UNIT 3: **PROBABILITY**

Let us assume a population formed by six students whose gender, height (in cm.) and time (min) required to arrive at university are the following (it is a three-dimensional random variable):

	gender	Height	Time
Individual 1	Male	165	45
Individual 2	Female	162	31
Individual 3	Male	146	24
Individual 4	Female	182	13
Individual 5	Male	175	35
Individual 6	Male	186	10

- We define the following events:
  A : height > 170
  B : gender = female
  C : height < 150
  D : height > 400
  E : time < 30

Probability of an event: "proportion of individuals among the population that verify the event".

As it is a proportion, the value is between 0 and 1 (0 - 100 if expressed in %)

PROBABILITY CALCULATIONS:

Applying the definition, the probability of one event is obtained by counting the number of individuals from the table that verify the event:

P(A) = 3/6 = 0.5 ; P(B) = 2/6 = 0.33

Complementary event:

$P(\overline{A}) \equiv P(A_C) \equiv P(\text{No} - A) = P(\text{height} \leq 170) = 3/6 = 0.5$ → $\boxed{P(\overline{A}) = 1 - P(A)}$

P(D) = 0 → <u>impossible</u> event (never happens)

$P(\overline{D}) = 1$ → <u>certain</u> event

- <u>Intersection of two events</u>:
(probability of occurrence of both events simultaneously)

$$P(A \cdot B) \equiv P(A \cap B) = P(\text{height} > 170 \text{ and gender} = \text{female}) = 1/6$$

Two events are <u>exclusive</u> if their intersection is an empty subset (impossible event).
In this example, A and C are exclusive because: P(C)=1/6 and P(A·C)=0.

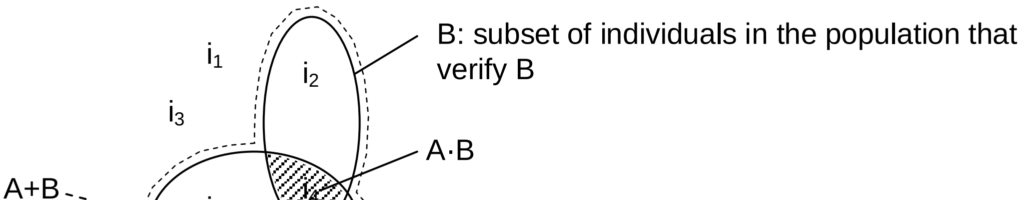
- <u>Union of events</u>:
(probability of occurrence of at least one of the two events)

$$P(A + B) \equiv P(A \cup B) = P(\text{height} > 170 \text{ or gender} = \text{female}) = 4/6$$

$\boxed{P(A + B) = P(A) + P(B) - P(A \cdot B)}$ Verification: 4/6 = 3/6 + 2/6 – 1/6

If two events are exclusive:

$P(A \cdot C)=0 \rightarrow P(A+C)= P(A) + P(C)$ Verification: $4/6 = 3/6 + 1/6$



Graphically we can check that:

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

In the case of 3 events, the probability of their union is calculated with the following formula:

$$P(A+B+E) = P(A) + P(B) + P(E) – P(A \cdot B) – P(A \cdot E) – P(B \cdot E) + P(A \cdot B \cdot E)$$

Verification: $5/6 = 3/6 + 2/6 + 3/6 - 1/6 - 2/6 - 1/6 + 1/6$

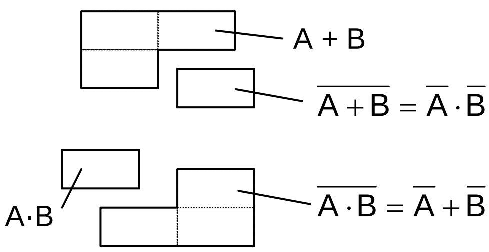
In the case of *n* events:

$$P(A_1+A_2+...+A_n) = \Sigma\left[P(A_i)\right] - \Sigma\left[P(A_i \cdot A_j)\right] + \Sigma\left[P(A_i \cdot A_j \cdot A_k)\right] + ... + (-1)^{n+1} \cdot \Sigma\left[P(A_1 \cdot ... \cdot A_n)\right]$$

DE MORGAN'S LAWS: $\overline{A+B} = \overline{A} \cdot \overline{B}$; $\overline{A \cdot B} = \overline{A} + \overline{B}$

Intuitive demonstration:

A	$A \cdot B$	$A \cdot \overline{B}$
$\overline{A}$	$\overline{A} \cdot B$	$\overline{A} \cdot \overline{B}$
	B	$\overline{B}$



EXAMPLE WITH TWO DICES

The sample space (**E**) is the set of values that the random variable can take. The following example illustrates this concept.

We have two dices perfectly symmetrical (the probability of obtaining each number is the same). We throw both at the same time and we make a note of the results:

	Dice A	Dice B
Throwing 1	2	4
Throwing 2	4	3
Throwing 3	6	1
Throwing 4	5	6
Throwing 5	1	2
...	...	...

It is a two-dimensional random variable. The cases or outcomes (pairs of values) that this random variable can take are: (1,1), (1,2), (1,3), (1,4), (1,5), (1,6), (2,1), ... (6,6). In total, $6 \cdot 6 = 36$. This set of possible results of this random experiment contains 36 cases and is denoted as **E** (sample space, which is finite). The population is the set of infinite pairs of values that could be obtained by throwing both dices. Each throwing of dices generates one individual of the population.

In order to calculate the probability as proportion of individuals in the population that verify a given event, we should conduct many throwings and, next, count how many verify that event. But this calculation is simplified if the following conditions are satisfied:

If the set **E** of cases is finite and the probability is the same for each case →
→ **probability** of one event ≡ proportion of throwings that verify that event =

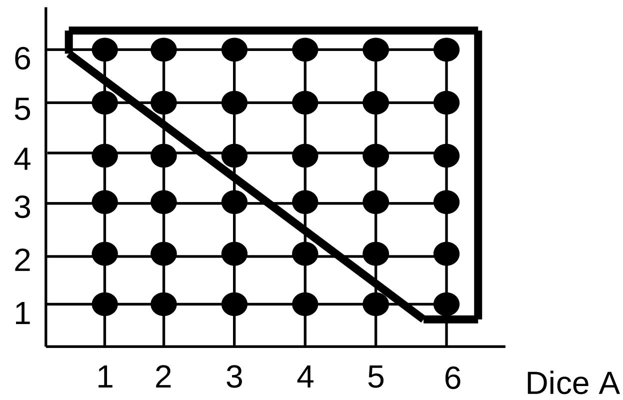
$$= \frac{\text{No. of cases of } \mathbf{E} \text{ favorable to that event}}{\text{No. of possible cases of } \mathbf{E}} \qquad \text{(Rule of Laplace)}$$

For example, considering the following events:

F : (sum of the results of both dices) > 6.
G : obtain at least one 5

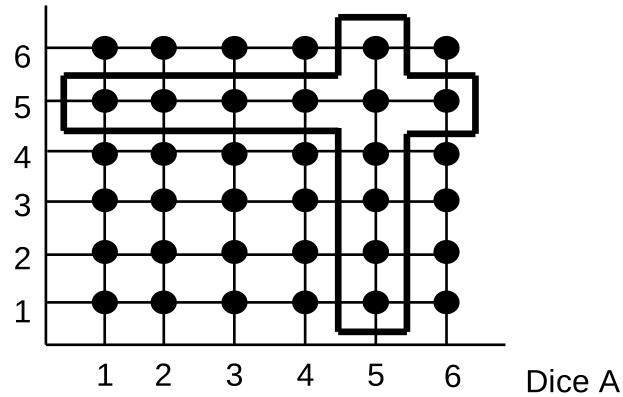
Dice B

P(F)=21/36 : there are 21 cases that satisfy event F (as indicated in the figure):



Dice B

P(G)=11/36: there are 11 cases that satisfy event G (as indicated in the figure):



EXAMPLE OF THE PIN:

Throwing a pin, it can fall with the tip upwards or downwards. We want to calculate the probability of falling with the tip upwards. Many people would say that the probability is ½. This intuitive calculation is based on the formula mentioned above: 1 favorable case / 2 possible outcomes. But that formula is only true if two conditions are satisfied:

- set **E** of finite cases (2 outcomes: tip upwards or downwards).
- The probability has to be the same for both cases: this condition is NOT satisfied, because the probability of falling with the tip upwards is not necessarily the same to the probability of falling with the tip downwards.

Therefore, the only way to calculate the probability is by generating a high number of individuals of the population (i.e. by throwing the pin many times) and counting afterwards how many times it fell with the tip upwards.

CONDITIONAL PROBABILITY

The probability of a certain event A is the proportion of individuals that verify that event in the set of N individuals of the population:

$$P(A) = \frac{\text{No. individuals that verify A}}{\text{Total No. of individuals in the population}}$$

The probability of occurrence of a given event A knowing that another event B (of probability not null) has occurred is represented by P(A/B) and is called *conditional probability of A given B*. It is calculated as the proportion of individuals that verify event A in the subset of individuals that verify B:

$$P(A/B) = \frac{\text{No. individuals that verify A and B}}{\text{No. individuals that verify B}} = \frac{\cdot 1/N}{\cdot 1/N} =$$

$$= \frac{\text{proportion individuals that verify A and B}}{\text{proportion individuals that verify B}} = \frac{P(A\cdot B)}{P(B)}$$

$$\boxed{\left.\begin{array}{l} P(A/B) = \dfrac{P(A\cdot B)}{P(B)} \\ P(B/A) = \dfrac{P(A\cdot B)}{P(A)} \end{array}\right\} \Rightarrow P(A\cdot B) = P(A/B)\cdot P(B) = P(B/A)\cdot P(A)}$$

Generalization ("Multiplicative Law"):

P(A·B·C) = P(A) · P(B·C / A) = P(A) · P(B/A) · P[(C/B) / A] = P(A) · P(B/A) · P(C / A·B)

- Verification of these proportions in the example of students:

P(B) = 2/6 → among the 6 students, 2 verify B (there are 2 women)
P(A·B) = 1/6 → among the 6 students, 1 verifies A and B (female and height > 170)
P(A/B) = 1/2 → among the two students who verify B (women), only one has a height >170

Therefore: $P(A/B) = \dfrac{P(A\cdot B)}{P(B)} \Rightarrow 1/2 = \dfrac{1/6}{2/6}$

- Verification with the example of the dices:

$$P(G/F) = \frac{P(F\cdot G)}{P(F)} \Rightarrow 9/21 = \frac{9/36}{21/36}$$

EXERCISE WITH SPANISH CARDS

In a Spanish pack of 40 cards (10 cups, 10 golden coins, 10 swords, 10 sticks), the following events are considered: A: obtain a cup ; B: obtain a 3.
Calculate: P(A) ; P(B) ; P(A+B) ; P(A·B) ; P(A/B) ; P(B/A)

INDEPENDENT EVENTS

- Let us assume that, instead of 6, the population is all university students of UPV.
- Let us assume that the distribution of time required to arrive at university is the same for men and women, which makes sense.
- Considering event E : time < 30 minutes,

$$\underbrace{\frac{\text{No. women with } t<30}{\text{No. women}}}_{P(E/B)\,=} = \frac{\text{No. men with } t<30}{\text{No. men}} = \underbrace{\frac{\text{No. students with } t<30}{\text{Total No. students}}}_{=\ P(E)}$$

When this condition is satisfied (and both events are not null), we say that events B and E are **independent**: time does not depend of gender, it is the same for men and women.

$$\text{Independent events} \Leftrightarrow \left| \begin{array}{l} P(B/A)=P(B) \\ P(A/B)=P(A) \end{array} \right. \Rightarrow P(A\cdot B)=P(A)\cdot P(B)$$

$$\text{Exclusive events} \Leftrightarrow P(A\cdot B)=0 \quad \rightarrow \quad P(A+B) = P(A) + P(B)$$

Generalization: if a set of N events $A_1, A_2,.., A_n$ are mutually independent,

$$P(A_1 \cdot A_2 \cdot ... \cdot A_n) = P(A_1) \cdot P(A_2) \cdot ... \cdot P(A_n)$$

In regression, we say that two variables are independent if they are not correlated (i.e. no trend observed in the scatterplot).

Time and height (both continuous variables) are expected to be independent (not correlated) because if we know the time that one student takes to arrive at university, it does not provide information about his/her height.

TOTAL PROBABILITY THEOREM (or partition theorem)

- If B is a given event.
- If E is a certain event → P(E)=1
- If A_1,... A_n is a set of *n* events mutually exclusive that comprise E:
  $P(A_i \cdot A_j)=0 \ (i\neq j) \rightarrow P(A_1+A_2+...+A_n) = P(A_1)+P(A_2)+... +P(A_n) = 1$

$$P(B) =^{(1)}= P[(A_1 + A_2 + ... + A_n)\cdot B] =^{(2)}= P(A_1\cdot B)+P(A_2\cdot B)+...+P(A_n\cdot B) =^{(3)}\Rightarrow$$

$$\boxed{P(B) = P(A_1)\cdot P(B/A_1) + P(A_2)\cdot P(B/A_2) + ... + P(A_n)\cdot P(B/A_n)}$$

BAYES' THEOREM

It allows the calculation of $P(A_k/B)$ (*a posteriori* probability) from $P(B/A_k)$ and $P(A_i)$ (*a priori* probability).

$$\boxed{P(A_k/B) = \frac{P(A_k \cdot B)}{P(B)} = \frac{P(A_k) \cdot P(B/A_k)}{P(B)} =^{(4)} = \frac{P(A_K) \cdot P(B/A_k)}{P(A_1) \cdot P(B/A_1) + \ldots + P(A_n) \cdot P(B/A_n)}}$$

(1): the occurrence probability of B and something that always occurs = P(B)
(2): for being mutually exclusive
(3): because $P(B/A_i) = \dfrac{P(A_i \cdot B)}{P(A_i)}$
(4): according to the Total Probability Theorem