P(A|B) can be found by dividing the probability of events A and B both occurring by the probability of event B:

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

$$P(\text{SARS–CoV2 IgG+ | Lombardy}) = \frac{P(\text{SARS–CoV2 IgG+ } and \text{ Lombardy})}{P(\text{Lombardy})}$$

Properties of probability: MULTIPLICATION RULE

To find the probability that event A occurs in one trial and event B occurs in another trial, multiply the probability of event A by the probability of event B, but be sure that the probability of event B is found by assuming that event A has already occurred.

$$P(A \text{ and } B) = P(A)P(B|A)=P(B)P(A|B)$$

$$P(\text{SARS−CoV2 IgG+ } and \text{ Lombardy}) = P(\text{Lombardy}) * P(\text{SARS−CoV2 IgG+ }|\text{Lombardy})$$

$$= 0.168 \quad * \quad 0.075$$

$$= \mathbf{0.013}$$

In Italy the probability to be in Lombardy and having SARS−CoV2 IgG+ is 1.3%.

SARS−CoV2	IgG+	IgG-	Tot
Lombardia	0.013	0.155	0.168
Other regions	0.012	0.820	0.832
Italy	0.025	0.975	1.000

Properties of probability: MULTIPLICATION RULE

P(A and B) = P(A)P(B|A)=P(B)P(A|B)

$$P(\text{SARS−CoV2 IgG+ } and \text{ Lombardy}) = P(\text{Lombardy}) * P(\text{SARS−CoV2 IgG+ }|\text{Lombardy})$$
$$= 0.168 * 0.075$$
$$\mathbf{= 0.013}$$

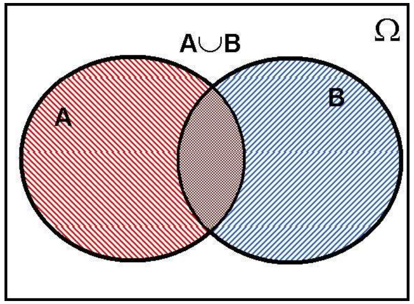
In Italy the probability to be in Lombardy and having SARS−CoV2 IgG+ is 1.3%.

SARS−CoV2	IgG+	IgG-	Tot
Lombardia	757 798	9 346 171	10 103 969
Other regions	748 165	49 386 388	50 134 553
Italy	1 505 963	58732559	60 238 522

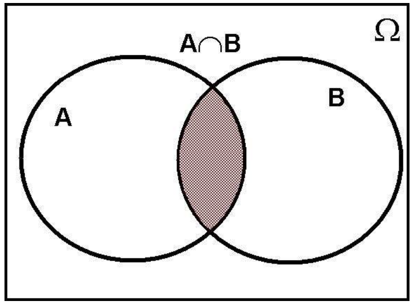
$$P(\text{SARS−CoV2 IgG+ } and \text{ Lombardy}) = {}^{757\ 798}/_{60\ 238\ 522} = 0.013$$

Summary

Union



Intersection



$\cup$ **= OR**

Addition rule

To find P(A or B), add the number of ways event A can occur and the number of ways event B can occur, but add in such a way that every outcome is counted only once. P(A or B) is equal to that sum, divided by the total number of outcomes in the sample space.

P(A or B) = P(A) + P(B) – P(A and B)

$\cap$ **= AND**

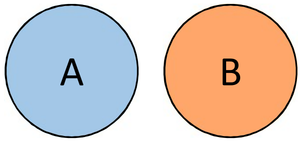
Moltiplication rule

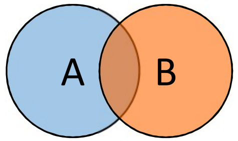
To find the probability that event A occurs in one trial and event B occurs in another trial, multiply the probability of event A by the probability of event B, but be sure that the probability of event B is found by assuming that event A has already occurred.

P(A and B) = P(A)P(B|A)=P(B)P(A|B)

Summary

Given two events A e B





$\cup$ = OR
Addition rule

<u>$A \cap B = \varnothing$</u>

- $P(A \cup B) = P(A) + P(B)$
- $P(A \cap B) = 0$

<u>$A \cap B \neq \varnothing$</u>

- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

  dove $P(A \cap B)$

$\cap$ = AND
Moltiplication rule

- $P(A \cap B) = P(A) \cdot P(B \mid A)$
  $= P(B) \cdot P(A \mid B)$

Se A e B sono indipendenti

- $P(A \cap B) = P(A) \cdot P(B)$