Minería de Datos para el Análisis de Big Data

Por: Carlos Carreño

ccarrenovi@Gmail.com

Abril,2021

Modulo 4 Teoría de la Probabilidad

- Fundamentos
- Regla de Laplace
- Intersección y unión de sucesos
- Unión de Sucesos
- Sucesos Incompatibles y Compatibles
- Experimentos compuestos
- Ley del producto
- Sucesos dependientes e independientes
- Intersección de sucesos
- Diagrama de árbol
- Leyes de Morgan

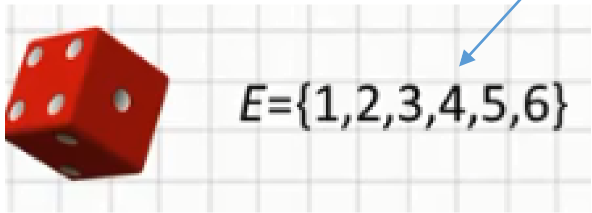
Fundamentos

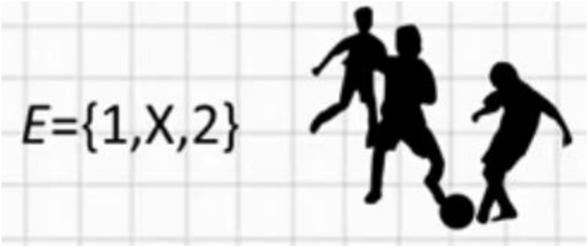
- Que es un experimento aleatorio?
  - Es aquel en el que no es posible predecir el resultado

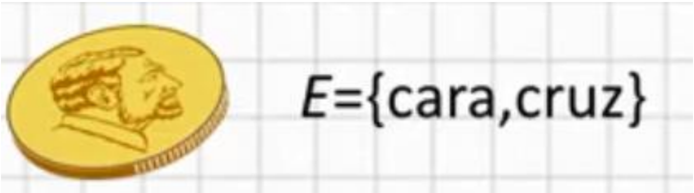
...

- Que es el **espacio muestral**?
- Es el conjunto de todos los resultados posibles

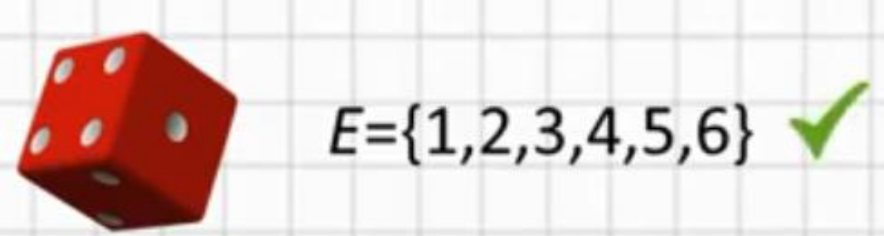
Sucesos Elementales







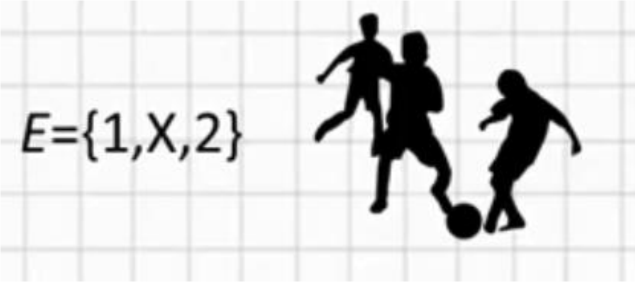
- Sucesos equiprobables
- Son aquellos que tienen la misma probabilidad de suceder



¿Siempre será un partido de futbol equiprobable?

¿Ganar la lotería es equiprobable?

G = {“gano”, “no gano”}



Regla de Laplace

- La probabilidad de que ocurra un suceso se calcula con la siguiente regla:

$$P(suceso) = \frac{n^{o}\ casos\ favorables}{n^{o}\ casos\ posibles}$$

$$P(5) = \frac{1}{6} \qquad P(>4) = \frac{2}{6} = \frac{1}{3}$$

$$P(3) = \frac{1}{6} \qquad P(múltiplo\ de\ 3) = \frac{2}{6} = \frac{1}{3}$$

$$P(par) = \frac{3}{6} = \frac{1}{2} \qquad P(primo) = \frac{3}{6}$$

Ejercicio: Probabilidades

- Probabilidades de Sacar un Carta en una Baraja

- Cual es la probabilidad de sacar un 3?
- Cual es la probabilidad de sacar un K de diamante?

- Cual es la probabilidad de sacar una bola roja

$$P(R)=\frac{2}{5}=0{,}4=40\%$$

Una probabilidad siempre esta en el intervalo cerrado [0,1]

$$\boxed{0\leq P(A)\leq 1}$$

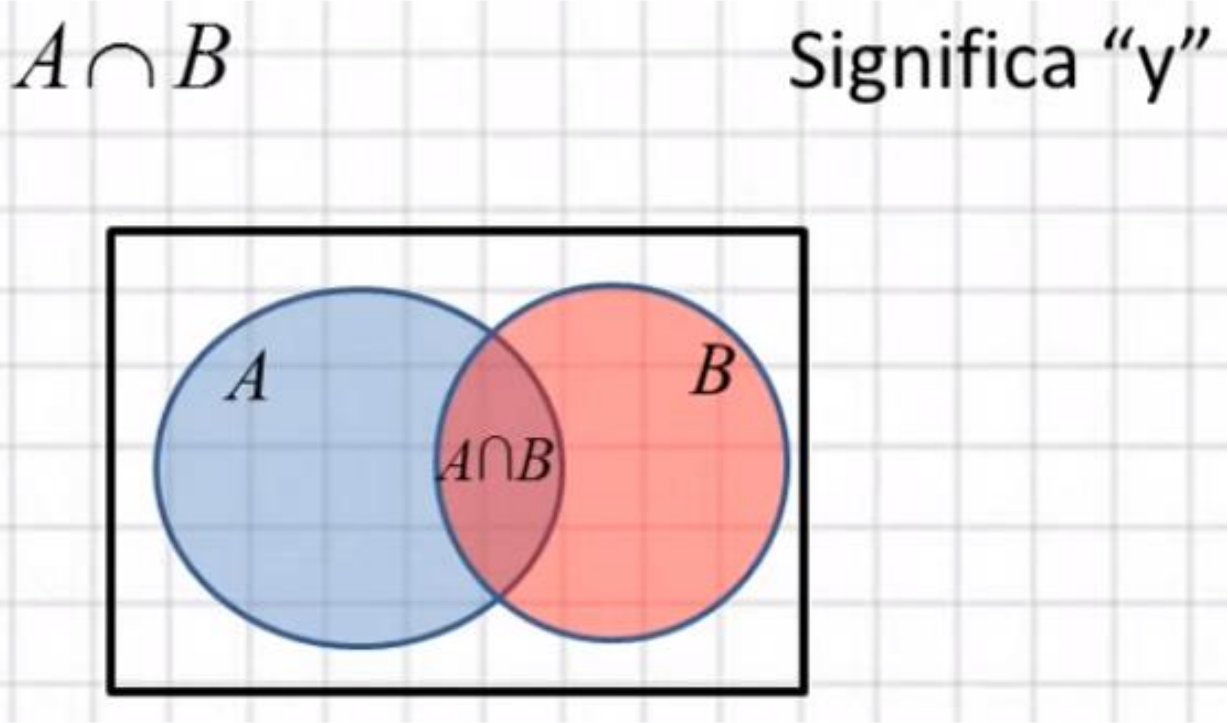
Probabilidad del suceso contrario

$$\boxed{P(\overline{A})=1-P(A)}$$

Cual es la probabilidad de no sacar una bola roja?

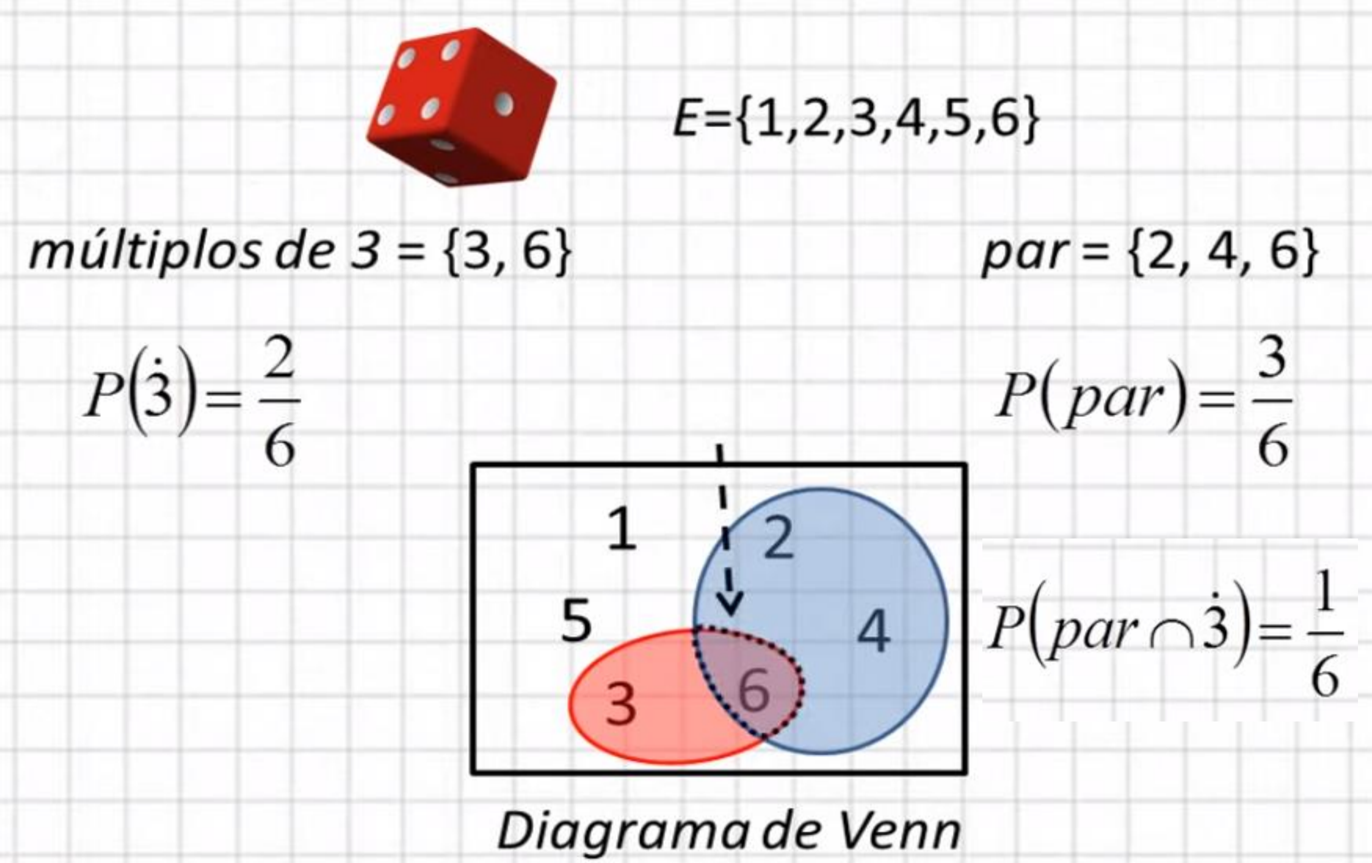
Intersección de sucesos

- La intersección de dos sucesos A y B es un nuevo suceso formado por todos los sucesos elementales comunes a A y B



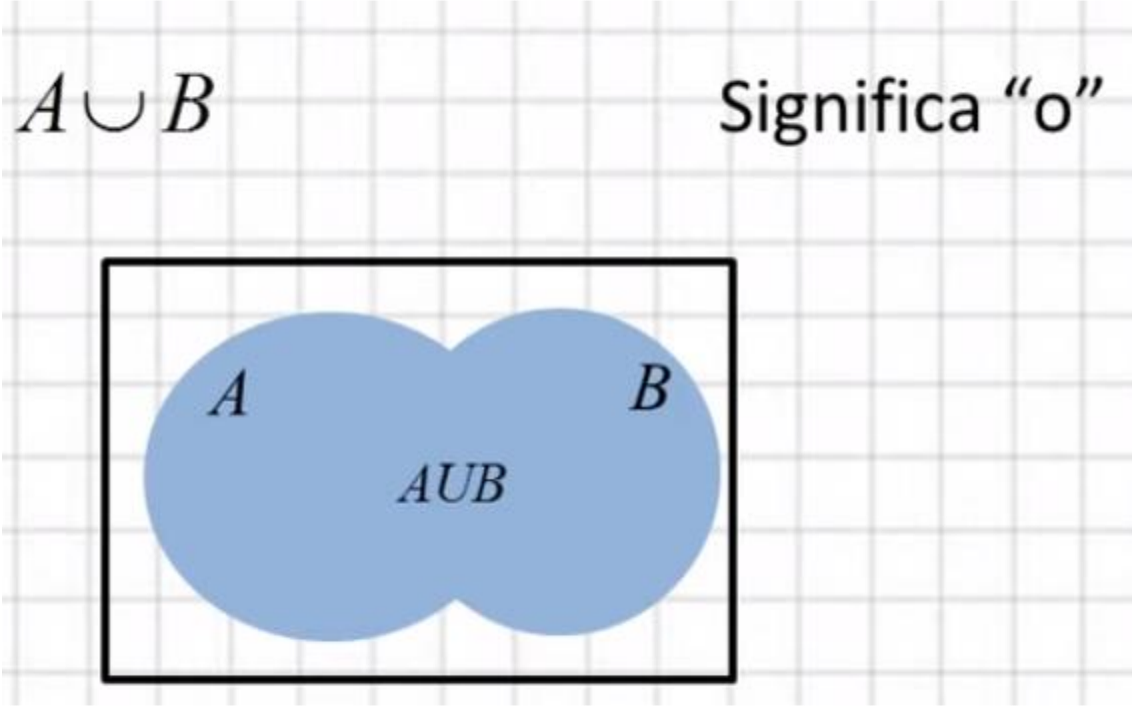
Ejemplo: Intersección de sucesos

- Cual es la probabilidad que aparezca un numero par y que sea múltiplo de 3



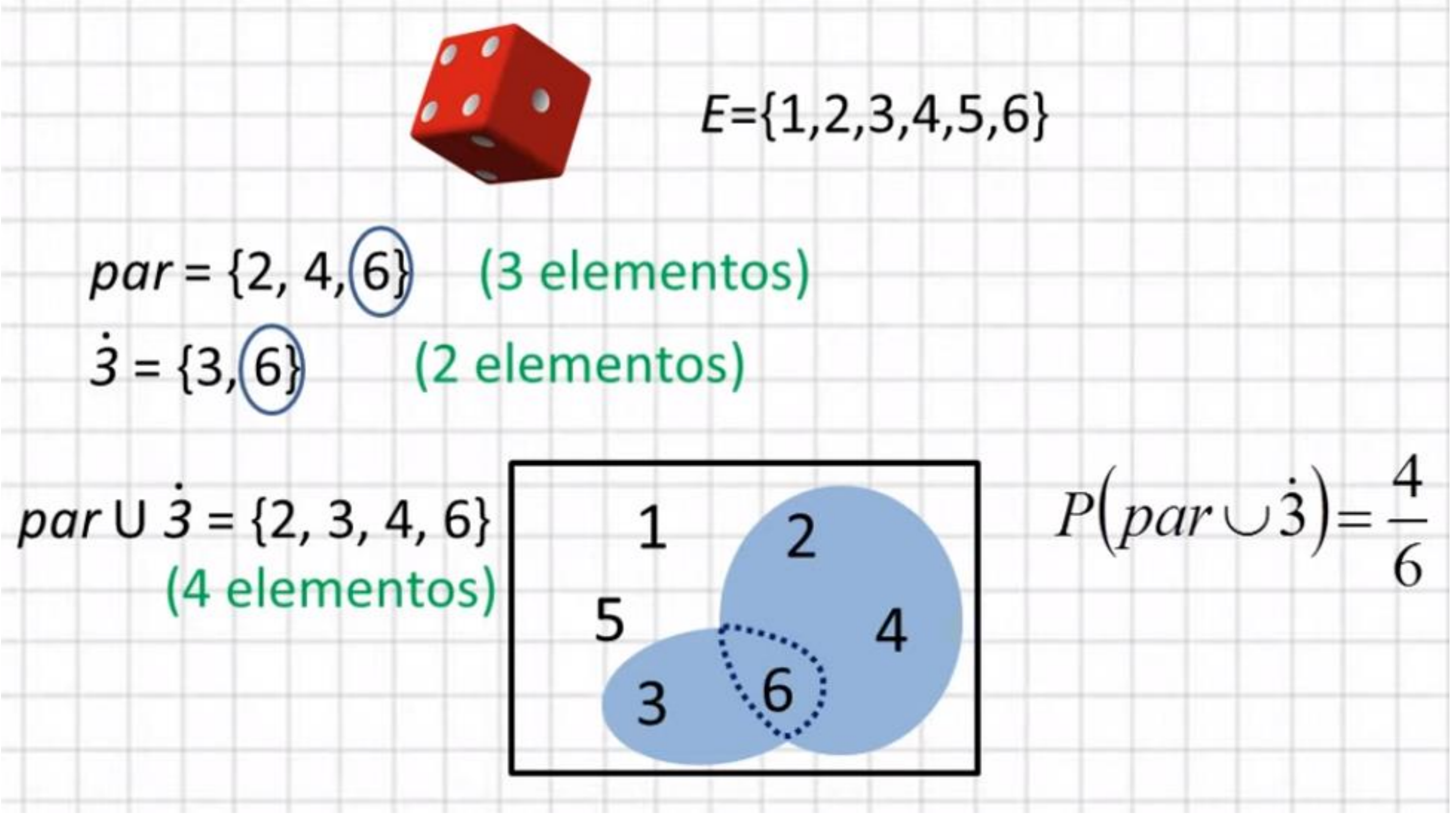
Unión de Sucesos

- La unión de sucesos A y B es un nuevo suceso formado por todos los sucesos elementales de A y B



Ejemplo: Unión de Sucesos

- Cual es la probabilidad que al lanzar un dado salga un numero par o múltiplo de 3



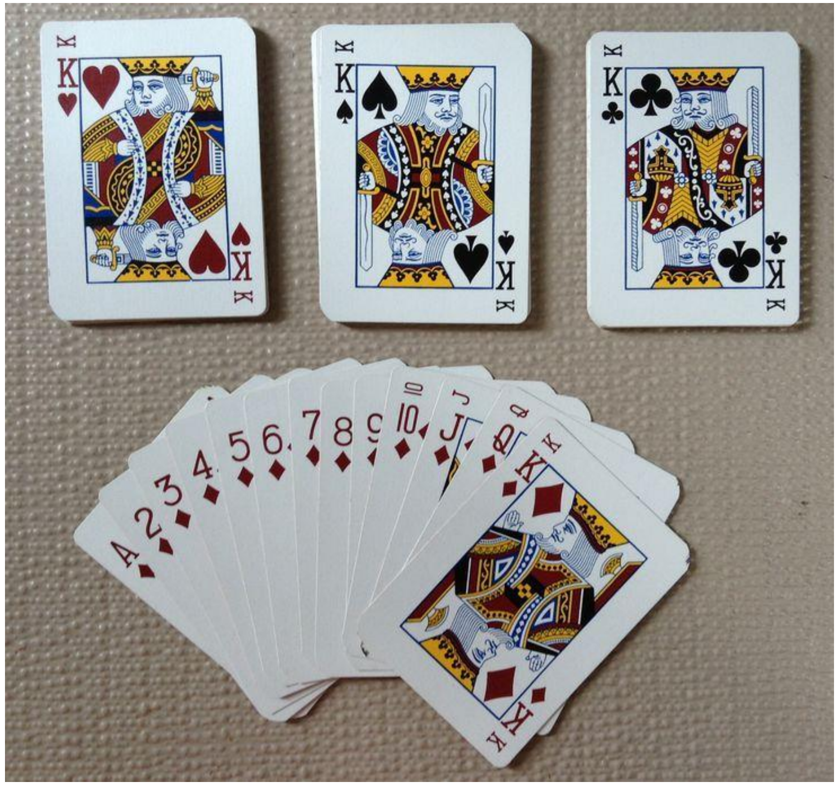
Teorema Fundamental de la Probabilidad

- En el calculo de la probabilidad de la unión, debo evitar contar los sucesos repetidos (intersección)

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

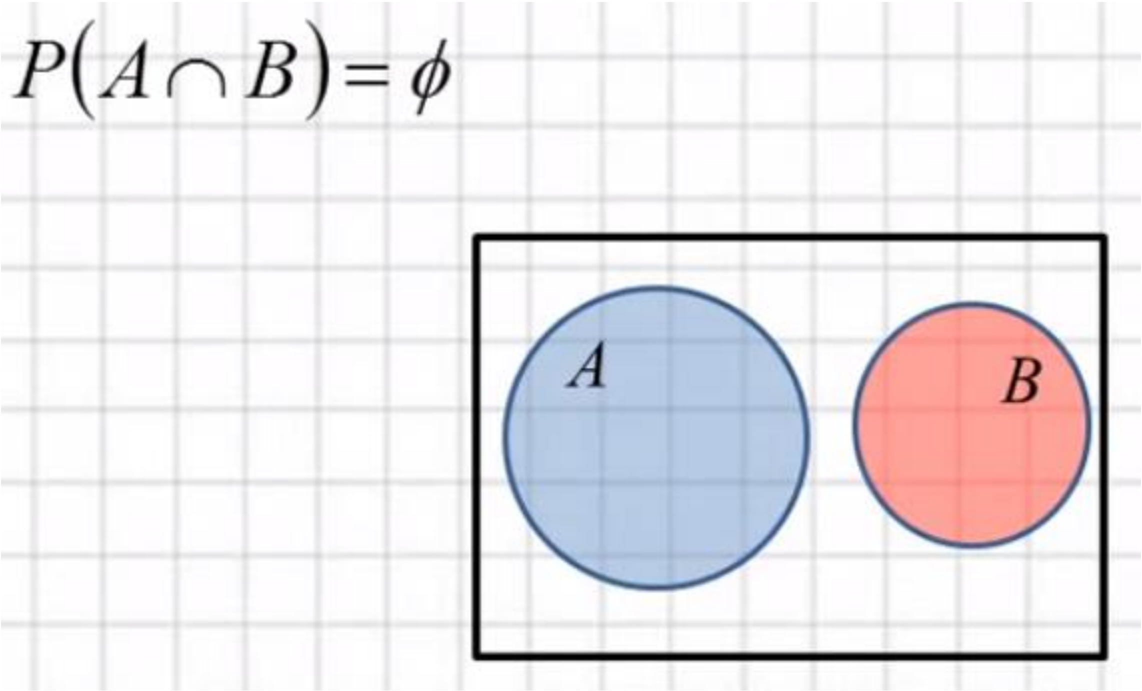
Ejercicio: Teorema de Probabilidad

- Cual es la probabilidad de sacar una carta de diamante o una carta con el numero 7 de una baraja?



Sucesos Incompatibles y Compatibles

- Dos sucesos A y B son incompatibles si no pueden ocurrir simultáneamente

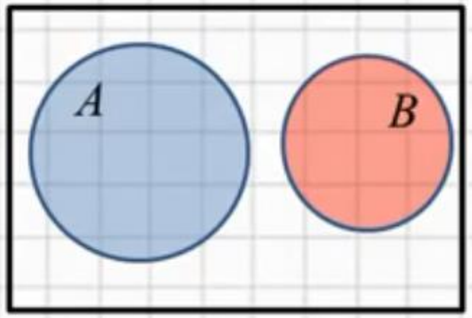


Ejemplo: Sacar una carta de una baraja que sea múltiplo de 5 y que contenga una figura.

...

- Sin dos sucesos son incompatibles el teorema de la probabilidad se modifica

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



$$P(A\cap B)=\phi$$

Únicamente en el caso de sucesos incompatibles $P(A\cup B)=P(A)+P(B)$

Experimentos Compuestos

- Son aquellos experimentos que están formados por varios experimentos simples

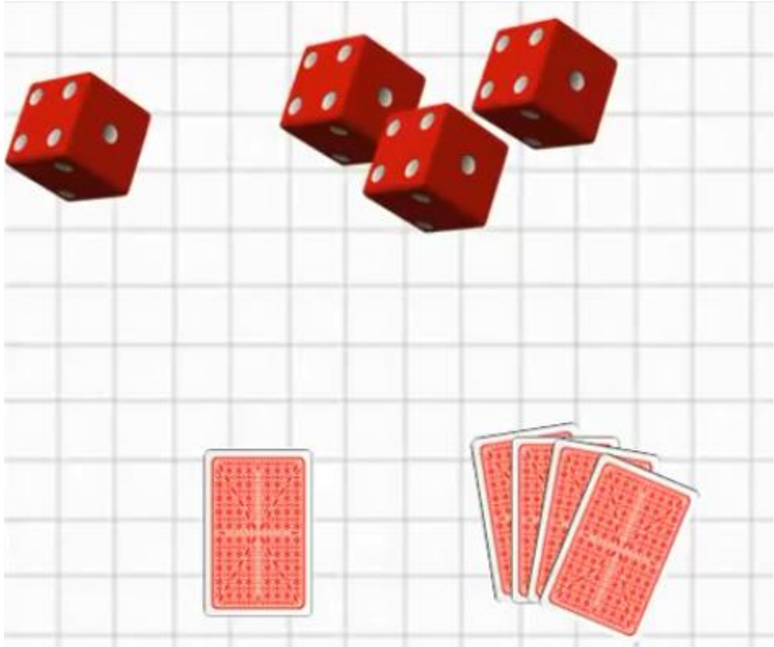


Diagrama Cartesiano

- Es una tabla de doble entrada que puede ser útil en algunos experimentos compuestos formados por dos experimentos simples

	C	+
C	CC	C+
+	+C	++

Espacio muestral

$$P(dos\,cruces) = \frac{1}{4} \qquad P(cruz, cara) = \frac{1}{4}$$

$$P\left(\begin{matrix} resultados \\ diferentes \end{matrix}\right) = \frac{2}{4} \qquad P\left(\begin{matrix} resultados \\ iguales \end{matrix}\right) = \frac{2}{4}$$

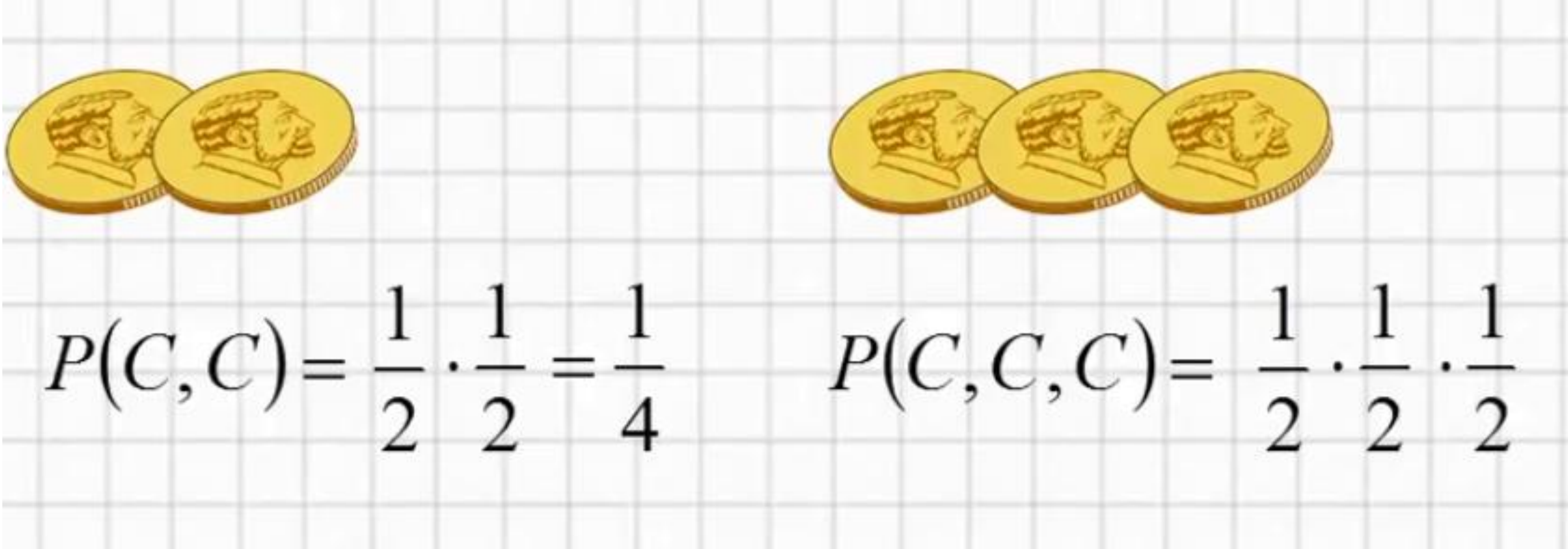
- Espacio muestral del lanzamiento de dos dados

$$P(\text{dos números iguales}) = \frac{6}{36}$$

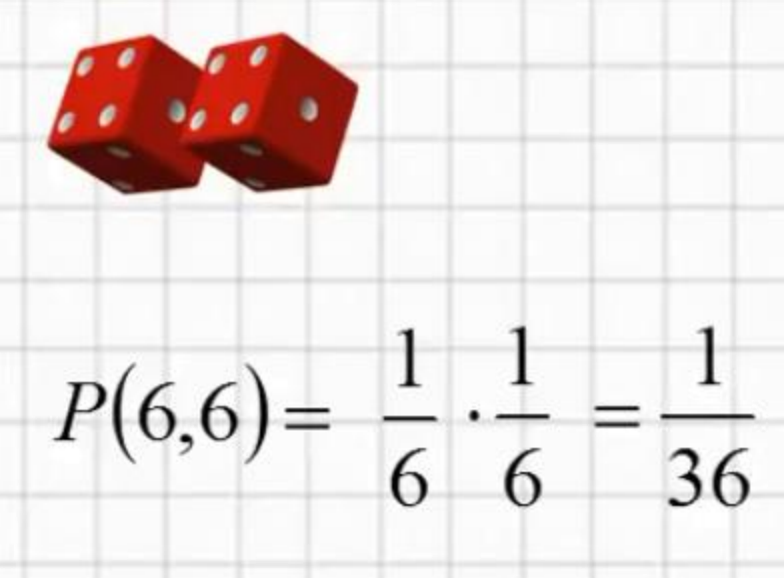
	1	2	3	4	5	6
1	11	12	13	14	15	16
2	21	22	23	24	25	26
3	31	32	33	34	35	36
4	41	42	43	44	45	46
5	51	52	53	54	55	56
6	61	62	63	64	65	66

Ley del Producto

- Para calcular la probabilidad de que varios sucesos ocurran de manera sucesiva, se multiplican sus probabilidades individuales.

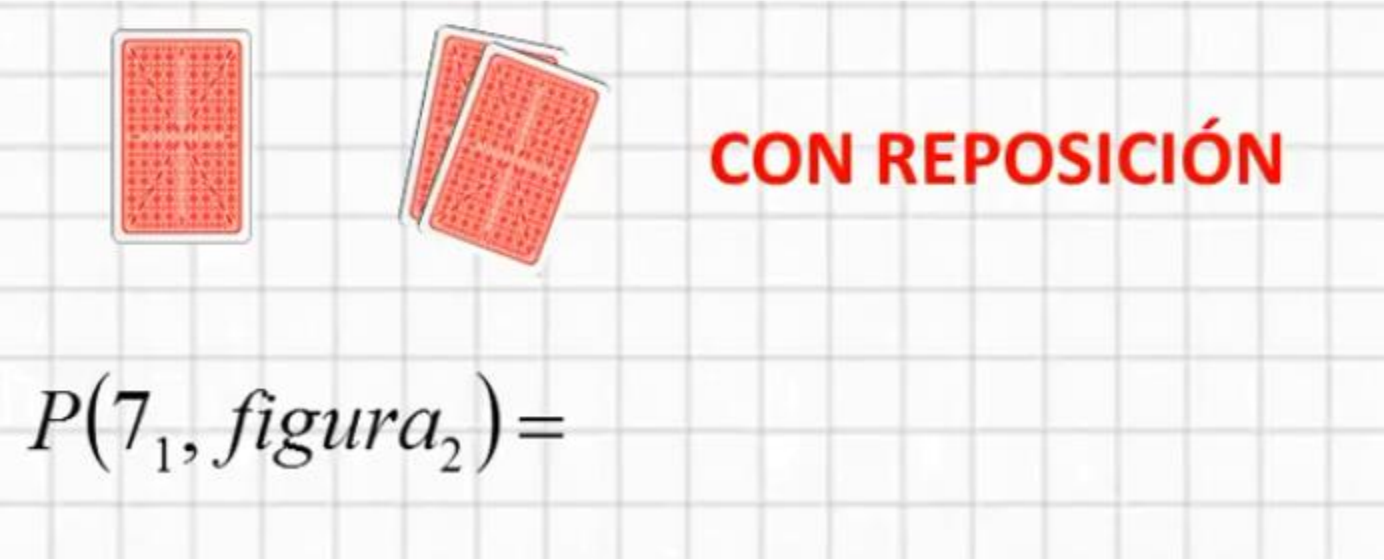


- Si lanzo dos dados, cual es la probabilidad que salgan dos 6

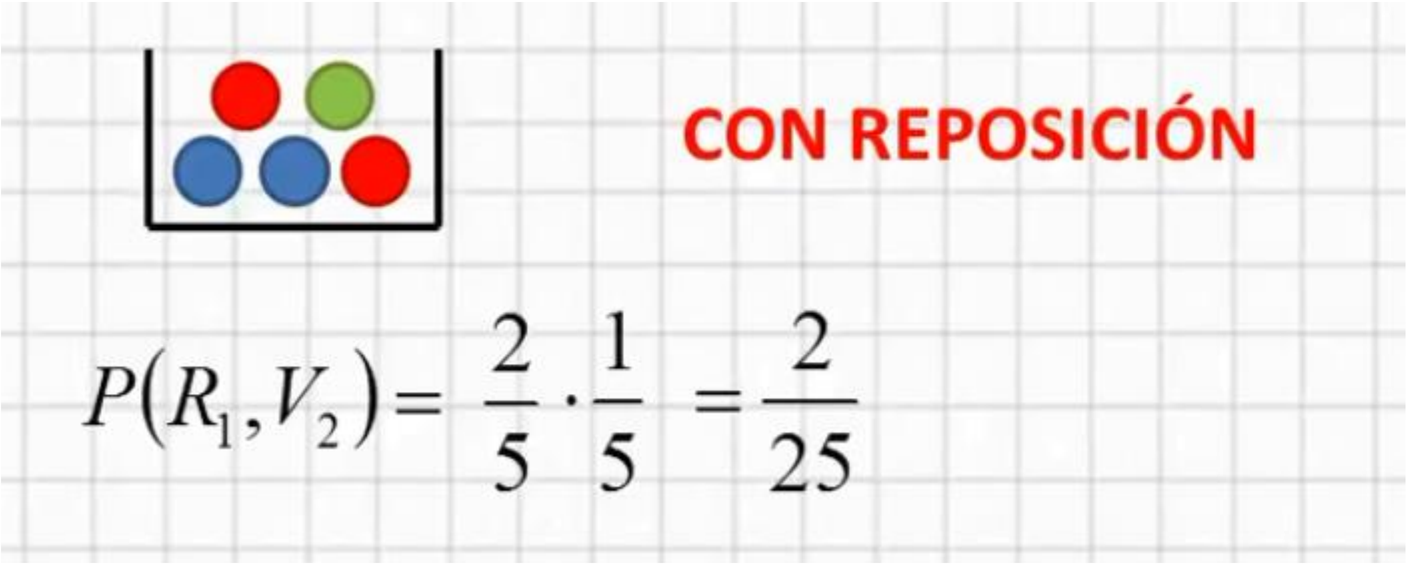


Ejercicio

- Cual es la probabilidad que de una baraja con reposición, saque un 7 y luego una figura



- Cual es la probabilidad de sacar una bola roja y luego una verde



Sucesos Dependientes e Independientes

- Dos sucesos A y B son **independientes** si la probabilidad de que uno de ellos ocurra **no depende** de que haya ocurrido el **otro**. En caso contrario son sucesos **dependientes**.

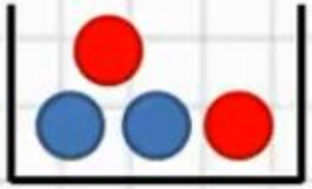
$P(B/A)$

"Probabilidad de B condicionado de A"

"Probabilidad de B sabiendo que ha ocurrido A"

Ejemplo: Sucesos dependientes

- Cual es la probabilidad de sacar un bola verde y una bola roja sin reposición



SIN REPOSICIÓN

$$P(R_2 / V_1) = \frac{2}{4}$$

$$P(V_1, R_2) = \frac{1}{5} \cdot \frac{2}{4} = \frac{2}{20} = \frac{1}{10}$$

Ejemplo: Sucesos Dependientes

- Probabilidad de sacar un rey de una baraja sabiendo que hemos sacado un rey previamente

$$P(dos\ \ reyes) = P(rey_1) \cdot P(rey_2 / rey_1)$$

Sucesos Independientes

- Cual es la probabilidad de sacar un 4 en el segundo dado sabiendo que en el primer dado salió un 4.

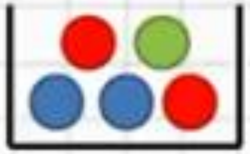
$$P(4_1,4_2) = P(4_1)\cdot P(4_2/4_1) = \frac{1}{6}\cdot\frac{1}{6} = \frac{1}{36}$$

Si dos sucesos A y B son **independientes:**

$$P(B/A) = P(B)$$

Intersección de Sucesos

- En un experimento compuesto, calcular la probabilidad de la intersección es calcular la probabilidad de que varios resultados ocurran sucesivamente



$$P(V_1 \cap R_2) = P(V_1) \cdot P(R_2 / V_1)$$

$$\frac{P(V_1 \cap R_2)}{P(V_1)} = P(R_2 / V_1)$$

- De forma general $\boxed{P(A \cap B) = P(A) \cdot P(B / A)}$

Probabilidad Condicionada

- La probabilidad de un suceso B habiendo ocurrido A previamente

$$\boxed{P(A \cap B) = P(A) \cdot P(B / A)}$$

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

- Cuando los sucesos son independientes

$$P(B/A) = P(B)$$

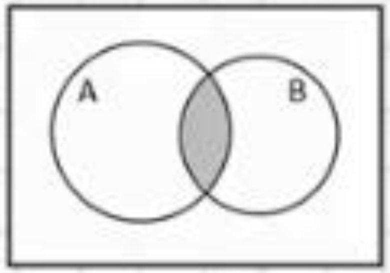
$$P(A \cap B) = P(A) \cdot P(B)$$

Resumen de Formulas

Intersección de sucesos

$$A \cap B$$

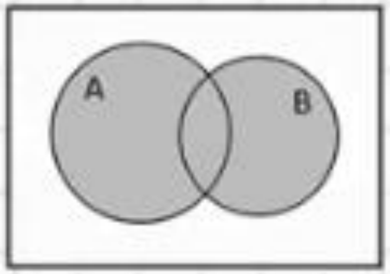
$$P(A \cap B) = P(A) \cdot P(B/A)$$



Si A y B son sucesos independientes:

Unión de sucesos

$$A \cup B$$

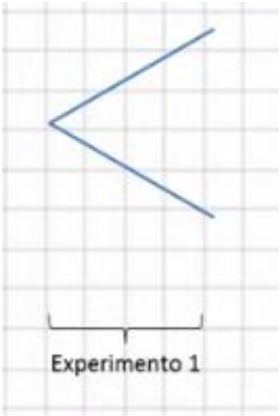


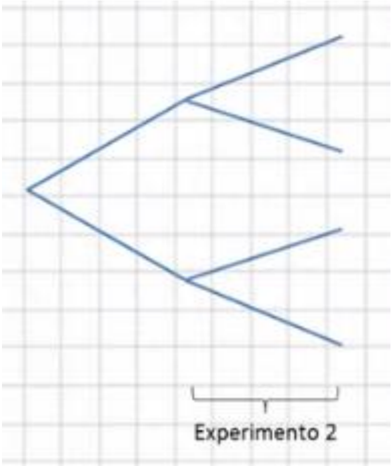
$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

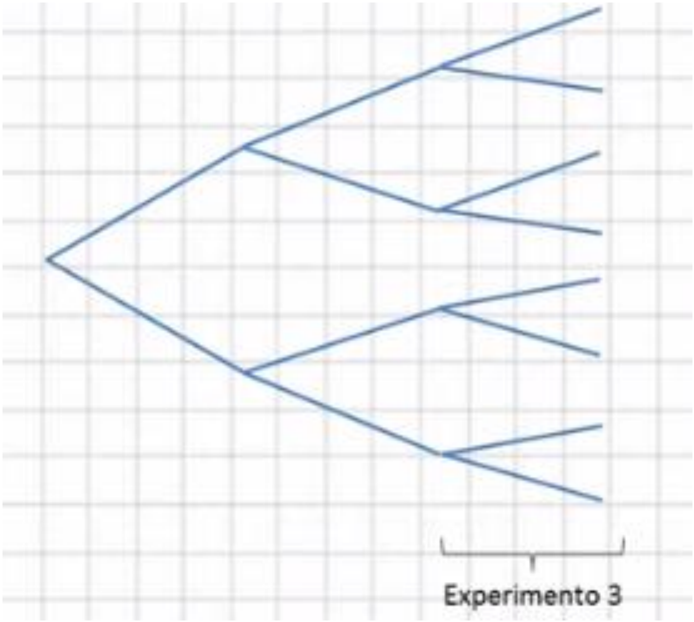
Si A y B son sucesos incompatibles: $P(A \cup B) = P(A) + P(B)$

Diagrama de Árbol

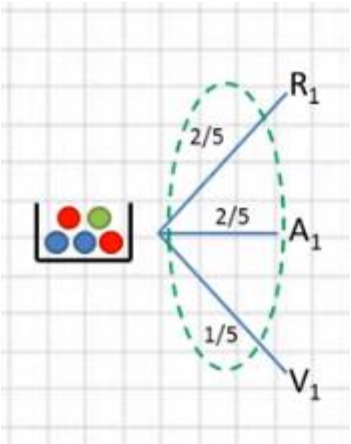
- Se forma dibujando las ramas que representan distintos resultados para cada experimento

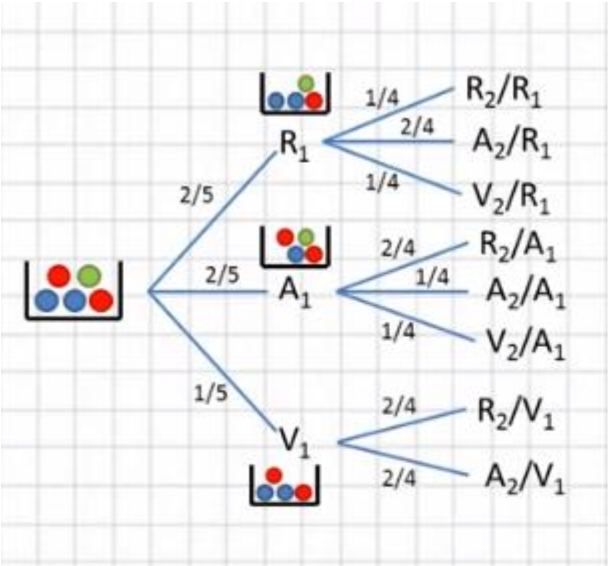


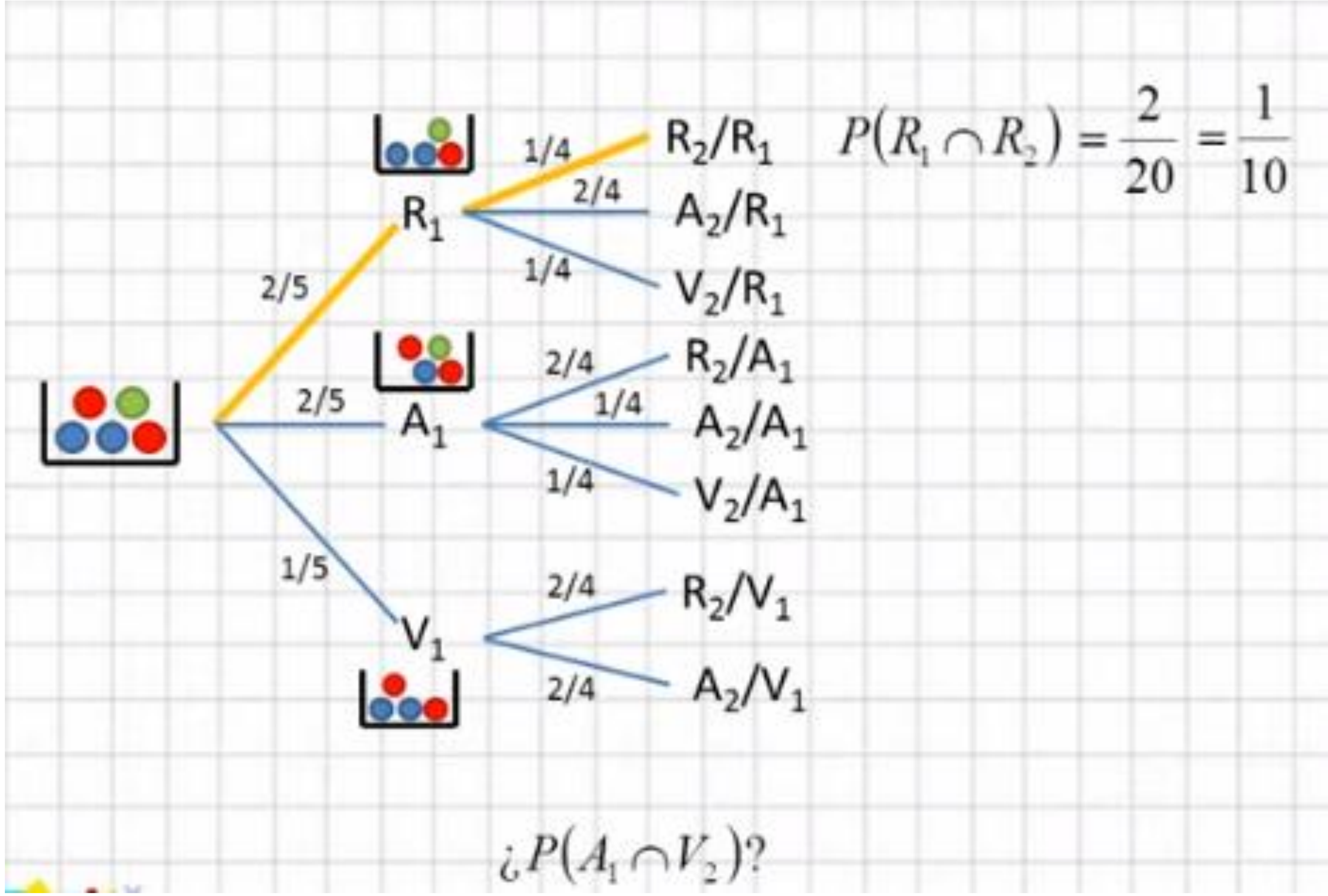




Ejemplo: Diagrama de árbol y la urna con bolas







1/4
R2/R1
$P(R_1 \cap R_2) = \frac{2}{20} = \frac{1}{10}$
2/4
R1
A2/R1
1/4
V2/R1
2/5
2/4
R2/A1
2/5
1/4
A1
A2/A1
1/4
V2/A1
1/5
2/4
R2/V1
V1
2/4
A2/V1
$¿P(A_1 \cap V_2)?$

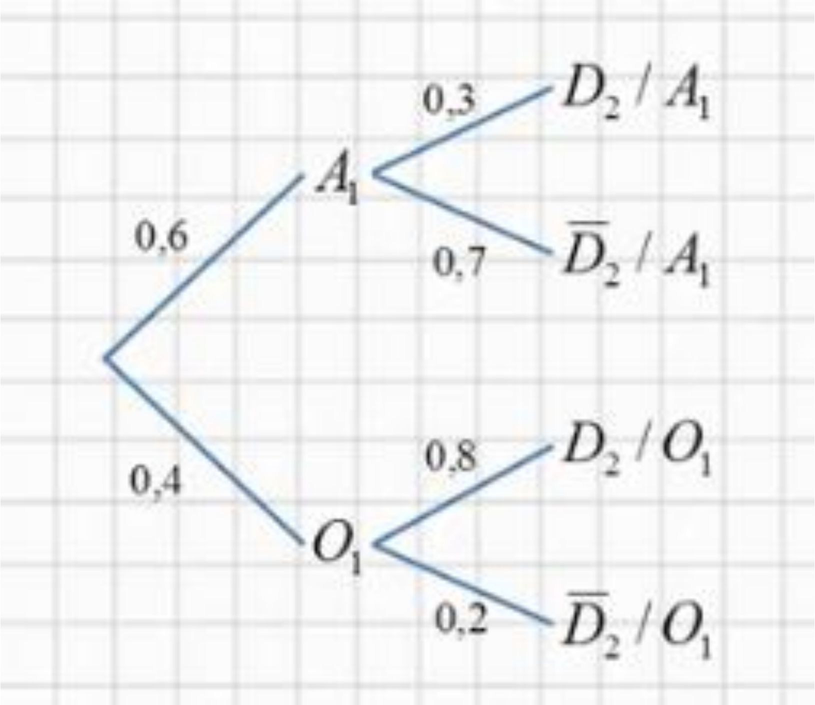
Diagrama de Árbol

- Para calcular la probabilidad de un determinado resultado, avanzamos por las ramas correspondientes
- Según avanzamos por un itinerario vamos multiplicando las probabilidades que encontremos
- Al ir multiplicando las probabilidades estamos calculando la intersección de sucesos

Ejemplo: Diagrama de Árbol

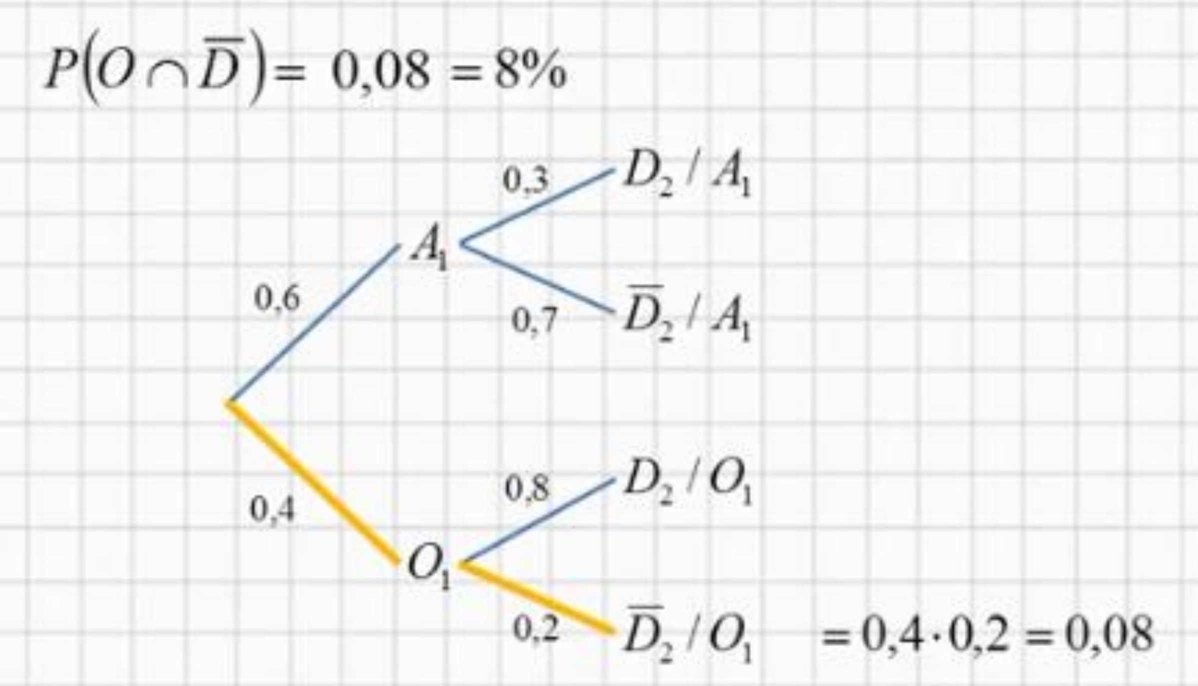
- En un determinado instituto el 60% del alumnado son chicas. De estas el 30% practica algún deporte. De los chicos, el 80% practica algún deporte.
- Si se escoge una persona al azar. ¿Qué probabilidad hay de que sea un chico que no practica deporte?
- Se escoge una persona al azar. ¿Qué probabilidad hay que sea deportista?
- Se escoge una persona al azar y resulta ser deportista. ¿Qué probabilidad hay de que sea una chica?

- Solución:



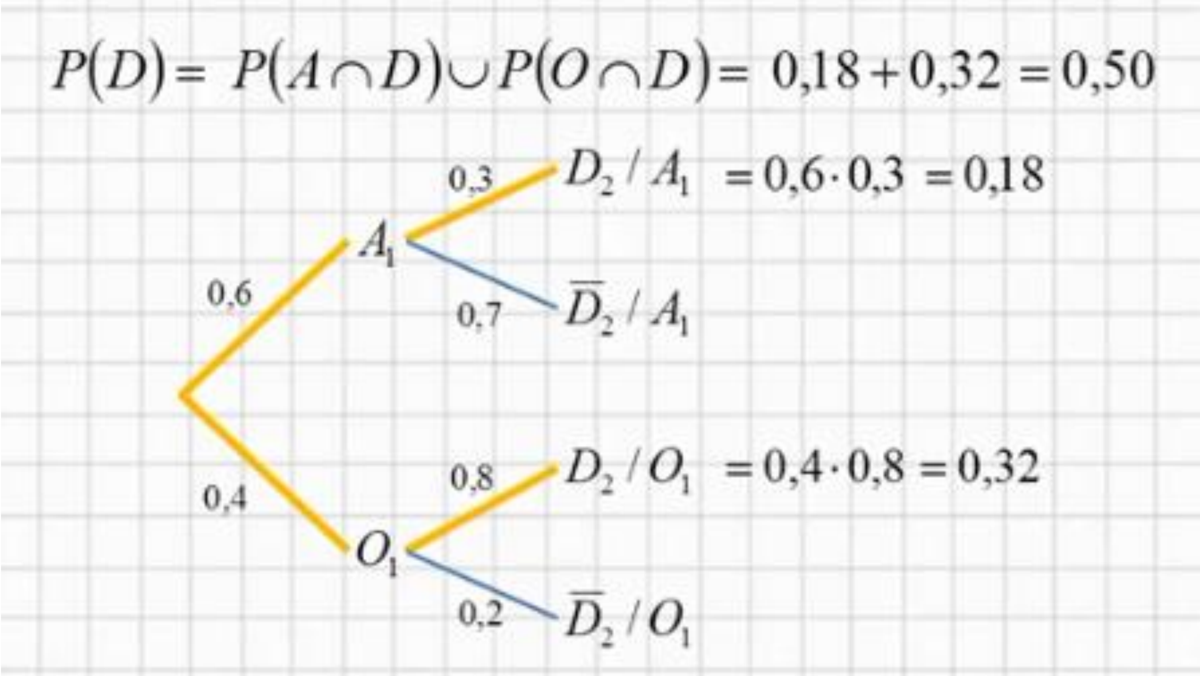
...

- Si se escoge una persona al azar. ¿Qué probabilidad hay de que sea un chico que no practica deporte?



...

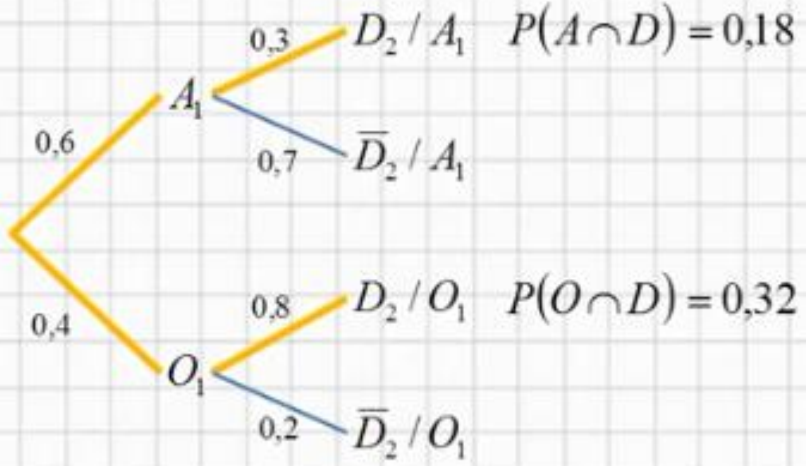
- Se escoge una persona al azar. ¿Qué probabilidad hay que sea deportista?



- Se escoge una persona al azar y resulta ser deportista. ¿Qué probabilidad hay de que sea una chica?

Recordando la regla de Laplace

$$P(A/D)=\frac{P(A\cap D)}{P(A\cap D)+P(O\cap D)}=\frac{0{,}18}{0{,}18+0{,}32}=0{,}36$$



Resumen Diagrama de Árbol

- Al avanzar por un itinerario multiplicamos las probabilidades y calculamos la intersección de sucesos
- Para calcular la probabilidad de un suceso que esta en varios itinerarios, sumamos las probabilidades en las que se encuentre, calculando la unión de sucesos.
- Calcular la probabilidad de un suceso, sabiendo que previamente ha ocurrido otro, es calcular la probabilidad condicionada

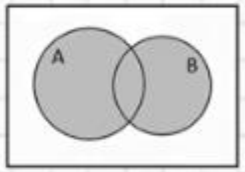
Leyes de Morgan

- Recordamos

$$P(suceso) = \frac{n^{\circ}\,casos\ favorables}{n^{\circ}\,casos\ posibles}$$

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

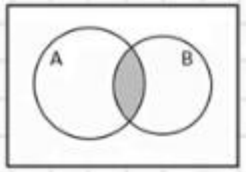
$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$



Si A y B son sucesos incompatibles: $P(A \cap B) = 0$

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

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



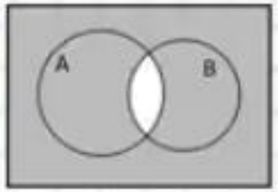
$$P(B / A) = \frac{P(A \cap B)}{P(A)}$$

Si A y B son sucesos independientes: $P(B / A) = P(B)$

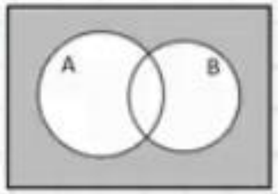
$$P(A \cap B) = P(A) \cdot P(B)$$

Leyes de Morgan

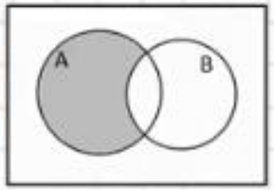
$$P(\overline{A}\cup\overline{B})=P(\overline{A\cap B})=1-P(A\cap B)$$



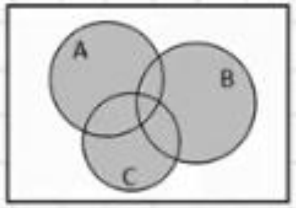
$$P(\overline{A}\cap\overline{B})=P(\overline{A\cup B})=1-P(A\cup B)$$



$$P(A\cap\overline{B})=P(A)-P(A\cap B)$$



$$P(A\cup B\cup C)=$$



$$P(A)+P(B)+P(C)-P(A\cap B)-P(A\cap C)-P(B\cap C)+P(A\cap B\cap C)$$

Muchos de los problemas se pueden resolver con formulas o con los diagramas de árbol

Ejemplo: Uso de Formulas

- Sean A y B dos sucesos independientes de un experimento aleatorio tal que P(A)=0.5 y P(no B)=0.8. Calcúlese

a) $P(A \cap B)$

$$P(A \cap B) = P(A) \cdot P(B) = 0{,}5 \cdot 0{,}2 = 0{,}1$$

b) $P(A \cup B)$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0{,}5 + 0{,}2 - 0{,}1 = 0{,}6$$

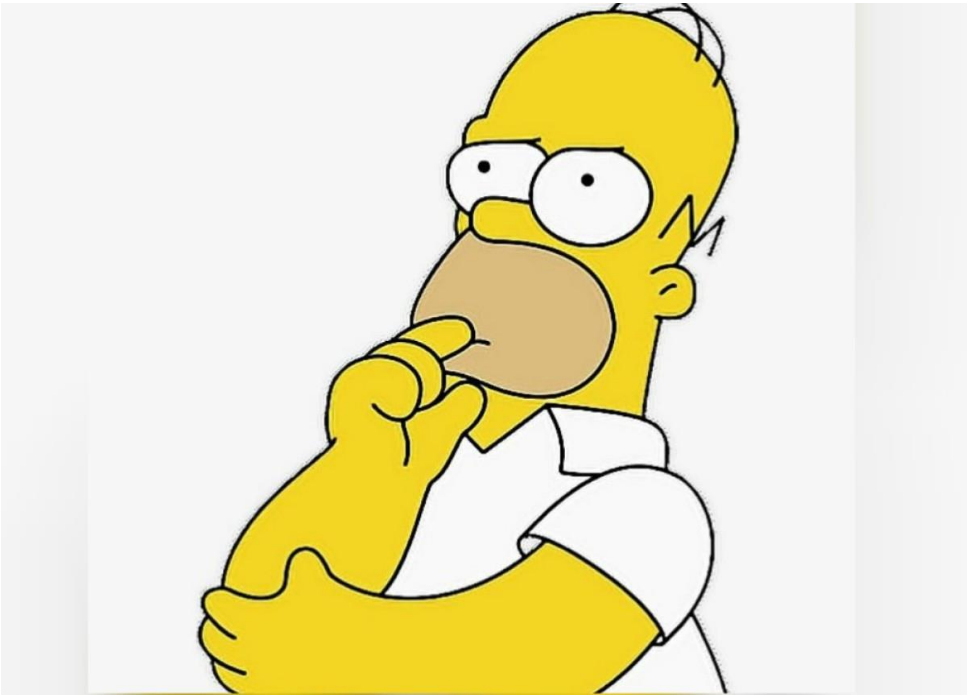
c) $P(\overline{A} / \overline{B})$

$$P(\overline{A} / \overline{B}) = \frac{P(\overline{A} \cap \overline{B})}{P(\overline{B})} = \frac{P(\overline{A \cup B})}{P(\overline{B})} = \frac{1 - P(A \cup B)}{P(\overline{B})}$$

$$= \frac{1 - 0{,}6}{0{,}8} = 0{,}5$$

Preguntas

- Alguna pregunta?



Ejercicios

- En cierto ensayo clínico se trata al 60% de pacientes con interferón y al 40% restante con ribavirina. Al cabo de unas semanas se observa una mejoría en el 43% de los pacientes tratados con interferón y el 71% de los tratados con ribavirina. Se toma un paciente al azar. Determínese la probabilidad de que:

A) Haya respondido favorablemente

B) Haya sido tratado con Interferón, si ha mejorado

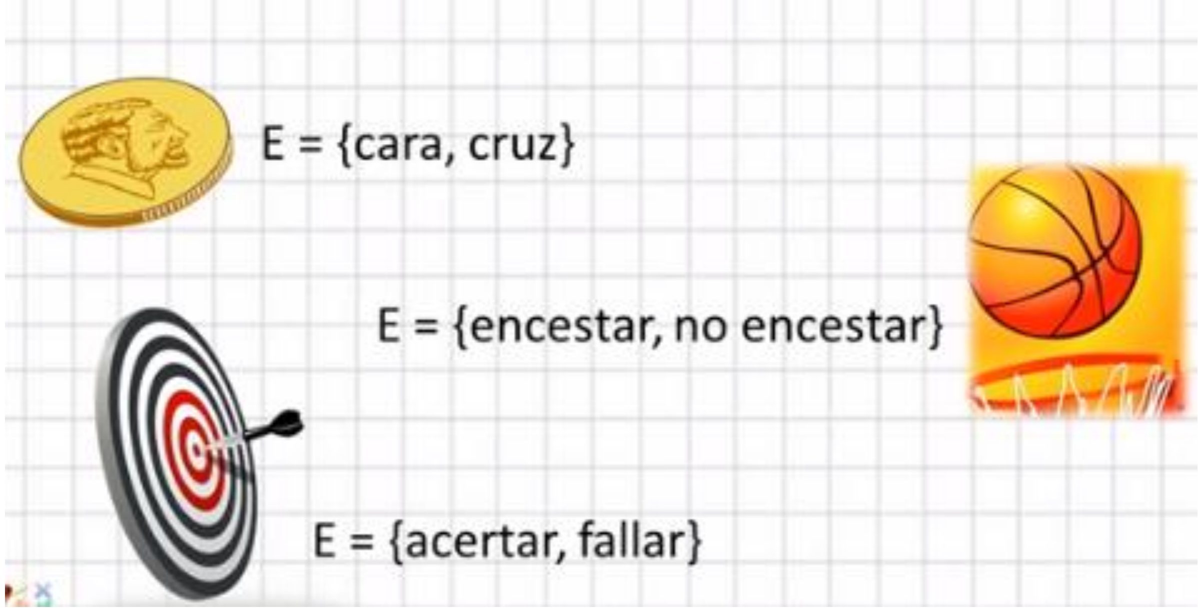
- Una empresa de paquetería tiene 3 tipos de furgonetas. El 25% de ellas tiene menos de 2 años de antigüedad, el 40% tiene una antigüedad entre 2 y 4 años y el resto tiene una antigüedad superior a 4 años. Las probabilidades de que una furgoneta se estropee son 0.01, 0.05 y 0.12 respectivamente según su antigüedad. Se escoge una furgoneta al azar. Calcúlese la probabilidad de que:

A) Se estropee

B) Tenga mas de 4 años, sabiendo que no se ha estropeado

Distribución Binomial

- Experimento aleatorio de Bernoulli, es cuando en el experimento solo existen dos resultados posibles (experimento dicotómico)

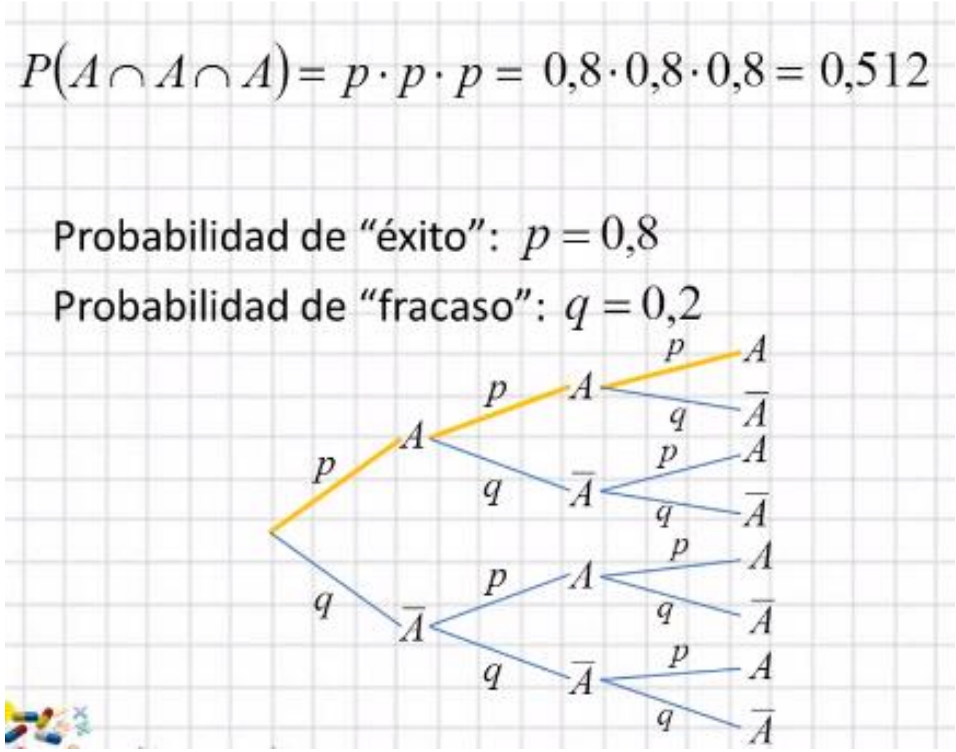


Probabilidad de “éxito”: p

Probabilidad de “fracaso”: $q = 1 - p$

Ejemplo: Distribución binomial

- Un jugador de baloncesto tiene un 80% de acierto en tiros libres. Si tira tres veces seguidas. ¿Cuál es la probabilidad que acierte los tres tiros?



...

- ¿Cual es la probabilidad de que falle tres tiros?

$$P(\overline{A} \cap \overline{A} \cap \overline{A}) = q \cdot q \cdot q = 0,2 \cdot 0,2 \cdot 0,2 = 0,008$$

- ¿Cual es la probabilidad de que falle solo el ultimo tiro?

$$P(A \cap A \cap \overline{A}) = p \cdot p \cdot q = 0,8 \cdot 0,8 \cdot 0,2 = 0,128$$

Ejercicio

- En juego de dados se gana si se sacan dos seis .¿Que probabilidad hay de sacarlos?

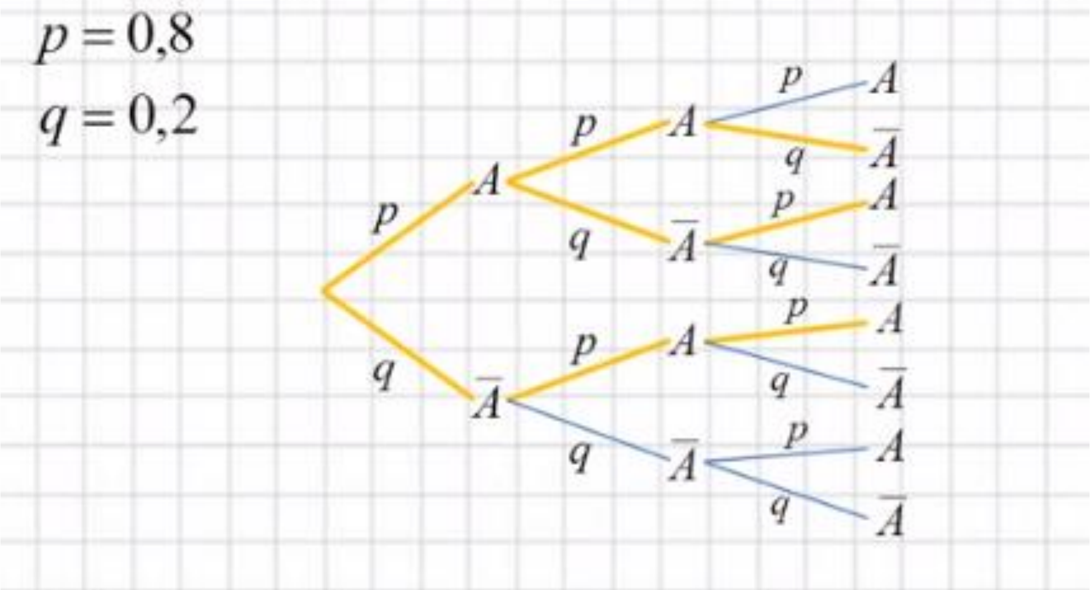
Probabilidad de "éxito": $p = 1/6$

Probabilidad de "fracaso": $q = 5/6$

$$P(6 \cap 6) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}$$

Distribución Binomial – Función de Probabilidad

- Un jugador de baloncesto tiene un 80% de aciertos en tiros libres. Si tira tres lanzamientos seguidos. ¿Cuál es la probabilidad de que acierte dos de los tres lanzamientos?



- Calculando la probabilidad de cada suceso

$$P(dos\ aciertos) = P(A \cap A \cap \overline{A}) \cup P(A \cap \overline{A} \cap A) \cup P(\overline{A} \cap A \cap A)$$

$$\left.\begin{aligned} P(A \cap A \cap \overline{A}) &= p \cdot p \cdot q = 0{,}8 \cdot 0{,}8 \cdot 0{,}2 = 0{,}128 \\ P(A \cap \overline{A} \cap A) &= p \cdot q \cdot p = 0{,}8 \cdot 0{,}2 \cdot 0{,}8 = 0{,}128 \\ P(\overline{A} \cap A \cap A) &= q \cdot p \cdot p = 0{,}2 \cdot 0{,}8 \cdot 0{,}8 = 0{,}128 \end{aligned}\right\} 3$$

$$P(dos\ aciertos) = 0{,}128 + 0{,}128 + 0{,}128 = 3 \cdot 0{,}128$$

...

- ¿Cuál seria la probabilidad de acertar dos canastas si hiciera cuatro lanzamientos?

$$\left.\begin{array}{ll} P(A\cap A\cap \overline{A}\cap \overline{A}) = 0,8\cdot 0,8\cdot 0,2\cdot 0,2 & = 0,0256 \\ P(A\cap \overline{A}\cap A\cap \overline{A}) & = 0,0256 \\ P(A\cap \overline{A}\cap \overline{A}\cap A) & = 0,0256 \\ P(\overline{A}\cap A\cap A\cap \overline{A}) & = 0,0256 \\ P(\overline{A}\cap A\cap \overline{A}\cap A) & = 0,0256 \\ P(\overline{A}\cap \overline{A}\cap A\cap A) & = 0,0256 \end{array}\right\} 6$$

$$P(dos\ aciertos) = 6\cdot 0,0256$$

- ¿Cuál seria la probabilidad de acertar dos canastas si hiciera cinco lanzamientos?

$$P(A \cap A \cap \bar{A} \cap \bar{A} \cap \bar{A}) = p \cdot p \cdot q \cdot q \cdot q$$

$$\left.\begin{array}{c} \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \end{array}\right\} ?$$

5 elementos

$p \cdot p \cdot q \cdot q \cdot q$

$p \cdot p \cdot q \cdot q \cdot q$

2 veces

$p \cdot p \cdot q \cdot q \cdot q$

3 veces

$$PR_5^{2,3} = \frac{5!}{2! \cdot 3!} = \binom{5}{2}$$

$$P(A \cap A \cap \bar{A} \cap \bar{A} \cap \bar{A}) = p \cdot p \cdot q \cdot q \cdot q$$

$$\left.\begin{array}{c} \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \end{array}\right\} 10$$

$$P(dos\ aciertos) = 10 \cdot p^2 \cdot q^3$$

Función de Probabilidad de la Distribución Binomial

$$B(n, p)$$

$$\boxed{P(x\ \acute{e}xitos) = \binom{n}{x} p^{x} \cdot q^{n-x}}$$

n: número de experimentos

p: probabilidad de éxito

Ejemplo: Función de Distribución Binomial

Un jugador de baloncesto tiene un 80% de acierto en tiros libres. Si tira **20** lanzamientos seguidos, ¿cuál es la probabilidad de que acierte **13** de los lanzamientos?

$$B(n, p) \quad B(20{,}0'8)$$

$$P(x=13)=\binom{20}{13}0{,}8^{13}\cdot 0{,}2^{7}=\frac{20!}{13!\cdot 7!}\cdot 0{,}8^{13}\cdot 0{,}2^{7}=0{,}0545$$

Ejemplo: Función de Distribución Binomial

Uno de cada diez yogures de una determinada marca tiene premio. Si compro un pack de 12 yogures, ¿cuál es la probabilidad de que me toquen 2 premios?

$$B(n, p) \quad B\left(12, \frac{1}{10}\right)$$

$$P(x=2) = \binom{12}{2}\left(\frac{1}{10}\right)^2 \cdot \left(\frac{9}{10}\right)^{10} = \frac{12!}{2! \cdot 10!} \cdot \left(\frac{1}{10}\right)^2 \cdot \left(\frac{9}{10}\right)^{10} = 0{,}23$$

Ejemplo: Función de Distribución Binomial

En un centro de ITV superan el control el 75% de los vehículos. Si en una mañana han acudido 60 vehículos:

¿Cuál es la probabilidad de que superen la inspección 40 vehículos?

$$B(n,p) \qquad B(60{,}0'75)$$

$$P(x=40) = \binom{60}{40} \cdot 0{,}75^{40} \cdot 0{,}25^{20} = \frac{60!}{40! \cdot 20!} \cdot 0{,}75^{40} \cdot 0{,}25^{20}$$

$$= 0{,}0383$$

Ejemplo: Función de Distribución Binomial

En un centro de ITV superan el control el 75% de los vehículos. Si en una mañana han acudido 60 vehículos:

¿Y cuál es la probabilidad de que superen la inspección 50 vehículos?

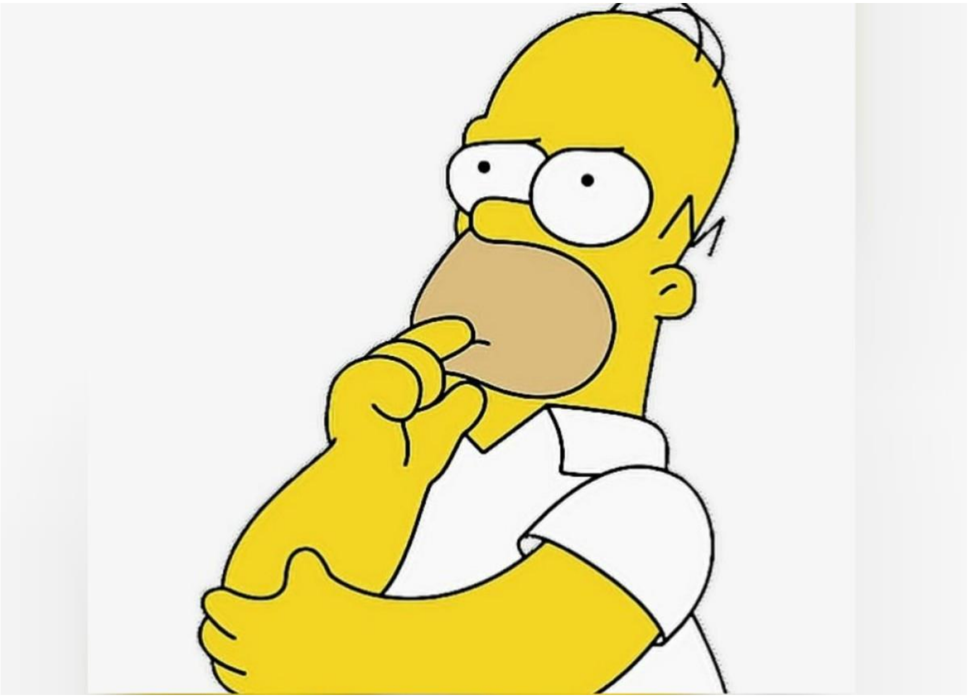
$$B(n,p) \qquad B(60{,}0'75)$$

$$P(x=50)=\binom{60}{50}\cdot 0{,}75^{50}\cdot 0{,}25^{10} = \frac{60!}{50!\cdot 10!}\cdot 0{,}75^{40}\cdot 0{,}25^{20}$$

$$=0{,}04$$

Preguntas

- Alguna pregunta?



Ejercicio:

- La probabilidad de que trabajador llegue puntual a su puesto de trabajo es ¾. Si se eligen tres trabajadores al azar, calcúlese la probabilidad de que **al menos** uno de ellos llegue puntual.

- Solución:

$$B(n,p) \quad B\left(3,\frac{3}{4}\right)$$

$$\text{P(al menos 1 llegue puntual)} = P(x=1)+P(x=2)+P(x=3)$$

$$P(x=1)=\binom{3}{1}\left(\frac{3}{4}\right)^1\cdot\left(\frac{1}{4}\right)^2=\frac{9}{64}$$

$$P(x=2)=\binom{3}{2}\cdot\left(\frac{3}{4}\right)^2\cdot\left(\frac{1}{4}\right)^1=\frac{27}{64}$$

$$P(x=3)=\binom{3}{3}\cdot\left(\frac{3}{4}\right)^3\cdot\left(\frac{1}{4}\right)^0=\frac{3}{4}\cdot\frac{3}{4}\cdot\frac{3}{4}=\frac{27}{64}$$

$$\text{P(al menos 1 llegue puntual)} = \frac{9}{64}+\frac{27}{64}+\frac{27}{64}=\frac{63}{64}$$

También puedes considerar que la probabilidad de que al menos uno llegue puntual es la probabilidad 1 – P(ninguno)

$$\boxed{P(al\ menos\ uno)=1-P(ninguno)}$$

Ejercicio: Resolver el ejercicio usando R Studio

- El 10% de los artículos producidos por una maquina son defectuosos, si elige una muestra aleatoria con reemplazo de 6 artículos.

A) determinar la probabilidad que dos artículos sean defectuosos

B) determinar la probabilidad que al menos un articulo sea defectuoso

C) determinar la probabilidad de que mas de cuatro artículos sean defectuosos