

Tema1. INCERTIDUMBRE Y PROBABILIDAD

- 1.1. Introducción. Azar, incertidumbre, selección aleatoria en una BD
 - 1.2. Probabilidad: concepción frecuentista y bayesiana. Indicadores de incertidumbre
 - 1.3. Probabilidad de sucesos. Axiomática. Propiedades.
 - 1.4. Probabilidad condicionada. Independencia estocástica. Teorema de Bayes
-

1.1 Introducción.

La palabra *azar* viene del árabe andalusí *azzahr* en el que significaba flores, de ahí viene también nuestra palabra *azahar* (la flor del naranjo). Originalmente provenía del árabe *zahr* que significaba dado. *Azar* viene a significar en nuestro idioma casualidad o fortuito cuando se aplica a un acontecimiento. El concepto en nuestro ámbito se relaciona con situaciones en las que en cierto experimento, prueba, medición o extracción o bien no se sabe cuál será el resultado o ha sido llevado a cabo por algún procedimiento que no ha considerado ningún criterio específico que nos ayude a conocer a priori el resultado.

A menudo hablaremos también de aleatorio (palabra de origen latino que significa “a suertes” y cuyo origen también se relaciona con los dados) o estocástico (que viene del griego y se relaciona con la Fortuna ser mitológico femenino completamente calva excepto por un largo pelo conocido como “la Oportunidad” al que el hombre debe asirse)

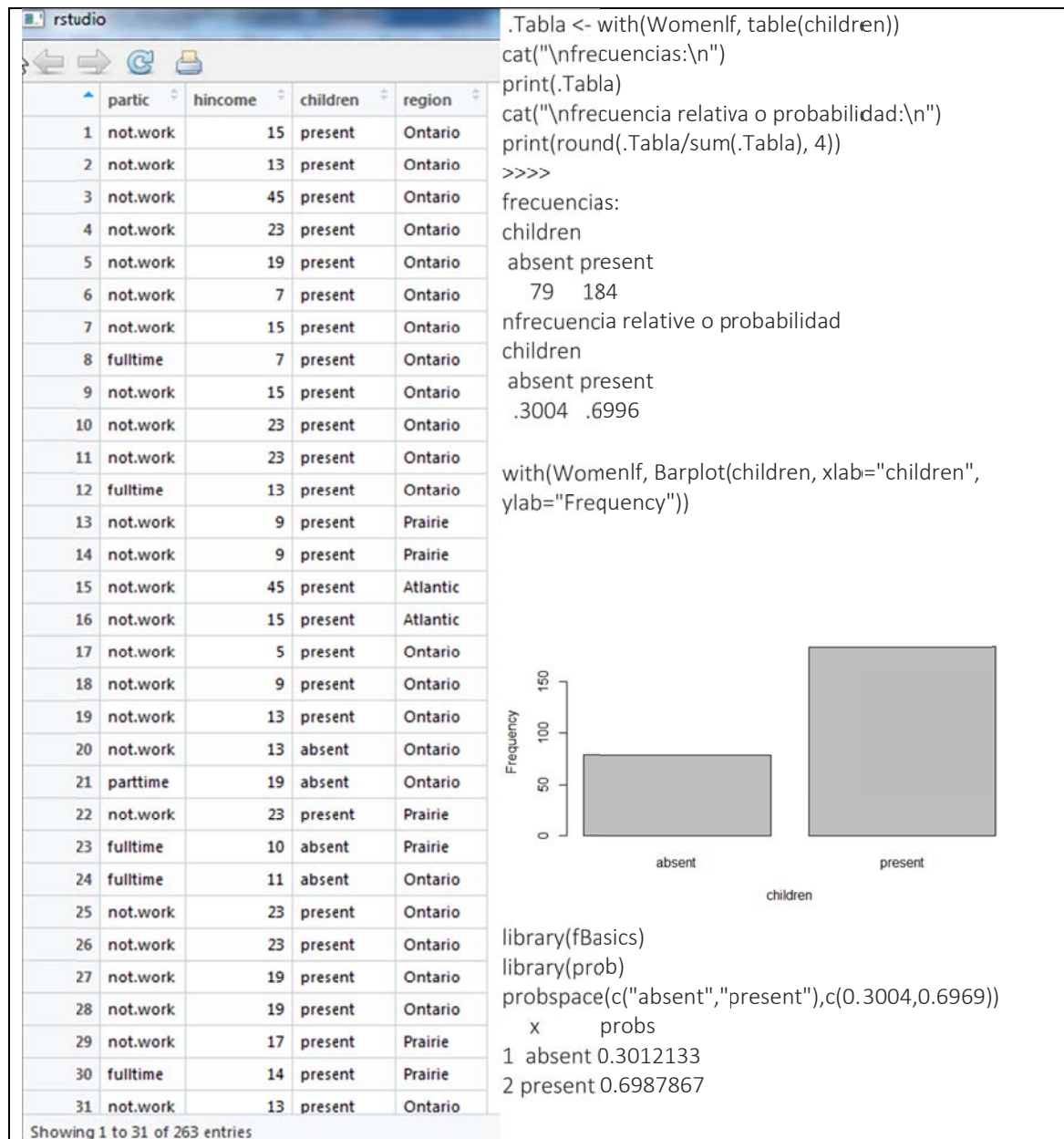
Hablaremos así de selecciones, extracciones o muestras aleatorias o al azar cuando se lleven a cabo selecciones de uno o varios individuos por algún procedimiento fortuito, azaroso.

La incertidumbre de los resultados a los que conlleva una acción aleatoria o azarosa suele intentar modelizarse mediante el concepto matemático de probabilidad. También usamos coloquialmente el concepto informal o no matemático de probabilidad para referirnos a la mayor o menor oportunidad que tenga un resultado u otro de darse en situaciones de este tipo o incluso en otras de distinta naturaleza aunque también inciertas. (p.ej.: esta semana no es probable que llueva).

Una situación típica de aleatoriedad donde se implica el azar y también el concepto de probabilidad es la selección aleatoria de individuos (registros, casos, etc.) de una base de datos.

Por ejemplo tenemos una pequeña base de datos llamada “Womenlf”, que está en el paquete de R “ carData” , que consiste en 263 individuos (mujeres casadas canadienses de las que se dispone de información de su inserción laboral, ingresos del marido en miles de \$ canadienses, presencia de hijos, y región de procedencia. Si seleccionamos al azar una mujer de esta base de datos los valores de estas variables (inserción laboral, ingresos del marido, hijos, y región) no podrán ser determinados a priori. Sin embargo es razonable pensar que cada valor de cada variable podrá darse con una probabilidad proporcional a la frecuencia con que se da en la base de datos.

```
data(Womenlf, package="carData")
Womenlf <- as.data.frame(Womenlf)
Womenlf
```



Ejercicio1.1: ¿Cómo serían la probabilidades para los distintos tipos de participación laboral?

Ejercicio1.2: Y para las distintas regiones.

Las funciones `sample()` y `runif()` de R generan números aleatorios , la primera de ellas números enteros aleatorios y la segundo números racionales aleatorios. La función `sample()` extrae una muestra aleatoria de cierto tamaño de entre un conjunto de números enteros y está en la librería `{base}` . La función `runif()` de la librería `{stats}` puede obtener un valor (racional) aleatorio de un cierto rango de la recta real (por defecto: $[0,1]$). Estrictamente, como veremos (Tema 3), obtiene uno o varios valores de una variable aleatoria que siga una distribución uniforme de probabilidad.

```
#sample(x, size, replace = FALSE, prob = NULL)
sample(1:20)
[1]  4  2  1  7 19 18 15 13  3 20  8 10 12 14 16 11 17  6  9  5
sample(1:20,2)
[1]  5  4
sample(1:20,21)
Error in sample.int(length(x), size, replace, prob) : cannot take a sample larger than the population when
sample(1:5,10,replace=TRUE)
[1] 1 5 5 5 3 1 2 1 1 3
sample(1:5,10,replace=TRUE)
[1] 2 1 2 3 1 4 1 4 3 4
sample(1:5,10,replace=TRUE,prob=c(.3,.2,.1,.2,.2))
[1] 4 1 1 5 1 2 1 1 2 3
```

Podemos utilizar la función `sample` para seleccionar para obtener una muestra aleatoria con reemplazamiento de tamaño 5 para el atributo `Children` de la base de datos `Womenlf`

```
children<-factor(c("ausencia","presencia"))
children
[1] ausencia presencia
Levels: ausencia presencia
sample (children,5, TRUE, c(.3012133,.6987867))
[1] presencia presencia ausencia ausencia presencia
Levels: ausencia presencia
sample (children,5, TRUE, c(.3012133,.6987867))
[1] presencia presencia ausencia ausencia
Levels: ausencia presencia
```

Alternativamente también podríamos cargar la base de datos y aplicar el muestreo con reemplazo a la variable (factor) `Children`

```
data(Womenlf, package="carData")
Womenlf <- as.data.frame(Womenlf)
sample (Womenlf$children,5, TRUE)
[1] present absent present present absent
```

Y para llevar a cabo una selección sin reemplazamiento:

```
sample (Womenlf$children,5, FALSE)
[1] present present present present present
Levels: absent present
```

Ejercicio 1.3. Llevar a cabo selecciones aleatorias semejantes a las de arriba con y sin reemplazamiento para los factores participación laboral y región y para la variable “ingresos del marido” (`hincom`)

1.2. Probabilidad: concepción frecuentista y bayesiana. Indicadores de incertidumbre.

La idea intuitiva de probabilidad y su definición más formal apuntan a una cierta medida de la oportunidad que tiene algo (que puede o no ocurrir) de llegar a producirse. Concebida así parece claro que pueden admitirse distintos matices y de hecho la interpretación de lo que realmente sea la probabilidad es un asunto controvertido desde el punto de vista, tanto científico como filosófico. Si entrar en demasiada profundidad y obviando algunas otras concepciones (objetivista y logicista p.ej.) vamos a dedicarle algunas líneas a los enfoques fundamentales del concepto de probabilidad.

Es un hecho intersubjetivamente admitido que si dada una colección de elementos de cualquier tipo (p.ej. una base de datos) la elección aleatoria de uno de ellos nos llevará a una situación en la que , con carácter general, todos coincidiremos en que :

Habrà mayor **probabilidad** de que el elemento elegido sea de un cierto tipo cuando mayor sea la **frecuencia** con que se da ese elemento en la colección y habrá una probabilidad tanto menor cuando la frecuencia sea menor.

Esa estrecha relación entre frecuencia y probabilidad lleva a pensar en algún tipo de relación entre una y otra. Sin embargo, el concepto de probabilidad implica una idea resbaladiza de “oportunidad” que no siempre puede asociarse a la extracción aleatoria de un elemento de una colección (p.ej. tiradas de dados, fenómenos meteorológicos, etc) y que implica también inseguridad o incertidumbre Por lo que no podemos asimilar, sin más, uno y otro concepto.

Una opción es considerar que la probabilidad es un límite de la frecuencia relativa cuando el número de realizaciones de un fenómeno aleatorio tiende a infinito.

Esta opción ha resultado muy fructífera en la práctica y ha supuesto la base filosófica de casi toda la inferencia estadística llamada “clásica”. Es, sin embargo, muy controvertida desde el punto de vista filosófico por dos razones principales, dejar fuera todas las situaciones intrínsecamente “irrepetibles” y por establecer cierta circularidad ontológica. Pero en esto, no vamos a entrar aquí.

Por otro lado, alternativamente, se ha ofrecido la opción de considerar que la probabilidad es una medida subjetiva sobre la confianza que un cierto sujeto deposita en la ocurrencia de cierto hecho, suceso, etc. esta es la base de la interpretación subjetivista, que cuando se exigen ciertos requisitos de racionalidad en el “hipotético” sujeto, permiten desarrollar una estadística igualmente válida.

Asumir una u otra postura ha llevado a desarrollos de las técnicas estadísticas diferentes que, a grandes rasgos, podemos denominar: “Inferencia estadística clásica” e “Inferencia estadística bayesiana”.

Junto con estas dos posturas y sin considerar que se trata de una tercera alternativa podemos tener en cuenta que enunciar una aserción o una hipótesis sobre cierto hecho cabe la posibilidad de que esta aserción tenga las posibilidades de resultar cierta o tautológica, imposible o falsa, o si se trata de un fenómeno no seguro, azaroso o aleatorio, meramente posible o contingente. De alguna manera estas situaciones están muy presentes en muy variados ámbitos y otra de las

funciones fundamentales del concepto de probabilidad va a ser dar cuenta de esta relativa "incertidumbre" que acompaña a estas situaciones.

1.3. Probabilidad de sucesos. Axiomática. Propiedades.

<https://www.uv.es/ceaces/base/probabilidad/simple.htm>

<https://www.uv.es/ceaces/pdf/prob1.pdf>

Vamos a generar distintos espacios de resultados (muestrales) y a trabajar con sucesos, y sus relaciones y funciones así como con la probabilidad de sucesos con R.

Un espacio de resultados, que en realidad no es más que un conjunto, puede obtenerse en R, como una lista o un vector y luego probabilizarlo adecuadamente:

```
require(prob)
#Podemos construir un un espacio de resultados y probabilizarlo y "armarlo" todo como un data frame
omega<- c("a","b","c","d","e")
probabilidades<- c(0.3,0.25,0.2,0.15, 0.1)
sum(probabilidades)
espacio.omega<-as.data.frame(cbind(omega,probabilidades))
view(espacio.omega)

# o bien podemos crear un espacio de resultados a través de la función probspace y aprovechar las
ventajas del paquete "prob"
espacio.omega <-probspace(omega, probs = probabilidades)
view(espacio.omega)
```

Dentro del paquete prob (vignette("prob"), para obtener la documentación) encontramos los más habituales espacios de resultados asociados a juegos de azar (lanzamiento de monedas, dados, extracciones de cartas, ruleta etc.) que suelen utilizarse para ilustrar el cálculo de probabilidades.

```
#monedas    tosscoin(número de monedas, probabilizado)
tosscoin(2)
tosscoin(3, makespace = TRUE)
#####
#dados      rolldie(número de dados, número de caras, probabilizado)
rolldie(2,makespace = TRUE)
rolldie(1,4,TRUE)
rolldie(3,10,TRUE)
#####
#cartas     cards(jokers, probabilizado)
cards()
cards(jokers=TRUE,makespace=TRUE)
#####
#ruleta     roulette(european,probabilizado) //european (disposición de números eu)
roulette()
roulette(european=TRUE,makespace=TRUE)
```

El espacio de resultados (espacio muestral) es el conjunto de resultados simples. Un suceso no es más que un subconjunto de Omega (una colección de resultados simples)

ejemplo: tirada de 2 monedas Omega es {(c,+),(c,c),(+,c),(+,+)} . Probabilidad de alguna cara.

Omega <- tosscoin(2, makespace = TRUE) Omega #esl suceso sacar alguna cara consistirá algunacara<-Omega[1:3,] algunacara Prob(algunacara)	toss1 toss2 probs 1 H H 0.25 2 T H 0.25 3 H T 0.25 > Prob(algunacara) [1] 0.75
--	---

Distintos sucesos en el experimento "Obtener una carta al azar"

omega<-cards(makespace=TRUE) subset(omega) #el propio suceso omega subset(omega, suit == "Heart") #suceso corazones corazones<-subset(omega, suit == "Heart") Prob(corazones) subset(omega, rank %in% 7:9) Prob(subset(omega, rank %in% 7:9))	 corazones rank suit probs 27 2 Heart 0.01923077 28 3 Heart 0.01923077 29 4 Heart 0.01923077 30 5 Heart 0.01923077 31 6 Heart 0.01923077 32 7 Heart 0.01923077 33 8 Heart 0.01923077 34 9 Heart 0.01923077 35 10 Heart 0.01923077 36 J Heart 0.01923077 37 Q Heart 0.01923077 38 K Heart 0.01923077 39 A Heart 0.01923077 Prob(corazones) [1] 0.25 subset(omega, rank %in% 7:9) ; Prob(subset(omega, rank %in% 7:9)) rank suit probs 6 7 Club 0.01923077 7 8 Club 0.01923077 8 9 Club 0.01923077 19 7 Diamond 0.01923077 20 8 Diamond 0.01923077 21 9 Diamond 0.01923077 32 7 Heart 0.01923077 33 8 Heart 0.01923077 34 9 Heart 0.01923077 45 7 Spade 0.01923077 46 8 Spade 0.01923077 47 9 Spade 0.01923077 [1] 0.2307692
---	---

#lanzamiento de 3 dados que sumen más de 16 subset(rolldie(3), X1+X2+X3 > 16) Prob(subset(rolldie(3,makespace=TRUE), X1+X2+X3 > 16))	subset(rolldie(3), X1+X2+X3 > 16) X1 X2 X3 180 6 6 5 210 6 5 6 215 5 6 6 216 6 6 6 > Prob(subset(rolldie(3,makespace=TRUE), X1+X2+X3 > 16)) [1] 0.01851852
--	--

Las relaciones y funciones entre sucesos : pertenencia, inclusión , unión, intersección, complementación etc. no son más que relaciones y funciones conjuntos y pueden también automatizarse con R :

<pre># función para evaluar la relación de pertenencia %in% x<-2:10 y<-8:15 y %in% x y %in% corazones\$rank</pre>	<pre>y %in% x [1] TRUE TRUE TRUE FALSE FALSE FALSE FALSE FALSE y %in% corazones\$rank [1] TRUE TRUE TRUE FALSE FALSE FALSE FALSE FALSE</pre>
<pre>all(y %in% x) all(x %in% corazones\$rank)</pre>	<pre>all(y %in% x) [1] FALSE > all(x %in% corazones\$rank) [1] TRUE</pre>
<pre># función para evaluar la relación de inclusión isin(X,y) #:: ¿está incluido en x el conjunto y? isin(x, y) y<-c(2,4,6) isin(x, y) y<-c(3,2,5) isin(x, y) y<-c(3,3,5) isin(corazones\$rank,y)</pre>	<pre>isin(x, y) [1] FALSE y<-c(2,4,6) isin(x, y) [1] TRUE y<-c(3,2,5) isin(x, y) [1] TRUE y<-c(3,3,5) isin(corazones\$rank,y) [1] TRUE</pre>
<pre>all(y %in% x) # verdadero porque todos los elementos de y #pertenecen a x isin(x,y) # falso porque y no está incluido en x ya que hay #una repetición del elemento 3 isin(x, c(3,4,5), ordered = TRUE) # podemos exigir el mismo orden isin(x, c(5,4,3), ordered = TRUE) # al cambiar el orden será falso</pre>	<pre>all(y %in% x [1] TRUE isin(x,y) # [1] FALSE isin(x, c(3,4,5), ordered = TRUE) [1] TRUE isin(x, c(5,4,3), ordered = TRUE) [1] FALSE</pre>
<pre>#union S <- cards(makespace = TRUE) A <- subset(S, suit == "Heart") C <- subset(S, suit == "Spade") B <- subset(S, rank %in% 7:9) union(A,B) union(A,C) Prob(union(A,C)); Prob(A) ; Prob(C) Prob(union(A,B)); Prob(A) ; Prob(B)</pre>	<pre>Prob(union(A,C)); Prob(A) ; Prob(C) [1] 0.5 [1] 0.25 [1] 0.25 Prob(union(A,B)); Prob(A); Prob(B) [1] 0.4230769 [1] 0.25 [1] 0.2307692</pre>
<pre># intersección intersect(A,B) Prob(intersect(A,B)) Prob(union(A,B)) ; Prob(A) + Prob(B)- Prob(intersect(A,B))</pre>	<pre>intersect(A,B) rank suit probs 32 7 Heart 0.01923077 33 8 Heart 0.01923077 34 9 Heart 0.01923077 Prob(intersect(A,B)) [1] 0.05769231 Prob(union(A,B)); Prob(A)+Prob(B)-Prob(intersect(A,B)) [1] 0.4230769 [1] 0.4230769</pre>

#Diferencia : dif(A,B): pertenece a A pero no a B == A & no(B) Prob(setdiff(A,B)) Prob(A);Prob(intersect(A,B));Prob(setdiff(A,B))	Prob(setdiff(A,B)) [1] 0.1923077 Prob(A);Prob(intersect(A,B)); Prob(setdiff(A,B)) [1] 0.25 [1] 0.05769231 [1] 0.1923077
--	---

1.4.Probabilidad condicionada. Independencia estocástica. Teorema de Bayes

<https://www.uv.es/ceaces/base/probabilidad/simple.htm>

<https://www.uv.es/ceaces/pdf/prob1.pdf>

Con R podemos trabajar la probabilidad condicionada con la función Prob , añadiendo el argumento “given = condición”

require(prob) S <- cards(makespace = TRUE) A <- subset(S, suit == "Heart") C <- subset(S, suit == "Spade") B <- subset(S, rank %in% 7:9) Prob(A,given=B) Prob(intersect(A,B)) Prob(B) Prob(intersect(A,B))/Prob(B)	Prob(A, given=B) [1] 0.25 Prob(intersect(A,B)) [1] 0.05769231 Prob(B) [1] 0.2307692 Prob(intersect(A,B))/Prob(B) [1] 0.25
--	--

S <- rolldie(2, makespace = TRUE) A <- subset(S, X1 == X2) B <- subset(S, X1 + X2 >= 8) Prob(A);Prob(B) Prob(union(A,B)) Prob(intersect(A,B)) Prob(A,given=B) Prob(B,given=A)	Prob(A) [1] 0.1666667 Prob(B) [1] 0.4166667 Prob(union(A,B)) [1] 0.5 Prob(intersect(A,B)) [1] 0.08333333 Prob(A, given=B) [1] 0.2 Prob(B, given=A) [1] 0.5
--	---

En el paquete prob, que estamos utilizando para aplicar los conceptos de la teoría elemental de la probabilidad no hay una función contruida específica para dar cuenta de la regla (o Teorema) de bayes . Sin embargo puede diseñarse ad-hoc un esquema de actuación según el script siguiente:

```
#Construimos un espacio inicial de probabilidades a partir
#de los sucesos de la partición a1,a2,a3,... an
# y de la asignación inicial de probabilidades / prob. iniciales/ prob.a priori
particion<- c("a1","a2","a3","a4") # con tantos elementos(sucesos) como sean necesarios
a.priori<-c(0.145,0.256,0.45,0.149)
# sum(a.priori) # comprobamos que suma 1
```

```
#Construimos el espacio muestral probabilizado#
distribucion.inicial<-probspace(particion, probs=a.priori)
distribucion.inicial

#consideramos un vector de verosimilitudes o
#probabilidades de un cierto suceso B condicionadas a cada uno de los sucesos de la partición
verosimilitud <-c(0.3,0.4,0.2,0.7)
#la distribución final vendrá dada por la regla de bayes.
# De forma intermedia podemos obtener las probabilidad de las intersecciones ai*b
intersecciones<-verosimilitud*a.priori
intersecciones
# y cada probabilidad final será la prob de la intersecciones/ suma prob.intersecc
a.posteriori<-intersecciones/sum(intersecciones)
a.posteriori
sum(a.posteriori)
distribucion.final<-probspace(particion,probs=a.posteriori)
distribucion.final

resumen<- as.data.frame(cbind(particion,a.priori,verosimilitud,intersecciones,a.posteriori))
resumen
```

Cuya salida , si lo ejecutamos será:

```
distribucion.inicial
  x probs
1 a1 0.145
2 a2 0.256
3 a3 0.450
4 a4 0.149
intersecciones
[1] 0.0435 0.1024 0.0900 0.1043
a.posteriori
[1] 0.1278660 0.3009994 0.2645503 0.3065844
sum(a.posteriori)
[1] 1
distribucion.final
  x probs
1 a1 0.1278660
2 a2 0.3009994
3 a3 0.2645503
4 a4 0.3065844

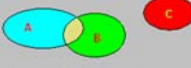
resumen
resumen
  particion a.priori verosimilitud intersecciones a.posteriori
1      a1      0.145           0.3         0.0435 0.127865961199295
2      a2      0.256           0.4         0.1024 0.300999412110523
3      a3       0.45           0.2          0.09 0.264550264550265
4      a4      0.149           0.7         0.1043 0.306584362139918
```

Obviamente, los elementos de la partición pueden variar en número y nombre y los valores de las probabilidades y verosimilitudes también pero el esquema será siempre el mismo.

CEA
REF [C] 1-100

test de teoría de la probabilidad I

2,2 10 2,2 2,2

- Si $P(A) = 0,2$ y $P(B) = 0,4$ conociendo que A y B son independientes entonces podemos asegurar que
 - ☐ $P(A \cap B) = 0,2$
 - ☐ $P(A \cap B) = 0,08$
 - ☐ A y B son disjuntos
- Siendo A, B y C sucesos con probabilidad no nula y cuya representación gráfica es :
 

entonces podemos afirmar

 - ☐ Los sucesos A y C son independientes
 - ☐ Los sucesos B y C son independientes
 - ☐ Los sucesos A y B pueden ser independientes
 - ☐ Los sucesos A y B no son independientes
- Si $P(A) = 0,2$ y $P(B) = 0,4$ conociendo que A y B son independientes entonces
 - ☐ $P(A \cup B) = 0,52$
 - ☐ $P(A \cup B) = 0,2$
 - ☐ $P(A \cap B) = 0,6$
- Siendo $P(A) = 0,5$ y $P(B) = 0,2$ y teniendo la siguiente representación gráfica:
 

Entonces podemos afirmar

 - ☐ $P(A \cap B) = 0,2$
 - ☐ $P(A \cup B) = 0,5$
 - ☐ $P(A \cap B) = 0,3$
- Siendo $P(A) = 0,7$ $P(B) = 0,4$ entonces
 - ☐ $P(A \cap B) = 0,21$
 - ☐ A y B son disjuntos
 - ☐ A y B son independientes
 - ☐ $A \cap B \neq \emptyset$

CULTURILLA CULTURETA.TIPS PARA FOMENTAR LA CURIOSIDAD CRÍTICA (VALOR UNIVERSITARIO EN DESUSO)

- Los primeros estudios serios de probabilidad surgen a partir de problemas relacionados con los juegos de azar y de apuestas. Antoine Gombard, **Caballero De Meré**, afamado jugador francés del siglo XVII les planteó algunos problemas relacionados con distintas apuestas a Fermat y Pascal .
http://matematicasentumundo.es/HISTORIA/historia_Pascal.htm
- El primer tratado sistemático de probabilidad es el Ars Conjectandi de Jacob Bernouilli , publicado póstumamente en 1713 en el sienta las bases tanto de la concepción frecuentista como de la logicista, y demuestra el teorema que lleva su nombre en el que prueba la convergencia de la frecuencia relativa a la probabilidad. <http://www.miscelaneamatematica.org/Misc60/6001.pdf>
- Flores, azar, fortuna, suerte y juegos de azar siempre han estado curiosamente entrelazados. Azar y azahar tienen un origen común, dados y suerte originalmente se decían igual en latín, lotería (el original lotto) se entronca con la “flor de loto”, el trébol de cuatro hojas es un símbolo de buena suerte. Algo que quizás deba investigarse más.
- Carmina Burana un oratorio compuesto Carl Orff en 1935-1936 está compuesto de piezas cuyas letras son poemas medievales de los goliardos, jóvenes clérigos y estudiantes vagabundos bastante vividores. La mayoría de ellas hacen referencia al vino, las mujeres y los juegos de dados. Especialmente destacable es la famosa obertura-epílogo “Oh, Fortuna”- “Fortuna Imperatrix Mundi”
<https://youtu.be/hp1VSSPvEao>