UNIDAD 3: **CÁLCULO DE PROBABILIDADES**

Supongamos una población formada por 6 alumnos de los que se conoce su sexo, estatura (en cm.) y tiempo (min.) que tardan en llegar a la universidad. Se trata de una variable aleatoria tridimensional cuyos datos son:

	Sexo	Estatura	Tiempo
Individuo 1	H	165	45
Individuo 2	M	162	31
Individuo 3	H	146	24
Individuo 4	M	182	13
Individuo 5	H	175	35
Individuo 6	H	186	10

- Se definen los siguientes sucesos:

A : estatura > 170
B : sexo = M (mujer)
C : estatura < 150
D : estatura > 400
E : tiempo < 30

Probabilidad de un suceso: "proporción de individuos de la población que verifican dicho suceso".

Al ser una proporción, su valor está comprendido entre 0 y 1 (0 y 100, si se expresa en %).

CÁLCULO DE PROBABILIDADES:

Aplicando la definición, la probabilidad de un suceso se obtiene contando los individuos de la tabla que verifican dicho suceso:

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

Suceso contrario:

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

P(D) = 0 → suceso imposible

$P(\overline{D}) = 1$ → suceso seguro

- Probabilidad del producto o intersección de dos sucesos:
(probabilidad de que se presenten ambos sucesos simultáneamente)

$P(A \cdot B) \equiv P(A \cap B) = P(estatura > 170 \;\; y \;\; sexo = M) = 1/6$

Se dice que dos sucesos son excluyentes si su intersección es el suceso imposible.
En el ejemplo, A y C son excluyentes, ya que: P(C)=1/6 y P(A·C)=0.

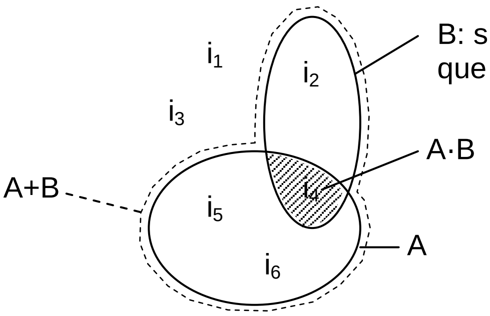
- Probabilidad de la suma o reunión de dos sucesos:
(probabilidad de que se verifique al menos uno de los dos sucesos)

$P(A + B) \equiv P(A \cup B) = P(estatura > 170 \;\; ó \;\; sexo = M) = 4/6$

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

Si dos sucesos son excluyentes:

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



Se comprueba gráficamente que:

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

En el caso de 3 sucesos, la probabilidad de la suma o reunión de ellos se calcula con la siguiente fórmula (la cual, como se comprueba, se cumple con los datos del ejemplo):

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

Comprobación: 5/6 = 3/6 + 2/6 + 3/6 – 1/6 – 2/6 – 1/6 + 1/6

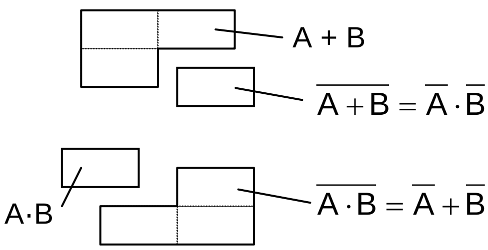
En el caso de *n* sucesos:

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

LEYES DE MORGAN: $\overline{A+B} = \overline{A} \cdot \overline{B}$; $\overline{A \cdot B} = \overline{A} + \overline{B}$

Demostración intuitiva:

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



EJEMPLO CON DOS DADOS

Se tienen dos dados perfectamente simétricos (la probabilidad de que salga cada cara es la misma). Se lanzan los dos a la vez y se anotan en una tabla los resultados:

	dado A	dado B
Lanzamiento 1	2	4
Lanzamiento 2	4	3
Lanzamiento 3	6	1
Lanzamiento 4	5	6
Lanzamiento 5	1	2
...	...	...

Se trata de una variable aleatoria bidimensional. Los casos (parejas de valores) que puede tomar esta variable aleatoria son: (1,1), (1,2), (1,3), (1,4), (1,5), (1,6), (2,1), ... (6,6). En total, 6·6=36. Este conjunto de los posibles resultados del experimento aleatorio contiene 36 casos y se denomina **E** (espacio muestral, que es finito). La población es el conjunto de infinitos pares de valores que se obtienen al lanzar repetidas veces los dos dados. Cada lanzamiento de los dos dados constituye un experimento aleatorio (ya que no se conoce *a priori* el resultado del experimento) y genera un individuo de la población.

Para calcular la probabilidad como proporción de individuos de la población que verifican un cierto suceso se tendría que realizar muchos lanzamientos y luego contar cuántos verifican dicho suceso. Pero este cálculo se simplifica si se cumplen las siguientes condiciones:

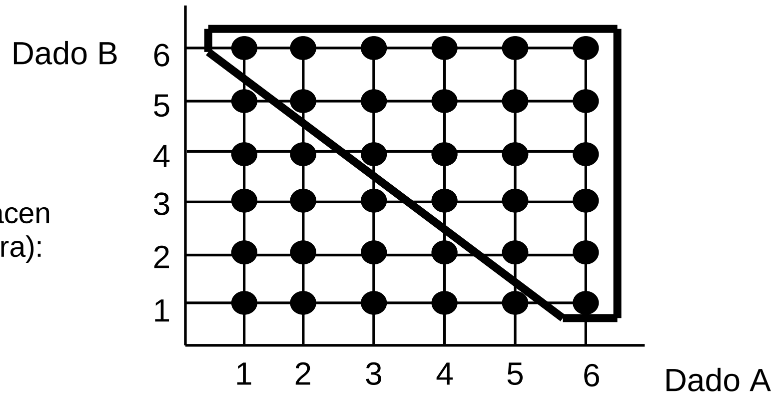
Si el conjunto **E** de casos es finito y la probabilidad es la misma para cada caso →
→ la **probabilidad** de un suceso ≡ proporción de lanzamientos que verifican dicho suceso =

$$= \frac{\text{nº de casos de } \mathbf{E} \text{ favorables a dicho suceso}}{\text{nº de casos posibles de } \mathbf{E}}$$

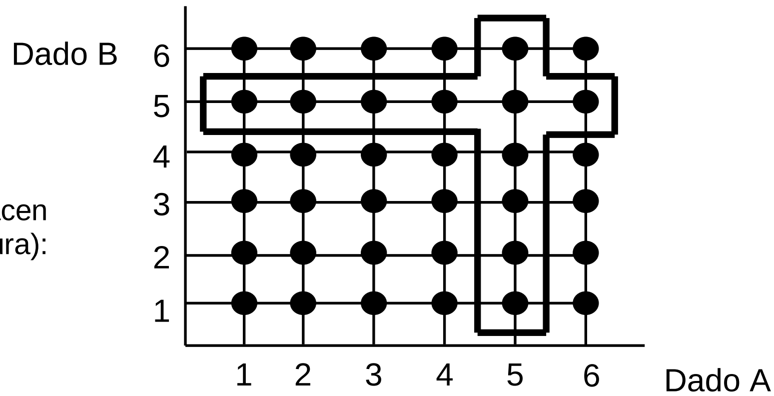
Por ejemplo, considérense los siguientes sucesos:

F : que la suma de los dos dados sea mayor de 6.
G : obtener al menos un 5

P(F)=21/36 : hay 21 casos que satisfacen el suceso F (como se indica en la figura):



P(G)=11/36: hay 11 casos que satisfacen el suceso G (como se indica en la figura):



EJEMPLO DE LA CHINCHETA:

Calcular la probabilidad de que al lanzar al aire una chincheta, caiga con la punta hacia arriba. A simple vista parece que la probabilidad es ½. En el fondo, este cálculo intuitivo se basa en la fórmula anterior: 1 caso favorable / 2 casos posibles. Sin embargo, la fórmula anterior solamente será cierta si se cumplen las dos condiciones:

- conjunto **E** de casos finito: se cumple (2 casos: que caiga hacia arriba o hacia abajo).
- que la probabilidad sea la misma para cada caso: NO se cumple, ya que la probabilidad de que caiga con la punta hacia arriba no tiene por qué ser la misma que la probabilidad de que caiga con la punta hacia abajo.

Por tanto, no hay más remedio que generar un número elevado de individuos de la población (lanzando la chincheta muchas veces) y contar después cuántas veces cae con la punta hacia arriba.

PROBABILIDAD CONDICIONAL

La probabilidad de un determinado suceso A es la proporción de individuos que verifican el suceso A en el conjunto de N individuos de la población:

$$P(A) = \frac{n^{o} \;\; individuos \;\; que \;\; verifican \;\; A}{n^{o} \;\; total \;\; de \;\; individuos \;\; de \;\; la \;\; población}$$

La probabilidad de que ocurra un suceso A sabiendo que ha ocurrido otro suceso B (de probabilidad no nula) se representa por P(A/B) y recibe el nombre de *probabilidad condicional de A dado B*. Se calcula como la proporción de individuos que verifican el suceso A en el subconjunto de individuos que verifican B:

$$P(A/B) = \frac{n^{o} \;\; individuos \;\; que \;\; verifican \;\; A \;\; y \;\; B}{n^{o} \;\; individuos \;\; que \;\; verifican \;\; B} = \frac{\cdot 1/N}{\cdot 1/N} =$$

$$= \frac{proporc. \;\; individuos \;\; que \;\; verifican \;\; A \;\; y \;\; B}{proporc. \;\; individuos \;\; que \;\; verifican \;\; B} = \frac{P(A \cdot B)}{P(B)}$$

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

Generalización (“Ley Multiplicativa”):

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)

- Comprobación de estas relaciones con el ejemplo de los alumnos:

P(B) = 2/6 → de los 6 alumnos, 2 cumplen B (hay 2 mujeres)
P(A·B) = 1/6 → de los 6 alumnos, 1 cumple A y B (ser mujer y de estatura > 170)
P(A/B) = 1/2 → de los 2 alumnos que cumplen B (mujeres) sólo una tiene una estatura > 170

Se cumple: $P(A/B) = \frac{P(A \cdot B)}{P(B)} \Rightarrow 1/2 = \frac{1/6}{2/6}$

- Comprobación en el ejemplo de los dados:

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

EJEMPLO CON UNA BARAJA

En una baraja española de 40 cartas (10 copas, 10 oros, 10 espadas y 10 bastos), se consideran los siguientes sucesos: A: obtener copa; B: obtener un tres.
Calcular: P(A); P(B); P(A+B); P(A·B); P(A/B); P(B/A).

SUCESOS INDEPENDIENTES

- Supongamos que en lugar de 6, la población son todos los universitarios de la UPV.
- Supongamos que la distribución de los datos de tiempo que se tarda en llegar a la universidad es la misma para chicos que chicas, con la misma media y varianza.
- Suponiendo el suceso E : tiempo < 30 minutos,

$$\frac{n^{o}\ \ mujeres\ \ que\ \ t<30}{n^{o}\ \ mujeres}=\frac{n^{o}\ \ hombres\ \ que\ \ t<30}{n^{o}\ \ hombres}=\frac{n^{o}\ \ universitarios\ \ que\ \ t<30}{n^{o}\ \ total\ \ universitarios}$$

$$P(E/B) = \qquad\qquad = P(E)$$

Cuando esta condición se cumple (siempre que ambos sucesos sean no nulos) se dice que los sucesos B y E son **independientes**: la distribución de los datos de tiempo no depende del sexo, es la misma para hombres que para mujeres.

Sucesos independientes $\Leftrightarrow \left| \begin{matrix} P(B/A)=P(B) \\ P(A/B)=P(A) \end{matrix} \right. \Rightarrow P(A\cdot B)=P(A)\cdot P(B)$

Sucesos excluyentes $\Leftrightarrow P(A\cdot B)=0$ → P(A+B) = P(A) + P(B)

Generalización: si un conjunto de N sucesos $A_1, A_2,.., A_n$ son mutuamente independientes,

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

En regresión se dice que dos variables son independientes si no están correlacionadas (es decir, si cuando aumenta una, la otra no tiende a aumentar ni a disminuir).

Las variables continuas tiempo y peso presumiblemente serán independientes (no correlacionadas) porque si sabemos el tiempo que un estudiante tarda en llegar a la universidad, esto no proporciona información sobre su estatura.

TEOREMA DE LA PROBABILIDAD TOTAL (o de la partición)

- Sea B un determinado suceso.
- Sea E un suceso seguro → P(E)=1
- Sean $A_1, ... A_n$ un conjunto de *n* sucesos mutuamente excluyentes en que se particiona E:
  $P(A_i \cdot A_j)=0\ (i \neq j)$ → $P(A_1+A_2+...+A_n) = P(A_1)+P(A_2)+... +P(A_n) = 1$

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

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

TEOREMA DE BAYES

Permite calcular $P(A_k/B)$ (probabilidad *a posteriori*) a partir de $P(B/A_k)$ y de $P(A_i)$ (probabilidad *a priori*).

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

(1): la probabilidad de que ocurra B y algo que siempre ocurre = P(B)
(2): al ser mutuamente excluyentes
(3): ya que $P(B/A_i) = \frac{P(A_i \cdot B)}{P(A_i)}$
(4): por el Teorema de la Probabilidad Total